BMJ open

National Guidelines – How evidence based are they?

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-004278
Article Type:	Research
Date Submitted by the Author:	17-Oct-2013
Complete List of Authors:	Baird, Alastair; The White House, Lawrence, James; Dumfries & Galloway Royal Infirmary, Research & Development Support Unit
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	General practice / Family practice
Keywords:	HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Clinical governance < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

National Guidelines – How evidence based are they?

Dr A Gordon Baird, MB ChB MRCOG FRCGP R&D General Practice Adviser, Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries DG1 4AP.

Dr James R Lawrence, BSc MD FRCP(Glas & Edin) R&D Clinical Lead, Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries DG1 4AP.

ABSTRACT

Background
National guidelines have become an integral part of UK general practice and aim to deliver quality and consistency within practices. As with any intervention, there are negative as well as positive consequences. Guideline effectiveness depends on the quality of evidence used.

Aim
To quantify and analyse the quality of evidence that is presented in national guidelines.

Design and setting
Levels of evidence used in all the current valid recommendations in the Scottish Intercollegiate Guideline Network (SIGN) guidelines were reviewed and statistically analysed.

Method
The data was collected from published guidelines available online to the public. A professional group selected by a national organisation develops each of these guidelines. Statistical analysis of the relationship between the number of guideline recommendations and the quality of evidence used in its recommendations was performed.

Result
A significant correlation between the number of recommendations in a guideline and the use of level D evidence was discovered.

Conclusion
Practice guidelines should be brief and based on scientific evidence. Paradoxically the longest guidelines have the highest proportion of recommendations based on the lowest level of evidence. Guideline developers should be more aware of the need for brevity and a stricter application of evidence-based principles could achieve this. The findings support calls for a review of how evidence is used and presented in guidelines.

Dr A Gordon Baird affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; any discrepancies from the study as planned have been explained.

The data is in the public domain; there were no ethical conflicts; there was no funding.

Article Summary

Article focus

Examines the hypothesis that larger guidelines rely disproportionately on poor evidence.

Key Messages

- One third of current national guidelines is supported only by case reports, case studies and expert opinion.
- Guidelines with large numbers of recommendations are more likely to use weak evidence.
- Guideline development groups appear to vary in their approach to offering recommendations.
- Guideline recommendations should be based on good evidence.
- Without good science scientific evidence, avoiding a recommendation or highlighting the need for research should be considered.

Strengths and limitations of the study

Strengths.

- This is the first objective evidence of inconsistencies in approach by national guideline developers
- This supports commentator suggestion that even without good evidence a group will prefer consensus.
- Adds to the current debate about how guidelines might be developed in the future

Limitations.

- The study is limited to only one set of national guidelines.
- Reasons for the differences in quality of evidence preferred by the guideline development groups is unclear

SIGN Guidelines – How evidence based are they?

INTRODUCTION

The Scottish Intercollegiate Guidelines Network (SIGN) was founded in 1993. It is a national body, professionally led and publicly funded. SIGN's founding principles proposed direct links between evidence and recommendations, offering a brief and succinct quick-reference guide for clinicians[1]. Guidelines anticipated presenting brief, evidence based clinical advice. They have developed into long and authoritative texts that are used by managers and politicians to inform policy. SIGN has no responsibility to consider cost-effectiveness and no direct input into the Quality Outcomes Framework (QOF). A formal arrangement between SIGN and the National Institute of Clinical Excellence (NICE) has existed from 2003.

The World Health Organisation (WHO) recognises that current recommendation categories may be ambiguous [2] and has encouraged guideline developers to use a different form of grading, including a category "Use only in the context of research" where doubt exists.

Guideline developers have conflict of interest policies reported as challenging to apply. Where doubt exists, groups of specialists may feel consensus more defensible than acknowledging uncertainty. [3]

Even with the best evidence, concerns are expressed about the relevance of guidelines in treating patients with multiple morbidities[4], and the emergence of the phenomenon of reversal[5,6], where established practice, sometimes evidence based, is shown to be sub-optimal or harmful. This study looks at the quality of evidence used for SIGN guidelines, and describes a significant trend for some groups to emphasise poorly evidence-backed recommendations.

METHODS

SIGN guidelines were accessed online. Guidelines that were “Current” and “Current 3-7 years. Some recommendations may be out of date.” in September 2013 were included. Those that had been "Withdrawn", "Recommendations being updated", "Need for update being considered" and those with no recommendations were excluded. Key recommendations and implementation recommendations were excluded.

SIGN guideline 50 describes an established process for developing guidelines.[7] This process is independent of this study, but is stated to be an objective process. SIGN guidelines have four grades of recommendation outlined in table 1. Table 2 describes the level of evidence supporting the recommendation grading.

The level of evidence used by each guideline was independently recorded by 3 investigators and errors resolved. A statistical analysis of the correlation between the proportion of level D evidence and the total number of recommendations was performed for the 42 guidelines.

RESULTS

The 42 guidelines consisted of 2559 pages, ranging from 26 to 161 (median 59.5) pages. The longest guideline, number 116 was 61 pages longer than the next largest. The number of recommendations per page ranged from 0.2 to 1.8 (median 0.7). The number of recommendations per guideline is presented in table 3.

Of the 1999 recommendations, 480 (24.0%) were level A, 491(24.6 %) were level B, 318 (15.9%) level C, and 710 (35.5%) level D. Thus 40.2% were poorly evidenced (C&D) and over a third (D) developed almost entirely on “expert opinion”. The number of level A recommendations per guideline ranged from 0-57 (median 9), level B from 2-62 (median 8.5) level C ranged from 0-26 (median 6) and D from 0-60 (median 14.5). 4 guidelines had no level A evidence.

The proportion of level D evidence increases with the number of recommendations made. This correlation is significant with Kendall's Tau=0.22 [approximate 95% confidence interval 0.008-0.45] p value =0.04, and Spearman rho=0.22 [approximate 95% CI 0.02-0.57] p value= .04.

DISCUSSION

This study reveals that expert groups who produce long guidelines rely on poor evidence more heavily than others. National guidelines are useful and important and there is a debate about how evidence is best presented. Guidleines define standards of care, help busy clinicians and allow managers and politicians to develop governance. An American study (using 3 not 4 levels of evidence) similarly found that 48% were "based on expert opinion, case studies, or standards of care." [8] Where patients are involved in clinical decisions, honestly declaring uncertainty has merit. In the absence of good scientific evidence, recommending a course of action without understanding the circumstances of the individual to whom it is applied seems both risky and, assuming patient choice, imbalanced.

This study did not examine why longer guidelines use poorer evidence. It has been postulated that there is security in "just doing what everyone else is doing – even if what everyone else is doing isn't very good."[3] Cloistered groups of experts may view their own opinion as more authoritative than science can support. Reliance on expert opinion has a poor track record. Blinded by certainty, expert groups defining established practice have perpetuated radical mastectomy instead of conservative surgery, Class 1C antiarrhythmics[9], pulmonary artery catheters in heart failure[10], electronic foetal monitoring in low risk pregnancies: even then practice can take a decade to reverse[11].

Even good evidence is subject to the phenomenon of reversal where new evidence contradicts current practice. Reversal can affect around 13-16% of publications [5,6]. This may partly explain why the implementation of even the most soundly evidence based national guidelines fails to improve outcome [12],[13],[14]. There is potential

harm[15,16] from guidelines in real clinical settings, for example increasing radiation dose without benefit[17] or increased risks of anticoagulation[18].

SIGN 116 (diabetes), is a notable exception. It is more than 50% larger than the next largest, 2.5 times longer than the average and yet uses the 4th lowest level D recommendations. There are a number of hypotheses why this group reports differently. The advisory committee to SIGN's English equivalent (NICE) informs Quality Outcomes Framework (QOF) policy. Diabetes is the largest clinical UK QOF indicator and is associated with substantial payment incentives. The need for objective evaluation of performance drives a use of surrogate outcomes without appropriate clinical endpoints. [19] Diabetes guidelines have suffered several noteworthy reversals. Examples include the recommendation of glycosylated haemoglobin reduction resulting in increased use of rosiglitazone (still mentioned in the current document) both associated with harm[20]including mortality.[21] Aspirin recommendations have also been changed from previous guidelines. Is it possible that the repeated use of surrogate outcomes arises from group dynamics driven by a powerful external agenda?

Many doctors whose expertise cross several guidelines[22,23], express concerns about guideline development groups. The inappropriate exclusion of disease groups from general population data is common. Smoking cessation advice for testicular cancer survivors is level C, although studies in the general population (without excepting specific disease groups) advises everyone to stop smoking. Overall smoking cessation was level D and C once each and B on three occasions. Using evidence in this way may imply group dysfunction. Differently constituted groups, or greater oversight might avoid problems.

In 1993, SIGN guidelines stated intention was to be evidence based, brief and succinct. Brevity increases value as a quick reference guide. Removing or reducing poorly evidenced recommendations would reduce size by more than a third overall and in some up to two thirds. 2559 pages is longer than the Oxford Textbook of Primary Medical Care. Evidence based medicine is described as "the use of

mathematical estimates of the risk of benefit and harm, derived from high-quality research on population samples, to inform clinical decision-making in the diagnosis, investigation or management of individual patients"[24]. Guidelines relevance to daily practice, the reliability of evidence and whether the application of evidence will improve outcomes are important questions.

These results may reflect how professional groups deal with uncertainty. If so, this is not good for individual patients faced with the same uncertainties (whether aware of it or not), nor is it good for scientists who actively seek unanswered questions by challenging established practice, an area in which medicine has a poor record from Semmelweis to the present day.

The finding of a significant increase of level D recommendations in larger guidelines has not happened by chance. A wider debate about how guideline groups can create greater clarity about the reliability of evidence used is needed.[25] Reducing the use of poorly evidenced recommendations has potential to create a shorter, more reliable and usable clinical support.

Acknowledgements.

Heather Barrington, Statistical Adviser; Bridget Bird, Administrative Assistant; Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries DG1 4AP.

