

routinely exaggerated and possible harms are minimised. More empirically grounded and theoretically informed ethical, social, and economic analyses of medical travel are greatly needed.

References

1. India medical tourism destination 2009: healthcare without borders. The first ever medical tourism exhibition and conference in Canada [Internet]. Toronto: IMTD; 2009 [cited 2010 Nov 30]. Available from: <http://www.imtd2009.com/>
2. Ramirez de Arellano AB. Patients without borders: the emergence of medical tourism. *Int J Health Serv*. 2007;37(1):193-8.
3. Jensen J. Silver City health tourism in the early twentieth century: a case study. *N M Hist Rev*. 2009;84:321-61.
4. Eggertson L. Wait-list weary Canadians seek treatment abroad. *CMAJ*. 2006 Apr 25; 174(9):1247.
5. Mudur G. Hospitals in India woo foreign patients. *BMJ*. 2004 Jun 5;328(7452):1338.
6. Bagchi S. Growth generates health care challenges in booming India. *CMAJ*. 2008 Apr 8;178(8):981-3.
7. Turner L. 'First World health care at third world prices': globalization, bioethics and medical tourism. *BioSocieties*. 2007;2:303-25.
8. The World Bank. HNP Stats- health financing [Internet]. The World Bank;(c)2010 [cited 2010 Nov 30]. Available from: [http://go.worldbank.org/3KMGKYPN90\[S2\]](http://go.worldbank.org/3KMGKYPN90[S2])
9. Chanda R. *Foreign investment in hospitals in India: status and implications* [Internet]. New Delhi:WHO; [date unknown, cited 2010 Nov 30]. Available from: http://whoindia.org/LinkFiles/Trade_Agreement_Foreign_investment-full_text.pdf
10. Thomas G., Krishnan S. Effective public-private partnership in healthcare: Apollo as a cautionary tale. *Indian J Med Ethics*. 2010 Jan-Mar;7(1):2-4.
11. Cortez N. Patients without borders: the emerging global market for patients and the evolution of modern health care. *Indiana Law J*. 2008;83:71-132.
12. Chanda R. Trade in health services. *Bull World Health Organ*. 2002 Feb;80(2):158-63.

GRADE the evidence before using the results in clinical practice

PRATHAP THARYAN

Director, South Asian Cochrane Network and Centre, Prof BV Moses and ICMR Centre for Advanced Research and Training in Evidence-Informed Healthcare, Christian Medical College, Vellore 632 002 Tamil Nadu INDIA e-mail: prathap@cmcvellore.ac.in

Abstract

Reports of clinical trials that do not describe the methods used to minimise the risk of bias, and reports that do not present results in a comprehensible and accurate manner, are unethical as they could lead to misleading conclusions, adverse health outcomes, and the inappropriate use of healthcare resources. The Grading of Recommendations: Assessment, Development, and Evaluation (GRADE) approach to framing healthcare recommendations provides a pragmatic approach to making summary evidence profiles of outcome-specific evaluations regarding the magnitude and precision of estimates of benefit and harms, and the overall quality of evidence from comparisons of healthcare interventions. In addition, contextual factors such as the balance between benefits, harms, and resource costs; baseline risks in different groups; inconveniences; varying values and preferences; and competing priorities and options, should ideally be extrapolated from these evidence profiles and other sources of evidence to determine the strength of recommendations regarding the use of an intervention.

Science and ethics: mutually inseparable

Selvan et al (1) attempt to demonstrate that the use of appropriate statistics in research reports of clinical trials could improve the understanding of clinicians and lay people of the clinical implications of research evidence, and accelerate their incorporation into clinical practice. Submitting their article to an ethics journal is appropriate because clinical trials, even those that are conducted according to the highest ethical standards, are unethical and wasteful if they do not yield results that are accurate and understandable, and can be trusted (2).

It is therefore necessary for ethicists and those who espouse the ethical conduct of clinical research to understand the importance of evaluating whether trial results are credible and clinically important before they are used.

Estimation of treatment effects: relative versus absolute effects

Selvan and colleagues rightly emphasise the importance of looking beyond p values in evaluating the significance of differences in outcomes between interventions in a clinical trial. P values are traditionally used to assess if the results are statistically significant. They tell us if the observed difference in the outcomes of interventions in clinical trials excludes the possibility of this being due to chance (or random error) by more than 95%, if the p value is less than 0.05. P values do not indicate if the observed difference in outcome is clinically important. Even if the difference is clinically important, they do not indicate how important this might be. Selvan et al ignore p values altogether and discuss, instead, the use of relative risks (RR) and relative risk reduction (RRR). These measure the relative magnitude of efficacy of one intervention over the other. More important, they can be used to derive the absolute risk reduction (ARR) and the numbers need to treat to benefit (NNTB) or harm (NNTH), measures of the actual numbers of people likely to benefit or be harmed by the intervention.

Uncertainties in effect estimates

These effect estimates would need to be presented with their 95% confidence intervals. The confidence interval (CI) is an estimate of uncertainty; it depicts the range of values for the RR,

RRR, ARR or NNT that could be expected 95% of the time if the experiment were repeated elsewhere. The CI must be presented for assurance that the most and least optimistic estimates of benefit or harm are still likely to be clinically important, if one were to extrapolate these results to other settings (3).

