

Clinicians' perceptions of reporting methods for back pain trials:

a qualitative study

Abstract

Background

How outcomes of clinical trials are reported alters the way treatment effectiveness is perceived: clinicians interpret the outcomes of trials more favourably when results are presented in relative (such as risk ratio) rather than absolute terms (such as risk reduction). However, it is unclear which methods clinicians find easiest to interpret and use in decision making.

Aim

To explore which methods for reporting back pain trials clinicians find clearest and most interpretable and useful to decision making.

Design and setting

Indepth interviews with clinicians at clinical practices/research centre.

Method

Clinicians were purposively sampled by professional discipline, sex, age, and practice setting. They were presented with several different summaries of the results of the same hypothetical trial. Each summary used a different reporting method, and the study explored participants' preferences for each method and how they would like to see future trials reported.

Results

The 14 clinicians interviewed (comprising GPs, manual therapists, psychologists, a rheumatologist, and surgeons) stated that clinical trial reports were not written with them in mind. They were familiar with mean differences, proportion improved, and numbers needed to treat (NNT), but unfamiliar with standardised mean differences, odds ratios, and relative risks (RRs). They found the proportion improved, RR, and NNT most intuitively understandable, and thought reporting between-group mean differences, RRs, and odds ratios could mislead.

Conclusion

Clinicians stated that additional reporting methods facilitate the interpretation of trial results, and using a variety of methods would make results easier to interpret in context and incorporate into practice. Authors of future back pain trials should report data in a format that is accessible to clinicians.

Keywords

back pain; data interpretation, statistical; outcome assessment (health care); primary care; qualitative research; randomised controlled trials; treatment outcome.

INTRODUCTION

How clinical trials are reported affects the way clinicians interpret effectiveness and make decisions. It is well established that the reporting of outcomes relatively, for example, using relative risk (RR) and other similar measures, leads to interventions being perceived as more effective than absolute reporting, where similar results are reported using measures of absolute difference between outcomes in different arms, such as the mean difference between outcomes in two arms.¹ This has been dubbed 'the framing effect'.¹ Several systematic reviews highlight the consistency of this phenomenon.²

In 2007, Covey systematically reviewed all studies involving people who make decisions about treatment, and performed a meta analysis.² Her review confirmed not only the presence of an effect, but its ubiquity across all consumers of trial reports. In marketing, the effects of therapeutics are framed in the most attractive ways possible. In randomised controlled trials (RCTs) it is ideal that authors report results in a way that accurately describes the effects of treatment, avoids misinterpretation, and provides a comprehensive description of which patients are likely to benefit. RCTs play a major role in informing clinicians' perception of effectiveness. As clinicians' perceptions influence clinical decision making, it is essential to explore their perceptions of different reporting methods.

It is challenging for clinicians who see

patients with low back pain to choose between effective therapies: larger high-quality trials commonly report, at best, small-to-medium effect sizes.³⁻⁶ This, combined with heterogeneous reporting of outcomes,⁷ complicates the process of decision making. In the US, the Food and Drug Administration (FDA) suggests that in cases where small differences exist, reporting the proportion of individuals who improve over a specified threshold may aid interpretation.⁸ This is facilitated following consensus on what constitutes an appropriate threshold above which individuals are deemed to have changed, for common low back pain outcome measures. Ostelo *et al* conducted a review of the measurement properties of back-specific outcome measurements, and emphasised the strengths and weaknesses of different methods for estimating minimally important change (MIC).⁹ Briefly, their findings show that approaches tend to involve using the statistical distributions of responses either to make decisions (distribution based), or to relate change scores to the global perceived change of the patient (anchor based). The authors point out that distribution-based methods are purely statistical measures of the magnitude of change and say little about clinical importance and that anchor-based methods, while providing information about clinical importance, do not take account of measurement precision. An expert panel considered all relevant clinimetric studies (in which either approach was taken) before

R Froud, PhD, senior research fellow; **D Carnes**, PhD, senior research fellow; **S Eldridge**, PhD professor of biostatistics, Centre for Primary Care and Public Health, Queen Mary University of London. **M Underwood**, MD, professor of primary care, Warwick Clinical Trials Unit, Warwick Medical School, Coventry.

Address for correspondence

Robert Froud, Centre for Primary Care and Public Health, Queen Mary University of London, Yvonne

Carter Building, 58 Turner Street, Whitechapel, London, E1 2AB.

E-mail: rj.froud@qmul.ac.uk

Submitted: 24 May 2011; **Editor's response:** 21 June 2011; **final acceptance:** 18 August 2011.

