

The relation between methods and recommendations in clinical practice guidelines for hypertension and hyperlipidemia

ATLE FRETHEIM, MD; JOHN W. WILLIAMS JR, MD, MHS; ANDREW D. OXMAN, MD; AND JEPH HERRIN, PhD
Oslo, Norway; Durham, North Carolina; and Charlottesville, Virginia

KEY POINTS FOR CLINICIANS

- Many guidelines address the same problems, often with conflicting recommendations.
- There is considerable variation regarding the methods the guideline developers use to make recommendations.
- Guideline developers who did not use rigorous methods appeared to make more aggressive recommendations for screening and treatment (ie, were more likely to promote interventions).

■ **OBJECTIVE** To assess the association between methods used to develop clinical practice guidelines and the recommendations that are made.

■ **STUDY DESIGN** Systematic review of clinical practice guidelines for hypertension or hyperlipidemia.

■ **OUTCOMES MEASURED** Two people independently appraised guideline methods by using 8 criteria and the aggressiveness of recommendations for treatment thresholds, initial drug selection, and screening.

■ **RESULTS** We identified 33 guidelines. Only 6 fulfilled 5 or more of the 8 criteria. For 5 of the criteria, fewer than 50% of the guidelines fulfilled those criteria. There was wide variation in recommendations for treatment thresholds, drug selection, and cholesterol screening. Guidelines that did not fulfill the criteria tended to suggest more aggressive recommendations than did guidelines that met the criteria. For 6 of the 8 criteria, guidelines published by specialty societies were less likely to fulfill them compared with guidelines not published by specialty societies.

■ **CONCLUSIONS** Guideline developers who did not use rigorous methods tended to promote intervening more aggressively for hypertension and hyperlipidemia.

■ **KEY WORDS** Practice guidelines; hypertension; hyperlipidemia; evidence-based medicine. (*J Fam Pract* 2002; 51:963-968)

Clinicians are inundated with clinical practice guidelines. Many guidelines address the same problems, often with conflicting recommendations. The disagreement concerning recommendations often hinges on how aggressive clinicians are in promoting interventions, eg, screening for prostate cancer or prescribe antibiotics for the treatment of sore throat. There is also variation with regard to the methods that guideline developers use to make recommendations. The traditional "GOBSAT"-technique ("Good Old Boys Sat At Table") has been criticized.¹ Several groups have proposed more systematic approaches for guideline development,²⁻⁶ and some databases, such as the Guideline Advisory Committee's Recommended Clinical Practice Guidelines, include only those guidelines that have been assessed for the rigor of the development process.⁷

There are strong logical arguments for developing guidelines systematically, eg, to ensure that they are based on current best evidence, to protect against bias, and to make the process transparent and open to criticism. However, guideline developers frequently do not adhere to such methods.⁸⁻¹¹ We explored the possible association between the methods used in guideline development and the recommendations given in those guidelines. We used guidelines for hypertension and hyperlipidemia in our study. Our hypothesis was that less rigorous methods would, on average, be associated with more aggressive recommendations in these guidelines.

From the Department of Health Services Research, Norwegian Directorate for Health and Social Welfare, Oslo, Norway (A.F., A.D.O.); the Department of Medicine, The Center for Health Services Research in Primary Care, Department of Veterans Affairs Medical Center and Duke University Medical Center, Durham, NC (J.W.W.); and Flying Buttress Associates, Charlottesville, VA (J.H.). Preliminary results were presented at a seminar on clinical practice guidelines hosted by the Norwegian Centre for Health Technology Assessment; Oslo, Norway; June 26, 2000. Some of the results also were presented at the Nordic Workshop on Evidence Based Health Care, hosted by the National Institute of Public Health; Oslo, Norway; May 2001. Address reprint requests to Atle Fretheim, MD, Department of Health Services Research, Norwegian Directorate for Health and Social Welfare, PO Box 8054 Dep, N-0031 Oslo, Norway. E-mail: atle.fretheim@shdir.no.