Conflict of interest; none

Table 1

GRADES OF RECOMMENDATION

A	At least one meta-analysis, systematic review, or RCT rated as 1++, and directly applicable to the target population; or A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results
B	A body of evidence including studies rated as 2++, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 1++ or 1+
C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 2++
D	Evidence level 3 or 4; or Extrapolated evidence from studies rated as 2+

Table 2

1++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1 -	-Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort studies
2+	High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2 -	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, eg case reports, case series
4	Expert opinion

Table 3

Main Recommendations

Number	Name	Pages	A	B	C	D	*Total*	%age D
133	Management of hepatitis C	57	20	24	7	52	***103***	50.5%
132	Long-term follow-up of survivors of childhood cancer	62	0	7	9	14	***30***	46.7%
131	Management of schizophrenia	64	10	19	3	15	***47***	31.9%
130	Brain injury rehabilitation in adults	68	0	14	7	8	***29***	27.6%
129	Antithrombotics: indication and management	68	25	11	6	19	***61***	31.1%
127	Management of perinatal mood disorders	47	0	5	6	15	***26***	57.7%
126	Diagnosis and management of colorectal cancer	56	11	19	15	29	***74***	39.2%
125	Management of atopic eczema in primary care	34	3	5	3	2	***13***	15.4%
124	Management of adult testicular germ cell tumours	63	6	6	9	21	***42***	50.0%
123	Management of early rheumatoid arthritis	27	3	7	2	0	***12***	0.0%
122	Prevention and management of venous thromboembolism	88	26	15	14	55	***110***	50.0%
121	Diagnosis and management of psoriasis and psoriatic arthritis in adults	65	11	16	6	26	***59***	44.1%
120	Management of chronic venous leg ulcers	46	5	3	4	7	***19***	36.8%
119	Management of patients with stroke: identification and management of dysphagia	42	0	6	4	20	***30***	66.7%
118	Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning	101	21	29	7	21	***78***	26.9%
117	Management of sore throat and indications for tonsillectomy	37	9	3	4	4	***20***	20.0%
116	Management of diabetes	161	57	62	23	16	***158***	10.1%
115	Management of Obesity	87	6	11	7	11	***35***	31.4%
114	Nonpharmaceutical management of depression	37	5	4	0	0	***9***	0.0%
113	Diagnosis and pharmacological management of Parkinson's disease	61	12	6	6	4	***28***	14.3%
112	Management of attention deficit and hyperkinetic disorders in children and young people	45	6	4	3	4	***17***	23.5%
111	Management of hip fracture in old people	49	10	9	8	14	***41***	34.1%
110	Early management of patients with a head injury	76	1	7	6	17	***31***	54.8%
109	Management of genital Chlamydia trachomatis infection	40	3	6	9	29	***47***	61.7%
108	Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention	100	42	27	18	14	***101***	13.9%
107	Diagnosis and management of headache in adults	81	17	16	9	34	***76***	44.7%
106	Control of pain in adults with cancer	71	5	7	3	19	***34***	55.9%
105	Management of acute upper and lower gastrointestinal bleeding	57	14	5	2	15	***36***	41.7%
103	Diagnosis and management of chronic kidney disease	50	9	6	4	3	***22***	13.6%
102	Management of invasive meningococcal disease in children and young people	46	1	4	6	26	***37***	70.3%
99	Management of cervical cancer	73	1	13	19	29	***62***	46.8%
97	Risk estimation and the prevention of cardiovascular disease	72	16	12	2	4	***34***	11.8%
96	Management of stable angina	59	13	10	3	11	***37***	29.7%
95	Management of chronic heart failure	55	9	12	1	1	***23***	4.3%
94	Cardiac arrhythmias and coronary heart disease	42	22	11	13	23	***69***	33.3%
93	Acute coronary syndromes	60	11	14	9	8	***42***	19.0%
91	Bronchiolitis in children	42	4	3	6	14	***27***	51.9%
90	Diagnosis and management of head and neck cancer	92	42	8	26	60	***136***	44.1%
89	Diagnosis and management of peripheral arterial disease	37	11	2	0	4	***17***	23.5%
88	Management of suspected bacterial urinary tract infection in adults	45	8	10	2	10	***30***	33.3%
87	Management of oesophageal and gastric cancer	70	3	26	23	28	***80***	35.0%
61	Investigation of postmenopausal bleeding	26	2	7	4	4	***17***	23.5%
	TOTAL	2559	480	491	318	710	1999	

Contributorship Statement

Both authors discussed the hypothesis that there was a disproportionate use of poor evidence in longer guidelines; both checked the raw data, and agreed on a statistical approach to discover whether the trend was significant or not. Both have been involved in writing and researching the evidence.

Funding.

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data Sharing.

Technical appendix, statistical code, and dataset available from the corresponding author, who will provide a permanent, citable and open access home for the dataset.

Competing interests

No competing interests

MeSH Headings

- Practice Guideline N04.761.700.350.650
- General Practice H02.403.340
- Evidence-Based Medicine H02.249.750

How this fits in.

Guidelines should encourage an evidence-based approach to clinical practice. Longer guidelines used significantly higher levels of poor evidence. WHO has proposed a different system of grading evidence. The effect of group behavior altering guideline development has been hypothesised. New research often challenges established clinical practice. Improving the quality of evidence, acknowledging uncertainty and shortening guideline length would make guidelines more clinically relevant and effective.

REFERENCES

1 Harbour R, Lowe G, Twaddle S. Scottish Intercollegiate Guidelines Network; the first 15 years (1993-2008). J R Coll Physicians Edinb 2011 Jun;41(2) 163-168

2 de Joncheere K, Hill S, Klazinga N, Mäkelä M, Oxman AD.The Clinical Guideline Programme of the National Institute for Health and Clinical Excellence (NICE) A review by the World Health Organization May 2006

3 Lenzer J. Why we can't trust clinical guidelines: BMJ 2013;346:f3830

4 Aylett V. Do geriatricians need guidelines? BMJ 2010 Sep;:341 :c5340

5 Prasad V, Cifu, A, Ioannidis JP. Reversals of Established Medical Practices: Evidence to Abandon Ship. JAMA. 2012 Jan4;307(1):37-38,

6 Prasad V, Gall V, Cifu A. The frequency of medical reversal. Arch Int Med. 2011 Oct 10;. 171(18): 1675-1676

7 http://www.sign.ac.uk/pdf/sign50.pdf

8 Tricoci P, Allen JM, Kramer JM, Califf RM, Smith Jr. SC. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA 2009 Feb25; 301 (8): 831-841

9 Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, Arensberg D, Baker A, Friedman L, Greene HL, Huther ML, Richardson DW. Mortality and morbidity in patients receiving encainide, flecainide, or placebo - The Cardiac Arrhythmia Suppression Trial. New Eng J Med. 1991 Mar21; 324 (12): 781-788

10 Binanay C, Califf RM, Hasselblad V, O'Connor CM, Shah MR, Sopko G, Stevenson LW, Francis GS, Leier CV, Miller LW Evaluation study of congestive heart failure and pulmonary artery catheterization effectiveness: the ESCAPE trial. JAMA. 2005 Oct 5;294(13):1625-33.

11 Tatsioni A, Bonitsis NG, Ioannidis JA. Persistence of Contradicted Claims in the Literature. JAMA. 2007 Dec5 ;298(21):2517-2526.

12 Anderson JE, McKenzie C, Singh N, Gajree N, Giles L, Sharma P, McDonald A. Compliance with the 62 day target does not improve long-term survival. Association of Coloproctology of Great Britain and Ireland Annual Meeting 2012 Dublin Ireland.

13 Kerr J, Smith R, Gray S, Beard D, Robertson CE. An audit of clinical practice in the management of head injured patients following the introduction of the Scottish Intercollegiate Guidelines Network (SIGN) recommendations. Emerg Med J.2005 Dec; 22(12):850-854

14 Caplan LR. How well does "evidence-based" medicine help neurologists care for individual patients? Revin Neurol Dis.2007;4(2):75-84

15 Hutchison G. Guidelines can harm patients too. BMJ (Clinical research ed.). 2012 Apr18;344: e2685.

16 Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines. Potential benefits, limitations, and harms of clinical guidelines. BMJ 1999 Feb20;318(7182):527-530

17 Miller L., Kidd A. Are sign guidelines key in the decision to CT? Post-head injury CT scanning within a paediatric population. Academic Emergency Medicine. Conference: 14th International Conference on Emergency Medicine, ICEM 2012 Dublin Ireland. Conference Publication: 19 (6) (pp 776), 2012. Date of Publication: June 2012.

18 Thomson R., Eccles M., Wood R., Chinn D.J. A cautionary note on data sources for evidence-based clinical decisions: Warfarin and stroke prevention. MedDecis Making. 2007;27 (4): 438-447

19 Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. BMJ. 2011 Dec28;343:d7995

20 Cohen D. Rosiglitazone: what went wrong? BMJ.2010 Sep; 341:c4848

21 Gerstein HC, Miller ME, Byington RP, Goff DC Jr., Bigger JT, Buse JB, Cushman WC, Genuth S, Ismail-Beigi F, Grimm RH Jr. Probstfield JL, Simons-Morton DG, Friedewald WT. Effects of intensive glucose lowering in type 2 diabetes. Action to Control Cardiovascular Risk in Diabetes Study Group. New Engl J Med. 2008 Jun12;358(24):2545-59

22 Rashidian A, Eccles MP, Russell I. Falling on stony ground? A qualitative study of implementation of clinical guidelines' prescribing recommendations in primary care. Health Policy.2008 Feb; 85 (2) 148-161

23 Guy S, Wardlaw JM. Who writes guidelines, and who should? ClinRadiol. 2002 Oct ;57 (10): 891-897.

24 Greenhalgh, Trisha. How To Read a Paper: The Basics of Evidence-Based Medicine. Wiley-Blackwell, fourth edition, 2010, p. 1

25 Zuiderent-Jerak T, Forland F, Macbeth F. Guidelines should reflect all knowledge, not just trials
BMJ 2012;345:e6702

TABLE 2

			Main Recommend		
Number	Name	Pages	A	B	
133	Management of hepatitis C	57	20	24	
132	Long-term follow-up of survivors of childhood cancer	62	0	7	
131	Management of schizophrenia	64	10	19	
130	Brain injury rehabilitation in adults	68	0	14	
129	Antithrombotics: indication and management	68	25	11	
127	Management of perinatal mood disorders	47	0	5	
126	Diagnosis and management of colorectal cancer	56	11	19	
125	Management of atopic eczema in primary care	34	3	5	
124	Management of adult testicular germ cell tumours	63	6	6	
123	Management of early rheumatoid arthritis	27	3	7	
122	Prevention and management of venous thromboembolism	88	26	15	
121	Diagnosis and management of psoriasis and psoriatic arthritis in adults	65	11	16	
120	Management of chronic venous leg ulcers	46	5	3	
119	Management of patients with stroke: identification and management of dysphagia	42	0	6	
118	Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning	101	21	29	
117	Management of sore throat and indications for tonsillectomy	37	9	3	
116	Management of diabetes	161	57	62	
115	Management of Obesity	87	6	11	
114	Nonpharmaceutical management of depression	37	5	4	
113	Diagnosis and pharmacological management of Parkinson's disease	61	12	6	
112	Management of attention deficit and hyperkinetic disorders in children and young people	45	6	4	
111	Management of hip fracture in old people	49	10	9	
110	Early management of patients with a head injury	76	1	7	
109	Management of genital Chlamydia trachomatis infection	40	3	6	
108	Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention	100	42	27	
107	Diagnosis and management of headache in adults	81	17	16	
106	Control of pain in adults with cancer	71	5	7	
105	Management of acute upper and lower gastrointestinal bleeding	57	14	5	
103	Diagnosis and management of chronic kidney disease	50	9	6	
102	Management of invasive meningococcal disease in children and young people	46	1	4	
99	Management of cervical cancer	73	1	13	
97	Risk estimation and the prevention of cardiovascular disease	72	16	12	
96	Management of stable angina	59	13	10	

95	Management of chronic heart failure	55	9	12	
94	Cardiac arrhythmias and coronary heart disease	42	22	11	
93	Acute coronary syndromes	60	11	14	
91	Bronchiolitis in children	42	4	3	
90	Diagnosis and management of head and neck cancer	92	42	8	
89	Diagnosis and management of peripheral arterial disease	37	11	2	
88	Management of suspected bacterial urinary tract infection in adults	45	8	10	
87	Management of oesophageal and gastric cancer	70	3	26	
61	Investigation of postmenopausal bleeding	26	2	7	
	TOTAL	2559	480	491	

TABLE 1.