©British Journal of General Practice

This is the full-length article (published online 27 Feb 2012) of an abridged version published in print. Cite this article as: **Br J Gen Pract 2012; DOI: 10.3399/bjgp12X630034.**

How this fits in

It is known that clinicians perceive treatments as more effective when results are presented using relative rather than absolute reporting methods, but it is unclear which reporting methods clinicians believe are most useful for informing decision making. This study explored clinicians' perceptions of reporting outcomes of back pain trials, and the clarity and usefulness of different methods. Clinicians stated that the interpretation of outcome data was problematic and that researchers should present data in a format that is more meaningful to trial end users.

reaching consensus on appropriate thresholds by which to judge individual improvement. However, while such consensus allows researchers to define individual improvements in back pain trials, there are numerous methods for reporting the results, and the clarity and usefulness of different methods, from the clinicians' perspective, needs to be explored.

Aims

The aims of this study were to explore the clarity and ease of interpretation of reporting methods for trials of low back pain, as perceived by clinicians; and explore how clinicians would prefer to see trials reported, and which methods they believe offer the most relevant and useful information for decision making.

METHOD

Clinicians were purposively sampled by age, experience, sex, discipline, and practice settings; they were all working in primary and/or secondary care, and were consulted by patients with back pain. These comprised GPs, manual therapists, neurosurgeons, orthopaedic surgeons, pain specialists, psychologists (clinical), and rheumatologists, working in Barts and the London NHS Trust, Tower Hamlets Primary Care Trust (PCT), Maidstone and Tunbridge Wells NHS Trust, and West Kent PCT, and/or private sectors in corresponding areas. Contact details were obtained from online practitioner lists, advertisements, and registers, and by directly visiting PCT/acute trusts. Invitations were by email or post. Clinicians were offered £20 worth of vouchers from a leading chain store, or a donation to charity, as an incentive to participate.

The study aimed to sample approximately 15 participants; based partially on a similar

study, it was believed that this would allow the inclusion of an adequate range of clinicians from different backgrounds and purposive sampling characteristics, and be conducive to reaching data saturation.^{10,11} Clinicians were invited to take part in in-depth interviews of no longer than 1-hour's duration, wherein their views on a range of different methods for reporting outcomes were explored. Five summary reports were developed describing a fictitious trial in which the primary outcome was change in score of the iBAQ (imaginary BACK pain Questionnaire): a hypothetical instrument measuring pain and disability. These were rescaled outcomes from the manual therapy arm of the UK Back Pain Exercise and Manipulation (BEAM) trial.¹²

Outcomes were presented using the most common reporting methods for low back pain, identified by an earlier systematic review:⁷ mean difference (with and without advice on MIC), absolute risk reduction, RR, odds ratio (ORs) and the number needed to treat (NNT; for improvements, and separately for 'benefit' — the NNT for, on average, both one patient to improve, and either one improvement or prevention of one deterioration).¹³ Table 1 summarises the characteristics of these presented reports.

A topic guide for semi-structured interviews was developed by clinical and research staff, and featured questions regarding the usefulness and clarity of methods used; a summary is provided in Table 2. At first, participants were blind to the fact that they were looking at the same data presented in the different ways, but at the end of the interview the participants were unblinded, to explore whether this changed their perceptions of the different reporting methods. All interviews were recorded and transcribed verbatim and in full.

Analysis

The 'framework' approach was taken to analyse the data.¹⁴ Researchers familiarised themselves with transcripts and developed a suitable framework of themes and subthemes in which to model the data. Selected transcripts were coded as a pilot exercise, then the framework was re-evaluated and revised. Once researchers had agreed that the framework was comprehensive, the remainder of the transcripts were coded. Throughout the process disagreements were discussed and, if necessary, resolved by arbitration.¹⁵ Anonymised data were discussed and triangulated with a qualitative group that meets regularly at Queen Mary University. QSR NVivo (version 7) facilitated the

Table 1. Characteristics of the presented summaries

Summary number	Reporting method(s) used	Reported results
1a	Difference in means and standardised mean difference	A net difference of 8 points (95% CI = 4 to 12), where standardised mean difference = 0.4
1b	Difference in means, with MIC ^a (for an individual) reported	A net difference of 8 points (95% CI = 4 to 12)
2	Numbers and proportion improved	125/287, or 44% (38 to 49), experienced an improvement \geq MIC in the physical behavioural praxis group, 62/256, or 24% (19 to 29), experienced an improvement \geq MIC ^a in the GP care group. The percentage difference was 19% (12 to 27)
3	Relative risk for an improvement \geq MIC ^a	1.8 (95% CI = 1.4 to 2.3)
4	Odds ratio for an improvement \geq MIC ^a	2.4 (95% CI = 1.7 to 3.5)
5a	Number needed to treat for an additional improvement of \geq MIC ^a	5.2 (95% CI = 3.7 to 8.7)
5b	Number needed to treat for either an additional improvement of \geq MIC ^a , or a deterioration of \leq MIC ^a prevented	5.1 (95% CI = 3.6 to 9.1)

