

Quality improvement in action

Evaluation of the methodological quality of the Health Protection Agency's 2009 guidance on neuraminidase inhibitors

Kevork Hopayian MD FRCGP
Honorary Senior Lecturer, Norwich Medical School, Norwich and GP, Leiston Surgery, Leiston, UK

Lucy Jackson MB ChB
Honorary Senior Lecturer, Norwich Medical School, Norwich and GP, Leiston Surgery, Leiston, UK

ABSTRACT

Background The Health Protection Agency (HPA) issued guidance advocating the prescription of neuraminidase inhibitors in July 2009 in response to a predicted pandemic of influenza. Although the contents of the guidance have been debated, the methodology has not.

Method The guidance was evaluated by two reviewers using a validated and internationally recognised tool for assessing guidelines, the Appraisal of Guidelines Research & Evaluation instrument (AGREE). This tool scores six domains independently of each other.

Results The guidance scored 61% for the domain scope and purpose and 54% for the domain clarity and presentation. By contrast, it scored only 31% for rigour of development due to poor linkage of its recommendations to evidence.

Conclusion The HPA should improve its performance in this domain to general practitioners in order to improve the credibility of its future guidance.

Keywords: antiviral agents, general practice, human influenza, practice guidelines, primary care

How this fits in with quality in primary care

What do we know?
The Health Protection Agency (HPA) may recommend the use of neuraminidase inhibitors during an influenza epidemic and did so in the guidance it issued preceding the winter of 2009. Systematic reviews of these drugs have raised doubts about their effectiveness.

What does this paper add?
The rigour by which the HPA guidance on the prescription of neuraminidase inhibitors in 2009 was poor. Future guidance should be developed more robustly if it is to have credibility.

Background

In July 2009, the Health Protection Agency (HPA) upgraded its response to the predicted pandemic of influenza A H1/N1 ('swine flu') from containment (phase 5) to treatment (phase 6). Phase 6 included the prescription of the neuraminidase inhibitors (NAIs) oseltamivir and zanamivir to suspected cases. The HPA provided guidance on the prescription of NAIs in the document, *Summary of Prescribing Guidance for the Treatment and Prophylaxis of Influenza-like Illness: Treatment Phase.*[1] The guidance gave as evidence for its recommendations an earlier publication from the Department of Health in response to the potential outbreak of avian flu A H5/N1, *Use of Antiviral Drugs in an Influenza Pandemic, Scientific Evidence Base*[2] and

the document from the European Medicines Association, *Assessment Report on Novel Influenza (H1N1) Outbreak*.[3]

Given the difficulty of predicting the scale of the spread of infection, the pressure of public expectations and the demand for effective communication, the HPA deserves credit for the speed with which it reacted to the rapidly evolving situation. It deserves credit also for its collaboration with primary care trusts and the Royal College of General Practitioners (RCGP). These organisations broadcasted the HPA's recommendations to general practitioners (GPs) who, together with NHS Direct, were responsible for prescriptions of NAIs. However, the validity of the recommendations have been challenged: a Cochrane systematic review raised doubts about the effectiveness and safety of NAIs on which the recommendations were based.[4] It concluded that the benefits of NAIs are modest; they shortened the duration of the illness by one day. However, there was no evidence that oseltamivir, the only NAI for which there was any data, reduced the rate of complications. Jefferson *et al* pointed out that the evidence on complication rates was probably affected by publication bias: of the relevant ten trials, only two had been published in a peer-reviewed journal. Publication bias is likely to overestimate the benefits of an intervention.[5] Even the modest benefit in reduction of duration of illness reported in the 2009 review is now in doubt. The review has been withdrawn by the authors as they have come to realise that most of the data were prone to publication bias and were unreliable. An updated review is expected in 2012 (Tom Jefferson, personal communication 17 December 2011). Although the content of the HPA's guidance has been challenged, its methodology has not been scrutinised to date. Methodology refers to the process of guideline development and presentation.

While the validity of the content of a guideline is judged by checking it against its evidence base, the quality of the guideline is judged by checking its methodology. Guideline users are not expected to check the evidence base for guidelines. To do so would defeat the object of accessing ready recommendations. However, guideline users can judge the credibility of a guideline and have a duty to do so.[6]

The aim of the study was to evaluate the methodological quality of the HPA's guidance on the prescribing of NAIs during the swine flu pandemic against a validated reference standard.

