

ORIGINAL ARTICLE

Guidelines in context of evidence

Eeva Ketola, Minna Kaila, Mari Honkanen

Qual Saf Health Care 2007;**16**:308–312. doi: 10.1136/qshc.2006.019752

Objective: In clinical practice guidelines, the quality of the available evidence is graded according to its reliability and quality. This study aimed to evaluate the quality of the available research evidence, using the levels of evidence, in the evidence summaries of 64 Finnish national evidence-based Current Care guidelines.

Design: Descriptive assessment.

Setting: Electronic web-based guidelines in Finland.

Main outcome measures: The proportions of evidence summaries with different levels of evidence (A–D).

Results: The 64 guidelines had a total of 2419 evidence summaries. Of these, 532 (22.0%) were evidence level A, 891 (36.8%) were evidence level B, 808 (33.4%) were evidence level C, and 188 (7.8%) were evidence level D. Most—that is, 81% of the level C and D evidence summaries dealt with diagnosis and treatment. Most of the evidence summaries pertained to treatment (58.2%) and diagnosis (22.4%). The sections on diagnosis and treatment represented 80% of all the level A and level B evidence, and 81% of all the level C and level D evidence.

Conclusions: There is adequate high-quality evidence (level A) to support only a fifth of the main statements of the 64 guidelines. This is most likely an optimistic estimate, since level D evidence often does not have an evidence summary. The guideline development groups find it easier to agree on recommendations based on level A and level B evidence.

See end of article for authors' affiliations

Correspondence to: Dr Eeva Ketola, Current Care Guidelines, Finnish Medical Society Duodecim, Helsinki, Finland; eeva.ketola@duodecim.fi

Accepted 8 April 2007

Quality is of pivotal importance in every healthcare system. Quality of care is a complex concept and consists of both subjective and more objective components.^{1 2} Knowledge is at the heart of good quality care, and according to Muir Gray³ it consists of three components: knowledge from research (evidence), knowledge from measurement of healthcare performance (statistics) and knowledge from experience (mistakes). We use published evidence to complement the silent knowledge passed on from previous generations of researchers. Without good research evidence, clinical decision making, be it diagnosis or treatment, is on shaky grounds.

Clinical practice guidelines have been widely adopted as a tool to improve quality of care.⁴ Such guidelines have been increasingly produced since the 1980s. Currently, most are produced in guideline programmes and aim to be explicitly evidence based. International collaboration has been established in this area,⁴ and efforts are underway to harmonise grading of the available evidence. However, as of now there are several systems.⁵ The basis of the grading systems is the evaluation of the validity of the research. It has been estimated that between 10% and 20% of healthcare decisions are based on high grades of evidence and only 50%, or even less, up to 15%, of medical treatment has been validated in clinical trials.⁶ However, a high grade does not necessarily mean clinically important.

In Finland, a national evidence-based guideline programme was established in 1994 under the auspices of the Finnish Medical Society Duodecim, Current Care (Käypä hoito; <http://www.kaypahoito.fi>).⁷ To ensure good methodology, a guideline developer's handbook has been available since the beginning of the programme. In January 2006, the collection included 64 Current Care guidelines covering a wide variety of clinical topics (table 1). These were supported by 2419 evidence summaries and graded recommendations. The present study aimed to provide an overview of the evidence for clinical decision making, using this collection of guidelines and the evidence summaries as the study material.

MATERIAL AND METHODS

The development process of a Current Care guideline is outlined in box 1 and follows the quality standards of the Appraisal of Guidelines, Research and Evaluation in Europe (AGREE) instrument.⁸ The Current Care board selects topics from suggestions made mostly by the specialist societies, lately with the help of a prioritising tool.⁹ The development group consists of clinical experts.

See p. 310 for definitions

The process begins with a literature search. Critical appraisal of the literature is based on criteria published by the Evidence Based Medicine Working Group.¹⁰ Depending on the quality and size of the original studies, the evidence base of the main statements is graded from A to D (table 2). The key statements are supported by evidence summaries. The Current Care consensus process is an informal one in which the guideline development group discusses the evidence in the context of the Finnish healthcare system. When there is lack of grade A or B evidence, and especially in the case of grade D evidence, this process can be tedious. The discussion is an iterative process at the end of which the actual recommendations are carefully worded. For more recent guidelines, the groups have been using computers to project the text on a screen for editing it together.