^aMinimally important change (MIC) = 25 points of a 100-point scale. Assumption: MIC for MIC for deterioration is equal in magnitude to MIC for improvement. These were rescaled outcomes from the manual therapy arm of the UK Back Pain Exercise and Manipulation (BEAM) trial.¹²

management of coded data. Following coding, matrices were formed in order to facilitate the identification of patterns and interrelationships, by searching between and within cases. Explanatory models were constructed to explain relationships between participants' views on reporting methods, their characteristics, and feelings toward the methods.

RESULTS

Fourteen participants were interviewed between April and July 2008. Interviews were stopped after the 14th interview as no novel themes were emerging and the sample was balanced for the a priori sampling framework. Sample characteristics and clinician specialties are shown in Table 3; all participants reported

receiving formal training in epidemiology/statistics as undergraduates.

Many of the views expressed by the 14 participants were similar. Although it was not an attribute for which the study purposively sampled, it was observed that clinicians who had been previously involved in research tended to recall better statistical/epidemiological concepts and appeared more inclined to evaluate methods critically and weight the impact the trial had on their practice. They also tended to have a research-based degree. Conversely, clinicians who trusted the author(s) had a less research-active past and were inclined to focus on conclusions (Figure 1). Clinicians stated that using simpler and more familiar reporting methods, including information regarding deteriorations and interpretive guidance, might better describe treatment effects and facilitate management. Table 4 shows the agreed coding framework headings and subheadings, and illustrative quotations typifying participants' comments described under the agreed framework subheadings follow.

Reading journal articles

The decision to read. Participants tended to base their decisions to read articles on their titles, abstracts, or summaries. A specific clinical question often motivated reading, although GPs and hospital-based physicians routinely read or 'skimmed' journals:

'I tend to go through the articles that are published, pick out the articles that affect me directly and then I tend to read the abstracts beforehand. If it looks interesting, I'll read the whole article.' (participant 14, orthopaedic surgeon)

Reading habits and thoroughness. The level of thoroughness with which participants read articles depended on the available time and perceived relevance of the article. Most scanned the article before committing to read it in its entirety; some only ever looked at specific sections:

'You read the abstract, look at the conclusion and then kind of go back and start again at the beginning, and work your way through, you know, depending on the time available and the level of interest it generates really.' (participant 3, osteopath)

Desire to critique. Participants could be divided into those who were critical of methods and those who were not. Those who critically evaluated methods, tended to

Table 2. Topic guide summary

Topic	Brief summary and example question
Introduction	Introductions, study brief, and consent taking
Responder background	Discipline, experience, training, article reading. For example: Are you confident about your own comprehension when reading journal articles?
Trial reports and clarity	Participant reads each of the trial reports and is asked about clarity, usefulness, and their overall impressions. For example: Do you find any of the methods particularly unclear?
Interpretation	Perceived interpretability is explored for each of the reporting methods. For example: Which methods conveyed the most useful information?
Preference	Preferences, and reasons for preferences for particular methods are explored. For example: Which reporting method(s) did you prefer, and why?
Relevance for clinical decision making	Which reports are most influential, most useful, and most comprehensive is explored. For example: Which report would result in you being most likely to recommend the intervention to your patients?
Closing the interview	Participant is thanked for her/his time, and asked if she/he would like to contribute anything else that has not been addressed

Table 3. Sample characteristics^a

Participant number	Specialty	Sex	Sector	Experience, years	Highest degree
1	Osteopathy	F	Private	10	PhD
2	Psychology	M	NHS	6	PhD
3	Osteopathy	M	Private	11	BSc
4	General practice	F	NHS	7	MSc
5	Osteopathy	M	Private	2	BSc
6	Chiropractic	M	Private	26	PhD
7	Psychology	F	NHS	11	DPsych
8	Neurosurgery	M	Both	18	MD
9	Physiotherapy	M	Both	15	PhD
10	Physiotherapy	F	NHS	7	MSc
11	Rheumatology	M	NHS	25	PhD
12	Physiotherapy	F	Private	21	PhD
13	General practice	F	NHS	28	MBBS
14	Orthopaedic surgery	M	Both	6	MBBS

