View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by University of Liverpool Repository

Editorial

How could we know if communication skills training needed *no* more evaluation? The case for rigour in research design

Peter Salmon[1] & Bridget Young[2]

Departments of [1]Psychological Sciences and [2]Health Services Research, University of Liverpool, Whelan Building, Brownlow Hill, Liverpool L69 3GB, UK.

Correspondence to:

Peter Salmon
Division of Clinical Psychology
University of Liverpool
Whelan Building
Brownlow Hill
Liverpool, L69 3GB

'phone +44 151 794 5531
fax +44 151 794 5537
email psalmon@liv.ac.uk

How could we know if communication skills training needed no more evaluation? The case for rigour in research design

An updated Cochrane review has confirmed the findings of previous reviews – that communication skills training (CST) can change oncology practitioners' behavior, but without clear evidence that patients benefit[1]. The authors conclude that more research needs to be done, and that this research must be of better quality than the studies they reviewed. Educational researchers also recognize the need for stronger evaluations, as Bylund et al recently illustrated in a report explicitly titled a 'robust' evaluation[2]. But what does robust evaluation mean in this context?

Most evaluations in PEC over recent years have, like Bylund et al's, emphasized statistical testing of effects on outcomes, often in the context of randomized trials, thereby aligning CST evaluation with the 'hierarchy of evidence' against which the efficacy of interventions across medicine is now routinely judged. CST evaluation is difficult, because it must contend with real-world constraints that exclude elements of quality research design that are common in clinical research. For instance, participants cannot be blinded to being trained. Similarly, randomization can be impractical where a training program is embedded across an institution[2]. Nevertheless, the standards by which quantitative evaluations of clinical interventions are now routinely scrutinized[3] provide pointers to what robust CST evaluation could look like. When judged against these standards, evaluations of CST that have been influential historically now look weak, and more recent evaluations lack essential elements of quality[4]. Here we identify ways in which such evaluation could improve if it is to become sufficiently robust to command authority more generally in clinical medicine – and to help reviewers tell when *enough* research of this kind has been done.

First, choose the primary outcome

The starting point for evaluation to reach current standards is a primary hypothesis that specifies a primary outcome; that is, a dependent variable on which success or failure of the intervention can be judged. The more statistical comparisons that are performed, the greater the risk that significant findings reflect random processes, so clinical trials normally specify only one or a very few primary outcomes[3]. Other outcomes are 'secondary'; that is, they answer questions other than 'did the intervention work?' Primary outcomes should be documented in

advance in a registered protocol to exclude any risk that outcomes are prioritized according to how they perform in data analysis. CST evaluation has not routinely distinguished between primary and secondary outcomes, or used pre-registered protocols. Girgis et al's report was a valuable exception in distinguishing primary from secondary outcomes[5]; however, since their primary outcome (emotional function) did not change, the inference should arguably be that intervention failed, notwithstanding the response of one secondary outcome (anxiety). When other studies have specified hypotheses, there have typically been multiple primary outcomes. For instance, the 'primary hypothesis' of a widely cited evaluation referred to eight outcome measures (each analyzed in multiple ways), only half of which responded[6]. In Bylund et al's report[2], evaluation was considered robust, in part, because of the range of variables measured at different levels, from learners' course evaluation to their patients' experience. Therefore, the primary research question entailed statistical comparisons of more than 100 outcomes, of which around half responded[2]. In another report, 13 out of 30 outcomes responded[7]; in another, seven out of 11[8]. Given that such studies routinely use the conventional significance criterion of $p<.05$, some findings probably reflect random variation.

CST evaluators do face a challenge in choosing primary outcomes because of the breadth of variables relevant to learning to communicate[9]. Outcomes can encompass learners' subjective experience, including satisfaction with training and confidence in communication, their performance of skills, and their patients' experience. While some of these outcomes have obvious face validity – learner satisfaction is a prerequisite for a training program to be viable – others are questionable. For instance, in a recent qualitative study, self-doubt and reflection on their communication were central in surgical oncologists' accounts of learning to communicate[10]. From that perspective, learners' confidence would be an ambiguous outcome: over-confidence might prematurely halt learning from practice.

Identify the theoretical framework

The bigger problem here is the need for theory that can give specific outcomes clear meaning[9, 11], and for better understanding of what practitioners need[12] and patients value[13]. Most evaluations implicitly or explicitly prioritize performance of communication skills as the most important outcomes. However, without theory that is sufficiently detailed to

distinguish the specific role of individual skills, findings can be hard to interpret. For instance, what should we make of a report that 'empathy' increased after CST but 'checking understanding' did not[6]? Both were predicted to increase, so did CST fail? Suppose, instead, that 'checking understanding' had increased but empathy had not. Would that be just as good (or poor) a result? Are the different outcomes substitutable, in that it does not matter which one(s) increase provided some do? Or are some outcomes more important than others; for instance, might 'checking understanding' need to increase in order that patients can benefit from increased 'empathy'? The challenge for researchers is to explain the underlying theory linking their intervention to patient benefit more precisely than they typically do[9]. Then they could derive a single primary outcome (or, at most, a very few). For instance, if their theory predicts that patients benefit from greater use of any of the measured skills, then a primary outcome might be the total of individual skills. If every skill is crucial, the primary outcome could be whether all are displayed in any consultation.

