Guideline evaluation: tricky business

Jeffrey D. Klausner

STD Prevention and Control Services Section, San Francisco Department of Public Health, San Francisco, California, USA.

Divisions of AIDS and Infectious Diseases, Department of Medicine, University of California, San Francisco, 1360 Mission Street, Suite #401, San Francisco, CA 94103, USA. Email: Jeff.Klausner@sfdph.org

How meaningfully to evaluate expert clinical guidelines for the management of specific diseases is an interesting question. Ideally, researchers would randomly allocate cases of the disease for which the different guidelines apply to management by clinicians adhering to the guidelines under study. Follow-up would occur and researchers would measure and compare outcomes. If the study were done well, we might be able to conclude which clinical management guideline was superior. Unfortunately, those types of studies are unlikely to be done and I am not sure such studies would be worth the cost of performing them. Different clinical management guidelines for the same disease usually have enough similarities and a few disputed or unresolved areas that expected differences in outcomes would be very minor and difficult to measure.

In lieu of experimentally evaluating guidelines, researchers have taken to describing how clinical guidelines meet published criteria for effective guideline development. The assumption is made that those criteria have clinical or external validity; that is, those criteria are associated with medical outcomes. In this issue of Sexual Health, authors from the UK have used the Cluzeau and AGREE instruments to evaluate national guidelines for the management of sexually transmitted diseases from the United States Centers for Disease Prevention and Control (CDC) and British Association for Sexual Health and HIV (BASHH). The 37-item Cluzeau and 23-item AGREE instruments are very similar, the AGREE having evolved from the Cluzeau.[1,2]

The CDC guidelines can be found at www.cdc.gov/STD/treatment and were most recently published in 2006. The BASHH guidelines are available at http://www.bashh.org/guidelines.asp and appear to be updated on a disease-specific basis. For example, the BASHH guideline for the management of genital tract infection with *Chlamydia trachomatis* was updated in 2006, whereas the BASHH guideline for the management of early syphilis was updated in 2002. The availability and dissemination of guidelines via the Internet offer the opportunity for focussed and timely updates such as the CDC recommendation to avoid fluoroquinolones in the treatment of *Neisseria gonorrhoeae* in the USA [http://www.cdc.gov/STD/treatment/2006/updated-regimens.htm] and the BASHH update about the availability of procaine penicillin [http://www.bashh.org/guidelines/penicillin_update_0306.pdf].

The guideline criteria used by Baird *et al.*[3] include various domains in the process of the guideline development with a focus on transparency or 'rigour of development' (clear reporting of funding, methodology, potential financial conflicts of interest, etc.), inclusiveness (involvement of stakeholders such as patients and clinical personnel), accountability (description of who exactly the authors are and their expertise) and process (disclosure of the writing and review schedule, dissemination activities). Each domain was equally weighted as the authors of the study created summary scores of each guideline characteristic.

Some consideration in the AGREE instrument was given to how the evidence-base was examined but attention to the type, level and strength of evidence was not evaluated.

That is, the guidelines do not score the evidence base used in the guidelines; what amount of evidence comes from randomised clinical trials, observational studies or expert opinion. That lack of attention to the quality of the evidence does not appear to be uncommon. One study reported that none of 24 appraisal tools of practice guidelines evaluated the clinical evidence base used to create the content of the guidelines the authors assessed.[4]

The authors find that the BASSH guidelines they evaluated – *which were developed in accordance with AGREE* – had higher summary and individual domain scores than the CDC guidelines in a similar topic. In the area of 'rigour' which might be consistent with the use of the available clinical evidence, the CDC guidelines consistently scored lower; however, that area include multiple measures related to the adequate articulation of the process for evaluating evidence rather than the quality of the evidence itself. The major differences between the guidelines were how each adhered to AGREE criteria regarding the issues of transparency, inclusiveness, accountability and process. Fortunately, the authors do not conclude superiority of one set of national guidelines over the other but allow the reader to infer his or herself that the guidelines with the higher score was superior. That logic is slippery at best and fallacious at worst. Given that the BASSH guidelines used the AGREE criteria as a framework should offer no surprise to the reader that when compared with guidelines that did not use the AGREE criteria one finds a higher score consistent with better adherence to predetermined criteria. In fact, one may be surprised as to why the BASSH guidelines did not score better and the CDC guidelines score as well as they did?