^aGeographical area of work is not reported to protect participants' identities.

way and that way. On the other hand, it's quite common in that the devil of most of these things is in the detail, and I think it [reading RCTs] does change the detail of how you practise.' (participant 8, neurosurgeon)

'I would have thought every time I do [look at the literature] that there'll be something which will change my practice; whether it's an RCT or not, is another matter. Could be a review, could be an observational study, could be a cross-section study ... I mean, RCTs are very important, but they don't work well for musculoskeletal medicine and, therefore, they're perhaps aren't as prevalent as they might be in other fields; sort of little brown tablets, for example ...' (participant 9, physiotherapist)

Current concerns

Statistical knowledge and recall. Participants were concerned that they and their peers may not recall enough of their statistics/epidemiology training to evaluate properly the methodology used, or to interpret results:

'... the vast majority of GPs are not going to have experience in medical statistics and, OK, we, I'm afraid, need to be told ... Well, I do!' (participant 13, GP)

Trust. Participants emphasised the role of trust in researchers and the journal. Some voiced suspicion, or mistrust, about the methods researchers use (especially statistical), and the conclusions at which they arrive:

'... any statistician would be able to run rings around me, and this is really why I rely on the editorial selection of the journal I depend on, you know? There's an element of trust in hoping that the editorial board weeds out these statistical shenanigans, if you like, because certainly it's ... you know, I'm aware where my weaknesses are and statistics is a potential weakness.' (participant 14, orthopaedic surgeon)

'I want to be spoon-fed, I want to know the bottom line and what is best for my patients. You see, I mean, perhaps naively ... well, I hope not naively; you know, I have a degree of trust in researchers who are doing this work, and I trust that your methods are water tight, they are trustworthy and actually the point ... the techniques, whatever you are using to get to that end point, are robust ... I know what I read, I can then translate into my practice.' (participant 13, GP)

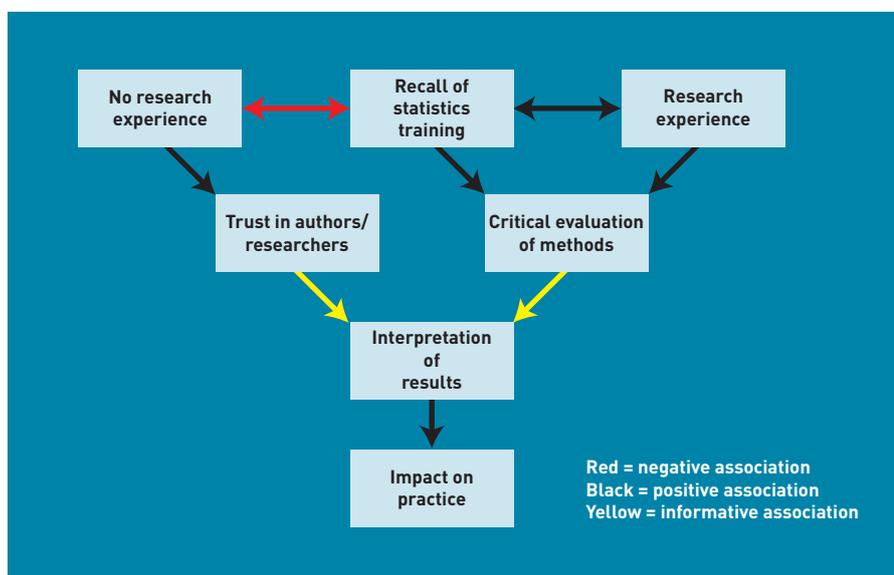


Figure 1. Associations between participant attributes and interpreting results. The figure shows how clinicians' levels of research experience and recall of statistics training might affect evaluation of research and trust placed in the authors'/researcher's conclusions.

disregard the articles they perceived to be of poor quality, or commented that the article had less impact on their practice:

'If I don't like the methods, I tend to stop! [slight laugh] If I do like the methods, I usually read the discussion and then look at the results to work out if it was a reasonable conclusion.' (participant 8, neurosurgeon)

Impact on practice. There were ambivalent feelings regarding the impact of trials on practice. There were concerns that RCTs were not sufficient sources of evidence: some felt practice behaviour should singularly be influenced by reviews. Participants generally felt some aspects of practice were changed by reading RCTs:

'Well, on the one hand it's rare; you know, I still take out discs and fuse spines and use this bit of kit, and manage patients in this