GRADES OF RECOMMENDATION

A	At least one meta-analysis, systematic review, or RCT rated as 1++, and directly applicable to the target population; or A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results
B	A body of evidence including studies rated as 2++, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 1++ or 1+
C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 2++
D	Evidence level 3 or 4; or Extrapolated evidence from studies rated as 2+

TABLE 2

Levels of Evidence

1++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1 -	-Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort studies
2+	High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2 -	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, eg case reports, case series
4	Expert opinion

TABLE 3
Results -Number of recommendations in each category.

				Main Recommendations					
Number	Number	Name	Pages	A	B	C	D	*Total*	%age D
133	133	Management of hepatitis C	57	20	24	7	52	***103***	50.5%
132	132	Long-term follow-up of survivors of childhood cancer	62	0	7	9	14	***30***	46.7%
131	131	Management of schizophrenia	64	10	19	3	15	***47***	31.9%
130	130	Brain injury rehabilitation in adults	68	0	14	7	8	***29***	27.6%
129	129	Antithrombotics: indication and management	68	25	11	6	19	***61***	31.1%
127	127	Management of perinatal mood disorders	47	0	5	6	15	***26***	57.7%
126	126	Diagnosis and management of colorectal cancer	56	11	19	15	29	***74***	39.2%
125	125	Management of atopic eczema in primary care	34	3	5	3	2	***13***	15.4%
124	124	Management of adult testicular germ cell tumours	63	6	6	9	21	***42***	50.0%
123	123	Management of early rheumatoid arthritis	27	3	7	2	0	***12***	0.0%
122	122	Prevention and management of venous thromboembolism	88	26	15	14	55	***110***	50.0%
121	121	Diagnosis and management of psoriasis and psoriatic arthritis in adults	65	11	16	6	26	***59***	44.1%
120	120	Management of chronic venous leg ulcers	46	5	3	4	7	***19***	36.8%
119	119	Management of patients with stroke: identification and management of dysphagia	42	0	6	4	20	***30***	66.7%
118	118	Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning	101	21	29	7	21	***78***	26.9%
117	117	Management of sore throat and indications for tonsillectomy	37	9	3	4	4	***20***	20.0%
116	116	Management of diabetes	161	57	62	23	16	***158***	10.1%
115	115	Management of Obesity	87	6	11	7	11	***35***	31.4%
114	114	Nonpharmaceutical management of depression	37	5	4	0	0	***9***	0.0%
113	113	Diagnosis and pharmacological management of Parkinson's disease	61	12	6	6	4	***28***	14.3%
112	112	Management of attention deficit and hyperkinetic disorders in children and young people	45	6	4	3	4	***17***	23.5%
111	111	Management of hip fracture in old people	49	10	9	8	14	***41***	34.1%
110	110	Early management of patients with a head injury	76	1	7	6	17	***31***	54.8%
109	109	Management of genital Chlamydia trachomatis infection	40	3	6	9	29	***47***	61.7%
108	108	Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention	100	42	27	18	14	***101***	13.9%
107	107	Diagnosis and management of headache in adults	81	17	16	9	34	***76***	44.7%
106	106	Control of pain in adults with cancer	71	5	7	3	19	***34***	55.9%
105	105	Management of acute upper and lower gastrointestinal bleeding	57	14	5	2	15	***36***	41.7%
103	103	Diagnosis and management of chronic kidney disease	50	9	6	4	3	***22***	13.6%
102	102	Management of invasive meningococcal disease in children and young people	46	1	4	6	26	***37***	70.3%
99	99	Management of cervical cancer	73	1	13	19	29	***62***	46.8%
97	97	Risk estimation and the prevention of cardiovascular disease	72	16	12	2	4	***34***	11.8%
96	96	Management of stable angina	59	13	10	3	11	***37***	29.7%
95	95	Management of chronic heart failure	55	9	12	1	1	***23***	4.3%
94	94	Cardiac arrhythmias and coronary heart disease	42	22	11	13	23	***69***	33.3%
93	93	Acute coronary syndromes	60	11	14	9	8	***42***	19.0%
91	91	Bronchiolitis in children	42	4	3	6	14	***27***	51.9%
90	90	Diagnosis and management of head and neck cancer	92	42	8	26	60	***136***	44.1%
89	89	Diagnosis and management of peripheral arterial disease	37	11	2	0	4	***17***	23.5%
88	88	Management of suspected bacterial urinary tract infection in adults	45	8	10	2	10	***30***	33.3%
87	87	Management of oesophageal and gastric cancer	70	3	26	23	28	***80***	35.0%
61	61	Investigation of postmenopausal bleeding	26	2	7	4	4	***17***	23.5%
			2559	480	491	318	710	1999	

BMJ open

Guidelines – is bigger better? A review of SIGN guidelines.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-004278.R1
Article Type:	Research
Date Submitted by the Author:	04-Dec-2013
Complete List of Authors:	Baird, Alastair; The White House, Lawrence, James; Dumfries & Galloway Royal Infirmary, Research & Development Support Unit
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	General practice / Family practice
Keywords:	HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Clinical governance < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SIGN Guidelines –analysis of evidence levels for their recommendations.

Alternatively

Guidelines – is bigger better? A review of SIGN national guidelines.

Dr A Gordon Baird, MB ChB MRCOG FRCGP R&D General Practice Adviser, Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries, Scotland, UK DG1 4AP.

Address for correspondence; Dr A Gordon Baird, The White House, Sandhead, Wigtownshire UK. DG9 9JA. Email gordon.baird@me.com. Phone/fax 01776830281

Dr James R Lawrence, BSc MD FRCP(Glas & Edin) R&D Clinical Lead, Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries, Scotland, UK DG1 4AP.

MeSH headings
General practice / Family practice
HEALTH SERVICES ADMINISTRATION & MANAGEMENT
Protocols & guidelines
Quality in health care
Clinical governance

ABSTRACT

Background

National guidelines have become an integral part of UK general practice and aim to effectively deliver quality and consistency in clinical practice. As with any intervention, there are negative as well as positive consequences. Guideline effectiveness depends on the quality of evidence used.

Aim

To quantify and analyse the quality of evidence that is presented in national guidelines.

Design and setting

Levels of evidence used in all the current valid recommendations in the Scottish Intercollegiate Guideline Network (SIGN) guidelines were reviewed and statistically analysed.

Method

The data was collected from published guidelines available online to the public. SIGN methodology entails a professional group selected by a national organisation to develop each of these guidelines. Statistical analysis of the relationship between the number of guideline recommendations and the quality of evidence used in its recommendations was performed.

Result

A significant correlation between the number of recommendations in a guideline and the use of level D evidence was discovered.

Conclusion

Practice guidelines should be brief and based on scientific evidence. Paradoxically the longest guidelines have the highest proportion of recommendations based on the lowest level of evidence. Guideline developers should be more aware of the need for brevity and a stricter application of evidence-based principles could achieve this. The findings support calls for a review of how evidence is used and presented in guidelines.

Article Summary

Article focus

Examines the hypothesis that larger guidelines rely disproportionately on poor evidence.

Key Messages

- One third of current national guidelines are supported only by case reports, case studies and expert opinion.
- Guidelines with large numbers of recommendations used a higher proportion of weak evidence.
- Guideline development groups appear to vary in their approach to offering recommendations.
- Guideline recommendations should be based on good evidence.
- Paucity of evidence should highlight topics for research.

Strengths and limitations of the study

Strengths.

- This is the first objective evidence of inconsistencies in approach by a national guideline developers
- This supports commentator suggestion that even without good evidence a group will prefer consensus.
- Adds to the current debate about how guidelines might be developed in the future

Limitations.

- The study is limited to only one set of national guidelines (SIGN).
- Reasons for the differences in quality of evidence preferred by the guideline development groups is unclear

Guidelines – is bigger better? A review of SIGN guidelines.

INTRODUCTION

The Scottish Intercollegiate Guidelines Network (SIGN) was founded in 1993. It is a national body, professionally led and publicly funded. SIGN's founding principles proposed direct links between evidence and recommendations, offering a brief and succinct quick-reference guide for clinicians[1]. Guidelines anticipated presenting brief, evidence based clinical advice. They have developed into long and authoritative texts used by managers and politicians to inform policy. SIGN has responsibility to consider cost-effectiveness and directly inputs to the Quality Outcomes Framework (QOF). A formal arrangement between SIGN and the National Institute of Clinical Excellence (NICE) has existed from 2003.

The World Health Organisation (WHO) recognises that current grades of recommendation (Table 1) may be ambiguous[2] and encourages guideline developers to use a system which includes a category "Use only in the context of research" where doubt exists.

Guideline developers have conflict of interest policies reported as challenging to apply. Where doubt exists, groups of specialists may feel consensus more defensible than acknowledging uncertainty. [3]

Even with the best evidence, concerns are expressed about the relevance of guidelines in treating patients with multiple morbidities[4], and the emergence of the phenomenon of reversal[5,6], where established practice, sometimes evidence based, is shown to be sub-optimal or harmful. This study looks at the quality of evidence used for SIGN guidelines, and describes a significant trend for some groups to emphasise poorly evidence-backed recommendations.

METHODS

SIGN guidelines were accessed online in September 2013. SIGN guidelines were chosen because they are internationally respected, the authors were familiar with

their format and they contribute to national government policy. Guidelines that were “Current” and “Current 3-7 years. Some recommendations may be out of date.” were included. Those that had been "Withdrawn", "Recommendations being updated", "Need for update being considered" and those with no recommendations were excluded.

SIGN guideline 50 clearly describes an established process for developing guidelines.[7] It explains how the process is planned, how it is implemented and by whom. This process is independent of this study, but is stated to be an objective process. SIGN guidelines have four grades of recommendation outlined in table 1. Table 2 describes the level of evidence SIGN uses to support the recommendation grading. SIGN guideline development groups vary in size depending on the scope of the topic under consideration, but generally comprise between 15 and 25 members. SIGN states they are aware of the many psychosocial factors, including the problems of overcoming professional hierarchies that can affect small group processes.