Electronic publication allows easy linking to the evidence base and wide dissemination. The important characteristic is the accessibility of the evidence summaries if more information on the topic is needed. Our most read guidelines in 2005 were the hypertension guideline (28 445 hits, 4.4% of all hits), lower back conditions guideline (24 615 hits, 3.8% of all hits) and schizophrenia guideline (21 455 hits, 3.4% of all hits). In all, the Current Care guidelines were read 640 434 times in 2005.

We examined all available 64 Current Care guidelines. The material for analysis was retrieved directly from the updated Current Care guideline XML database (accessed 7 February 2006). The guidelines were listed on the basis of the topics and then further by the sections in each guideline (epidemiology, prevention, diagnosis, treatment, rehabilitation, screening, recommended organisational level of care). Then all the evidence summaries were listed and classified by their topic

Table 1 The Current Care guideline topics included in the present study and distribution of their evidence summaries

Year of publication (latest update)	Title	Evidence grade (%)			
		A	B	C	D
	Central nervous system				
2005	Prolonged epileptic seizure	0	14	52	33
2003	Brain injuries and post-traumatic states following a brain injury in adults	15	42	33	10
2002	Migraine	8	60	24	8
2002	Diagnosis and pharmacotherapy of multiple sclerosis	25	53	29	3
	Eye disorders				
2002	Surgical treatment of refractive errors	13	38	44	6
2002	Glaucoma	32	38	24	6
2005	Cataracts in adults	29	53	12	6
	Surgery				
2000	Spinal cord injury	2	16	70	12
2003	Treatment of tibial fracture in adults	0	38	31	31
2003	Lower extremity venous insufficiency	8	46	42	4
2006	Treatment of hip fracture	39	25	26	11
2005	Benign prostatic hyperplasia	35	44	16	6
	Public health/general practice				
2002	Neck pain	3	27	52	18
2001	Lower back conditions	8	43	41	8
2003	Rheumatoid arthritis	28	35	34	2
2002	Obesity in adults	28	36	32	4
2000	Osteoporosis	8	62	31	0
2005	Treatment of alcohol misuse	38	37	18	8
2005	Hypertension	43	32	26	0
2002	Smoking, nicotine addiction and interventions for cessation	48	33	14	5
2004	Dyslipidaemias	48	48	4	0
	Emergency care/cardiology				
2005	Treatment of severe sepsis in adults	7	38	44	12
2002	Resuscitation	4	60	32	4
2004	Vein thrombosis and pulmonary embolism	36	36	29	0
2003	Coronary event: unstable angina pectoris and cardiac infarction without ST elevation—risk assessment and treatment	43	38	19	0
2005	Atrial fibrillation	54	32	9	6
2000	Diagnosis of cardiac infarction	18	68	14	0
	Mental health				
2002	Eating disorders in children and adolescents	5	19	54	23
2001	Schizophrenia	18	14	46	23
2004	Depression	48	29	21	2
	Respiratory diseases				
2000	Diagnosis and treatment of asthma	14	49	35	2
2003	Chronic obstructive pulmonary disease	27	46	26	0
	Cancer				
2005	Skin melanoma	12	35	50	3
2002	Diagnosis and follow-up of breast cancer	17	42	42	0
2001	Ovarian cancer	22	42	28	9
2002	Oral cancer	19	48	29	5
2003	Prostate cancer	11	61	25	2
2001	Lung cancer	39	36	26	0
2002	Treatment of breast cancer	42	39	18	0
	Digestive system disorders				
2003	Safe use of anti-inflammatory analgesics	21	32	40	8
2005	Coeliac disease	30	23	47	0
2001	Endoscopic examinations of the colon	9	47	38	6
2005	Treatment of Crohn disease	26	33	37	5
2001	Endoscopic retrograde cholangiopancreatography	6	56	39	0
2001	Endoscopic examination of the oesophagus, stomach and duodenum (gastroscopy)	19	44	28	8
	Infections				
2002	Bacterial skin infections	6	11	44	39
2000	Urinary tract infections	0	33	67	0
1999	Pharyngitis	28	10	48	14
2000	Acute bronchial infection	0	36	38	25
2004	Sinusitis	17	42	33	8
2005	Fungal infections of the skin, hair and nails: sampling, diagnosis and response praxis	10	50	35	5
2002	Diagnosis and treatment of <i>Helicobacter pylori</i> infection	14	47	31	7
2004	Acute otitis media	22	44	16	18
	Reproductive health				
2005	Postcoital contraception	10	20	50	20
2004	Extrauterine pregnancy	15	20	50	15
2001	Induced abortion	26	22	26	26
2005	Evaluation and treatment of menorrhagia	48	23	19	10
2000	Corticosteroid treatment in patients at risk of premature labour	43	40	7	11

