

ARTICLE

Key concepts that people need to understand to assess claims about treatment effects

Astrid Austvoll-Dahlgren¹, Andrew D. Oxman², Iain Chalmers³, Allen Nsangi⁴, Claire Glenton², Simon Lewin^{2,5}, Angela Morelli², Sarah Rosenbaum⁶, Daniel Semakula⁴ and Nelson Sewankambo⁴

¹ Norwegian Knowledge Centre for the Health Services, Prevention, Health promotion, and Organisation Unit, Oslo, Norway

² Norwegian Knowledge Centre for the Health Services, Global Health Unit, Oslo, Norway

³ James Lind Initiative, Oxford, UK

⁴ Makerere University College of Health Sciences, New Mulago Hospital Complex, Kampala, Uganda

⁵ Health Systems Research Unit, South African Medical Research Council, Cape Town, South Africa

⁶ Norwegian Knowledge Centre for the Health Services, Communication Unit, BOKS 7004 St.Olavsplass, 0130 Oslo, Norway

Keywords

Evidence based medicine; health literacy; user involvement.

Correspondence

Astrid Austvoll-Dahlgren, Norwegian Knowledge Centre for the Health Services, Prevention, Health promotion, and Organisation Unit, BOKS 7004 St. Olavsplass, 0130 Oslo, Norway.
Tel: +47-46-400406;
Fax: +47-46-400406;
Email: astrid.austvoll-dahlgren@nokc.no

Funding

Fair Tests of Treatments project is funded by the Norwegian Research Council.

Declaration of Competing Interests.

The authors of this paper declare that they have no conflict of interest.

Received 6 June 2014; accepted for publication 25 January 2015.

doi: 10.1111/jebm.12160

Abstract

Objective: People are confronted with claims about the effects of treatments and health policies daily. Our objective was to develop a list of concepts that may be important for people to understand when assessing claims about treatment effects.

Methods: An initial list of concepts was generated by the project team by identifying key concepts in literature and tools written for the general public, journalists, and health professionals, and consideration of concepts related to assessing the certainty of evidence for treatment effects. We invited key researchers, journalists, teachers and others with expertise in health literacy and teaching or communicating evidence-based health care to patients to act as the project's advisory group.

Results: Twenty-nine members of the advisory group provided feedback on the list of concepts and judged the list to be sufficiently complete and organised appropriately. The list includes 32 concepts divided into six groups: (i) Recognising the need for systematic reviews of fair tests, (ii) Judging whether a comparison of treatments is fair comparison, (iii) Understanding the role of chance, (iv) Considering all the relevant fair comparisons, (v) Understanding the results of fair comparisons of treatments, (vi) Judging whether fair comparisons of treatments are relevant.

Conclusion: The concept list provides a starting point for developing and evaluating resources to improve people's ability to assess treatment effects. The concepts are considered to be universally relevant, and include considerations that can help people assess claims about the effects of treatments, including claims that are found in mass media reports, in advertisements and in personal communication.

Introduction

People are confronted with claims about the effects of treatments and health policies daily. These claims are made both in mass media and in personal communication. They may be well-intentioned or driven by commercial or other special interests. Many of these claims are biased, inaccurate, or unsubstantiated (1–4). People who act on unsubstantiated claims and who fail to use treatments and policies supported by re-

liable evidence may suffer unnecessarily (or cause others to suffer unnecessarily) and waste resources. This problem has particularly serious implications for people in low-income countries that have few resources to waste. To avoid these problems, people need to be able to make informed choices about health care (5, 6). This requires “health literacy”; i.e., *cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good*

health p. 10, (7). In particular, people need to be able to assess and use information about the effects of healthcare interventions in order to promote and maintain their own and their families' health. They also need these skills to be able to participate in shaping the policies that determine how health care is delivered, financed and governed. However, studies suggest that people's health literacy skills are generally poor, and that they commonly do not understand or apply key concepts that are essential for appraising these claims (8–14).

A number of resources have been developed to improve people's ability to assess claims about treatment effects, but these have not been evaluated consistently and they frequently focus on a specific concept, such as randomization (13–22). Among these resources are several tools for communicating and appraising claims about treatment effects (23–29). There are also several books that have been written with the objective of improving people's ability to assess treatment effects (30–33). However, these resources focus on different, albeit overlapping sets of concepts and use different terminology. As a starting point for developing and evaluating resources to improve people's ability to assess treatment effects, we have developed a list of concepts that are essential for doing this. This work was undertaken as the first step in the *Informed Healthcare Choices* project that aims to develop and evaluate resources to teach children and adults how to assess claims about treatment effects in low-income countries.

Our objective was to develop a list of concepts that may be important for people to understand when assessing claims about treatment effects. By treatment, we mean any type of healthcare intervention. This includes changes in health behavior, screening and other preventive interventions, any type of therapeutic intervention, rehabilitation, and public health and health system interventions that are targeted at groups of people.

In a later stage of the project, we plan to develop resources that can help people using mass media and school children in low-income countries understand and make use of these concepts. Although our ultimate aim is to reach specific target groups in low-income countries, our first objective has been to develop a list of concepts that are universally relevant. In developing the list, we have aimed to use plain language, to avoid jargon and to provide easily understood explanations for each concept. However, the list should not be regarded as a final product but as a starting point for developing further resources.

Methods

In developing the list we applied pragmatic, but explicit criteria, for identifying concepts in a systematic, transparent, and iterative process, including involving potential end users and

experts within the field. Similar methodologies have been used for managing the processes of research priorities, and policy and guideline development (34–36).

Identifying concepts

In the first step of the process, members of the *Informed Healthcare Choices* project group generated an initial list of concepts by identifying key concepts in the existing literature. For this review, we compiled a database of contemporary resources for improving people's ability to understand, use and communicate research evidence about treatment effects. The project includes researchers in four countries (Uganda, the UK, Australia, and Norway) with diverse backgrounds and experience in different research areas, including; research methodology, health literacy, development and evaluation of effective strategies for communicating research evidence, and teaching evidence-based health care to professionals (including journalists and policy makers) and patients.