When one actually looks at the guidelines and compares clinical management recommendations, one finds multiple similarities and a few potentially important differences. For example, in the management of genital chlamydial infection

the BASSH and CDC guidelines recommend similar therapy – doxycycline 100 mg orally twice daily or azithromycin 1 g orally once – and short-term follow-up in pregnant women to perform a test of cure. The recommendation for follow-up in infected patients is markedly different however. The CDC guidelines recommend repeat screening at 3–4 months and the BASSH guidelines offer no such recommendation for repeat screening. Both guidelines recommend the use of epidemiological treatment in recent sex partners, but the BASSH guidelines make no mention of patient-delivered partner therapy or what is increasingly known at expedited partner therapy, whereas the CDC guidelines acknowledge the safety and benefit of expedited partner therapy and sanction its use. Thus, at the patient level the rigorous application of the CDC management guideline could result in a decreased frequency of repeat infection (through the increased likelihood of partner treatment) and at the public health level a reduced prevalence of infection (through repeated screening and treatment among positives). The BASSH guidelines' failure to include repeat testing and partner therapy other than epidemiological treatment upon presentation of a partner to a clinical setting are evidenced-based practices that were not assessed in the guideline evaluation performed by Baird *et al*.[3]

A second example of differences in clinical recommendations between the guidelines is in the management of early syphilis. The BASSH guidelines define early syphilis as syphilis acquired in the previous 2 years, whereas the CDC definition of early syphilis is syphilis acquired in the previous year. The BASSH guidelines recommend the use of daily injections of procaine penicillin G for 10 days for the treatment of early syphilis, whereas the CDC guidelines recommend a single dose injection of penicillin G benzathine. A further difference between the guidelines is in the treatment of HIV-infected patients with syphilis. The authors of the BASSH guidelines recommend treating all HIV-infected patients with syphilis presumptively for neurosyphilis with a 17–21-day regimen of procaine penicillin G injections plus oral probenecid, whereas in the CDC guidelines authors recommend the same single dose of penicillin G benzathine as in HIV-uninfected patients with early syphilis. The same study that the BASSH guideline authors cite to justify treating all HIV-infected patients with syphilis for neurosyphilis based on a rate of treatment failure of 18% at 6 months among those treated with standard non-neurosyphilis penicillin regimens, the CDC guideline authors cite as evidence of no difference in clinical outcomes between HIV-infected patients with syphilis treated with and without regimens for neurosyphilis.[5] Again, the AGREE criteria fail to address and identify those key differences that certainly could impact the clinical outcomes of patients with early syphilis.

In summary the authors of the US–British guideline evaluation study found statistically significant differences in the format, structure and reporting of the text of the guidelines when evaluated using two different but highly related standardised criteria. Importantly the British guidelines were developed utilising one of those evaluation instruments as a guide. While perhaps those findings are epistemologically meaningful, they are epidemiologically meaningless. To judge guidelines, one must look at the content, clinical relevance and the use and interpretation of the evidence base. One must compare like to like and what is meaningful. Perhaps some day evaluation criteria can be externally validated – demonstrated to truly measure something of clinical or public health importance – at this point, however, we cannot conclude the superiority of US or British guidelines based on the currently utilised evaluation methods.

Conflicts of interest

References

1 Cluzeau FA, Littlehons P, Grimshaw JM, Feder G, Moran SE. Development and application of generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999; 11: 21–8. doi: 10.1093/intqhc/11.1.21

2 The AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003; 12: 18–23. doi: 10.1136/qhc.12.1.18

3 Baird A, Olarinde O, Talbot M. An evaluation, using two assessment instruments, of the American and British national guidelines for the management of sexually transmissible and genital infections. *Sex Health* 2007; 4: 255–60. doi: 10.1071/SH07048

4 Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers D. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficity. *Int J Qual Health Care* 2005; 17: 235–42. doi: 10.1093/intqhc/mzi027

5 Rolfs RT, Joesoef MR, Hendershot EF, Rompalo AM, Augenbraun MH, Chiu M, *et al*. A randomized trial of enhanced therapy for early syphilis in patients with and without human immunodeficiency virus infection. The Syphilis and HIV Study Group. *N Engl J Med* 1997; 337: 307–14. doi: 10.1056/NEJM199707313370504