Methods

Reference standard: AGREE tool

The Appraisal of Guidelines Research & Evaluation instrument, AGREE, is a tool for assessing the methodological quality of clinical guidelines.[7] It has been validated[8] and is widely used.[9–15] It analyses the rigour and transparency with which guidelines have been developed, thereby providing both guideline developers and users the means to gauge how much the guideline inspires confidence in the recommendations. It comprises 23 items divided into six domains; scope and purpose, stakeholder involvement, rigour of development, clarity and presentation, applicability and editorial independence. Each item consists of a statement to which the reviewer can award a score from 1 to 4, in which 1 = *strongly disagree* and 4 = *strongly agree*. Although AGREE was replaced by the updated AGREEII in 2010,[16] we evaluated the HPA guidance according to the standards in existence at the time.

Documents examined

The HPA guidance *Summary of Prescribing Guidance for the Treatment and Prophylaxis of Influenza-like Illness: Treatment Phase*,[1] described itself as a summary so we requested the full document from the HPA. However, the HPA replied that this was the only advice document for distribution (email communication, Public Information Office, HPA – communications, 13 August 2009). In accordance with the AGREE principles, we obtained the evidence[2,3] on which the guidance was based to judge how well the recommendations were linked to the evidence.

Analysis

Two appraisers applied the AGREE instrument independently. A standardised score for each domain was calculated in accordance with the AGREE method. The agreement between the two appraisers was quantified using a weighted kappa calculation. Finally, as required by the AGREE instrument, we each gave an overall assessment on whether we would recommend the guidance.

The present study was confined to analysis of documents. It cannot gauge the rigour of guideline development that might have occurred internally within the HPA. However, our method and the AGREE instrument reflect the real situation of health providers who can only judge guidelines by the documents available to them.[17]

Results

Only two of the six domains attracted scores of >50% (Table 1). Domain 3, the extent to which evidence has been sought, included and linked to recommendations, scored lower. To scrutinise further whether this was a failure to report links that did exist or reflected an absence of links, we searched the supporting documents.[2,3] It transpired that although they had been referred to as the 'scientific evidence base', neither of them was a systematic review. Therefore, we could not judge whether evidence existed to support these recommendations without performing a systematic literature search ourselves. The weighted kappa value for agreement between the reviewers was 0.41. Although there is no absolute rule regarding the interpretation of kappa, consensus takes this to indicate moderately strong agreement.[18] Visual inspection of the individual scores reveals that differences between the two reviewers were no greater than 1 in all but one case (Table 2).

Discussion

Strengths and limitations of the study

The AGREE tool requires the exercise of judgement. This can raise concerns about the possibility of bias in the reviewers. We therefore state our starting positions. Both authors are GPs working at the same surgery. KH is the immunisation lead for the Leiston surgery and directed the surgery's response to the crisis. Aware of the controversy over the benefits of NAIs, he nevertheless complied with the HPA guidance. He brought to the study experience of having been a member of a Guideline Review Panel for the National Institute for Health and Clinical Excellence (NICE) and of lecturing on guideline development at the University of East Anglia. LJ had no experience of guideline development appraisal and describes herself as generally accepting of and adherent to guidelines issued by authoritative bodies.

The AGREE manual states that between two and four appraisers may be used. Circumstances restricted us to two reviewers. It could be argued that having two reviewers with such differing perspectives, as in our case, leads to a balance between potential personal biases, since the final score in AGREE is the average between the appraisers. This may be as good as if not better than having four appraisers with equivalent perspectives.

Implications of the results

The most striking feature of the analysis is the contrast between the higher scores attained for clarity of purpose and presentation, on the one hand, and the lower scores for transparency of the rigour of development, on the other hand. While the clarity of purpose and presentation mean that users of the guideline would find its relevance and application easy to comprehend,

Table 1 Standardised scores for each domain of the AGREE tool

Domain	Description	Score (%)
Scope and purpose	Assessment of the clarity with which the aims of the guideline, the clinical questions asked and the target population were described	61
Stakeholder involvement	Assesses the extent to which users of the guideline have been consulted or participated in guideline development, including piloting the guideline	13
Rigour of development	Evaluation of the extent to which evidence has been sought, included and linked to recommendations. Also scores for external review of product and date for review	31
Clarity and presentation	Relates to the format of the guideline and its clarity, specifically whether recommendations are unambiguous and key recommendations are clearly presented	54
Applicability	Concerned with the organisational and cost implications of the guidance	6
Editorial independence	Assessment of independence from the funding body and acknowledgement of possible conflicts of interest	0

Table 2 Scores for the AGREE tool by individual appraiser

AGREE Item	Appraiser 1	Appraiser 2	Average
1	3	2	2.5
2	2	3	2.5
3	3	4	3.5
4	1	1	1
5	1	1	1
6	2	3	2.5
7	1	1	1
8	1	1	1
9	1	2	1.5
10	2	1	1.5
11	3	4	3.5
12	2	3	2.5
13	1	1	1
14	2	3	2.5
15	3	4	3.5
16	2	1	1.5
17	2	3	2.5
18	2	4	3
19	1	1	1
20	1	1	1
21	2	1	1.5
22	1	1	1
23	1	1	1

the lack of transparency means that the reader is less able to judge the credibility of the guideline. It could be argued that this deficiency does not matter because in an emergency 'getting the message out' is more important than 'proving the point'. We do not accept this, believing that the credibility of a guideline is important in all situations, otherwise practitioners may be ambivalent in their commitment.