Table 4. Coding framework themes and subthemes

Themes	Subthemes					
Reading journal articles	The decision to read	Reading habits and thoroughness	Desire to critique	Impact on practice		
Current concerns	Statistical knowledge and recall	Trust	Interpretation			
Attitudes towards reporting methods	Mean difference	Standardised mean difference	Proportion improved	Relative risk	Odds ratio	Number needed to treat
Preferences	Familiarity	Clarity and intuitiveness	Usefulness	Interpretation (perception of effect)	Considering deterioration	
Future reporting of trials	Simplicity	Target audience	Relaying information to patients	Using a standardised set of methods	Explanations or interpretive guides	

Interpretation. Participants stated that current reporting methods are difficult to understand. This was especially felt to be a problem with low back pain RCTs, because of difficulty classifying subgroups compounding difficulty in interpreting outcomes:

'The biggest problem I have with back pain trials is who are they? What's chronic? And, you know, there is a world of difference in patients that have chronic back pain. I see an endless diet of chronic back pain and it's not one condition ... For instance, if you do a trial of facet joint injections and you do it outside of the context of exercise-based therapy, they're rubbish. But there are a group of patients who can't exercise, who you can ... I mean no exercise works, who you can get to exercise by giving them some injections. So the decomposition of the treatment package and then the application of randomised trials to it has been a huge backward step ... When I go to the shelf, it's not that I have some statistical problem with these studies, it's that I have a problem with back pain patients.' (participant 8, neurosurgeon)

Attitudes towards reporting methods

Mean difference. Although participants were familiar with mean difference, they found it difficult to interpret in the absence of MIC guidance:

'I'd need to know what is a clinically meaningful change on the iBAQ. It [the between-group difference of 8 points] could be completely irrelevant or it could be highly relevant. It's meaningless.' (participant 12, physiotherapist)

Providing MIC guidance improved participants' perceptions of the usefulness of the mean difference and led to them interpreting the intervention as less effective because the between-group difference was less than the MIC for an individual. However, some were concerned that group changes might mask individual changes.

'So it kind of doesn't look quite as impressive a finding when you read it like that! [8 against the MIC threshold of 25] [laughter] Yeah, clinically, so, you know, pretty much it's three times less than it would be in order to show a clinically significant improvement really ... I think it just puts the size of that moderate benefit in a bit more context.' (participant 3, osteopath)

Standardised mean difference. Participants thought that the standardised mean difference, although unfamiliar, was a good guide of treatment effect that facilitated comparisons between trials:

'I think it's quite nice to have kind of a rough pointer as to whether it's a small, medium, or large thing.' (participant 5, osteopath)

Proportion improved. Participants found the proportion of individual improvements or absolute risk reduction easy to understand. They felt it portrayed the effects of the intervention in a clinically relevant and useful way, facilitated comparisons, and did not require specialist statistical knowledge. It drew participants' attention to the proportion who did not improve:

'I find that this is understandable. And it can't be spun.' (participant 8, neurosurgeon)

'What about the other 64 [%]? You know, what happened to the other people who didn't ... ?' (participant 2, psychologist)

Relative risk. Participants thought the RR was intuitive, useful, easy to understand, and allowed them to relate results to individuals. Some commented that it was not as clear as the proportion improved. Many were unfamiliar with it, or felt that others would be unfamiliar:

'I would probably have to go and Google RR.' (participant 1, osteopath)

'If I explain to my dad that one in five people are going to get benefit, he'd go, "That's rubbish!" ... Whereas if you said to him, "Twice as likely to get benefit" he'd go, "That's great!"' (participant 4, GP)

Odds ratio. ORs were generally disliked; participants tended to compare them to RRs, which they found intuitively understandable. Participants felt ORs had the potential to mislead, as these might be interpreted as RRs. There was concern about presenting in relative, as opposed to absolute, terms:

'... because it's a derivative ratio, you lose ... well, you're not sure what the original odds were. So you lose track of it quite quickly ... statistically, it may well have meaning, but actually in terms of judging improvement, I'm not sure it really helps.' (participant 5, osteopath)

'OR and RR? Well, they seem to be very similar to me.' (participant 6, chiropractor)

Number needed to treat. Participants liked the NNT and found that it was familiar and intuitive, and thought it was difficult to manipulate, that it allowed comparisons between therapies, and that it was a concept that patients could understand:

'I guess what would really help me, because I know there's lots of things we do in general practice, like when we give statins, and I'm sure the NNTs for those are a whole lot higher ... It's certainly interesting, particularly in terms of when you're looking at things like cost effectiveness and rationing.' (participant 4, GP)

Participants thought the NNT for benefit (either an improvement gained or a deterioration prevented),¹³ was unfamiliar and difficult to understand. Those who understood it felt it gave a more comprehensive picture of treatment effects, but thought others may not understand it without guidance.