For example, the relative risk (RR) of 0.67 in Table 1 of Selvan et al's paper indicates a relative risk reduction (RRR) of 0.33. This number is derived by subtracting the RR from 1 (an RR of 1 would indicate no difference in the effects of Drug A over B). There is a 33% reduction in the risk of an adverse outcome (or increase in the risk of a good outcome) with Drug A compared to Drug B.

If the 95% confidence intervals (CI) for the RR were 0.1 to 0.7, this would still indicate that Drug A was preferable to Drug B as the relative risk reduction could be as little as 30% (1 minus 0.7 expressed as a percentage) or as much as 90% (1 minus 0.9 expressed as a %). In other words, in the best case and worst case scenarios, the outcomes with Drug A would still be clinically important compared to Drug B. If the CI were 0.1 to 3.9, as is the case here, there would be considerable uncertainty regarding the benefits with Drug A; the results could suggest a 10 % reduction in risk (1 minus 0.1 expressed as a %) but also the possibility of a 290% increase in the risk of harm (3.9 minus 1 expressed as a %) compared to Drug B. This indicates significant imprecision of the effect estimate compared to the first example where the confidence intervals indicated greater precision. The precision around the effect estimate in the first example would have been even greater had the CI ranged from 0.6 to 0.7 (30% to 40% relative risk reduction with intervention A compared to intervention B).

The absolute risk reduction in the example by Selvan et al is 0.01% (the difference in the risk of the event with both drugs). That is, Drug A would benefit 1 person more than Drug B. This is not an impressive achievement and the confidence intervals of -0.04% to 0.07% around this estimate suggests that Drug A could benefit 4 fewer people or benefit 7 more people than Drug B, leaving one even more uncertain as to its true effects.

Improving transparency, accountability and applicability

Three additional points made by Selvan and colleagues that address transparency and accountability in trial reports and their applicability are: 1) the need to report the results of randomised controlled trials (RCTs) in accordance with the CONSORT guidelines and their many extensions (www.consort-statement.org) to ensure that the trial used methods that are likely to yield valid results; 2) the need to prospectively register important details of a trial's methods in a publicly accessible trials register to prevent or detect selective reporting and to detect publication and other reporting biases; and 3) the need to supplement the results of explanatory trials with pragmatic or real world trials (efficacy versus effectiveness).

Standards for reporting trials and the validity of results

CONSORT incorporates many aspects of trial design that increase the internal validity of a clinical trial, or the possibility that the methods used minimised the risk of bias so that we can have greater confidence that the results are as close to the "truth" as is possible. The endorsement of these reporting standards by the editors of medical journals in many parts of the world has increased confidence in the results of published trials. However, editors of many Indian medical journals have not supported this initiative adequately and therefore the internal validity (or confidence in an unbiased and accurate result) of many trials published in Indian medical journals is suspect (4).

Preventing and detecting reporting biases and improving study design

The Clinical Trials Registry- India (CTR-I; www.ctri.in) has seen a dramatic increase in the number of trials prospectively registered since June 15, 2009 following the directive from the Drugs Controller General of India that trials of all new drug applications must be registered in the CTR-I before enrolment of the first participant. This increase in trials registration is largely due to compliance by industry-sponsored trials; investigator-initiated trials need to follow suit, if transparency and accountability are to be better served. However, the template provided by the CTR-I (derived from the CONSORT statement), to encourage disclosure of important aspects of trial design that have an impact on internal validity, has resulted in the better design of trials, as evidenced from protocols registered in the CTR-I (5). This will, hopefully, lead to more transparent reporting of these methods when these trials are published and could increase our confidence in their results, compared to our uncertainty regarding the validity of results in trial reports currently published in Indian medical journals (4).

Efficacy versus effectiveness

An important aspect of effectiveness is the ability to generalise the results of efficacy and safety, generated by explanatory trials in well-controlled experimental settings, to the real world. In actual clinical practice (and in pragmatic or "real-world" trials), patients may have multiple co-morbidities and are not excluded from care, unlike in an explanatory clinical trial; newer interventions are compared to standard treatments and not placebos; and outcomes are (or ought to be) measured not in terms of statistical significance but in terms of clinical importance to the patient and to carers. The NNT is useful in understanding the effects of interventions in absolute terms, but if one has to be confident that the results are internally valid, likely to be beneficial, and applicable to the people normally seen in clinical practice, a summary of the evidence from the trial, or, more importantly, from a systematic review and meta-analysis of the results of all relevant trials of the intervention, would be required before research results could be reliably translated to clinical practice. This evidence summary needs to include an assessment of the impact that the methods of the trial(s) have on the confidence that one can place in the results; it must also be contextualised for the clinical situation in which one proposes to use the intervention.