This ambivalence was demonstrated by our response to the final question, a global assessment, asking the raters whether they would recommend the guideline. From our experience of using the AGREE tool, we would have answered no, but we had to admit that in practice our sense of duty towards a national policy would have led us to adhere to the guidelines.

It could be argued that it is unfair to evaluate the HPA guidance as we have done because not all aspects of the AGREE tool are relevant to it. In particular the score of 0 in domain 6, financial independence, may seem irrelevant because the HPA is a public body without financial interests. Also domain 2, stakeholder involvement, may be considered by some to be relatively insignificant in an emergency, but the opposition to the guidelines by those who had to apply them or advise on them suggests otherwise.[19] However, a strength of the AGREE tool is that there is no summative score. Each domain is marked separately so we are able to consider performance in each domain independently. Therefore, it is of particular concern that performance was poor in the domain relating to evidence which plays a large part in determining the

credibility of a guideline. We would suggest that the HPA improve the presentation of its guidance by adhering to the reporting guidelines in AGREE, at least for those elements that the target audience might legitimately question. These are scope and purpose, rigour of development (especially linking recommendations to evidence), clarity and presentation, and applicability. Given the doubts now being raised about the evidence on which the guidance was based, GPs in future might be less willing to adhere to the HPA guidance unless transparency is improved.

REFERENCES

1 Health Protection Agency. *Summary of Prescribing Guidance for the Treatment and Prophylaxis of Influenza-like Illness: treatment phase v1.6*. London: Health Protection Agency, 2009.
2 Department of Health. *Use of Antiviral Drugs in an Influenza Pandemic: scientific evidence base*. www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_077277.pdf. 2007 (accessed 16 February 2012).
3 European Medicines Agency. *CHMP Assessment Report on Novel Influenza (H1N1) Outbreak Tamiflu (Oseltamivir) Relenza (Zanamivir)*. London: European Medicines Agency, 2009.
4 Jefferson T, Jones M, Doshi P and Del Mar C. Neuraminidase inhibitors for preventing and treating influenza in healthy adults: systematic review and meta-analysis. *BMJ* 2009;339:b5106.
5 Jefferson T, Doshi P, Thompson M and Heneghan C. Ensuring safe and effective drugs: who can do what it takes? *BMJ* 2011;342:c7258.
6 Hopayian K. *Making Your Practice Evidence-based. A self-study guide for primary care*. London: RCGP Publications, 2010.
7 The AGREE Collaboration. *AGREE Instrument Training Manual*. www.agreecollaboration.org/pdf/aitraining.pdf. 2003 (accessed 11 August 2011).
8 The AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality and Safety in Health Care* 2003;12(1):18–23.
9 Appleyard TL, Mann CH and Khan KS. Guidelines for the management of pelvic pain associated with endometriosis: a systematic appraisal of their quality. *British Journal of Obstetrics and Gynaecology* 2006;113(7):749–57.
10 Boluyt N, Lincke CR and Offringa M. Quality of evidence-based pediatric guidelines. *Pediatrics* 2005;115(5):1378–91.
11 Burgers JS, Fervers B, Haugh M *et al.* International assessment of the quality of clinical practice guidelines in oncology using the Appraisal of Guidelines and Research and Evaluation Instrument. *Journal of Clinical Oncology* 2004;22(10):2000–7.
12 Cates JR, Young DN, Bowerman DS and Porter RC. An independent AGREE evaluation of the occupational medicine practice guidelines. *Spine Journal* 2006;6(1):72–7.
13 Harpole LH, Kelley MJ, Schreiber G, Toloza EM, Kolimaga J and McCrory DC. Assessment of the scope and quality of clinical practice guidelines in lung cancer. *Chest* 2003;123(Suppl 1):7S–20S.
14 Hurdowar A, Graham ID, Bayley M, Harrison M, Wood-Dauphinee S and Bhogal S. Quality of stroke rehabilitation clinical practice guidelines. *Journal of Evaluation in Clinical Practice* 2007;13(4):657–64.
15 Navarro Puerto MA, Ibarluzea IG, Ruiz OG *et al.* Analysis of the quality of clinical practice guidelines on established ischemic stroke. *International Journal of Technology Assessment in Health Care* 2008;24(3):333–41.
16 Brouwers MC, Kho ME, Browman GP *et al.* AGREE II: advancing guideline development, reporting and evaluation in health care. *Canadian Medical Association Journal* 2010;182(18): E839–42.
17 Brouwers M, Kho ME, Browman GP *et al.* AGREE II: Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. www.agreecollaboration.org/instrument (accessed 12 August 2011).
18 Altman DG. *Practical Statistics for Medical Research*. London: Chapman and Hall, 1997.
19 English PM, Carroll K, Majeed A, Sundkvist T, Millership S and Chambers S. A/H1N1 flu. Policy on antiviral drugs needs to be revised [Letter]. *BMJ* 2009;339:b2728.