We extracted all potentially relevant concepts for inclusion in the list. A majority of the concepts were identified from *Testing Treatments* (32), a book written for the general public and providing an overview of key concepts that people may need to understand to assess claims about treatment effects (32). This was supplemented by reviewing other books that were written for the general public (30, 37), checklists for the general public, journalists, and health professionals (23, 25, 27, 38), and consideration of concepts related to assessing the certainty of evidence for treatment effects (39).

We discussed and revised this list of concepts iteratively and grouped the concepts into coherent, logical categories. We included concepts that children or adults with a primary school education and basic reading and numeracy skills could potentially learn to understand and apply, with the objective of helping them to assess whether:

- The basis for a claim is reliable; i.e. whether it is based on fair comparisons of treatments (treatment comparisons designed to minimise the risk of errors).
- The results of fair comparisons are relevant to them.
- Additional information is needed to assess the reliability and relevance of claims about treatments and, if so, what information is needed.

We excluded concepts that were judged to be redundant or rarely important when assessing claims about treatment effects. We also excluded concepts that were only relevant to other judgments, such as concepts related to the ethics of research or shared decision-making. Although it may be important for people to understand such concepts, they are not crucial for assessing claims about treatment effects.

This process resulted in a draft list of 34 concepts, accompanied by a short explanation and text describing for each

Table 1 Quantitative summary of feedback from the advisory group

Question	Responses (N = 25)		
	No	Unclear	Yes
1. Are there concepts included that should not be?	No = 16	Unclear = 4	Yes = 5
2. Are the concepts described and explained in a way that you think could be understood by someone without a research background?	Yes = 2	Unclear = 13	No = 10
3. Are there important concepts that are missing?	No = 14	Unclear = 5	Yes = 6
4. Are the concepts organized in a logical way?	Yes = 19	Unclear = 2	No = 4

concept the implications for assessing claims about treatment effects.

Consultation with end users and experts

Evidence on group composition and consultation processes in guideline development concludes that there is no “gold standard” best method (35, 40). Despite the methodological uncertainties, approaches are encouraged which ensure that all participants are heard, and which adopt transparent, explicit criteria (40). Decisions about who should be invited to take part in such processes are generally based on logic: there is little evidence to provide guidance on how to organise such groups (35). However, group composition has been found to influence the impact on conclusions. Considering this, and based on logical arguments and experience, it is recommended that panels and advisory groups are broadly composed, including people from different disciplines and with content expertise, as well as stakeholders and end users (35).

Based on purposeful sampling and suggestions by the members of the *Informed Healthcare Choices* project, we compiled a list of people who might agree to act as members of the project’s advisory group. It was our goal to include experts within relevant research areas (such as health literacy and teaching or communicating evidence-based health care), and end users such as journalists, health educators, and teachers, across different country settings (including low-income contexts). The purposes of the advisory group were to provide feedback on the draft list of concepts and to suggest potentially relevant resources.

We circulated the draft list of 34 concepts to members of this advisory group for feedback. We also sent them a feedback form, asking them to answer four questions using the response options “yes,” “no,” or “uncertain,” together with any open-ended comments:

- Are concepts included that should not be?
- Are the concepts described and explained in a way that you think could be understood by someone without a research background?
- Are there important concepts that are missing?
- Are the concepts organised in a logical way?

Feedback from the advisory group was summarised in two ways. First, we summarised the group’s responses to the four questions in a quantitative overview. Second, we compiled thematically all open-ended comments and suggestions for revisions in a descriptive overview, sorted by associated question in the feedback form (1 to 4) and related concept(s). This procedure was conducted by one member of the project team, and quality checked by two other members.

Results

Twenty-nine members of the advisory group provided feedback on the list of concepts, of which four group members provided general feedback without filling out the feedback form. A quantitative summary of their feedback can be found in Table 1. Most members found that all of the concepts were relevant, that there were few if any important concepts missing and that the concepts were organised in a logical way. However, most felt that the concepts were not explained in a way that could be understood by someone without a research background or training in these concepts. This was in part due to a misunderstanding by some of the members that the list was intended as a resource for people with a primary school education rather than as a syllabus or inventory of concepts for which resources are needed to help people to understand and apply these concepts. Several of the people who responded negatively to the first question, explained this in relationship to people’s ability to understand the concept (the second question).

We addressed all comments and suggestions for revisions. Based on this feedback we revised and reorganised the list of concepts. All concepts that were identified as being unclear or not easily understood were rewritten and the format of the descriptions and explanations were standardised across concepts (Appendix 1). Suggestions for changing the organisation of concepts included reordering how some of the concepts were grouped, adding a brief introduction to each group of concepts, improving the consistency of how the groups of concepts were labeled, and improving the consistency of how the concepts were described and explained.

Nine suggestions for additional concepts were made. We considered each of these suggestions for new concepts and

determined that all of these were either redundant or partly covered by other concepts, and so were incorporated in the latter where appropriate. In response to the feedback, we also merged six concepts into three concepts that were perceived as similar or redundant. This resulted in 32 concepts in the following groups:

- Recognising the need for fair comparisons of treatments.
- Judging whether a comparison of treatments is a fair comparison.
- Understanding the role of chance.
- Considering all the relevant fair comparisons.
- Understanding the results of fair comparisons of treatments.
- Judging whether fair comparisons of treatments are relevant.

The quantitative summary and descriptive overview of the open-ended comments and suggestions for revisions of the feedback from the advisory group, together with changes made based on this feedback, were presented to project group, which reviewed and commented on the revised list of concepts. Only minor changes to the wording of some of the concepts were suggested by the members of the project group. These were incorporated and the final draft was edited by a professional copy editor to ensure that the concepts were clearly labeled and explained in plain language. To make explicit the target audience and use of the list, an introductory text was added to accompany the list. We also added a glossary of terms that required further explanation (Appendix 1).