METHODS

Inclusion criteria

We defined clinical guidelines as recommendations intended to assist health professionals and patients in making decisions for specific clinical circumstances. To be included a guideline had to address at least 1 of the following issues: threshold for drug treatment of essential hypertension in primary prevention, threshold for drug treatment of hyperlipidemia in primary prevention, or identification of the target population for cholesterol screening. Guidelines were excluded if they did not clearly identify the panel responsible for developing the guideline, identify a sponsoring organization, or include a reference list. We excluded textbooks, editorials, and commentaries. Review articles that were prepared and published as background documents were included with the relevant clinical practice guidelines. We included guidelines published after 1992. When multiple versions were available from the same organization, we used the most recent version.

Search strategy

We searched MEDLINE from 1992 through February 2000 by using *hypertension*, *blood pressure*, *hyperlipidemia*, or *cholesterol* as the key term, limited to *practice guidelines* as publication type. In addition, we searched databases of guidelines maintained by several groups around the world. One author (A.F.) reviewed all the citations and reference lists of relevant guidelines, retrieved potentially relevant guidelines, and selected guidelines for inclusion.

Guideline development methods

We used 8 criteria to rate the methodologic quality of the guidelines (Table 1). The criteria were adapted from a guideline appraisal instrument that is being developed and tested by a group of European researchers (the Agree/Biomed collaboration).¹² It is based on a British *Appraisal Instrument for Clinical Guidelines*¹³ that has been tested for its validity and reliability and has been characterized as “the most well developed to date” in a recent study.¹⁴

TABLE 1

Criteria used to appraise guideline-development methods

Criterion	Standard for fulfillment
Main outcomes identified	
Is there an explicit statement of the main outcomes considered when developing the guideline?	Explicit statement of the main outcomes considered in developing the guidelines
Key stakeholders involved	
Are the essential stakeholders involved in the development group?	Inclusion of all “essential” stakeholders (generalist physicians, specialists, and methodologists)*
Systematic search and selection	
Has a systematic search for evidence been carried out and are criteria for inclusion and exclusion specified?	Search specifying all relevant databases or described “electronic databases” <i>or</i> Inclusion–exclusion criteria defined, at least briefly
Recommendations linked to evidence	
Is there an explicit link between the evidence and the recommendations given?	Grading of strength of recommendations or level of evidence
Benefits and risks considered	
Have the health benefits, side effects, and risks been considered?	Any quantitative or qualitative weighing of benefits and harms that is incorporated into formulating recommendations
Resources/costs	
Has the impact on resources been considered?	Economic analysis or qualitative consideration of cost issues that is linked explicitly to the formulation of the recommendations
No industry influence	
Is the guideline developed without funding or influence from the pharmaceutical industry?	No financial support from a pharmaceutical company and no involvement on the panel
Conflicts of interest stated	
Is there an explicit statement of conflict of interest?	Statement of potential conflicts of interest for panel members

*Essential stakeholders are generalist physicians, specialists, and methodologists. Optional stakeholders are patients, policy makers/health administrators, nurses, pharmacists, economists, and other physicians.

Two authors (A.F. and J.W.W.) evaluated each guideline independently. For analytical purposes, all criteria were dichotomized (Table 1). Because fewer than 50% of the guidelines reported sufficient information to determine stakeholder involvement, we supplemented information on authors through an Internet search.

Aggressiveness of recommendations

To grade the aggressiveness of the treatment recommendations, we evaluated the threshold for initiating pharmacologic treatment of hypertension (low = aggressive), threshold for initiating pharmacologic treatment of hyperlipidemia (low = aggressive), first-line antihypertensive drug (all drugs = aggressive), and the number of persons eligible for cholesterol screening (high = aggressive).