The level of evidence used by each examined guideline was independently enumerated by 3 investigators and discrepancies resolved. They discounted any duplication implicit in text-embedded key recommendations and also implementation recommendations. A statistical analysis of the correlation between the proportion of level D evidence and the total number of recommendations was performed for the 42 guidelines.

RESULTS

The 42 guidelines consisted of 2559 pages (including references), ranging from 26 to 161 (median 59.5) pages. The longest guideline, number 116 was 61 pages longer than the next largest. The number of recommendations per page ranged from 0.2 to 1.8 (median 0.7). The number of recommendations per guideline is presented in table 3.

Of the 1999 recommendations, 480 (24.0%) were level A, 491(24.6 %) were level B, 318 (15.9%) level C, and 710 (35.5%) level D. Thus 51.4% were poorly evidenced

(C&D) and over a third (D) developed almost entirely on "expert opinion". The number of level A recommendations per guideline ranged from 0-57 (median 9), level B from 2-62 (median 8.5) level C ranged from 0-26 (median 6) and D from 0-60 (median 14.5). 4 guidelines had no level A evidence.

The proportion of level D evidence increases with the number of recommendations made. This correlation is significant with Kendall's Tau=0.22 [approximate 95% confidence interval 0.008-0.45] p value =0.04, and Spearman rho=0.22 [approximate 95% CI 0.02-0.57] p value= .04.

DISCUSSION

This study reveals that expert groups who produce long guidelines rely on poor evidence more heavily than others. While this study only looks at SIGN, this study highlights a problem that has escaped national guideline developers, a wide range of professionals and the public to whom these guidelines are applied. National guidelines are useful and important and there is a debate about how evidence is best presented. Guidelines define standards of care, help busy clinicians and allow managers and politicians to develop governance. An American study (using 3 not 4 levels of evidence) similarly found that 48% were "based on expert opinion, case studies, or standards of care."[8]; we show comparable results for current SIGN guidelines. Where patients are involved in clinical decisions, honestly declaring uncertainty has merit. In the absence of good scientific evidence, recommending a course of action without understanding the circumstances of the individual to whom it is applied seems both risky and, assuming a right to patient choice, unwarranted. Other guidelines that use high levels of poor evidence should evaluate the proportion of poorly evidenced recommendations and seek explanations for such trends.

This study did not examine why longer guidelines use poorer evidence. Cloistered groups of experts may view their own opinion as more authoritative than science can support. It has been postulated that there is security in "just doing what everyone else is doing – even if what everyone else is doing isn't very good."[3]

Reliance on expert opinion has a poor track record. Blinded by certainty, expert groups defining established practice have, in the past, perpetuated radical mastectomy instead of conservative surgery, Class 1C antiarrhythmics[9], pulmonary artery catheters in heart failure[10], electronic foetal monitoring in low risk pregnancies: even then practice can take a decade to reverse[11].

Even good evidence is subject to the phenomenon of reversal where new evidence contradicts current practice. Reversal can affect around 13-16% of publications [5,6]. This may partly explain why the implementation of even the most soundly evidence based national guidelines fails to improve outcome [12],[13],[14]. There is potential harm[15,16] from guidelines in real clinical settings, for example increasing radiation dose without benefit[17] or increased risks of anticoagulation[18].

SIGN 116 (diabetes), is a notable exception. It is more than 50% larger than the next largest, 2.5 times longer than the average and yet uses the 4th lowest level D recommendations. There are a number of hypotheses why this group reports differently. SIGN guidelines inform Quality Outcomes Framework (QOF) policy. Diabetes is the largest clinical UK QOF indicator and is associated with substantial payment incentives. The need for objective evaluation of performance drives a use of surrogate outcomes without appropriate clinical endpoints. [19] Diabetes guidelines have suffered several noteworthy reversals. Examples include the recommendation of glycosylated haemoglobin reduction resulting in increased use of rosiglitazone (still mentioned in the current document) both associated with harm[20]including mortality. [21] Aspirin recommendations have also been changed from previous guidelines. Is it possible that the repeated use of surrogate outcomes arises from group dynamics driven by a powerful external agenda?

Many doctors whose expertise cross several guidelines[22,23], express concerns about guideline development groups. The inappropriate exclusion of disease groups from general population data is common. Smoking cessation advice for testicular cancer survivors is level C, although studies in the general population (without excepting specific disease groups) advises everyone to stop smoking. Overall smoking cessation

was level D and C once each and B on three occasions. Using evidence in this way may imply group dysfunction. Differently constituted groups, or greater oversight might avoid problems.

In 1993, SIGN guidelines stated intention was to be evidence based, brief and succinct. Brevity increases value as a quick reference guide. Removing or reducing poorly evidenced recommendations would reduce size by more than a third overall and in some up to two thirds. The two volumes Oxford Textbook of Primary Medical Care (2005) is a relatively brief 1420 pages, more than a thousand less than the 2559 pages of guidelines. Evidence based medicine is described as "the use of mathematical estimates of the risk of benefit and harm, derived from high-quality research on population samples, to inform clinical decision-making in the diagnosis, investigation or management of individual patients"[24]. Guidelines relevance to daily practice, the reliability of evidence and whether the application of evidence will improve outcomes are important questions.

These results may reflect how professional groups deal with uncertainty. If so, this is not good for individual patients faced with the same uncertainties (whether aware of it or not), nor is it good for scientists who actively seek unanswered questions by challenging established practice, an area in which medicine has a poor record from Semmelweis to the present day.

The finding of a significant increase of level D recommendations in larger guidelines has not happened by chance. A wider debate about how guideline groups can create greater clarity about the reliability of evidence used is needed.[25] Reducing the use of poorly evidenced recommendations has potential to create a shorter, more reliable and usable clinical support. The GRADE working group was formed in 2000.[26] SIGN proposed a move to a new grading system in 2001.[27] Whether the changes in process at present being considered will resolve the challenges that underpin the inconsistencies we have outlined remains to be seen.

Table 1

GRADES OF RECOMMENDATION

A	At least one meta-analysis, systematic review, or RCT rated as 1++, and directly applicable to the target population; or A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results
B	A body of evidence including studies rated as 2++, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 1++ or 1+
C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 2++
D	Evidence level 3 or 4; or Extrapolated evidence from studies rated as 2+

Table 2

1++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1 -	-Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort studies
2+	High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2 -	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, eg case reports, case series
4	Expert opinion

Table 3

Main Recommendations

Number	Name	Pages	A	B	C	D	*Total*	%age D
133	Management of hepatitis C	57	20	24	7	52	***103***	50.5%
132	Long-term follow-up of survivors of childhood cancer	62	0	7	9	14	***30***	46.7%
131	Management of schizophrenia	64	10	19	3	15	***47***	31.9%
130	Brain injury rehabilitation in adults	68	0	14	7	8	***29***	27.6%
129	Antithrombotics: indication and management	68	25	11	6	19	***61***	31.1%
127	Management of perinatal mood disorders	47	0	5	6	15	***26***	57.7%
126	Diagnosis and management of colorectal cancer	56	11	19	15	29	***74***	39.2%
125	Management of atopic eczema in primary care	34	3	5	3	2	***13***	15.4%
124	Management of adult testicular germ cell tumours	63	6	6	9	21	***42***	50.0%
123	Management of early rheumatoid arthritis	27	3	7	2	0	***12***	0.0%
122	Prevention and management of venous thromboembolism	88	26	15	14	55	***110***	50.0%
121	Diagnosis and management of psoriasis and psoriatic arthritis in adults	65	11	16	6	26	***59***	44.1%
120	Management of chronic venous leg ulcers	46	5	3	4	7	***19***	36.8%
119	Management of patients with stroke: identification and management of dysphagia	42	0	6	4	20	***30***	66.7%
118	Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning	101	21	29	7	21	***78***	26.9%
117	Management of sore throat and indications for tonsillectomy	37	9	3	4	4	***20***	20.0%
116	Management of diabetes	161	57	62	23	16	***158***	10.1%
115	Management of Obesity	87	6	11	7	11	***35***	31.4%
114	Nonpharmaceutical management of depression	37	5	4	0	0	***9***	0.0%
113	Diagnosis and pharmacological management of Parkinson's disease	61	12	6	6	4	***28***	14.3%
112	Management of attention deficit and hyperkinetic disorders in children and young people	45	6	4	3	4	***17***	23.5%
111	Management of hip fracture in old people	49	10	9	8	14	***41***	34.1%
110	Early management of patients with a head injury	76	1	7	6	17	***31***	54.8%
109	Management of genital Chlamydia trachomatis infection	40	3	6	9	29	***47***	61.7%
108	Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention	100	42	27	18	14	***101***	13.9%
107	Diagnosis and management of headache in adults	81	17	16	9	34	***76***	44.7%
106	Control of pain in adults with cancer	71	5	7	3	19	***34***	55.9%
105	Management of acute upper and lower gastrointestinal bleeding	57	14	5	2	15	***36***	41.7%
103	Diagnosis and management of chronic kidney disease	50	9	6	4	3	***22***	13.6%
102	Management of invasive meningococcal disease in children and young people	46	1	4	6	26	***37***	70.3%
99	Management of cervical cancer	73	1	13	19	29	***62***	46.8%
97	Risk estimation and the prevention of cardiovascular disease	72	16	12	2	4	***34***	11.8%
96	Management of stable angina	59	13	10	3	11	***37***	29.7%
95	Management of chronic heart failure	55	9	12	1	1	***23***	4.3%
94	Cardiac arrhythmias and coronary heart disease	42	22	11	13	23	***69***	33.3%
93	Acute coronary syndromes	60	11	14	9	8	***42***	19.0%
91	Bronchiolitis in children	42	4	3	6	14	***27***	51.9%
90	Diagnosis and management of head and neck cancer	92	42	8	26	60	***136***	44.1%
89	Diagnosis and management of peripheral arterial disease	37	11	2	0	4	***17***	23.5%
88	Management of suspected bacterial urinary tract infection in adults	45	8	10	2	10	***30***	33.3%
87	Management of oesophageal and gastric cancer	70	3	26	23	28	***80***	35.0%
61	Investigation of postmenopausal bleeding	26	2	7	4	4	***17***	23.5%
	TOTAL	2559	480	491	318	710	1999	

Acknowledgements.

Heather Barrington, Statistical Adviser; Bridget Bird, Administrative Assistant; Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries DG1 4AP.
Dr A Gordon Baird affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; any discrepancies from the study as planned have been explained.

Contributorship Statement

Both authors discussed the hypothesis that there was a disproportionate use of poor evidence in longer guidelines; both checked the raw data, and agreed on a statistical approach to discover whether the trend was significant or not. Both have been involved in writing and researching the evidence.

Funding.
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data Sharing.
Technical appendix, statistical code, and dataset available from the corresponding author, who will provide a permanent, citable and open access home for the dataset.