Members of the advisory group were sent a summary of their feedback, our response to this feedback, and the resulting list of concepts.

Discussion

We have developed a list of concepts that we believe people need to be able to understand and apply in order to assess claims about the effects of treatments. We have used plain language to describe each concept and have strived to organise the concepts in a logical way. Feedback from people with diverse backgrounds and expertise suggests that the list of concepts is sufficiently complete and is organised appropriately.

There are many approaches to managing group processes (35, 40). Consequently, there might be equally good ways to label, explain and organise the concepts. In the development of our list of concepts, we drew on a wide range of resources, obtained input from a diverse group of people systematically, and used an iterative, transparent process. The result of this process is a list of concepts that the people we consulted, including journalists, teachers, health professionals, and researchers, generally found to be comprehensive and sensible.

A limitation of our list is that many of the concepts may not be easily understood and applied by people without a research background or training in these concepts (8–13, 41–43). However, the list is not intended as a direct resource for adults or children without appropriate training. Instead, the list is intended as a syllabus or inventory of concepts for which we will develop resources.

Researchers exploring peoples' understanding of the effects of treatments are faced with two main challenges. First, there exists no consensus on the conceptualisation and key concepts critical to assessing the effects of treatments. Second, studies mapping or evaluating people's understanding of the effects of treatments have not been measured consistently, and are characterised by differences in terminology and parallel discourses. Such inconsistencies are attributable to some extent to different research areas and disciplines being responsible for studies that have often focused on more general competencies, such as health literacy studies, or more specific concepts such as the understanding of risk or randomization (13–20, 44–46). To help address these needs, we developed the concept list to provide a starting point for developing and evaluating resources to improve people's ability to assess treatment effects. In the next stages of this project, we will prioritise these concepts and develop and evaluate resources to help people understand and apply these concepts in low-income countries. We will involve teachers and journalists familiar with our target audiences to do this. In addition, we are developing a measurement instrument to assess people's ability to understand and apply the concepts. The list and measurement tool will be made available as online resources, where we will invite people to comment on the concepts, offer examples to illustrate them, and suggest questions to test people's understanding of the concepts.

We are not aware of any previous studies that have generated a similar list of concepts. Since we extracted relevant concepts from resources aimed at helping lay people to understand and apply these concepts (30, 32, 37, 39), our list includes all of the concepts that we considered relevant from those resources. However, we have excluded some concepts that were outside our scope, including concepts related to the ethics of fair comparisons of treatments and some concepts related to shared decision-making.

Conclusion

People are confronted daily with claims about the effects of treatments. To assess these claims and assess which to believe and which not to believe, people need to be able to understand and apply the basic concepts that are essential for assessing the reliability of such claims. We have generated a list of such concepts, some of which will require the

development of resources to help people understand and apply them. The list of concepts includes considerations that can help people assess claims about the effects of treatments, including claims that are found in mass media reports, in advertisements, and in personal communication. We believe that it might be possible for children or adults with a primary school education to learn to understand and apply all or most of these concepts, and that doing so might help them to assess claims about treatment effects and make well-informed choices.

Funding

Informed Healthcare Choices project is funded by the Norwegian Research Council.

Declaration of Competing Interests

The authors of this paper declare that they have no conflict of interest.

Strengths and limitations of this study

Researchers exploring peoples' understanding of the effects of treatments are faced with two main challenges; there exists no consensus on the conceptualisation and key concepts critical to assessing the effects of treatments and studies mapping or evaluating peoples understanding of the effects of treatments have not been measured consistently. As a starting point for developing and evaluating resources to improve people's ability to assess treatment effects, we have developed a list of key concepts that people need to understand to assess claims about treatment effects. There might be equally good ways to label, explain and organise the concepts. However, in the development of the list we drew on a wide range of resources, systematically obtained input from a diverse group of people and used an iterative and transparent process. The result of this process is a list of concepts that the people we consulted, including journalists, teachers, health professionals and researchers, generally found to be comprehensive and sensible.

References

- Lewis M, Orrock P, Myers S. Uncritical reverence in CM reporting: assessing the scientific quality of Australian news media reports. *Health Sociology Review* 2010; 19(1): 57–72.
- Glenton C, Paulsen E, Oxman A. Portals to Wonderland? Health portals lead confusing information about the effects of health care. *BMC Medical Informatics and Decision Making* 2005; 5: 7.
- Wolfe R, Sharp L, Laipsky M. Content and design attributes of antivaccination web sites. *JAMA* 2002; 287(24): 3245–48.
- Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J, et al. Coverage by the news media of the benefits and risks of medications. *New England Journal of Medicine* 2000; 342(22): 1645–50.
- Zarcadoolas C, Pleasant A, Greer D. *Advancing Health Literacy: A Framework for Understanding and Action*. 1st edition. San Francisco, CA: Jossey-Bass; 2006: p. 368.
- Coulter A. *The Autonomous Patient: Ending Paternalism in Medical Care*. London: TSO; 2002: p. 128.
- Nutbeam D. *Health promotion glossary*. *World Health Organization*, 1998 17.01.2012.
- Woloshin S, Schwartz L, Byram S, Sox H, Fischhoff B, Welch H. Women's understanding of the mammography screening debate. *Archives of Internal Medicine* 2000; 160(10): 1434–40.
- Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? *Social Science & Medicine* 2007; 64(9): 1853–62
- Weinstein N. What does it mean to understand a risk? Evaluating risk comprehension. *Journal of the National Cancer Institute Monographs* 1999; 25: 15–20.
- Lloyd A. The extent of patients' understanding of the risk of treatments. *Quality in Health Care* 2001;10(Suppl 1): i14–i18.
- Tickner S, Leman P, Woodcock A. Factors underlying suboptimal childhood immunisation. *Vaccine* 2006; 24(49–50): 7030–36.
- Robinson E, Kerr C, Stevens A, Lilford R, Braunholtz D, Edwards S, et al. Lay public's understanding of equipoise and randomisation in randomised controlled trials. Research Support, Non-U.S. Gov't. NHS R&D HTA Programme, 2005 Mar. Report No. 1366-5278 (Linking) Contract No. 8.
- Berkman N, Sheridan S, Donahue K, Halpern D, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Annals of Internal Medicine* 2011; 155(2): 97–107.
- Dickersin K, Braun L, Mead M, Millikan R, Wu AM, Pietenpol J, et al. Development and implementation of a science training course for breast cancer activists: Project LEAD (leadership, education and advocacy development). *Health Expect* 2001; 4(4): 213–20.
- Milne R, Oliver S. Evidence-based consumer health information: developing teaching in critical appraisal skills. *International Journal for Quality in Health Care* 1996; 8(5): 439–45.
- Berger B, Steckelberg A, Meyer G, Kasper J, Muhlhauser I. Training of patient and consumer representatives in the basic competencies of evidence-based medicine: a feasibility study. *BMC Medical Education* 2010; 10: 16.
- Jordan J, Osborne R, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *Journal of Clinical Epidemiology* 2011; 64(4): 366–79.
- D'Eath M, Barry MM, Sixsmith J. *Rapid Evidence Review of Interventions for Improving Health Literacy*. Stockholm: ECDC, 2012.
- Akl E, Oxman A, Herrin J, Vist GE, Terrenato I, Sperati F, et al. Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database of Systematic Reviews* 2011; (3): CD006776.