The thresholds for treatment of hypertension and hyperlipidemia were categorized with 4 clinical scenarios. By applying the scenarios we could determine the recommended thresholds for drug treatment among the various guidelines. We chose these specific scenarios to illustrate the existing variation in recommendations among the guidelines:

1. A 50-year-old man without a high-risk profile of cardiovascular disease.
2. A 50-year-old man with a high-risk profile of cardiovascular disease.
3. A 70-year-old man without a high-risk profile of cardiovascular disease.
4. A 70-year-old man with a high-risk profile of cardiovascular disease.

High-risk profile was defined separately for treatment of hypertension and treatment of hyperlipidemia. For hypertension, the high-risk scenario was a patient who smoked and was hyperlipidemic (total cholesterol > 310 mg/dL). For hyperlipidemia, the high-risk scenario was a patient who smoked and was hypertensive (blood pressure > 160/100 mm Hg). We assumed that lifestyle interventions had been attempted. For simplicity we used only the systolic value for blood pressure. We dichotomized the aggressiveness of thresholds for lipid-lowering treatment rather than calculating averages because many guidelines gave recommendations other than actual cholesterol values; eg, no treatment or familial hypercholesterolemia. Guidelines were considered nonaggressive if, for 2 or more of the scenarios, the threshold was set at 310 mg/dL or higher or was specified as familial hypercholesterolemia.

Guidelines recommending all common antihypertensive drugs were considered aggressive, and the ones suggesting more restrictive recommendations were considered nonaggressive. In a few guidelines this recommendation depended on the patient's age. For simplicity we examined the recommendations for 50-year-olds. We graded aggressiveness of recommendations on cholesterol

screening by estimating the proportion of the general adult population (in Norway) who would be candidates for screening or by case finding per year, if the guidelines were fully implemented.

Analysis

We qualitatively and quantitatively examined the relation between fulfillment of a methodologic criterion and the aggressiveness of recommendations. The power of our statistical analyses was limited by the available sample size. For hypertension, we averaged the treatment threshold for the 4 clinical scenarios within each guideline and compared the overall mean between guidelines meeting and not meeting the criterion. For the threshold to treat hyperlipidemia and for first-line therapy for hypertension, the degree of aggressiveness was dichotomized. We used the Fisher exact test to calculate *P* values for the association between the proportion of guidelines fulfilling a methodologic criterion and whether the recommendation was classified as aggressive. For cholesterol screening we found the mean yearly proportion of the adult population eligible for screening among guidelines fulfilling a methodologic criterion, and compared this with the mean for guidelines that did not fulfill the criterion. In addition to examining the association between methods and recommendations, we examined whether the level of stakeholder involvement or sponsorship by specialty societies was associated with fulfillment of the methodologic criteria. We included generalist physicians, specialists, and methodologists as "essential stakeholders" and patients, policy makers/health administrators, nurses, pharmacists, economists, and other physicians as "optional stakeholders."

RESULTS

We found 12 clinical guidelines for managing hypertension, 12 for hyperlipidemia, 5 for cholesterol screening, and 4 general guidelines for the prevention of coronary heart disease that met our inclusion criteria (references are available from the authors). Because each guideline was appraised according to the 8 methodologic criteria, we ended up with 264 appraisals (8 criteria applied to each of the 33 guidelines). There were 28 disagreements (11%), all of which were easily resolved by discussion. As expected, there was variation among the guidelines regarding fulfillment of methodologic criteria and the aggressiveness of recommendations.

Most guidelines did not meet the majority of the methodologic criteria (Table 2). Only 6 of the 33 guidelines met 5 or more of the 8 criteria. The threshold to start antihypertensive treatment varied in systolic blood pressure from 140 to 180 mm Hg for each of the 4 clinical scenarios we applied to the guidelines (Table 3). For 3 of the scenarios, the threshold to treat hyperlipidemia ranged from