Competing interests
No competing interests

MeSH Headings

- Practice Guideline N04.761.700.350.650
- General Practice H02.403.340
- Evidence-Based Medicine H02.249.750

How this fits in.
Guidelines should encourage an evidence-based approach to clinical practice. Longer guidelines used significantly higher levels of poor evidence. WHO has proposed a different system of grading evidence. The effect of group behavior altering guideline development has been hypothesised. New research often challenges established clinical practice. Improving the quality of evidence, acknowledging uncertainty and shortening guideline length would make guidelines more clinically relevant and effective.

REFERENCES

1 Harbour R, Lowe G, Twaddle S. Scottish Intercollegiate Guidelines Network; the first 15 years (1993-2008). J R Coll Physicians Edinb 2011 Jun;41(2) 163-168

2 de Joncheere K, Hill S, Klazinga N, Mäkelä M, Oxman AD.The Clinical Guideline Programme of the National Institute for Health and Clinical Excellence (NICE) A review by the World Health Organization May 2006

3 Lenzer J. Why we can't trust clinical guidelines: BMJ 2013;346:f3830

4 Aylett V. Do geriatricians need guidelines? BMJ 2010 Sep;:341 :c5340

5 Prasad V, Cifu, A, Ioannidis JP. Reversals of Established Medical Practices: Evidence to Abandon Ship. JAMA. 2012 Jan4;307(1):37-38,

6 Prasad V, Gall V, Cifu A. The frequency of medical reversal. Arch Int Med. 2011 Oct 10;. 171(18): 1675-1676

7 SIGN 50: A guideline developer's handbook http://www.sign.ac.uk/pdf/sign50.pdf

8 Tricoci P, Allen JM, Kramer JM, Califf RM, Smith Jr. SC. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA 2009 Feb25; 301 (8): 831-841

9 Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, Arensberg D, Baker A, Friedman L, Greene HL, Huther ML, Richardson DW. Mortality and morbidity in patients receiving encainide, flecainide, or placebo - The Cardiac Arrhythmia Suppression Trial. New Eng J Med. 1991 Mar21; 324 (12): 781-788

10 Binanay C, Califf RM, Hasselblad V, O'Connor CM, Shah MR, Sopko G, Stevenson LW, Francis GS, Leier CV, Miller LW Evaluation study of congestive heart failure and pulmonary artery catheterization effectiveness: the ESCAPE trial. JAMA. 2005 Oct 5;294(13):1625-33.

11 Tatsioni A, Bonitsis NG, Ioannidis JA. Persistence of Contradicted Claims in the Literature. JAMA. 2007 Dec5 ;298(21):2517-2526.

12 Anderson JE, McKenzie C, Singh N, Gajree N, Giles L, Sharma P, McDonald A. Compliance with the 62 day target does not improve long-term survival. Association of Coloproctology of Great Britain and Ireland Annual Meeting 2012 Dublin Ireland.

13 Kerr J, Smith R, Gray S, Beard D, Robertson CE. An audit of clinical practice in the management of head injured patients following the introduction of the Scottish Intercollegiate Guidelines Network (SIGN) recommendations. Emerg Med J.2005 Dec; 22(12):850-854

14 Caplan LR. How well does "evidence-based" medicine help neurologists care for individual patients? Revin Neurol Dis.2007;4(2):75-84

15 Hutchison G. Guidelines can harm patients too. BMJ (Clinical research ed.). 2012 Apr18;344: e2685.

16 Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines. Potential benefits, limitations, and harms of clinical guidelines. BMJ 1999 Feb20;318(7182):527-530

17 Miller L., Kidd A. Are sign guidelines key in the decision to CT? Post-head injury CT scanning within a paediatric population. Academic Emergency Medicine. Conference: 14th International Conference on Emergency Medicine, ICEM 2012 Dublin Ireland. Conference Publication: 19 (6) (pp 776), 2012. Date of Publication: June 2012.

18 Thomson R., Eccles M., Wood R., Chinn D.J. A cautionary note on data sources for evidence-based clinical decisions: Warfarin and stroke prevention. MedDecis Making. 2007;27 (4): 438-447

19 Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. BMJ. 2011 Dec28;343:d7995

20 Cohen D. Rosiglitazone: what went wrong? BMJ.2010 Sep; 341:c4848

21 Gerstein HC, Miller ME, Byington RP, Goff DC Jr., Bigger JT, Buse JB, Cushman WC, Genuth S, Ismail-Beigi F, Grimm RH Jr. Probstfield JL, Simons-Morton DG, Friedewald WT. Effects of intensive glucose lowering in type 2 diabetes. Action to Control Cardiovascular Risk in Diabetes Study Group. New Engl J Med. 2008 Jun12;358(24):2545-59

22 Rashidian A, Eccles MP, Russell I. Falling on stony ground? A qualitative study of implementation of clinical guidelines' prescribing recommendations in primary care. Health Policy.2008 Feb; 85 (2) 148-161

23 Guy S, Wardlaw JM. Who writes guidelines, and who should? ClinRadiol. 2002 Oct ;57 (10): 891-897.

24 Greenhalgh, Trisha. How To Read a Paper: The Basics of Evidence-Based Medicine. Wiley-Blackwell, fourth edition, 2010, p. 1

25 Zuiderent-Jerak T, Forland F, Macbeth F. Guidelines should reflect all knowledge, not just trials BMJ 2012;345:e6702

26 GRADE working group http://www.gradeworkinggroup.org/about_us.htm

27 Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ. 2001 Aug 11;323(7308):334-6.

SIGN Guidelines –analysis of evidence levels for their recommendations.

Alternatively

Guidelines – is bigger better? A review of SIGN national guidelines.

~~National Guidelines – How evidence based are they?~~

Dr A Gordon Baird, MB ChB MRCOG FRCGP R&D General Practice Adviser, Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries Scotland, UK DG1 4AP.

Address for correspondence; Dr A Gordon Baird, The White House, Sandhead, Wigtownshire UK. DG9 9JA. Email gordon.baird@me.com. Phone/fax 01776830281

Dr James R Lawrence, BSc MD FRCP(Glas & Edin) R&D Clinical Lead, Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries Scotland, UK DG1 4AP.

Formatted: Justified

MeSH headings
General practice / Family practice
HEALTH SERVICES ADMINISTRATION & MANAGEMENT
Protocols & guidelines
Quality in health care
Clinical governance

Formatted: Justified

ABSTRACT

Background

National guidelines have become an integral part of UK general practice and aim to deliver quality and consistency ~~within practices~~in clinical practice. As with any intervention, there are negative as well as positive consequences. Guideline effectiveness depends on the quality of evidence used.

Aim

To quantify and analyse the quality of evidence that is presented in national guidelines.

Design and setting

Levels of evidence used in all the current valid recommendations in the Scottish Intercollegiate Guideline Network (SIGN) guidelines were reviewed and statistically analysed.

Method

The data was collected from published guidelines available online to the public. ~~A~~ SIGN methodology entails a professional group selected by a national organisation develops each of these guidelines. Statistical analysis of the relationship between the number of guideline recommendations and the quality of evidence used in its recommendations was performed.

Result

A significant correlation between the number of recommendations in a guideline and the use of level D evidence was discovered.

Conclusion

Practice guidelines should be brief and based on scientific evidence. Paradoxically the longest guidelines have the highest proportion of recommendations based on the lowest level of evidence. Guideline developers should be more aware of the need for brevity and a stricter application of evidence-based principles could achieve this. The findings support calls for a review of how evidence is used and presented in guidelines.

Dr A Gordon Baird affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; any discrepancies from the study as planned have been explained.

The data is in the public domain; there were no ethical conflicts; there was no funding.

Article Summary

Article focus

Examines the hypothesis that larger guidelines rely disproportionately on poor evidence.

Key Messages

- One third of current national guidelines ~~is~~are supported only by case reports, case studies and expert opinion.
- Guidelines with large numbers of recommendations are more likely to use weak evidence.
- Guideline development groups appear to vary in their approach to offering recommendations.
- Guideline recommendations should be based on good evidence.
- Without good science scientific evidence, avoiding a recommendation or highlighting the need for research should be considered.

Strengths and limitations of the study

Strengths.

- This is the first objective evidence of inconsistencies in approach by national guideline developers
- This supports commentator suggestion that even without good evidence a group will prefer consensus.
- Adds to the current debate about how guidelines might be developed in the future

Limitations.

- The study is limited to only one set of national guidelines.
- Reasons for the differences in quality of evidence preferred by the guideline development groups is unclear

Formatted: Left

Guidelines – is bigger better? A review of SIGN guidelines.
~~SIGN Guidelines – How evidence based are they?~~

INTRODUCTION

The Scottish Intercollegiate Guidelines Network (SIGN) was founded in 1993. It is a national body, professionally led and publicly funded. SIGN's founding principles proposed direct links between evidence and recommendations, offering a brief and succinct quick-reference guide for clinicians[1]. Guidelines anticipated presenting brief, evidence based clinical advice. They have developed into long and authoritative texts ~~that are~~ used by managers and politicians to inform policy. SIGN has responsibility to consider cost-effectiveness and directly inputs to the Quality Outcomes Framework (QOF). A formal arrangement between SIGN and the National Institute of Clinical Excellence (NICE) has existed from 2003.

~~SIGN has no responsibility to consider cost-effectiveness and no direct input into the Quality Outcomes Framework (QOF). A formal arrangement between SIGN and the National Institute of Clinical Excellence (NICE) has existed from 2003.~~

The World Health Organisation (WHO) recognises that current grades of recommendation ~~categories~~ (Table 1) may be ambiguous [2] and has encouraged guideline developers to use a different form of grading, including a category "Use only in the context of research" where doubt exists.

Guideline developers have conflict of interest policies reported as challenging to apply. Where doubt exists, groups of specialists may feel consensus more defensible than acknowledging uncertainty. [3]

Even with the best evidence, concerns are expressed about the relevance of guidelines in treating patients with multiple morbidities[4], and the emergence of the phenomenon of reversal[5,6], where established practice, sometimes evidence based,

is shown to be sub-optimal or harmful. This study looks at the quality of evidence used for SIGN guidelines, and describes a significant trend for some groups to emphasise poorly evidence-backed recommendations.

METHODS

SIGN guidelines were accessed online in September 2013. SIGN guidelines were chosen because they are internationally respected, the authors were familiar with their format and they contribute to national government policy. Guidelines that were "Current" and "Current 3-7 years. Some recommendations may be out of date." ~~in September 2013~~ were included. Those that had been "Withdrawn", "Recommendations being updated", "Need for update being considered" and those with no recommendations were excluded. Key recommendations and implementation recommendations were excluded.