21. Ndebele PM, Wassenaar D, Munalula E, Masiye F. Improving understanding of clinical trial procedures among low literacy populations: an intervention within a microbicide trial in Malawi. *BMC Medical Ethics*, 2012; 13: 29.
22. McCormack L, Sheridan S, Lewis M, Boudewyns V, Melvin CL, Kistler C, et al. Communication and dissemination strategies to facilitate the use of health-related evidence. Evidence Report/Technology Assessment No. 213. <http://www.effectivehealthcare.ahrq.gov/reports/final.cfm>. 2013.
23. Charnock D. *The DISCERN Handbook: Quality Criteria for Consumer Health Information on Treatment Choices*. Ltd. ORMP, editor1998.
24. Trevena L, Davey H, Barratt A, Butow P, Caldwell P. A systematic review on communicating with patients about evidence. *Journal of Evaluation in Clinical Practice* 2006; 12(1): 13-23.
25. Oxman AD, Guyatt GH, Cook DJ, Jaeschke R, Heddle N, Keller J. An index of scientific quality for health reports in the lay press. *Journal of Clinical Epidemiology* 1993; 46(9): 987-1001.
26. Austvoll-Dahlgren A, Danielsen S, Opheim E, Bjørndal A, Reinart LM, Flottorp S, et al. Development of a complex intervention to improve health literacy skills. *Health Information and Libraries Journal* 2013; 30(4): 278-93.
27. Moynihan R. Tipsheet for reporting on drugs, devices and medical technologies 2004; September. <http://www.commonwealthfund.org/Publications/Other/2004/Sep/Tipsheet-For-Reporting-on-Drugs-Devices-and-Medical-Technologies.aspx>.
28. Schwartz LM, Woloshin S, Welch HG. Using a drug facts box to communicate drug benefits and harms: two randomized trials. *Annals of Internal Medicine* 2009; 150(8): 516-27.
29. Stacey D, Légaré F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews* 2014; 1: CD001431.
30. Irwig L, Irwig J, Trevena L, Sweet M. *Smart Health Choices: Making Sense of Health Advice*. Hammersmith, London: 2008; 242.
31. Woloshin S, Schwartz L, Welch H. *Know Your Chances: Understanding Health Statistics*. London: University of California Press, 2008; 142.
32. Evans I, Thornton H, Chalmers I, Glasziou P. *Testing Treatments: Better Research for Better Healthcare*. 2nd edn. London: Pinter & Martin, 2011. <http://www.testingtreatments.org/new-edition/>.
33. Fischhoff B, Brewer NT, Downs JS. *Communicating Risks and Benefits: An Evidence-Based User's Guide*. Silver Spring: U.S. Food and Drug Administration, 2011. <http://www.fda.gov/AboutFDA/ReportsManualsForms/Reports/ucm268078.htm>.
34. Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 2. Priority setting. *Health Research Policy and Systems* 2006; 4: 14.
35. Fretheim A, Schunemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 3. Group composition and consultation process. *Health Research Policy and Systems* 2006; 4: 15.
36. Cliff J, Lewin S, Woelk G, Fernandes B, Mariano A, Sevene E, et al. Policy development in malaria vector management in Mozambique, South Africa and Zimbabwe. *Health Policy Plan* 2010; 25(5): 372-83.
37. Woloshin S, Schwartz LM, Moncur M, Gabriel S, Tosteson AN. Assessing values for health: numeracy matters. *Medical Decision Making* 2001; 21(5): 382-90.
38. Guyatt GH, Ray J, Gibson N, Cook DJ, Ashpole B, Baines C, et al. A Journalist's Guide for Health Stories. *American Medical Writers Association Journal* 1999; 14(1): 32-41.
39. Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. *Journal of Clinical Epidemiology* 2011; 64(4): 380-82.
40. Fretheim A, Schunemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 5. Group processes. *Health Research Policy and Systems* 2006; 4: 17.
41. Schwartz L, Woloshin S, Black W, Welch H. The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine* 1997; 127(11): 966-72.
42. Peterson G, Aslani P, Williams KA. How do consumers search for and appraise information on medicines on the Internet? A qualitative study using focus groups. *Journal of Medical Internet Research* 2003; 5(4): e33.
43. Stead M, Eadie D, Gordon D, Angus K. "Hello, hello - it's English I speak!" a qualitative exploration of patients' understanding of the science of clinical trials. *Journal of Medical Ethics* 2005; 31(11): 664-69.
44. Horsley T, Hyde C, Santesso N, Parkes J, Milne R, Stewart R. Teaching critical appraisal skills in healthcare settings. *Cochrane Database of Systematic Reviews* 2011; (11): CD001270.
45. Austvoll-Dahlgren A, Bjørndal A, Odgaard-Jensen J, Helseth S. Evaluation of a Web Portal for Improving Public Access to Evidence-Based Health Information and Health Literacy Skills: A Pragmatic Trial. *PLoS One* 2012; 7(5): e37715.
46. Car J, Lang B, Colledge A, Ung C, Majeed A. Interventions for enhancing consumers' online health literacy. *Cochrane Database of Systematic Reviews* 2011; (6): CD007092.