TABLE 2
Variability in fulfillment of methodologic criteria

Criterion	Fulfilled, n (%)	Not fulfilled, n (%)	No information, n (%)
Main outcomes identified	10 (30)	23 (70)	0 (0)
Key stakeholders involved	21 (64)	9 (27)	3 (9)
Systematic search and selection	7 (21)	26 (79)	0 (0)
Recommendations linked to evidence	10 (30)	23 (70)	0 (0)
Benefits and risks considered	21 (64)	12 (36)	0 (0)
Resources/costs	14 (42)	19 (58)	0 (0)
No industry influence	23 (70)	2 (6)	8 (24)
Conflicts of interest stated	4 (12)	29 (88)	0 (0)

a total cholesterol value of 190 mg/dL to more than 310 mg/dL (Table 3). Fifteen guidelines gave recommendations for first-line therapy for hypertension. Three recommended thiazides only; 6 recommended thiazides and β-blockers; 1 recommended thiazides, β-blockers, and angiotensin-converting enzyme inhibitors; and 5 recommended all commonly used drugs. Recommendations for cholesterol screening ranged from no screening to testing the entire adult population every 2 to 5 years.

Associations between recommendations and methodologic criteria

The threshold to treat hypertension did not seem to be associated with fulfillment of methodologic criteria. Differences in recommendations for first-line drugs for hypertension were not strongly associated with any of the criteria. Although not statistically significant, there was a trend for guidelines to recommend all commonly available drugs when methodologic criteria were not met (Table W1, available on the JFP web site: <http://www.jfponline.com>).

For all but 1 quality criteria (main outcomes identified), fulfilling the criteria tended to be associated with a higher threshold to treat hyperlipidemia. Similarly, guidelines meeting quality criteria tended to give less aggressive recommendations for cholesterol screening than did guidelines not fulfilling the criteria. The criterion on stakeholder involvement was the exception, but this criterion was fulfilled by all but 1 of the guidelines (Table 4).

Stakeholder involvement and sponsorship by specialty societies

Guidelines that involved major stakeholders in the development

process tended to fulfill the methodologic criteria to a greater extent than did guidelines that did not (Table W2, available on the JFP web site: <http://www.jfponline.com>). Nine of the 33 guidelines were sponsored by specialty societies. These fulfilled the methodologic criteria less often than did other guidelines (Table 5).

DISCUSSION

We found that nonadherence to rigorous methods when developing guidelines for hypertension

and hyperlipidemia tends to be associated with more aggressive recommendations. We are not aware of other studies that have investigated the relation between methods and recommendations in clinical practice guidelines. The relatively small number of guidelines that met our inclusion criteria limited the power of our analyses, which rarely reached the conventional level of statistical significance ($P < .05$). Notwithstanding this lack of statistical significance, we believe our qualitative assessment of the pattern we found has practical importance.

Many articles have assessed the methodologic quality of clinical practice guidelines with the use of similar criteria, all these studies found poor adherence to recommendations for guideline development.⁸⁻¹¹ Grilli and colleagues found that “the quality of reporting of practice guidelines produced by specialty societies fell short of acceptable methodology” for the 431 guidelines they assessed.^{10(p104)} Shaneyfelt and colleagues found no difference in methodologic rigor between guidelines published by specialty societies and those

TABLE 3
Variability in guideline recommendations*

Age, clinical scenario	BP threshold (mm Hg) to treat hypertension				
	140	150	160	170	180
50 y, low risk	4 (25)	1 (6)	5 (31)	4 (25)	2 (13)
50 y, high risk	9 (56)	2 (13)	3 (19)	1 (6)	1 (6)
70 y, high risk	6 (38)	1 (6)	5 (31)	2 (13)	2 (13)
70 y, low risk	8 (50)	1 (6)	5 (31)	1 (6)	1 (6)
Age, clinical scenario	Cholesterol threshold (mg/dL) to treat hyperlipidemia				
	<230	≤230 < 270	≤270 < 310	≥310†	
50 y, low risk	3 (19)	3 (19)	1 (6)	9 (56)	
50 y, high risk	6 (38)	8 (50)	0	2 (13)	
70 y, high risk	5 (31)	3 (19)	2 (13)	3 (19)	
70 y, low risk	7 (44)	5 (31)	1 (6)	3 (19)	

*Data are presented as number (%) of patients.