Year of publication (latest update)	Title	Evidence grade (%)			
		A	B	C	D
	Children				
2005	Obesity in children	10	17	59	14
2004	Food allergy in children	4	35	57	4
2003	Headache in children	0	50	50	0
	Miscellaneous				
2001	Investigation of child sexual abuse	0	36	64	0
2002	Desensitisation	10	30	47	13
2004	Appropriate treatment of medical problems associated with Down syndrome	19	29	36	17

and level of evidence (A–D). We also retrieved the evidence summaries linked to every Current Care guideline and classified these according to the sections of the guideline to which they were linked. Here we report the actual numbers of evidence summaries within these classifications and the percentages. We correlated the number of evidence summaries with the length (in pages) of the Current Care guideline.

RESULTS

There were 2419 evidence summaries for the 64 guidelines. There were 532 level A, 891 level B, 808 level C and 188 level D evidence summaries. The distribution of the evidence summaries (A–D) of all the guidelines is shown in fig 1. The sections on treatment and diagnosis had the most evidence (all levels from A to D). The section on treatment contained 58.2% of all evidence and diagnosis had 22.4% (fig 2). Level A and B evidence represented 58.8% of all evidence, with the sections on diagnosis and treatment containing 80% of all level A and level B evidence. Figure 3 provides a breakdown of all the level A and level B evidence according to the sections. The dyslipidaemias guideline had the greatest percentage of level A and B evidence ($n = 23$; 95.7%). Figure 4 depicts the distribution of the evidence summaries graded C or D. The sections on diagnosis and treatment represented 81% of all level C and D evidence.

The number of printed pages (PDF layout, without references) of the guideline varied from four pages (corticosteroid treatment in patients at risk of premature labour) to 25 pages (atrial fibrillation), with a mean of 12.1 pages. The number of evidence summaries in a guideline ranged from eight in six pages (acute bronchitis) to 96 in 19 pages (rheumatoid arthritis). Evidence summaries are published only in the electronic format. The spinal cord injury guideline had a total of 67 evidence summaries. This guideline had the greatest percentage of level C and D evidence ($n = 55$). The guideline on schizophrenia had the greatest proportion of level D evidence ($n = 11$; 23%).

DISCUSSION

The main result of the present study is that, strictly speaking, only 22% of the key statements of the 64 guidelines were supported by high-grade evidence (level A). This is probably an overestimate, since there seems to be a relative lack of level D evidence summaries. The sections on diagnosis (30%) and treatment (51%) had the greatest proportion of level C and D evidence. Current Care guidelines are meant to provide recommendations on the diagnosis and treatment of a condition, however, areas such as pathophysiology, rehabilitation and prevention are also included. The end result is a comprehensive guideline on a clinical condition with recommendations supported by various levels of evidence. That the level of evidence is “only” C or D does not indicate that the recommendation is clinically less important. On the contrary, important clinical decisions have to be made despite the level of evidence.

The use of only one set of guidelines is both a strength and a limitation of the present study. The methodology and especially grading of the evidence are based on the same handbook. Some updating has been done, but the basic rules remain the same as at the beginning of the programme. The contents of the guidelines are similar, with a basic set of sections that form the core. On the other hand, although English translations of the most recent guideline abstracts have been available since 2004, the guidelines are only available in Finnish. It is therefore difficult for guideline developers in other countries to evaluate them.