Appendix 1

Assessing claims about treatments effects: Key concepts that people need to understand

There are endless claims about treatments in the mass media, advertisements and everyday personal communication. Some are true and some are false. Many are unsubstantiated. We do not know whether they are true or false. Unsubstantiated claims about the effects of treatments are often wrong. Consequently, people who believe and act on these claims suffer unnecessarily and waste resources by doing things that do

not help and might be harmful, and by not doing things that do help.

We have prepared a list of key concepts that they can use to assess claims about the effects of a treatment (any action intended to improve health), including whether:

- The basis for a claim is reliable; i.e. whether it is based on fair comparisons of treatments (treatment comparisons designed to minimise the risk of errors)
- The results of fair comparisons are relevant to them
- Additional information is needed to assess the reliability and relevance of claims about treatments and, if so, what information is needed

The list serves as a syllabus for identifying the resources needed to help people understand and apply the concepts.

Effective treatments can prevent health problems, save lives and improve quality of life. However, nature is a great healer and people often recover from illness without treatment. Likewise, some health problems may get worse despite treatment, or treatment may actually make things worse. For these reasons, knowledge of the natural course of illness should be the starting point for making informed decisions about treatments.

We have written the concepts and explanations in plain language. However, some of these concepts may be unfamiliar and difficult to understand. We did not design the list as a teaching tool. It is a framework, or starting point, for teachers, journalists and other intermediaries for identifying and developing resources (such as longer explanations, examples, games and interactive applications) to help people to understand and apply the concepts.

The concepts in this list are universally relevant. However, our focus has been on concepts that are important and possible to teach or communicate to children and poorly educated adults living in low-income countries. The list includes 32 concepts, divided into 6 groups:

- Recognising the need for fair comparisons of treatments
- Judging whether a comparison of treatments is a fair comparison
- Understanding the role of chance
- Considering all the relevant fair comparisons
- Understanding the results of fair comparisons of treatments
- Judging whether fair comparisons of treatments are relevant

1. Recognising the need for fair comparisons of treatments

Well-informed treatment decisions require reliable information. Not all claims about the effects of treatments are reliable.

Concepts	Explanations	Implications
1.1 Treatments may be harmful	People often exaggerate the benefits of treatments and ignore or downplay potential harms. However, few effective treatments are 100% safe.	Always consider the possibility that a treatment may have harmful effects.
1.2 Personal experiences or anecdotes (stories) are an unreliable basis for assessing the effects of most treatments	People often believe that improvements in a health problem (e.g. recovery from a disease) was due to having received a treatment. Similarly, they might believe that an undesirable health outcome was due to having received a treatment. However, the fact that an individual got better after receiving a treatment does not mean that the treatment caused the improvement, or that others receiving the same treatment will also improve. The improvement (or undesirable health outcome) might have occurred even without treatment.	Claims about the effects of a treatment may be misleading if they are based on stories about how a treatment helped individual people, or if those stories attribute improvements to treatments that have not been assessed in systematic reviews of fair comparisons.
1.3 A treatment outcome may be associated with a treatment, but not <i>caused</i> by the treatment	The fact that a treatment outcome (i.e. a potential benefit or harm) is associated with a treatment does not mean that the treatment <i>caused</i> the outcome. For example, people who seek and receive a treatment may be healthier and have better living conditions than those who do not seek and receive the treatment. Therefore, people receiving the treatment might <i>appear</i> to benefit from the treatment, but the difference in outcomes could be because of their being healthier and having better living conditions, rather than because of the treatment.	Unless other reasons for an association between an outcome and a treatment have been ruled out by a fair comparison, do not assume that the outcome was <i>caused</i> by the treatment.
1.4 Widely used treatments or treatments that have been used for a long time are not necessarily beneficial or safe	Treatments that have not been properly evaluated but are widely used or have been used for a long time are often assumed to work. Sometimes, however, they may be unsafe or of doubtful benefit.	Do not assume that treatments are beneficial or safe simply because they are widely used or have been used for a long time, unless this has been shown in systematic reviews of fair comparisons of treatments.
1.5 New, brand-named, or more expensive treatments may not be better than available alternatives	New treatments are often assumed to be better simply because they are new or because they are more expensive. However, they are only very slightly likely to be better than other available treatments. Some side effects of treatments, for example, take time to appear and it may not be possible to know whether they will appear without long term follow-up.	A treatment should not be assumed to be beneficial and safe simply because it is new, brand-named or expensive.
1.6 Opinions of experts or authorities do not alone provide a reliable basis for deciding on the benefits and harms of treatments	Doctors, researchers, patient organisations and other authorities often disagree about the effects of treatments. This may be because their opinions are not always based on systematic reviews of fair comparisons of treatments.	Do not rely on the opinions of experts or other authorities about the effects of treatments, unless they clearly base their opinions on the findings of systematic reviews of fair comparisons of treatments.
1.7 Conflicting interests may result in misleading claims about the effects of treatments	People with an interest in promoting a treatment (in addition to wanting to help people), such as making money, may promote treatments by exaggerating benefits and ignoring potential harmful effects. Conversely, people may be opposed to a treatment for a range of reasons, such as cultural practices.	Ask if people making claims that a treatment is effective have conflicting interests. If they have conflicting interests, be careful not to be misled by their claims about the effects of treatments.
1.8 Increasing the amount of a treatment does not necessarily increase the benefits of a treatment and may cause harm	Increasing the dose or amount of a treatment (e.g. how many vitamin pills you take) often increases harms without increasing beneficial effects.	If a treatment is believed to be beneficial, do not assume that more of it is better.