SIGN guideline 50 clearly describes an established process for developing guidelines.[7] It explains how the process is planned, how it is implemented and by whom. This process is independent of this study, but is stated to be an objective process. SIGN guidelines have four grades of recommendation outlined in table 1. Table 2 describes the level of evidence SIGN uses to support the recommendation grading. SIGN guideline development groups vary in size depending on the scope of the topic under consideration, but generally comprise between 15 and 25 members. SIGN states they are aware of the many psychosocial factors, including the problems of overcoming professional hierarchies that can affect small group processes.~~SIGN guideline 50 describes an established process for developing guidelines.[8] This process is independent of this study, but is stated to be an objective process. SIGN guidelines have four grades of recommendation outlined in table 1. Table 2 describes the level of evidence supporting the recommendation grading.~~

The level of evidence used by each examined guideline was independently enumerated by 3 investigators and discrepancies resolved~~used by each guideline was independently recorded by 3 investigators and errors resolved~~. They discounted any duplication implicit in text-embedded key recommendations and also implementation ~~A~~recommendations. A statistical analysis of the correlation between the proportion of level D evidence and the total number of recommendations was performed for the 42 guidelines.

RESULTS

The 42 guidelines consisted of 2559 pages(including references), ~~,~~ ranging from 26 to 161 (median 59.5) pages. The longest guideline, number 116 was 61 pages longer than the next largest. The number of recommendations per page ranged from 0.2 to 1.8 (median 0.7). The number of recommendations per guideline is presented in table 3.

Of the 1999 recommendations, 480 (24.0%) were level A, 491(24.6 %) were level B, 318 (15.9%) level C, and 710 (35.5%) level D. Thus 51.4% ~~40.2%~~ were poorly evidenced (C&D) and over a third (D) developed almost entirely on “expert opinion”. The number of level A recommendations per guideline ranged from 0-57 (median 9), level B from 2-62 (median 8.5) level C ranged from 0-26 (median 6) and D from 0-60 (median 14.5). 4 guidelines had no level A evidence.

The proportion of level D evidence increases with the number of recommendations made. This correlation is significant with Kendall’s Tau=0.22 [approximate 95% confidence interval 0.008-0.45] p value =0.04, and Spearman rho=0.22 [approximate 95% CI 0.02-0.57] p value= .04.

DISCUSSION

This study reveals that expert groups who produce long guidelines rely on poor evidence more heavily than others. While this study only looks at SIGN, this study highlights a problem that has escaped national guideline developers, a wide range of professionals and the public to whom these guidelines are applied. National

guidelines are useful and important and there is a debate about how evidence is best presented. ~~Guidleines~~Guidelines define standards of care, help busy clinicians and allow managers and politicians to develop governance. An American study (using 3 not 4 levels of evidence) similarly found that 48% were "based on expert opinion, case studies, or standards of care." [9] we show comparable results for current SIGN guidelines. Where patients are involved in clinical decisions, honestly declaring uncertainty has merit. In the absence of good scientific evidence, recommending a course of action without understanding the circumstances of the individual to whom it is applied seems both risky and, assuming a right to patient choice, unwarranted~~imbalanced~~.

Other guidelines that use high levels of poor evidence should evaluate the proportion of poorly evidenced recommendations and seek explanations for such trends.

This study did not examine why longer guidelines use poorer evidence. Cloistered groups of experts may view their own opinion as more authoritative than science can support. It has been postulated that there is security in "just doing what everyone else is doing – even if what everyone else is doing isn't very good."[3] ~~Cloistered groups of experts may view their own opinion as more authoritative than science can support.~~ Reliance on expert opinion has a poor track record. Blinded by certainty, expert groups defining established practice have perpetuated radical mastectomy instead of conservative surgery, Class 1C antiarrhythmics[10], pulmonary artery catheters in heart failure [11], electronic foetal monitoring in low risk pregnancies: even then practice can take a decade to reverse[12].

Even good evidence is subject to the phenomenon of reversal where new evidence contradicts current practice. Reversal can affect around 13-16% of publications [5,6]. This may partly explain why the implementation of even the most soundly evidence based national guidelines fails to improve outcome [13],[14],[15]. There is potential harm[16,17] from guidelines in real clinical settings, for example increasing radiation dose without benefit[18] or increased risks of anticoagulation[19].

SIGN 116 (diabetes), is a notable exception. It is more than 50% larger than the next largest, 2.5 times longer than the average and yet uses the 4th lowest level D recommendations. There are a number of hypotheses why this group reports differently. SIGN guidelines inform Quality Outcomes Framework (QOF) policy~~The advisory committee to SIGN's English equivalent (NICE) informs Quality Outcomes Framework (QOF) policy~~. Diabetes is the largest clinical UK QOF indicator and is associated with substantial payment incentives. The need for objective evaluation of performance drives a use of surrogate outcomes without appropriate clinical endpoints. [20] Diabetes guidelines have suffered several noteworthy reversals. Examples include the recommendation of glycosylated haemoglobin reduction resulting in increased use of rosiglitazone (still mentioned in the current document) both associated with harm[21]including mortality.[22] Aspirin recommendations have also been changed from previous guidelines. Is it possible that the repeated use of surrogate outcomes arises from group dynamics driven by a powerful external agenda?

Many doctors whose expertise cross several guidelines[23,24], express concerns about guideline development groups. The inappropriate exclusion of disease groups from general population data is common. Smoking cessation advice for testicular cancer survivors is level C, although studies in the general population (without excepting specific disease groups) advises everyone to stop smoking. Overall smoking cessation was level D and C once each and B on three occasions. Using evidence in this way may imply group dysfunction. Differently constituted groups, or greater oversight might avoid problems.

In 1993, SIGN guidelines stated intention was to be evidence based, brief and succinct. Brevity increases value as a quick reference guide. Removing or reducing poorly evidenced recommendations would reduce size by more than a third overall and in some up to two thirds. The two volumes Oxford Textbook of Primary Medical Care (2005) is a relatively brief 1420 pages, more than a thousand less than the 2559 pages of guidelines.~~2559 pages is longer than the Oxford Textbook of Primary Medical Care.~~ Evidence based medicine is described as "the use of mathematical

estimates of the risk of benefit and harm, derived from high-quality research on population samples, to inform clinical decision-making in the diagnosis, investigation or management of individual patients"[25]. Guidelines relevance to daily practice, the reliability of evidence and whether the application of evidence will improve outcomes are important questions.

These results may reflect how professional groups deal with uncertainty. If so, this is not good for individual patients faced with the same uncertainties (whether aware of it or not), nor is it good for scientists who actively seek unanswered questions by challenging established practice, an area in which medicine has a poor record from Semmelweis to the present day.

The finding of a significant increase of level D recommendations in larger guidelines has not happened by chance. A wider debate about how guideline groups can create greater clarity about the reliability of evidence used is needed.[26] Reducing the use of poorly evidenced recommendations has potential to create a shorter, more reliable and usable clinical support. The GRADE working group was formed in 2000.[27] SIGN proposed a move to a new grading system in 2001.[28] Whether the changes in process at present being considered will resolve the challenges that underpin the inconsistencies we have outlined remains to be seen.

Acknowledgements.

Heather Barrington, Statistical Adviser; Bridget Bird, Administrative Assistant; Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries DG1 4AP.

Conflict of interest; none

Table 1

GRADES OF RECOMMENDATION	
A	At least one meta-analysis, systematic review, or RCT rated as 1++, and directly applicable to the target population; or A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results
B	A body of evidence including studies rated as 2++, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 1++ or 1+
C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 2++
D	Evidence level 3 or 4; or Extrapolated evidence from studies rated as 2+

Table 2

1++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1 -	-Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort studies
2+	High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2 -	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, eg case reports, case series
4	Expert opinion

Table 3

Main Recommendations

Number	Name	Pages	A	B	C	D	*Total*	%age D
133	Management of hepatitis C	57	20	24	7	52	***103***	50.5%
132	Long-term follow-up of survivors of childhood cancer	62	0	7	9	14	***30***	46.7%
131	Management of schizophrenia	64	10	19	3	15	***47***	31.9%
130	Brain injury rehabilitation in adults	68	0	14	7	8	***29***	27.6%
129	Antithrombotics: indication and management	68	25	11	6	19	***61***	31.1%
127	Management of perinatal mood disorders	47	0	5	6	15	***26***	57.7%
126	Diagnosis and management of colorectal cancer	56	11	19	15	29	***74***	39.2%
125	Management of atopic eczema in primary care	34	3	5	3	2	***13***	15.4%
124	Management of adult testicular germ cell tumours	63	6	6	9	21	***42***	50.0%
123	Management of early rheumatoid arthritis	27	3	7	2	0	***12***	0.0%
122	Prevention and management of venous thromboembolism	88	26	15	14	55	***110***	50.0%
121	Diagnosis and management of psoriasis and psoriatic arthritis in adults	65	11	16	6	26	***59***	44.1%
120	Management of chronic venous leg ulcers	46	5	3	4	7	***19***	36.8%
119	Management of patients with stroke: identification and management of dysphagia	42	0	6	4	20	***30***	66.7%
118	Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning	101	21	29	7	21	***78***	26.9%
117	Management of sore throat and indications for tonsillectomy	37	9	3	4	4	***20***	20.0%
116	Management of diabetes	161	57	62	23	16	***158***	10.1%
115	Management of Obesity	87	6	11	7	11	***35***	31.4%
114	Nonpharmaceutical management of depression	37	5	4	0	0	***9***	0.0%
113	Diagnosis and pharmacological management of Parkinson's disease	61	12	6	6	4	***28***	14.3%
112	Management of attention deficit and hyperkinetic disorders in children and young people	45	6	4	3	4	***17***	23.5%
111	Management of hip fracture in old people	49	10	9	8	14	***41***	34.1%
110	Early management of patients with a head injury	76	1	7	6	17	***31***	54.8%
109	Management of genital Chlamydia trachomatis infection	40	3	6	9	29	***47***	61.7%
108	Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention	100	42	27	18	14	***101***	13.9%
107	Diagnosis and management of headache in adults	81	17	16	9	34	***76***	44.7%
106	Control of pain in adults with cancer	71	5	7	3	19	***34***	55.9%
105	Management of acute upper and lower gastrointestinal bleeding	57	14	5	2	15	***36***	41.7%
103	Diagnosis and management of chronic kidney disease	50	9	6	4	3	***22***	13.6%
102	Management of invasive meningococcal disease in children and young people	46	1	4	6	26	***37***	70.3%
99	Management of cervical cancer	73	1	13	19	29	***62***	46.8%
97	Risk estimation and the prevention of cardiovascular disease	72	16	12	2	4	***34***	11.8%
96	Management of stable angina	59	13	10	3	11	***37***	29.7%
95	Management of chronic heart failure	55	9	12	1	1	***23***	4.3%
94	Cardiac arrhythmias and coronary heart disease	42	22	11	13	23	***69***	33.3%
93	Acute coronary syndromes	60	11	14	9	8	***42***	19.0%
91	Bronchiolitis in children	42	4	3	6	14	***27***	51.9%
90	Diagnosis and management of head and neck cancer	92	42	8	26	60	***136***	44.1%
89	Diagnosis and management of peripheral arterial disease	37	11	2	0	4	***17***	23.5%
88	Management of suspected bacterial urinary tract infection in adults	45	8	10	2	10	***30***	33.3%
87	Management of oesophageal and gastric cancer	70	3	26	23	28	***80***	35.0%
61	Investigation of postmenopausal bleeding	26	2	7	4	4	***17***	23.5%
	TOTAL	2559	480	491	318	710	1999	

Contributorship Statement

Both authors discussed the hypothesis that there was a disproportionate use of poor evidence in longer guidelines; both checked the raw data, and agreed on a statistical approach to discover whether the trend was significant or not. Both have been involved in writing and researching the evidence.