Decide on a target level of change

Having identified a primary outcome, the trialist's next task is to decide by how much the intervention should change for it to be judged successful. In CST research, this could mean specifying, for example, what level of patient satisfaction indicates a good consultation, or what level of performance of specific skills is desirable. Girgis et al were, again, unusual in specifying a target level of improvement (in patients' emotional function)[5]. Specifying targets requires attention to starting levels. In Bylund et al's report, baseline patient satisfaction with many aspects of their doctors' communication exceeded 4.5 on a 5-point scale, so might have been adequate from the start[2]. Pre-training counts of skills are even harder to interpret because communication theory lacks clear benchmarks for their performance. For the same reason, target outcomes cannot usually be specified. When CST trialists seek, instead, to increase skill performance from whatever is the starting level, they adopt the implicit theory that 'the more the better'. Taken to the limit, the target consultation style would therefore be one filled with talk that displays the measured skills, whether or not it achieves the clinical work that communication skill scoring systems often exclude. Similarly, without theory about the importance of confidence and self-doubt in learning communication, the implicit theory is typically that total confidence is better than a level that might leave learners inclined to reflect and question their

performance.

Accepting the need to define a target level of change allows trialists to address a related element of quality in design – clinical significance. That is, what extent of change in the primary outcome will matter in practice? Unfortunately, while CST trials report statistical significance, they only rarely refer to clinical significance. Some, cautiously, warn that statistically significant findings might not be clinically meaningful[2] or, incautiously, suggest that non-significant ones might be. However, embracing clinical significance will mean a more systematic approach to calibrating outcomes in the context of theory that goes beyond 'more is better'.

Choose the sample size

Having chosen a primary outcome and having specified a clinically significant level of change, trialists can decide how many participants will be needed to be confident of a fair test of training. That is, they can ensure adequate power. Unfortunately, CST trials routinely neglect power. In this respect, also, Girgis et al's report was an exception, in linking their intended sample size to the target level of change in the primary outcome[5]. Without knowing that a study was adequately powered, the reader cannot know what to make of failure to find an effect. Were there too few participants; or was the prediction flawed? It might be thought that power no longer matters because meta-analyses periodically combine data from single studies into large data-sets with enhanced power. But under-powered trials are poor ones, increasing the risk of reporting bias whereby trials with few or no significant effects remain unpublished. Moreover, even CST meta-analyses are limited in power, given the diverse measures that different trials use. Considerations of power were not within the remit of the recent Cochrane review[1], but confidence in each of its meta-analyses depends on whether sufficient participants were available to them. Samples for tests of different outcomes ranged from 202-844 participants. These are large by the standards of CST research but might not be large enough, particularly if trialists were to become more ambitious in targeting patient-centred outcomes such as morbidity or health-related behaviour rather than practitioners' performance of the skills they were taught. Studies linking practitioners' communication to reduced morbidity in primary care, or to smoking cessation, have sample sizes more characteristic of epidemiology than of the small-scale studies that predominate in research into communication in oncology[14, 15].

Is more research needed, and what kind of research?

We can now return to our starting point: the recommendation, in response to unconvincing evidence for the value of CST, that 'more research needs to be done'[1]. More research that disregards standards of design that are now routine across clinical evaluation risks adding to the mountain of 'research waste' that squanders resources and distorts clinical science because of poor design, inadequate methodology and partial or biased reporting[16]. By contrast, if CST trialists were to address the sequence of design decisions that begin with identifying a primary outcome, they might ultimately be able to reach more informative conclusions than 'more research needs to be done'. They might even be able to tell when *no* more research of this kind needs to be done!

Of course, evaluations that have multiple outcomes, and that are under-powered for some or all of these, can be valuable as exploratory or developmental studies. They can help refine interventions, shape complex interventions like CST, or prepare for definitive evaluation[11]. They can test theory, such as in Meunier et al's study of the role of physiological arousal in implementing learned skills[17]. In particular, measuring outcomes across different levels of effect, as Bylund et al[2] and Pehrson et al[8] did, is essential if we are to understand processes that mediate between teaching and practice. But it seems that, often, CST evaluators have conflated exploratory and definitive evaluation; based on findings from designs that are essentially exploratory, they have claimed the authority associated with definitive evaluation to advocate roll-out of CST programmes as proven elements of evidence-based practice.

Arguably, conflating exploratory and definitive evaluation belies an enduring conflation of moral and evidential reasoning in justifying CST[18]. That is, belief that CST is inherently valuable coexists with desire to show that is effective. CST educators therefore face a broader challenge than simply to design better evaluations based on statistical testing of effects on primary outcomes. They can consider to what extent their claims to the value of CST should be justified by evidence that it improves outcomes, or by moral claims to its status as an essential element of patient-centered clinical training. To the extent that educators want to claim that CST improves outcomes they will, indeed, need to adopt the rigorous design standards to which clinical trialists are held. By contrast, to the extent that they seek to justify CST on ethical grounds, evidence of effects on outcomes become less important; after all, the view that

practitioners should be taught subjects such as clinical ethics is not ‘evidence based’. However, this view would bring different obligations regarding evaluation of CST. In particular, educators would be less concerned with standardizing and promoting existing practice. Instead, rigorous evaluation from this perspective would mean embracing the debate, diversity and innovation that is more characteristic of the medical humanities[13, 19]. PEC welcomes evaluations of CST that are robust from whichever perspective.