†Includes the recommendations “no treatment” and “familial hypercholesterolemia.” BP, blood pressure.

TABLE 4

Guidelines for hyperlipidemia: relation between adherence to methodologic criteria and recommendations given*

Criterion	Guidelines recommending a low treatment-threshold [‡]		
	Criterion fulfilled	Criterion not fulfilled	P
Main outcomes identified	3/4 (75)	7/12 (58)	1.00 [†]
Key stakeholders involved	5/10 (50)	3/4 (75)	.58 [†]
Systematic search and selection	0/3	10/13 (77)	.036 [†]
Recommendations linked to evidence	1/3 (33)	9/13 (69)	.52 [†]
Benefits and risks considered	4/10 (40)	6/6 (100)	.034 [†]
Resources/costs	4/8 (50)	6/8 (75)	.61 [†]
Conflicts of interest stated	0/2	10/14 (71)	.13 [†]
Population to screen annually [§]			
Main outcomes identified	8 (0–21)	11 (5–17)	3 (–8.1 to 14) [¶]
Key stakeholders involved	11 (4.8–16)	1 [†]	–10 [¶]
Systematic search and selection	4 (0–9.8)	12 (6.2–18)	8 (–2.3 to 18) [¶]
Recommendations linked to evidence	6 (0–13)	11 (5.3–18)	5 (–5.5 to 16) [¶]
Benefits and risks considered	8 (1.2–15)	14 (6.0–21)	6 (–4.1 to 15) [¶]
Resources/costs	6 (0.9–12)	14 (5.8–22)	8 (–1.2 to 16) [¶]
Conflicts of interest stated	0	12 (6.6–16)	12 (–1.8 to 25) [¶]

*The criterion on industry influence is not included because all the guidelines either fulfilled the criterion or provided insufficient information to assess if the criterion was met

[†]P values assessed with the Fisher exact test.

[‡]Guidelines in which the threshold to treat is less than 310 mg/dL for 3 or more of the clinical scenarios described in the text. Data are presented as proportion (%).

[§]Data are presented as percentage (95% confidence interval).

[¶]One guideline did not fulfill this criterion, so confidence intervals could not be calculated.

[¶]Difference in percentage (95% confidence interval).

published by others but decided that methodologic criteria frequently were not met.¹¹ We also found that methodologic criteria frequently were not met, and that they were met less often for guidelines sponsored by specialty societies than for those sponsored by other groups.

Stakeholder involvement, as we have defined it, was closely related to panel composition, which has been examined by others. For example, a link was found between panel composition and ratings of the appropriateness of procedures. Those who used a given procedure were more likely to rate it as appropriate than were those who did not use it.^{15,16} Murphy and coworkers found that “members of a specialty are more likely to advocate techniques that involve their specialty.”^{17(p37)} Savoie and colleagues, in their critical appraisal of guidelines for cholesterol testing, found that “the greater the involvement of clinical experts in the development process of the clinical practice guidelines, the less the recommendations reflected the research evi-

dence.”^{19(p76)} This is consistent with our finding that broader stakeholder involvement was associated with methodologic criteria being met more often.

In our study, guideline developers that did not use rigorous methods appeared more likely to promote aggressive intervention. This may be true for guidelines for conditions other than hypertension and hyperlipidemia. However, guideline developers also may introduce biases toward less aggressive recommendations, eg, purchasers of health services. The degree to which bias is likely and even the direction sometimes may be difficult to predict.

The quality among the guidelines we assessed was not associated with year of publication or the country where the guidelines were developed. The 6 guidelines fulfilling 5 or more of the quality criteria were not published more recently. The countries of origin for these 6 guidelines were Australia, Canada, France, the

United Kingdom, and the United States.

There are strong logical reasons for users of guidelines to consider the methods used by guideline developers. Given the extent of disagreement among guidelines, it is necessary for users to understand the basis of those recommendations.