Preferences

Familiarity. Participants were confident about interpreting outcomes when a familiar reporting was used. They felt unfamiliar with standardised mean differences, ORs, and RRs:

'And the fact that you know that loads of the papers are presented like this [NNT] ... And I know that, you know, when things change, change has got to start somewhere, but ...

you know, that would be my preferred one, still, just because it's really familiar and I know I'll understand it immediately.' (participant 4, GP)

Clarity and intuitiveness. Participants felt the clearest and most intuitive methods were the proportion improved, NNT, and RR. A perceived advantage of the proportion improved was that it did not rely on knowledge recall:

'But this [proportion improved] is much clearer, I would say, and I would be happier. I mean, you can tell that I've been able to deal with the three so far. But I'm happier dealing even with this than that [mean difference with MIC advice], and that's better than the first one [mean difference].' (participant 11, rheumatologist)

Usefulness. Participants stated that mean differences were necessary and useful to the extent of knowing that the distributions of group scores differed beyond chance. A minority felt the point estimate of the mean difference was not particularly useful. Beyond this, participants felt that all other methods — generally with the exception of ORs — were useful additions that facilitated interpretation:

'I think using a combination of these kind of methods of reporting would be useful ... The mean difference, so it gives you, you know, a complete overview of the whole thing added up and averaged out. So on average, you get a small to moderate benefit [standardised mean difference], which is fine; that might give you a sort of general understanding, but in order to do that I think you then have to say actually what that means in reality, is that for, you know, some patients there is going to be a significant improvement (clinically). And then looking at the proportion and numbers improving added more meat to the bones really.' (participant 3, osteopath)

Interpretation (perception of effect). After being unblinded to the fact that the different methods all described the same results, participants expressed surprise at the extent to which their perceptions varied by reporting method. They stated that using relative terms, such as RRs or ORs, portrayed the treatment more favourably. Absolute terms, such as the difference in proportion improved or NNT, were perceived to have better face validity as they were less subject to misinterpretation. Opinions varied as to whether the reported NNT of approximately five represented an

effective or ineffective treatment. Presenting results using mean difference, especially with the MIC value, was thought to portray the treatment in weak light. Participants thought ORs, RRs, and mean differences with MIC values could be misleading because MIC pertains to an individual (as opposed to minimally important difference, which pertains to groups):

'So is this presenting the same results all the time? [slight laugh] It's amazing, isn't it? How you can present the same results with such a completely different slant on it! Mmn, mmn. Lies, damn lies and statistics!' (participant 13, GP; reaction to RR after reading mean difference and proportion improved summaries)

'... it's just fascinating how it can totally alter how effective you think something is! [slight laugh] Because I'm kind of ... I think if I were to see that ... Oh, God ... You know to get three people well, I'm going to have to see 18 people ...' (participant 7, psychologist)

Considering deterioration. Participants stated that it was important to consider deteriorations, and that these should be described separately using proportional differences, or the number needed to harm:

'I think that's one of the key elements that's missing in a lot of the current stuff that I read.' (participant 2, psychologist)

Future reporting of trials

Simplicity. Participants stated that reporting methods should be kept as simple as possible:

'I would say that there are, in many spheres, lots of issues about how journals get their message over to people in a user-friendly way! Rather than one which is just academic. Particularly if you're a clinician because what you're trying to look for is user-friendly information, rather than wanting to be bogged down by academic arguments. You want to see how these things relate to practice and that's what we're trying to do, rather than having lots of academic information without being able to link that to practice.' (participant 2, psychologist)

Target audience

Participants felt the level of statistical knowledge of the typical clinician reading the report should be borne in mind by authors:

'I think it all depends who you're presenting the data to and the language they understand. So, it says something and it conveys something that's meaningful, but if I was presenting this [RR] to a whole load of GPs, they would all understand that. If I was presenting it to a load of physios here, they wouldn't necessarily understand that. It's not ... they're not so familiar with it.' (participant 12, physiotherapist)

'Unfortunately, you know, you are not going to turn any GP particularly on by vast reams of ... You're just not! You know, we haven't got the time or the inclination. And, you know, it has got to be bullet points and a bottom line.' (participant 13, GP)

Relaying information to patients.