Funding.
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data Sharing.
Technical appendix, statistical code, and dataset available from the corresponding author, who will provide a permanent, citable and open access home for the dataset.

Competing interests
No competing interests

MeSH Headings

- Practice Guideline N04.761.700.350.650
- General Practice H02.403.340
- Evidence-Based Medicine H02.249.750

How this fits in.
Guidelines should encourage an evidence-based approach to clinical practice. Longer guidelines used significantly higher levels of poor evidence. WHO has proposed a different system of grading evidence. The effect of group behavior altering guideline development has been hypothesised. New research often challenges established clinical practice. Improving the quality of evidence, acknowledging uncertainty and shortening guideline length would make guidelines more clinically relevant and effective.

REFERENCES

1 Harbour R, Lowe G, Twaddle S. Scottish Intercollegiate Guidelines Network; the first 15 years (1993-2008). J R Coll Physicians Edinb 2011 Jun;41(2) 163-168

2 de Joncheere K, Hill S, Klazinga N, Mäkelä M, Oxman AD.The Clinical Guideline Programme of the National Institute for Health and Clinical Excellence (NICE) A review by the World Health Organization May 2006

3 Lenzer J. Why we can't trust clinical guidelines: BMJ 2013;346:f3830

4 Aylett V. Do geriatricians need guidelines? BMJ 2010 Sep;:341 :c5340

5 Prasad V, Cifu, A, Ioannidis JP. Reversals of Established Medical Practices: Evidence to Abandon Ship. JAMA. 2012 Jan4;307(1):37-38,

6 Prasad V, Gall V, Cifu A. The frequency of medical reversal. Arch Int Med. 2011 Oct 10;. 171(18): 1675-1676

7 SIGN 50: A guideline developer's handbook http://www.sign.ac.uk/pdf/sign50.pdf

8 http://www.sign.ac.uk/pdf/sign50.pdf

9 Tricoci P, Allen JM, Kramer JM, Califf RM, Smith Jr. SC. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA 2009 Feb25; 301 (8): 831-841

10 Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, Arensberg D, Baker A, Friedman L, Greene HL, Huther ML, Richardson DW. Mortality and morbidity in patients receiving encainide, flecainide, or placebo - The Cardiac Arrhythmia Suppression Trial. New Eng J Med. 1991 Mar21; 324 (12): 781-788

11 Binanay C, Califf RM, Hasselblad V, O'Connor CM, Shah MR, Sopko G, Stevenson LW, Francis GS, Leier CV, Miller LW Evaluation study of congestive heart failure and

pulmonary artery catheterization effectiveness: the ESCAPE trial. JAMA. 2005 Oct 5;294(13):1625-33.

12 Tatsioni A, Bonitsis NG, Ioannidis JA. Persistence of Contradicted Claims in the Literature. JAMA. 2007 Dec5 ;298(21):2517-2526.

13 Anderson JE, McKenzie C, Singh N, Gajree N, Giles L, Sharma P, McDonald A. Compliance with the 62 day target does not improve long-term survival. Association of Coloproctology of Great Britain and Ireland Annual Meeting 2012 Dublin Ireland.

14 Kerr J, Smith R, Gray S, Beard D, Robertson CE. An audit of clinical practice in the management of head injured patients following the introduction of the Scottish Intercollegiate Guidelines Network (SIGN) recommendations. Emerg Med J.2005 Dec; 22(12):850-854

15 Caplan LR. How well does "evidence-based" medicine help neurologists care for individual patients? Revin Neurol Dis.2007;4(2):75-84

16 Hutchison G. Guidelines can harm patients too. BMJ (Clinical research ed.). 2012 Apr18;344: e2685.

17 Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines. Potential benefits, limitations, and harms of clinical guidelines. BMJ 1999 Feb20;318(7182):527-530

18 Miller L., Kidd A. Are sign guidelines key in the decision to CT? Post-head injury CT scanning within a paediatric population. Academic Emergency Medicine. Conference: 14th International Conference on Emergency Medicine, ICEM 2012 Dublin Ireland. Conference Publication: 19 (6) (pp 776), 2012. Date of Publication: June 2012.

19 Thomson R., Eccles M., Wood R., Chinn D.J. A cautionary note on data sources for evidence-based clinical decisions: Warfarin and stroke prevention. MedDecis Making. 2007;27 (4): 438-447

20 Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. BMJ. 2011 Dec28;343:d7995

21 Cohen D. Rosiglitazone: what went wrong? BMJ.2010 Sep; 341:c4848

22 Gerstein HC, Miller ME, Byington RP, Goff DC Jr., Bigger JT, Buse JB, Cushman WC, Genuth S, Ismail-Beigi F, Grimm RH Jr. Probstfield JL, Simons-Morton DG, Friedewald WT. Effects of intensive glucose lowering in type 2 diabetes. Action to Control Cardiovascular Risk in Diabetes Study Group. New Engl J Med. 2008 Jun12;358(24):2545-59

23 Rashidian A, Eccles MP, Russell I. Falling on stony ground? A qualitative study of implementation of clinical guidelines' prescribing recommendations in primary care. Health Policy.2008 Feb; 85 (2) 148-161

24 Guy S, Wardlaw JM. Who writes guidelines, and who should? ClinRadiol. 2002 Oct ;57 (10): 891-897.

25 Greenhalgh, Trisha. How To Read a Paper: The Basics of Evidence-Based Medicine. Wiley-Blackwell, fourth edition, 2010, p. 1

26 Zuiderent-Jerak T, Forland F, Macbeth F. Guidelines should reflect all knowledge, not just trials
BMJ 2012;345:e6702

TABLE 2

			Main Recommendations				
Number	Name	Pages	A	B	C	D	***Total***
133	Management of hepatitis C	57	20	24	7	52	***103***
132	Long-term follow-up of survivors of childhood cancer	62	0	7	9	14	***30***
131	Management of schizophrenia	64	10	19	3	15	***47***
130	Brain injury rehabilitation in adults	68	0	14	7	8	***29***
129	Antithrombotics: indication and management	68	25	11	6	19	***61***
127	Management of perinatal mood disorders	47	0	5	6	15	***26***
126	Diagnosis and management of colorectal cancer	56	11	19	15	29	***74***
125	Management of atopic eczema in primary care	34	3	5	3	2	***13***
124	Management of adult testicular germ cell tumours	63	6	6	9	21	***42***
123	Management of early rheumatoid arthritis	27	3	7	2	0	***12***
122	Prevention and management of venous thromboembolism	88	26	15	14	55	***110***
121	Diagnosis and management of psoriasis and psoriatic arthritis in adults	65	11	16	6	26	***59***
120	Management of chronic venous leg ulcers	46	5	3	4	7	***19***
119	Management of patients with stroke: identification and management of dysphagia	42	0	6	4	20	***30***
118	Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning	101	21	29	7	21	***78***
117	Management of sore throat and indications for tonsillectomy	37	9	3	4	4	***20***
116	Management of diabetes	161	57	62	23	16	***158***
115	Management of Obesity	87	6	11	7	11	***35***
114	Nonpharmaceutical management of depression	37	5	4	0	0	***9***
113	Diagnosis and pharmacological management of Parkinson's disease	61	12	6	6	4	***28***
112	Management of attention deficit and hyperkinetic disorders in children and young people	45	6	4	3	4	***17***
111	Management of hip fracture in old people	49	10	9	8	14	***41***
110	Early management of patients with a head injury	76	1	7	6	17	***31***
109	Management of genital Chlamydia trachomatis infection	40	3	6	9	29	***47***
108	Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention	100	42	27	18	14	***101***
107	Diagnosis and management of headache in adults	81	17	16	9	34	***76***
106	Control of pain in adults with cancer	71	5	7	3	19	***34***
105	Management of acute upper and lower gastrointestinal bleeding	57	14	5	2	15	***36***
103	Diagnosis and management of chronic kidney disease	50	9	6	4	3	***22***
102	Management of invasive meningococcal disease in children and young people	46	1	4	6	26	***37***
99	Management of cervical cancer	73	1	13	19	29	***62***
97	Risk estimation and the prevention of cardiovascular disease	72	16	12	2	4	***34***

96	Management of stable angina	59	13	10	3	11	***37***
95	Management of chronic heart failure	55	9	12	1	1	***23***
94	Cardiac arrhythmias and coronary heart disease	42	22	11	13	23	***69***
93	Acute coronary syndromes	60	11	14	9	8	***42***
91	Bronchiolitis in children	42	4	3	6	14	***27***
90	Diagnosis and management of head and neck cancer	92	42	8	26	60	***136***
89	Diagnosis and management of peripheral arterial disease	37	11	2	0	4	***17***
88	Management of suspected bacterial urinary tract infection in adults	45	8	10	2	10	***30***
87	Management of oesophageal and gastric cancer	70	3	26	23	28	***80***
61	Investigation of postmenopausal bleeding	26	2	7	4	4	***17***
	TOTAL	2559	480	491	318	710	1999

27 GRADE working group http://www.gradeworkinggroup.org/about_us.htm

28 Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ. 2001 Aug 11;323(7308):334-6.