TABLE 5

Relation between specialty society sponsorship and fulfillment of methodologic criteria*

Criterion	Guidelines fulfilling criterion		
	Sponsored by specialty society	Not sponsored by specialty society	P [†]
Main outcomes identified	6/9 (67)	4/24 (17)	.010
Key stakeholders involved [‡]	2/9 (22)	19/21 (90)	.001
Systematic search and selection	1/9 (11)	6/24 (25)	.64
Recommendations linked to evidence	2/9 (22)	8/24 (33)	.69
Benefits and risks considered	3/9 (33)	18/24 (75)	.044
Resources/costs	3/9 (33)	11/24 (46)	.70
No industry influence [‡]	2/4 (50)	21/21 (100)	.02
Conflicts of interest stated	2/9 (22)	2/24 (8)	.30

*Data are presented as proportion (%).

[†]Assessed with the Fisher exact test.

[‡]We did not take into account the guidelines for which we had insufficient information to assess whether the criterion was met.

This is only possible if guideline developers employ systematic methods and explicitly report the methods that were used. Our study provides empirical support of skepticism toward guidelines that have been developed without employing systematic methods.

ACKNOWLEDGMENTS We thank Signe Flottorp and Lena Nordheim for helping with the appraisal of non-English guidelines and Jonathan Lomas and Brian Hutchison who helped develop the idea for this study.

REFERENCES

1. Miller J, Petrie J. Development of practice guidelines. *Lancet* 2000; 355:82-3.
2. Eddy D. *A Manual for Assessing Health Practices and Designing Practice Policies: The Explicit Approach*. Philadelphia: American College of Physicians; 1992.
3. Scottish Intercollegiate Guidelines Network. *An Introduction to SIGN Methodology for the Development of Evidence-Based Clinical Guidelines*. Vol 39. Edinburgh: Scottish Intercollegiate Guidelines Network; 1999.
4. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. *Clinical guidelines: developing guidelines*. *BMJ* 1999; 318:593-6.
5. *Guidelines for Clinical Practice. From Development to Use*. Washington, DC: National Academy Press; 1992.
6. Canadian Medical Association. *Quality of Care Program: The Guidelines for Canadian Clinical Practice*. Ottawa: Canadian Medical Association; 1993.
7. Guideline Advisory Committee. *Recommended clinical practice guidelines*. Available at: <http://www.gacguidelines.ca/aboutGAC.html>. Accessed April 4, 2002.
8. Gibson P. Asthma guidelines and evidence-based medicine. *Lancet* 1993; 342:1305.
9. Savoie I, Kazanjian A, Bassett K. Do clinical practice guidelines reflect research evidence? *J Health Serv Res Policy* 2000; 5:76-82.
10. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet* 2000; 355:103-6.
11. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999; 281:1900-5.
12. St George's Hospital Medical School Health Care Evaluation Unit. *Appraisal of guidelines research & evaluation*. Available at: <http://www.sghms.ac.uk/depts/phs/hceu/biomed.htm>. Accessed December 12, 2000.
13. Cluzeau FA, Littlejohns P, Grimshaw J, Feder G. *Appraisal Instrument for Clinical Guidelines*. London: St George's Hospital Medical School; 1997.
14. Graham ID, Calder LA, Hébert PC, Carter AO, Tetroe JM. A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care* 2000; 16:1024-38.
15. Herrin J, Etchason JA, Kahan JP, Brook RH, Ballard DJ. Effect of panel composition on physician ratings of appropriateness of abdominal aortic aneurysm surgery: elucidating differences between multispecialty panel results and specialty society recommendations. *Health Policy* 1997; 42:67-81.
16. Coulter I, Adams A, Shekelle P. Impact of varying panel membership on ratings of appropriateness in consensus panels: a comparison of a multi- and single disciplinary panel. *Health Serv Res* 1995; 30:577-91.
17. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998; 2(3): i-iv, 1-88.

JFP