Participants felt a good reporting method should be clear and simple to understand for clinicians and be conducive to easily relaying information to patients. Those felt easiest to convey to patients were RR, NNT and the proportion of patients who improve or deteriorate:

'I think NNT is a very effective way of presenting data, and presenting it to Joe Public; far more effective than bloomin' confidence intervals and all the rest of it! It's something that they can understand ... I mean, the way I do it is I had 10 patients sitting in the waiting room, I would have to give the treatment to all 10 of them for one of them to benefit.' (participant 13, GP)

Using a standardised set of methods.

Participants felt using a standardised set of methods would give a more balanced view of the treatment effect and allow clinicians to choose the reporting method with which they are most familiar:

'... They all tell different bits of the story. And you're giving me drip, drip. And I'll say well this gives you an additional and additional so that lends itself to the fact that there is no warning lights sort of thing of anything that you did.' (participant 12, physiotherapist)

Explanations or interpretive guides.

Participants believed that it may sometimes be appropriate to include guidance on how to interpret outcome data given a specific reporting method:

'... If what you just said to me was a little line [that is, including a sentence in the report on how to interpret the reporting method] or something like that, then that would be brilliant.' (participant 2, psychologist)

DISCUSSION

Summary

Results suggest that some clinicians were not confident about interpreting the results of trials and felt trial reports were not written with them in mind. Different reporting methods uniquely contributed to interpretation of treatment effect. It was thought that using a variety, or a standardised 'suite' of methods, would prevent erroneous portrayal of effectiveness and facilitate interpretation. Basic guidance on how to interpret unusual reporting methods may be useful, although clinicians' preferences were for methods that do not require specialist knowledge to interpret results correctly.

Strengths and limitations

Previous research in this area has focused on defining which reporting methods have the greatest framing effects or explored how doctors might improve their communication of risk to patients. This paper elucidates how clinicians who see patients with back pain would like to see trial results reported, providing insight for authors of future trials into their audience.

Clinicians were offered £20 of high street vouchers as a 'thank you' for their participation. This may be seen as incentivising participants, and there is an argument that this can lead to unrepresentative samples. However, this may be of less importance in qualitative research where inference is not the intent, and when clinicians are the subjects of the research. There may also be an argument that by not offering any reward, it is also possible to procure an unrepresentative sample.

More clinicians with research degrees were recruited, which may be due to selection effects: clinicians with more interest in research were more likely to participate and have a research degree. Selection bias is not as critical as it would be if the research were trying to make inferences to a wider population. However, it may mean that understanding is actually worse in the wider population. The MIC threshold presented with the mean difference was derived from individual changes; these may not be a good proxy for important differences at a population level. However, important population-level estimates have not been established, and comparisons of mean differences with individual-specific MIC thresholds is evident in the literature. While participants did not question the validity of the iBAQ questionnaire, there is increasing concern

that the similar patient-reported outcome measures used in back pain trials may not measure aspects that are important to patients.¹⁶⁻¹⁸ For clinical trial interpretation to be maximised, this needs to be a focus of future research.

Comparison with existing literature

In 1998, Edwards *et al* explored the views of primary care professionals about communication of standardised risk.¹⁹ Their participants believed that standardising language would be useful between professionals, but that flexibility needed to be retained for conveying risk to patients. In 1999, Edwards *et al* subsequently piloted a range of complementary risk-communication tools in simulated general practice.²⁰ Participants felt data were often not in a digestible or relevant form for the practising doctor, or that doctors do not have sufficient time to access them. This was thought to be compounded by patients accessing the internet and presenting unfamiliar information. Doctors also felt that many data were biased, especially those originating from pharmaceutical companies. In addition, RRs were felt to be misleading. The results of the present study are congruent with these findings, but specific to reporting of trials on back pain. The work more thoroughly explores why clinicians favour or dislike particular methods and the perceived clarity and usefulness of each of these methods.

Implications for practice and research

These results suggest that authors need to re-evaluate their reporting of outcomes. There may be advantages in using simple outcomes based on the number of individuals who improve over a specified threshold, which do not require specialist knowledge. It is not suggested that such reporting replaces traditional methods, rather that these are complementary and will aid interpretation of treatment effects. These results have informed a Delphi study,²¹ from which a consensus statement for future reporting of back pain trials is recommended.

This study showed that a group of clinicians who see patients with low back pain felt clinical trials are difficult to interpret and not written with them in mind. Clinicians thought that presenting a standardised set of reporting methods, including methods based on individual improvements would, by aiding understanding at the clinician level, facilitate transition of the research into practice, hopefully improving patient care.