It seems that the Finnish guideline groups have a preference for level A and B evidence summaries. They probably find it easiest to make recommendations that are based on evidence levels A and B. Burgers¹¹ recently stated that high-quality guidelines are based on evidence as well as a broad consensus of opinions, which facilitates the acceptance and effective use of the guideline by the target group. There is, in particular, probably a relative lack of level D summaries, since these should consist of an outline of the consensus reached by the development group on a topic with clearly little evidence. In the

Table 2 Grading of the evidence in the Current Care guidelines

Level	Description
A	Strong research-based evidence (multiple, relevant, high-quality studies with homogeneous results—eg, two or more randomised controlled trials or a systematic review with clearly positive results)
B	Moderate evidence (eg, one randomised controlled trial, or multiple adequate studies)
C	Limited research-based evidence (eg, controlled prospective studies)
D	No evidence (eg, retrospective studies, or the consensus reached in the absence of good quality evidence)

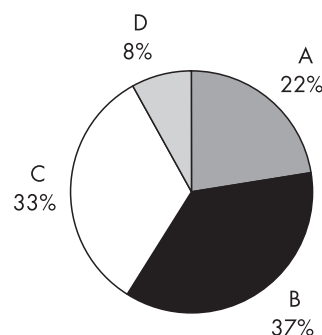


Figure 1 Percentages of the evidence summaries in the 64 Current Care guidelines by evidence levels A–D.

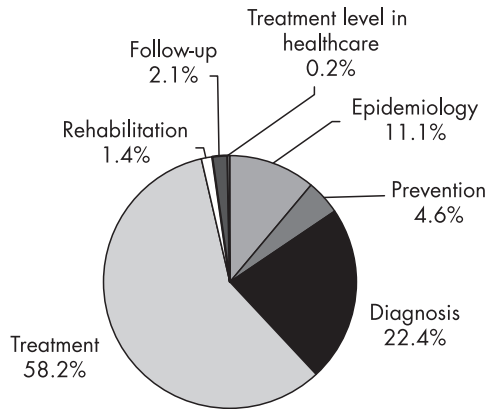


Figure 2 Distribution of the evidence summaries (n=2419) according to the guideline sections.

earlier guidelines, level D key statements were just indicated by the letter D without an accompanying evidence summary, but this practice has since changed. Now the level of evidence has to be supported by an evidence summary. There are probably some key statements or recommendations that will need to be supported by evidence level D summaries, and therefore the proportion of level D evidence is underestimated in the present overview. The Current Care handbook states that only the most central recommendations in the guideline should be supported by evidence summaries. Therefore level D evidence might easily be supported by just giving the reference—for example, an overview.

Another source of bias may be that because the Current Care guidelines are ideologically evidence based, the groups may be tending to draw up summaries that are graded at least C. For this study, we did not analyse the key recommendations that should be backed by an evidence summary. It is an interesting point whether the evidence summaries, especially evidence level D (and C) in clinical practice guidelines could be used as an important source of research questions. Since the guidelines are developed by clinicians, these questions may have direct clinical relevance, and this reasonable notion will shortly be explored in our guideline material.

We did not systematically analyse how often an evidence summary was referred to in the guidelines. However, on the basis of preliminary scanning, this seems exceptional. One of the basic rules is that the evidence summary should only answer one question. According to our experience, the groups abide by this rather well.

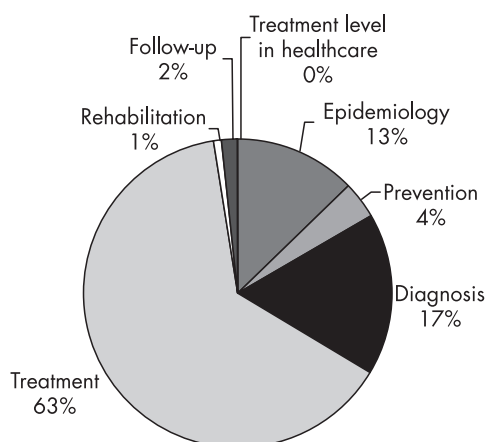


Figure 3 Distribution of the level A and B evidence summaries (n=1423) according to the sections in the guidelines.