FUNDING

None.

ETHICAL APPROVAL

This was a study of published texts so ethical approval was not required.

PEER REVIEW

Not commissioned; externally peer reviewed.

CONFLICTS OF INTEREST

The clinical workload of KH and LJ is affected by the HPA guidelines.

ADDRESS FOR CORRESPONDENCE

Kevork Hopayian, Leiston Surgery, Main St, Leiston IP16 4ES, UK. Email: k.hopayian@btinternet.com

Received 28 September 2011
Accepted 14 January 2012

Appendix: AGREE instrument for guideline appraisal

Instructions

Response scale

Each item is rated on a four-point scale ranging from 4 *Strongly Agree* to 1 *Strongly Disagree*, with two mid-points: 3 *Agree* and 2 *Disagree*. The scale measures the extent to which a criterion (item) has been fulfilled.

- If you are confident that the criterion has been fully met then you should answer *Strongly Agree*.
- If you are confident that the criterion has not been fulfilled at all or if there is no information available then you should answer *Strongly Disagree*.
- If you are unsure that a criterion has been fulfilled, for example, because the information is unclear or because only some of the recommendations fulfil the criterion, then you should answer *Agree* or *Disagree*, depending on the extent to which you think the issue has been addressed.

Comments

There is a box for comments next to each item. You should use this box to explain the reasons for your responses. For example, you may *Strongly Disagree* because the information is not available, the item is not applicable or the methodology described in the information provided is unsatisfactory. Space for further comments is provided at the end of the instrument.

Calculating domain scores

Domain scores can be calculated by summing all the scores of the individual items in a domain and by standardising the total as a percentage of the maximum possible score for that domain.

Note: The six domain scores are independent and should not be aggregated into a single quality score. Although the domain scores may be useful for comparing guidelines and will inform the decision as to whether or not to use or to recommend a guideline, it is not possible to set thresholds for the domain scores to mark a 'good' or 'bad' guideline.

Overall assessment

A section for overall assessment is included at the end of the instrument. This contains a series of options: *Strongly recommend*, *Recommend* (with provisos or alterations), *Would not recommend* and *Unsure*.

The overall assessment requires the appraiser to make a judgement as to the quality of the guideline, taking each of the appraisal criteria into account.

Criteria

I Scope and purpose

1. The overall objective(s) of the guideline is (are) specifically described.
2. The clinical question(s) covered by the guideline is (are) specifically described.
3. The patients to whom the guideline is meant to apply are specifically described.

II Stakeholder involvement

4. The guideline development group includes individuals from all the relevant professional groups.
5. The patients' views and preferences have been sought.
6. The target users of the guideline are clearly defined.
7. The guideline has been piloted among target users.

III Rigour of development

8. Systematic methods were used to search for evidence.
9. The criteria for selecting the evidence are clearly described.
10. The methods used for formulating the recommendations are clearly described.
11. The health benefits, side-effects and risks have been considered in formulating the recommendations.
12. There is an explicit link between the recommendations and the supporting evidence.

13. The guideline has been externally reviewed by experts prior to its publication.
14. A procedure for updating the guideline is provided.

IV Clarity and presentation

15. The recommendations are specific and unambiguous.
16. The different options for management of the condition are clearly presented.
17. Key recommendations are easily identifiable.
18. The guideline is supported with tools for application.

V Applicability

19. The potential organisational barriers in applying the recommendations have been discussed.
20. The potential cost implications of applying the recommendations have been considered.
21. The guideline presents key review criteria for monitoring and/or audit purposes.

VI Editorial independence

22. The guideline is editorially independent from the funding body.
23. Conflicts of interest of guideline development members have been recorded.