Concepts	Explanations	Implications
1.9 Earlier detection of disease is not necessarily better	People often assume that early detection of disease leads to better outcomes. However, screening people to detect disease is only helpful if two conditions are met. First, there must be an effective treatment. Second, people who are treated before the disease becomes apparent must do better than people who are treated after the disease becomes apparent. Screening tests can be inaccurate (e.g. misclassifying people who do not have disease as having disease). Screening can also cause harm by labelling people as being sick when they are not and because of side effects of the tests and treatments.	Do not assume that early detection of disease is worthwhile if it has not been assessed in systematic reviews of fair comparisons between people who were screened and people who were not screened.
1.10 Hope or fear can lead to unrealistic expectations about the effects of treatments	Hope can be a good thing, but sometimes people in need or desperation hope that treatments will work and assume they cannot do any harm. Similarly, fear can lead people to use treatments that may not work and can cause harm. As a result, they may waste time and money on treatments that have never been shown to be useful, or may actually cause harm.	Do not assume that a treatment is beneficial or safe, or that it is worth whatever it costs, simply because you hope that it might help.
1.11 Beliefs about how treatments work are not reliable predictors of the actual effects of treatments	Treatments that should work in theory often do not work in practice, or may turn out to be harmful. An explanation of how or why a treatment might work does not prove that it works or that it is safe.	Do not assume that claims about the effects of treatments based on an explanation of how they might work are correct if the treatments have not been assessed in systematic reviews of fair comparisons of treatments.
1.12 Large, dramatic effects of treatments are rare	Large effects (where everyone or nearly everyone treated experiences a benefit or a harm) are easy to detect without fair comparisons, but few treatments have effects that are so large that fair comparisons are not needed.	Claims of large effects are likely to be wrong. Expect treatments to have moderate, small or trivial effects, rather than dramatic effects. Do not rely on claims of small or moderate effects of a treatment, which are not based on systematic reviews of fair comparisons of treatments.

2. Judging whether a comparison of treatments is a fair comparison

Well-informed treatment decisions require fair comparisons of treatments; i.e. comparisons designed to minimise the risk of errors. Not all comparisons of treatments are fair comparisons.

Concepts	Explanations	Implications
2.1 Evaluating the effects of treatments requires appropriate comparisons	If a treatment is not compared to something else, it is not possible to know what would happen without the treatment, so it is difficult to attribute outcomes to the treatment.	Always ask what the comparisons are when considering claims about the effects of treatments. Claims that are not based on appropriate comparisons are not reliable.
2.2 Apart from the treatments being compared, the comparison groups need to be similar (i.e. 'like needs to be compared with like')	If people in the treatment comparison groups differ in ways other than the treatments being compared, the apparent effects of the treatments might reflect those differences rather than actual treatment effects. Differences in the characteristics of the people in the comparison groups might result in estimates of treatment effects that appear either larger or smaller than they actually are. A method such as allocating people to different treatments by assigning them random numbers (the equivalent of flipping a coin) is the best way to ensure that the groups being compared are similar in terms of both measured and unmeasured characteristics.	Be cautious about relying on the results of non-randomized treatment comparisons (for example, if the people being compared chose which treatment they received). Be particularly cautious when you cannot be confident that the characteristics of the comparison groups were similar. If people were <i>not</i> randomly allocated to treatment comparison groups, ask if there were important differences between the groups that might have resulted in the estimates of treatment effects appearing either larger or smaller than they actually are.
2.3 People's experiences should be counted in the group to which they were allocated	Randomized allocation helps to ensure that the treatment and comparison groups share similar characteristics. However, people sometimes do not receive or take the allocated treatments. The characteristics of such people often differ from those who do. Therefore, excluding from the analysis people who did not receive the allocated treatment may mean that like is no longer being compared with like.	Be cautious about relying on the results of treatment comparisons if patients' outcomes are not counted in the group to which they were allocated. For example, in a comparison of surgery and drug treatments, people who die while waiting for surgery should be counted in the surgery group, even though they did not receive surgery.
2.4 People in the groups being compared need to be cared for similarly (apart from the treatments being compared)	Apart from the treatments being compared, people in the treatment comparison groups should otherwise receive similar care. If, for example, people in one group receive more attention and care than people in the comparison group, differences in outcomes could be due to differences in the amount of attention each group received rather than due to the treatments that are being compared. One way of preventing this is to keep providers unaware ("blind") of which people have been allocated to which treatment.	Be cautious about relying on the results of treatment comparisons if people in the groups that are being compared were not cared for similarly (apart from the treatments being compared). The results of such comparisons could be misleading.
2.5 If possible, people should <i>not</i> know which of the treatments being compared they are receiving	People in a treatment group may experience improvements (for example, less pain) because they <i>believe</i> they are receiving a better treatment, even if the treatment is not actually better (this is called a placebo effect), or because they behave differently (due to knowing which treatment they received, compared to how they otherwise would have behaved). If individuals know that they are receiving (they are not "blinded" to) a treatment that they believe is better, some or all of the apparent effects of the treatment may be due either to a placebo effect or because the recipients behaved differently.	Be cautious about relying on the results of treatment comparisons if the participants knew which treatment they were receiving, this may have affected their expectations or behaviour. The results of such comparisons could be misleading.