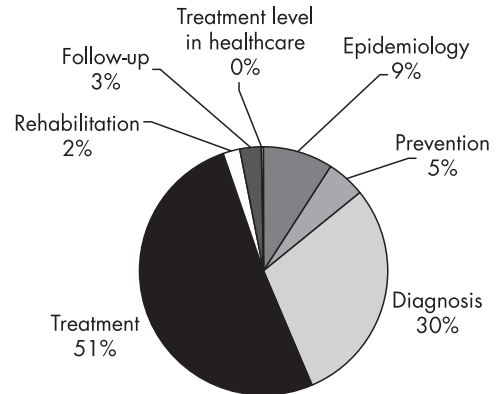


Figure 4 Distribution of the level C and D evidence summaries (n=1089) according to the sections in the guidelines.

Use of guidelines may measure one component of the organisation's maturity (Maturity Matrix¹²). The other components of the Maturity Matrix are clinical records, audit of clinical performance, clinician access to clinical information, prescribing, practice-based organisational meetings, sharing

Box 1: Outline of the development process of a Current Care guideline

1. A topic is suggested (most commonly by a specialist medical society)
2. The topic is selected (by the Current Care board)
3. The working group assembles (chairperson, editor, other members)
4. Critical appraisal training is arranged for the group
5. Systematic search of the literature is carried out (by an experienced medical librarian)
6. Evidence summaries drawn up and then the key clinical statements based on the available research (by the clinical experts—the main body of work)
7. Guideline is written based on the evidence summaries
8. Draft guideline is circulated for critical comments
9. Guideline is published:
 - Internet (<http://www.kaypahoito.fi>)
 - Evidence Based Medicine Guideline (Finnish version)
 - Also available via a widely used health portal (<http://www.terveysportti.fi>)
 - *Medical Journal Duodecim*
 - An interview and summary-based version for the lay public in the number one health magazine in Finland (*Hyvä Terveys*)
 - In addition, other health professionals outline the guidelines from their perspective in their respective journals
10. Updating
 - Minor updates as substantial evidence accumulates
 - Major updates at set intervals of about 3 years
11. Publication in electronic format and a summary of substantial changes in the *Medical Journal Duodecim*. The abstract of a new guideline is translated into English and published on the website.

information with patients and patient feedback systems. The grading of maturity increases as the level of competence increases. The Current Care guidelines are well distributed and disseminated to all professionals and practically all healthcare organisations via a professional health portal (Terveysportti) and they are also freely available on the internet (<http://www.kaypahoito.fi>). Thus the guidelines are incorporated into clinical information systems, which underline the importance of the quality of the guidelines. Care pathways and the 2005 implemented national criteria for referral to elective care (<http://www.stm.fi>) are based on guidelines whenever possible, so the evidence is integrated in the core healthcare.

The quality of evidence is relevant to guideline implementation. According to Dutch researchers, compliance with guidelines is better if the evidence base is solid.¹³ The implementation of a guideline is also facilitated by the quality of the guideline itself—that is, its readability and directness.¹⁴ One cornerstone of good-quality guidelines is supporting the most central recommendations with evidence summaries. These also serve as a message from the guideline development group to the audience, highlighting the importance of specific areas of the guideline, which therefore are supported by the evidence summaries. The aim of using the evidence and the guidelines is to improve the quality of healthcare. The existing evidence is not always applicable or even directly relevant to clinical practice. Pertaining to guidelines, a tool has been developed to aid the Current Care board in making decisions about new guideline topics, to better ensure their clinical relevance and importance to the national healthcare system.⁹ Another project has been launched with the aim of developing an interactive electronic decision support (EDS) based on the guidelines, which, along with the electronic patient record (EPR), will support clinical decision making.¹⁵ There are still many obstacles, such as motivating clinicians to use the EPR actively in a structured way and resolution of data protection issues. However, in the long run, when the electronic guidelines, the EDS and the EPR, are in place, it will become normal practice to collect and analyse clinical outcomes data online. This system may also enable us to improve the relevance of our research questions.