Funding

Barts and the London Charity (reference 522/660).

Provenance

Freely submitted; externally peer reviewed.

Competing interests

Martin Underwood was a member of the UK BEAM trial team. The other authors have stated that there are no competing interests.

Acknowledgements

Thanks are due to the study participants.

Discuss this article

Contribute and read comments about this article on the Discussion Forum: <http://www.rcgp.org.uk/bjgp-discuss>

REFERENCES

1. McGettigan P, Dianne S, O'Connell K, *et al*. The effects of information framing on the practices of physicians. *J Gen Intern Med* 1999; **14(10)**: 633–642.
2. Covey J. A meta-analysis of the effects of presenting treatment benefits in different formats. *Med Decis Making* 2007; **27(5)**: 638–654.
3. UK BEAM Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. *BMJ* 2004; **329(7479)**: 1377–1381.
4. Witt CM, Jena S, Selim D, *et al*. Pragmatic randomized trial evaluating the clinical and economic effectiveness of acupuncture for chronic low back pain. *Am J Epidemiol* 2006; **164(5)**: 487–496.
5. Hay EM, Mullis R, Lewis M, *et al*. Comparison of physical treatments versus a brief pain-management programme for back pain in primary care: a randomised clinical trial in physiotherapy practice. *Lancet* 2005; **365(9476)**: 2024–2030.
6. Lamb SE, Hansen Z, Lall R, *et al*. Basic Skills Training trial Investigators. Group cognitive behavioural treatment for low-back pain in primary care: a randomised controlled trial and cost-effectiveness analysis. *Lancet* 2010; **375(9718)**: 916–923.
7. Froud R. *Improving interpretation of patient-reported outcomes in low back pain trials*. PhD thesis. London: Barts and the London School of Medicine and Dentistry, Queen Mary University of London, 2010.
8. US Department of Health and Human Services FDA Center for Drug Evaluation and Research, US Department of Health and Human Services FDA Center for Biologics Evaluation and Research and US Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. *Health Qual Life Outcomes* 2006; **4(79)**: 17, 20, 26.
9. Ostelo R, Deyo R, Stratford P, *et al*. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 2008; **33(1)**: 90–94.
10. Carnes D. *Understanding and measuring chronic musculoskeletal pain in the community using self-completed pain drawings*. PhD thesis. London: Centre for Health Sciences, Queen Mary University, 2006.
11. Pope C, Ziebland S, Mays N. Qualitative research in health care: analysing qualitative data. *BMJ* 2000; **320(7227)**: 114–116.
12. UK BEAM Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: cost effectiveness of physical treatments for back pain in primary care. *BMJ* 2004; **329(7479)**: 1381.
13. Froud R, Eldridge S, Lall R, Underwood M. Estimating number needed to treat from continuous outcomes in randomised controlled trials: methodological challenges and worked example using data from the UK Back Pain Exercise and Manipulation (BEAM) trial. *BMC Med Res Meth* 2009; **9**: 35.
14. Ritchie J, Lewis J. *Qualitative research practice*. Thousand Oaks, CA: Sage Publications, 2003.
15. Patton M. *Qualitative research evaluation and methods*. 3rd edn. Thousand Oaks, CA: Sage, 2002.
16. Foster N, Dziedzic K, van der Windt D, *et al*. Research priorities for non-pharmacological therapies for common musculoskeletal problems: nationally and internationally agreed recommendations. *BMC Musculoskelet Disord* 2009; **10**: 3.
17. Hush J, Refshauge K, Sullivan G, *et al*. Recovery: what does this mean to patients with low back pain? *Arthritis Rheum* 2009; **61(1)**: 124–131.
18. Mullis R, Barber J, Lewis M, Hay E. ICF core sets for low back pain: do they include what matters to patients? *J Rehabil Med* 2007; **39**: 353–357.
19. Edwards A, Matthews E, Pill R, Bloor M. Communication about risk: the responses of primary care professionals to standardizing the 'language of risk' and communication tools. *Fam Pract* 1998; **15(4)**: 301–307.
20. Edwards A, Elwyn G, Gwyn R. General practice registrar responses to the use of different risk communication tools in simulated consultations: a focus group study. *BMJ* 1999; **319(7212)**: 749–752.
21. Froud R, Eldridge S, Kovacs F, *et al*. Reporting outcomes of back pain trials: a modified Delphi study. *Eur J Pain* 2011 **15(10)**: 1068–1074. Epub 2011 May 18.