Concepts	Explanations	Implications
2.6 Outcomes should be measured in the same way (fairly) in the treatment groups being compared	If an outcome is measured differently in two comparison groups, differences in that outcome may be due to <i>how</i> the outcome was measured rather than <i>because</i> of the treatment received by people in each group. For example, if outcome assessors believe that a particular treatment works and they know which patients have received that treatment, they may be more likely to observe better outcomes in those who have received the treatment. One way of preventing this is to keep outcome assessors unaware ("blind") of which people have been allocated to which treatment. This is less important for "objective" outcomes like death than for "subjective" outcomes like pain.	Be cautious about relying on the results of treatment comparisons if outcomes were not measured in the same way in the different treatment comparison groups. The results of such comparisons could be misleading.
2.7 It is important to measure outcomes in <i>everyone</i> who was included in the treatment comparison groups	People in a treatment comparison who are not followed up to the end of the study may have worse outcomes than those who are. For example, they may have dropped out because the treatment was not working or because of side effects. If those people are excluded, the findings of the study may be misleading.	Be cautious about relying on the results of treatment comparisons if many people were lost to follow-up, or if there was a big difference between the comparison groups in the percentages of people lost to follow-up. The results of such comparisons could be misleading.

3. Understanding the role of chance

Well-informed treatment decisions require information about the risk of being misled by the play of chance.

Concepts	Explanations	Implications
3.1 Small studies in which few outcome events occur are usually not informative and the results may be misleading	When there are only few outcome events, differences in outcome frequencies between the treatment comparison groups may easily have occurred by chance and may mistakenly be attributed to differences between the treatments.	Be cautious about relying on the results of treatment comparisons with few outcome events may be misleading. The results of such comparisons could be misleading.
3.2 The use of p-values to indicate the probability of something having occurred by chance may be misleading; confidence intervals are more informative	The observed difference in an outcome is the best estimate of how effective or safe a treatment is (or would be, if the comparison were made in many more people). However, because of the play of chance, the true difference may be larger or smaller. The confidence interval is the range within which the true difference is likely to lie, after taking into account the play of chance. Although a confidence interval (margin of error) is more informative than a p-value, the latter is often reported. P-values are often misinterpreted to mean that treatments have or do not have important effects.	Understanding a confidence interval may be necessary to understand the reliability of an estimated treatment effect. Whenever possible, consider confidence intervals when assessing estimates of treatment effects. Do not be misled by p-values.
3.3 Saying that a difference is statistically significant or that it is not statistically significant can be misleading	Statistical significance is often confused with importance. The cut-off for considering a result as statistically significant is arbitrary, and statistically non-significant results can be either informative (showing that it is very unlikely that a treatment has an important effect) or inconclusive (showing that the effects of the treatment are uncertain).	Claims that results were significant or non-significant usually mean that they were not <i>statistically</i> significant or non-significant. This is not the same as important or not important. Do not be misled by such claims.

4. Considering all of the relevant fair comparisons

Well-informed treatment decisions require systematic reviews of the evidence. Non-systematic summaries can be misleading.

Concepts	Explanations	Implications
4.1 The results of single comparisons of treatments can be misleading	A single comparison of treatments rarely provides conclusive evidence and results are often available from other comparisons of the same treatments. These other comparisons may have different results or may help to provide more reliable and precise estimates of the effects of treatments.	The results of single comparisons of treatments can be misleading.
4.2 Reviews of treatment comparisons that do not use systematic methods can be misleading	Reviews that do not use systematic methods may result in biased or imprecise estimates of the effects of treatments because the selection of studies for inclusion may be biased or the methods may result in some studies not being found. In addition, the appraisal of some studies may be biased, or the synthesis of the results of the selected studies may be inadequate or inappropriate.	Whenever possible, use systematic reviews of fair comparisons rather than non-systematic reviews of fair comparisons of treatments to inform your decisions.
4.3 Well done systematic reviews often reveal a lack of relevant evidence, but they provide the best basis for making judgements about the certainty of the evidence	The certainty of the evidence (the extent to which the research provides a good indication of the likely effects of treatments) can affect the treatment decisions people make. For example, someone might decide not to use or to pay for a treatment if the certainty of the evidence is low or very low. How certain the evidence is depends on the fairness of the comparisons, the risk of being misled by the play of chance, and how directly relevant the evidence is. Systematic reviews provide the best basis for these judgements and should report an assessment of the certainty of the evidence based on these judgements.	When using the findings of systematic reviews to inform your decisions, always consider the degree of certainty of the evidence.

5. Understanding the results of fair comparisons of treatments

Well-informed treatment decisions require information about the size of effects. Research results may be presented in misleading ways.

Concepts	Explanations	Implications
5.1 Treatments usually have beneficial <i>and</i> harmful effects	Because treatments can have harmful effects as well as beneficial effects, decisions should be informed by the balance between the benefits and harms of treatments. Costs also need to be considered.	Always consider the trade-offs between the potential benefits of treatments and the potential harms and costs of treatments.
5.2 Relative effects of treatments alone can be misleading	Relative effects (e.g. the ratio of the probability of an outcome in one treatment group compared with that in a comparison group) are insufficient for judging the importance of the difference (between the probabilities of the outcome). A relative effect may give the impression that a difference is larger than it actually is when the likelihood of the outcome is small to begin with. For example, if a treatment reduces the probability of getting a stroke by 50% but also has harms, and your risk of getting a stroke is 2 in 100, receiving the treatment is likely to be worthwhile. If, however, your risk of getting a stroke is 2 in 10,000, then receiving the treatment is unlikely to be worthwhile even though the <i>relative</i> effect is the same.	Always consider the absolute effects of treatments – that is, the difference in outcomes between the treatment groups being compared. Do not make a treatment decision based on relative effects alone.
5.3 Average differences between treatments can be misleading	For outcomes that are measured on a scale (e.g. weight or pain) the difference between the average in a treatment group and the average in a comparison group may not make it clear how many people experienced a big enough change (e.g. in weight or pain) for them to notice it, or that they would regard as important.	When outcomes are measured on a scale, it cannot be assumed that everyone has experienced the average effect of a treatment.