References

[1] P.M. Moore, S. Rivera, G.A. Bravo-Soto, C. Olivares, T.A. Lawrie, Communication skills training for healthcare professionals working with people who have cancer, Cochrane Database Syst. Rev. 7 (2018) CD003751.
[2] C.L. Bylund, S.C. Banerjee, P.A. Bialer, R. Manna, T.T. Levin, P.A. Parker, E. Schofield, Y. Li, A. Bartell, A. Chou, S.A. Hichenberg, M. Dickler, D.W. Kissane, A rigorous evaluation of an institutionally-based communication skills program for post-graduate oncology trainees, Patient Educ. Couns. 101(11) (2018) 1924-1933.
[3] K.F. Schulz, D.G. Altman, D. Moher, t.C. Group, CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials, Trials 11(1) (2010) 32.
[4] P. Salmon, B. Young, Is clinical communication the one area of clinical oncology that needs no new ideas?, Med. Educ. 51(12) (2017) 1291-1293.
[5] A. Girgis, J. Cockburn, P. Butow, D. Bowman, P. Schofield, E. Stojanovski, C. D'Este, M.H.N. Tattersall, C. Doran, J. Turner, Improving patient emotional functioning and psychological morbidity: Evaluation of a consultation skills training program for oncologists, Patient Educ. Couns. 77(3) (2009) 456-462.
[6] L. Fallowfield, V. Jenkins, V. Farewell, J. Saul, A. Duffy, R. Eves, Efficacy of a Cancer Research UK communication skills training model for oncologists: a randomised controlled trial, Lancet 359(9307) (2002) 650-6.
[7] J. Gorniewicz, M. Floyd, K. Krishnan, T.W. Bishop, F. Tudiver, F. Lang, Breaking bad news to patients with cancer: A randomized control trial of a brief communication skills training module incorporating the stories and preferences of actual patients, Patient Educ. Couns. 100(4) (2017) 655-666.
[8] C. Pehrson, S.C. Banerjee, R. Manna, M.J. Shen, S. Hammonds, N. Coyle, C.A. Krueger, E. Maloney, T. Zaider, C.L. Bylund, Responding empathically to patients: Development, implementation, and evaluation of a communication skills training module for oncology nurses, Patient Educ. Couns. 99(4) (2016) 610-616.
[9] D.W. Kissane, C.L. Bylund, S.C. Banerjee, P.A. Bialer, T.T. Levin, E.K. Maloney, T.A. D'Agostino, Communication Skills Training for Oncology Professionals, J. Clin. Oncol. 30(11) (2012) 1242-1247.
[10] N. Mendick, B. Young, C. Holcombe, P. Salmon, How do surgeons think they learn about communication? A qualitative study, Med. Educ. 49(4) (2015) 408-16.
[11] P. Craig, P. Dieppe, S. Macintyre, S. Michie, I. Nazareth, M. Petticrew, G. Medical

Research Council, Developing and evaluating complex interventions: the new Medical Research Council guidance, Brit. Med. J. (Clinical research ed.) 337 (2008) a1655-a1655.
[12] P. Salander, M. Sandström, A Balint-inspired reflective forum in oncology for medical residents: Main themes during seven years, Patient Educ. Couns. 97(1) (2014) 47-51.
[13] P. Salmon, B. Young, A new paradigm for clinical communication: critical review of literature in cancer care, Med. Educ. 51(3) (2017) 258-268.
[14] J.L. Jackson, K. Kroenke, The effect of unmet expectations among adults presenting with physical symptoms, Ann. Intern. Med. 134(9 Pt 2) (2001) 889-97.
[15] L. Stead, G. Bergson, T. Lancaster, Physician advice for smoking cessation, Cochrane Database Syst. Rev. (2) (2008) CD000165.

[16] P. Glasziou, I. Chalmers, Research waste is still a scandal-an essay by Paul Glasziou and Iain Chalmers, Brit. Med. J. 363 (2018) k4645.
[17] J. Meunier, I. Merckaert, Y. Libert, N. Delvaux, A.M. Etienne, A. Lienard, I. Bragard, S. Marchal, C. Reynaert, J.L. Slachmuylder, D. Razavi, The effect of communication skills training on residents' physiological arousal in a breaking bad news simulated task, Patient Educ. Couns. 93(1) (2013) 40-7.
[18] P. Salmon, B. Young, Core assumptions and research opportunities in clinical communication, Patient Educ. Couns. 58(3) (2005) 225-234.
[19] F. Stiefel, C. Bourquin, Communication in oncology: now we train - but how well?, Ann. Oncol. 27(9) (2016) 1660-3.