We have used the available Current Care guideline material to describe the available proportions of the evidence levels A–D for guideline development. **The present results confirm the previous estimates that about 10–20% of clinical decisions are based on good evidence.** The next step will be to analyse more carefully the grade C and D evidence summaries and whether these can be used to highlight the gaps in our evidence base, and thereby guide the questions for researchers to answer next.

AUTHORS' CONTRIBUTIONS

All authors contributed to the study design and interpretation of the data, carried out the analyses and prepared the draft manuscript. All

authors participated in revising the manuscript for important intellectual content and approved the version to be published.

Authors' affiliations

Eeva Ketola, Finnish Medical Society Duodecim, Helsinki, Finland
Minna Kaila, Department of Pediatrics, Tampere University Hospital, and Finnish Office for Health Technology Assessment at STAKES, Finland
Mari Honkanen, Current Care, Finnish Medical Society Duodecim, Helsinki, Finland

No ethics approval was required for this study.

Competing interests: None.

EK is the Editor-in Chief and MH is Editor of Current Care Guidelines.

REFERENCES

- 1 **Shojania KG**, McDonald KM, Wachter RM, *et al.* *Closing the quality gap: a critical analysis of quality improvement strategies, volume 1—series overview and methodology.* AHRQ Publication No. 04-0051-1. Rockville, MD: Agency for Healthcare Research and Quality, August 2004. <http://www.ahrq.gov/downloads/pub/evidence/pdf/qualgap1/qualgap1.pdf> (accessed 3 June 2006).
- 2 **Staniszewska SH**, Henderson L. Patients' evaluations of the quality of care: influencing factors and the importance of engagement. *J Adv Nurs* 2005;**49**:530–7.
- 3 **Muir Gray JA. *Evidence-based health care: how to make health policy and management decisions.* Edinburgh: Churchill Livingstone, 2001.**
- 4 **Ollenschläger G**, Marshall C, Qureshi S, *et al.* Improving the quality of health care: using international collaboration to inform guideline programmes by founding the Guidelines International Network (G-I-N). *Qual Saf Health Care* 2004;**13**:455–60.
- 5 **Atkins D**, Eccles M, Flottorp S, GRADE Working Group, *et al.* Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. The GRADE Working Group. *BMC Health Serv Res* 2004;**4**:38.
- 6 **Millenson MM. *Beyond the managed care backlash. Medicine in the information age.* PPI Policy Report No 1. Washington, DC: Progressive Policy Institute, 1997.**
- 7 **Winell K**, Kaila M, Mäkelä M. Finnish Current Care guidelines now target tobacco cessation. *Suom Lääkäril* 2003;**58**:2983–4.
- 8 **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.
- 9 **Ketola E**, Toropainen E, Kaila M, *et al.* Prioritising guideline topics—development and evaluation of a practical tool. *J Eval Clin Pract* 2006 (in press).
- 10 Guyatt G, Drummond R, eds. *User's guide to the medical literature. Essentials of evidence-based clinical practice.* USA: AMA Press, 2002.
- 11 **Burgers JS.** Guideline quality and content: are they related? *Clin Chem* 2006;**52**:3–4.
- 12 **Elwyn G**, Rhydderch M, Edwards A, *et al.* Assessing organisational development in primary medical care using group assessment: the Maturity Matrix. *Qual Saf Health Care* 2004;**13**:287–94.
- 13 **Grol R**, Dalhuijsen J, Thomas S, *et al.* Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998;**317**:858–61.
- 14 **Michie S**, Lester K. Words matter: increasing the implementation of clinical guidelines. *Qual Saf Health Care* 2005;**14**:367–70.
- 15 **Komulainen J**, Kunnamo I, Nyberg P, *et al.* Developing an evidence-based medicine decision support system integrated with EPRs utilizing standard data elements. In: Ten Teije A, Miksch S, Lucas P, eds. *Proceedings of the workshop AI techniques in healthcare: evidence-based guidelines and protocols.* 28 August – 1 September 2006, Riva del Garda, Italy.