6. Judging whether fair comparisons of treatments are relevant

Well-informed treatment decisions require relevant information. The results of fair comparisons may not be relevant to you

Concepts	Explanations	Implications
6.1 Fair comparisons of treatments should measure outcomes that are important	Patients, professionals and researchers may have different views about which outcomes are important. Studies often measure outcomes, such as heart rhythm irregularities, as surrogates for important outcomes, like death after heart attack. However, the effects of treatments on surrogate outcomes often do not provide a reliable indication of the effects on outcomes that are important.	Do not be misled by surrogate outcomes.
6.2 Fair comparisons of treatments in animals or highly selected groups of people may not be relevant	Systematic reviews of studies that only include animals or a selected minority of people are unlikely to provide results that are relevant to most people.	Results of systematic reviews of studies in animals or highly-selected groups of people may be misleading.
6.3 The treatments evaluated in fair comparisons may not be relevant or applicable	A fair comparison of the effects of a surgical procedure done in a specialised hospital may not provide a reliable estimate of the effects and safety of the same procedure performed in other settings. Similarly, comparing a new drug to a drug or dose that is not commonly used (and which may be less effective or safe than those in common use) would not provide a good estimate of how the new drug compares to what is commonly done.	Be aware that your circumstances may be sufficiently different that the results of systematic reviews of fair comparisons of treatments may not apply to you.
6.4 Results for a selected group of people <i>within</i> fair comparisons can be misleading	Comparisons of treatments often report results for a selected group of participants in an effort to assess whether the effect of a treatment is different for different types of people (e.g. men and women or different age groups). These analyses are often poorly planned and reported. Most differential effects suggested by these 'subgroup results' are likely to be due to the play of chance and are unlikely to reflect true differences.	Findings based on results for subgroups of people within a treatment comparison may be misleading.

Glossary

Absolute effects	Absolute effects are differences between outcomes in the groups being compared. For example, if 10% (10 per 100) experience an outcome in one of the treatment comparison groups and 5% (5 per 100) experience that outcome in the other group, the absolute effect is $10\% - 5\% =$ a 5% difference.
Allocation	Allocation is the assignment of participants in comparisons of treatments to the different treatments (groups) being compared.
Association	Association is a relationship between two attributes, such as using a treatment and experiencing an outcome.
Average difference	The average difference is used to express treatment differences for continuous outcomes, such as weight, blood pressure or pain measured on a scale. It is the difference between the average value for an outcome measure (for example kilograms) in one group and that in a comparison group.
Certainty of the evidence	The certainty of the evidence is an assessment of how good an indication a systematic review provides of the likely effect of a treatment; i.e. the likelihood that the effect will be substantially different from what the studies found (different enough that it might affect a decision). Judgements about the certainty of the evidence are based on factors that reduce the certainty (risk of bias, inconsistency, indirectness, imprecision and publication bias) and factors that increase the certainty.
Chance	In the context of comparisons of treatments, chance is the occurrence of differences between comparison groups that are not due to treatment effects or bias. The play of chance (random error) can lead to incorrect conclusions about treatment effects if too few outcomes occur in studies.
Confidence interval	A confidence interval is a statistical measure of a range within which there is a high probability (usually 95%) that the actual value lies. Wide intervals indicate lower confidence; narrow intervals greater confidence.
Fair comparison	Fair comparisons of treatments are comparisons designed to minimize the risk of systematic errors (biases) and random errors (resulting from the play of chance).
Outcome	An outcome is a potential benefit or harm of a treatment measured in a treatment comparison. An outcome measure is how the outcome is measured in a study.
P-value	A p-value is the probability (ranging from zero to one) that the results observed in a study (or results more extreme) could have occurred by chance if in reality there were no treatment differences.
Placebo	A placebo is a treatment that does not contain active ingredients, which has been designed to be indistinguishable from the active treatment being assessed.
Placebo effect	A measurable, observable, or felt improvement in health or behaviour not attributable to the treatment administered.
Probability	Probability is the chance or risk of something, such as an outcome, occurring. See Risk
Relative effects	Relative effects are ratios. For example, if the probability of an outcome in the treatment group is 10% (10 per 100) and the probability of that outcome in a comparison group is 5% (5 per 100), the relative effect is $10/5 = 0.50$.
Reliable	The reliability of a claim or evidence about a treatment effect is the extent to which it is dependable or can be trusted. It should be noted that reliability often has a different meaning in the context of research, which is the degree to which results obtained by a measurement procedure can be replicated.
Risk	Risk is the probability of an outcome occurring. See Probability
Scale	A scale is an instrument for measuring or rating an outcome with a potentially infinite number of possible values within a given range, such as weight, blood pressure, pain or depression.
Statistical significance	Statistical significance is a difference that is unlikely (below a specified level of confidence – typically 5%) to be explained by the play of chance.
Study	A study is an investigation that uses specified methods to evaluate something. Different types of studies can be used to evaluate the effects of treatments. Some are more reliable than others.
Subgroup	A subgroup is a subdivision of a group of people; a distinct group within a group. For example, in studies or systematic reviews of treatment effects, questions are often asked about whether there are different effects for different subgroups of people in the studies, such as women and men, or people of different ages.
Surrogate outcomes	Surrogate outcomes are outcome measures that are not of direct practical importance but are believed to reflect outcomes that are important. For example, blood pressure is not directly important to patients but it is often used as an outcome in studies because it is a risk factor for stroke and heart attacks.
Systematic review	A systematic review is a summary of research evidence (studies) that uses systematic and explicit methods to summarise the research. It address a clearly formulated question using a structured approach to identify, select, and critically appraise relevant studies, and to collect and analyse data from the studies that are included in the review.
Theory	A theory is a supposition or a system of ideas intended to explain something.
Treatment	A treatment is any intervention (action) intended to improve health, including preventive, therapeutic and rehabilitative interventions and public health or health system interventions.
Treatment comparison	Treatment comparisons are studies of the effects of treatments.