Grading the overall quality of the evidence

The internationally approved Grading of Recommendations: Assessment, Development, and Evaluation (GRADE) approach to developing guidelines separates the quality of evidence from the strength of recommendations (6, 7). It acknowledges that the confidence one can place in the effects of an intervention is determined not only by the magnitude of treatment effects but also by the overall quality of the methods used to evaluate its efficacy and safety, and that this is likely to vary for each outcome assessed. In this grading system, RCTs are graded as providing evidence of high quality but the evidence can be downgraded for various reasons starting with limitations in study design. These limitations, that could affect the reliability of outcome estimates, occur due to a number of reasons. Improper generation of the random sequence, and inadequate concealment of allocation of participants to intervention arms, could lead to selection bias, where the trial participants in the two arms may differ in prognosis at baseline. Insufficient blinding of participants, outcome assessors and care providers could lead to performance and detection bias, particularly for subjectively reported, as opposed to objectively evaluated, outcomes. Incomplete data reported for outcomes or for participants, and selective reporting of outcomes, could result in reporting and other biases such as those due to conflicts of interest, or specific to types of RCT designs (6).

The evidence from a trial, or the pooled data from a systematic review of trials, could also be downgraded because the results are imprecise. Data from trials with only a few participants are likely to result in imprecise effect estimates (wide confidence intervals) that leave us uncertain as to the true effects of the intervention. The evidence may be further downgraded if the trial, or trials, provided indirect rather than direct evidence of effects. This happens when the trial excludes participants likely to be seen in clinical practice (infants, children, women, older people, the more severely ill, etc). This can also happen when surrogate outcomes are chosen rather than what one actually is hoping to achieve with the intervention (evaluating lowering of blood sugars in the short-term with an anti-diabetic drug, rather than using the complications caused by diabetes in the long term as the outcome of interest). Further downgrading could occur if the results across the trials are not consistent and if this inconsistency is substantial and cannot be explained by differences in study methods or differences in participant characteristics. Finally, the evidence may be downgraded for evidence of publication bias, where trials with unfavourable results are not published at all, or are published in less easily accessible journals, and are therefore unlikely to have been included in the body of evidence.

Evidence from observational studies (cohort, case-control studies, etc) are assumed to be of low quality but may be upgraded if there is a very large magnitude of effect, if there are no biases due to confounding that explain the magnitude of effects, and if a dose-response gradient can be demonstrated. The overall quality of evidence after considering all of the above may range from high, moderate, and low to very low.

High quality evidence is convincing and restores confidence in the robustness of the results while low quality evidence is unconvincing and reposes little confidence in the results.

Framing strong and weak recommendations

These summary evidence profiles, created by this pragmatic, explicit and sequential approach, are then ideally discussed by a multidisciplinary panel of relevant stakeholders. This panel incorporates judgments about the underlying values and preferences between management options and outcomes. This includes judgments about: 1) the importance of the outcome that one is trying to achieve or prevent; 2) the magnitude of treatment effect and the uncertainties in the estimates of likely benefit and risk; 3) the balance between risks and benefits, as well as between health-benefits and resource costs; 4) the inconvenience and the burdens of therapy; 5) the baseline risks of developing the outcome for different patient groups; 6) alternative interventions or strategies and other priorities; and 7) the varying values of people that are likely to affect their use of the intervention. This evaluation is made before grading the strength of recommendations (a strong or a weak recommendation) and formulating guidelines on the use of the intervention (see <http://www.gradeworkinggroup.org> for more details on the process).

Conclusion

The summary evidence profiles and the final grading of the strength of recommendations and formulation of guidelines thus incorporate issues critical to the confidence that one can place in the effects of interventions. They are of greater relevance to clinical implementation than the results of the intervention alone, expressed in relative or absolute terms.

Declaration of interest: *The author is a contributor to the Cochrane Collaboration, an organisation that prepares, maintains and disseminates the results of systematic reviews of the effects of interventions and that uses the GRADE approach to create summary of findings tables and evidence profiles.*

References

1. Selvan MS, Subbian S, Cantor SB, Rodriguez A, Smith ML, Walsh GL. Ethics of transparency in research reports. *Indian J Med Ethics*. 2011 Jan-Mar; 8(1):
2. Tharyan P. Don't just do it, do it right. Evidence for better health in low and middle income countries. *Ceylon Med J*. 2010 Mar; 55(1):1-4.
3. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.1 [Internet]. [place unknown]: The Cochrane Collaboration; 2008 [cited 2010 August 31]. Available from: www.cochrane-handbook.org
4. Tharyan P, Premkumar TS, Mathew V, Barnabas JP, Manuelraj. Editorial policy and the reporting of randomized controlled trials: a survey of instructions for authors and assessment of trial reports in Indian Medical Journals (2004-2005). *Natl Med J India*. 2008 Mar; 21(2):62-8.
5. Tharyan P. Prospective registration of clinical trials in India: strategies, achievements & challenges. *JEBM*. 2009; 1(2):19-28.
6. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ; GRADE Working Group. What is "quality of evidence" and why is it important to clinicians? *BMJ*. 2008 May 3; 336 (7651):995-8.
7. Ansari MT, Tsertsvadze A, Moher D. Grading quality of evidence and strength of recommendations: a perspective. *PLoS Med*. 2009 Sep; 6(9): e1000151. doi:10.1371/journal.pmed.1000151