Table 1

A	At least one meta-analysis, systematic review, or RCT rated as 1++, and directly applicable to the target population; or A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results
B	A body of evidence including studies rated as 2++, directly applicable to the target population, and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 1++ or 1+
C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 2++
D	Evidence level 3 or 4; or Extrapolated evidence from studies rated as 2+

Table 2

1++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1 -	-Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort studies
2+	High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2 -	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, eg case reports, case series
4	Expert opinion

Table 3 – for editing

Number	Number	Name	Pages	Main Recommendations A	B	C	D	***Total***	%age D
133	133	Management of hepatitis C	57	20	24	7	52	***103***	50.5%
132	132	Long-term follow-up of survivors of childhood cancer	62	0	7	9	14	***30***	46.7%
131	131	Management of schizophrenia	64	10	19	3	15	***47***	31.9%
130	130	Brain injury rehabilitation in adults	68	0	14	7	8	***29***	27.6%
129	129	Antithrombotics: indication and management	68	25	11	6	19	***61***	31.1%
127	127	Management of perinatal mood disorders	47	0	5	6	15	***26***	57.7%
126	126	Diagnosis and management of colorectal cancer	56	11	19	15	29	***74***	39.2%
125	125	Management of atopic eczema in primary care	34	3	5	3	2	***13***	15.4%
124	124	Management of adult testicular germ cell tumours	63	6	6	9	21	***42***	50.0%
123	123	Management of early rheumatoid arthritis	27	3	7	2	0	***12***	0.0%
122	122	Prevention and management of venous thromboembolism	88	26	15	14	55	***110***	50.0%
121	121	Diagnosis and management of psoriasis and psoriatic arthritis in adults	65	11	16	6	26	***59***	44.1%
120	120	Management of chronic venous leg ulcers	46	5	3	4	7	***19***	36.8%
119	119	Management of patients with stroke: identification and management of dysphagia	42	0	6	4	20	***30***	66.7%
118	118	Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning	101	21	29	7	21	***78***	26.9%
117	117	Management of sore throat and indications for tonsillectomy	37	9	3	4	4	***20***	20.0%
116	116	Management of diabetes	161	57	62	23	16	***158***	10.1%
115	115	Management of Obesity	87	6	11	7	11	***35***	31.4%
114	114	Nonpharmaceutical management of depression	37	5	4	0	0	***9***	0.0%

Table 3 – for editing

113	113	Diagnosis and pharmacological management of Parkinson's disease	61	12	6	6	4	***28***	14.3%
112	112	Management of attention deficit and hyperkinetic disorders in children and young people	45	6	4	3	4	***17***	23.5%
111	111	Management of hip fracture in old people	49	10	9	8	14	***41***	34.1%
110	110	Early management of patients with a head injury	76	1	7	6	17	***31***	54.8%
109	109	Management of genital Chlamydia trachomatis infection	40	3	6	9	29	***47***	61.7%
108	108	Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention	100	42	27	18	14	***101***	13.9%
107	107	Diagnosis and management of headache in adults	81	17	16	9	34	***76***	44.7%
106	106	Control of pain in adults with cancer	71	5	7	3	19	***34***	55.9%
105	105	Management of acute upper and lower gastrointestinal bleeding	57	14	5	2	15	***36***	41.7%
103	103	Diagnosis and management of chronic kidney disease	50	9	6	4	3	***22***	13.6%
102	102	Management of invasive meningococcal disease in children and young people	46	1	4	6	26	***37***	70.3%
99	99	Management of cervical cancer	73	1	13	19	29	***62***	46.8%
97	97	Risk estimation and the prevention of cardiovascular disease	72	16	12	2	4	***34***	11.8%
96	96	Management of stable angina	59	13	10	3	11	***37***	29.7%
95	95	Management of chronic heart failure	55	9	12	1	1	***23***	4.3%
94	94	Cardiac arrhythmias and coronary heart disease	42	22	11	13	23	***69***	33.3%
93	93	Acute coronary syndromes	60	11	14	9	8	***42***	19.0%
91	91	Bronchiolitis in children	42	4	3	6	14	***27***	51.9%
90	90	Diagnosis and management of head and neck cancer	92	42	8	26	60	***136***	44.1%
89	89	Diagnosis and management of peripheral arterial disease	37	11	2	0	4	***17***	23.5%
88	88	Management of suspected bacterial urinary tract infection in adults	45	8	10	2	10	***30***	33.3%

Table 3 – for editing

87	87	Management of oesophageal and gastric cancer	70	3	26	23	28	***80***	35.0%
61	61	Investigation of postmenopausal bleeding	26	2	7	4	4	***17***	23.5%
		TOTAL	2559	480	491	318	710	1999	

Scottish Intercollegiate Guideline Network–analysis of evidence levels for their recommendations.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-004278.R2
Article Type:	Research
Date Submitted by the Author:	13-Jan-2014
Complete List of Authors:	Baird, Alastair; The White House, Lawrence, James; Dumfries & Galloway Royal Infirmary, Research & Development Support Unit
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	General practice / Family practice
Keywords:	HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Clinical governance < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Scottish Intercollegiate Guideline Network–analysis of evidence levels for their recommendations.

Alternatively

Guideline recommendations are more better? A review of Scottish Intercollegiate Guideline Network national guidelines.

Dr A Gordon Baird, MB ChB MRCOG FRCGP R&D General Practice Adviser, Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries, Scotland, UK DG1 4AP.

Address for correspondence; Dr A Gordon Baird, The White House, Sandhead, Wigtownshire UK. DG9 9JA. Email gordon.baird@me.com. Phone/fax 01776830281

Dr James R Lawrence, BSc MD FRCP(Glas & Edin) R&D Clinical Lead, Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries, Scotland, UK DG1 4AP.

MeSH headings
General practice / Family practice
HEALTH SERVICES ADMINISTRATION & MANAGEMENT
Protocols & guidelines
Quality in health care
Clinical governance

ABSTRACT

Background

National guidelines have become an integral part of UK general practice and aim to effectively deliver quality and consistency in clinical practice. As with any intervention, there are negative as well as positive consequences. Guideline effectiveness depends on the quality of evidence used.

Aim

To quantify and analyse the quality of evidence that is presented in national guidelines.

Design and setting

Levels of evidence used in all the current valid recommendations in the Scottish Intercollegiate Guideline Network (SIGN) guidelines were reviewed and statistically analysed.

Method

The data was collected from published guidelines available online to the public. SIGN methodology entails a professional group selected by a national organisation to develop each of these guidelines. Statistical analysis of the relationship between the number of guideline recommendations and the quality of evidence used in its recommendations was performed.

Result

A significant correlation between the number of recommendations in a guideline and the use of level D evidence was discovered.

Conclusion

Practice guidelines should be brief and based on scientific evidence. Paradoxically the longest guidelines have the highest proportion of recommendations based on the lowest level of evidence. Guideline developers should be more aware of the need for brevity and a stricter application of evidence-based principles could achieve this. The findings support calls for a review of how evidence is used and presented in guidelines.

Article Summary

Article focus

Examines the hypothesis that larger guidelines rely disproportionately on poor evidence.

Key Messages

- One third of current national guidelines are supported only by case reports, case studies and expert opinion.
- Guidelines with large numbers of recommendations used a higher proportion of weak evidence.
- Guideline development groups appear to vary in their approach to offering recommendations.
- Guideline recommendations should be based on good evidence.
- Paucity of evidence should highlight topics for research.

Strengths and limitations of the study

Strengths.

- This is the first objective evidence of inconsistencies in approach by a national guideline developer
- This supports commentator suggestion that even without good evidence a group will prefer consensus.
- Adds to the current debate about how guidelines might be developed in the future

Limitations.

- The study is limited to only one set of national guidelines (SIGN).
- Reasons for the differences in quality of evidence preferred by the guideline development groups is unclear

Guidelines – is bigger better? A review of SIGN guidelines.

INTRODUCTION

The Scottish Intercollegiate Guidelines Network (SIGN) was founded in 1993. It is a national body, professionally led and publicly funded. SIGN's founding principles proposed direct links between evidence and recommendations, offering a brief and succinct quick-reference guide for clinicians [1]. Guidelines anticipated presenting brief, evidence based clinical advice. They have developed into long and authoritative texts used by managers and politicians to inform policy. A formal arrangement between SIGN and the National Institute of Clinical Excellence (NICE) has existed from 2003. Both have responsibility to consider cost-effectiveness and input to the Quality Outcomes Framework (QOF).

The World Health Organisation (WHO) recognises that current grades of recommendation (Table 1) may be ambiguous[2] and encourages guideline developers to use a system which includes a category "Use only in the context of research" where doubt exists.

Guideline developers have conflict of interest policies reported as challenging to apply. Where doubt exists, groups of specialists may feel consensus more defensible than acknowledging uncertainty. [3]

Even with the best evidence, concerns are expressed about the relevance of guidelines in treating patients with multiple morbidities[4], and the emergence of the phenomenon of reversal[5,6], where established practice, sometimes evidence based, is shown to be sub-optimal or harmful. This study looks at the quality of evidence used for SIGN guidelines, and describes a significant trend for some groups to emphasise poorly evidence-backed recommendations.

METHODS

SIGN guidelines were accessed online in September 2013. SIGN guidelines were chosen because they are internationally respected, the authors were familiar with their format and they contribute to national government policy. Guidelines that were “Current” and “Current 3-7 years”. Some recommendations may be out of date.” were included. Those that had been "Withdrawn", "Recommendations being updated", "Need for update being considered" and those with no recommendations were excluded.

SIGN guideline 50 clearly describes an established process for developing guidelines.[7] It explains how the process is planned, how it is implemented and by whom. This process is independent of this study, but is stated to be an objective process. SIGN guidelines have four grades of recommendation outlined in table 1. Table 2 describes the level of evidence SIGN uses to support the recommendation grading. SIGN guideline development groups vary in size depending on the scope of the topic under consideration, but generally comprise between 15 and 25 members. SIGN states they are aware of the many psychosocial factors, including the problems of overcoming professional hierarchies that can affect small group processes.

3 investigators (JRL, AGB, ABB) independently enumerated the level of evidence used by each guideline. They discounted any duplication implicit in text-embedded key recommendations and also implementation recommendations. There were no discrepancies. A statistical analysis of the correlation between the proportion of level D evidence and the total number of recommendations was performed for the 42 guidelines.

RESULTS

The 42 guidelines consisted of 2559 pages (including references), ranging from 26 to 161 (median 59.5) pages. The longest guideline, number 116 was 61 pages longer than the next largest. The number of recommendations per page ranged from 0.2 to 1.8 (median 0.7). The number of recommendations per guideline is presented in table 3.

Of the 1999 recommendations, 480 (24.0%) were level A, 491(24.6 %) were level B, 318 (15.9%) level C, and 710 (35.5%) level D. Thus 51.4% were poorly evidenced (C&D) and over a third (D) developed almost entirely on "expert opinion". The number of level A recommendations per guideline ranged from 0-57 (median 9), level B from 2-62 (median 8.5) level C ranged from 0-26 (median 6) and D from 0-60 (median 14.5). 4 guidelines had no level A evidence.

The proportion of level D evidence increases with the number of recommendations made. This correlation is significant with Kendall's Tau=0.22 [approximate 95% confidence interval 0.008-0.45] p value =0.04, and Spearman rho=0.22 [approximate 95% CI 0.02-0.57] p value= .04.

DISCUSSION

This study reveals that expert groups who produce long guidelines rely on poor evidence more heavily than others. While this study only looks at SIGN, this study highlights a problem that has escaped national guideline developers, a wide range of professionals and the public to whom these guidelines are applied. National guidelines are useful and important and there is a debate about how evidence is best presented. Guidelines define standards of care, help busy clinicians and allow managers and politicians to develop governance. An American study (using 3 not 4 levels of evidence) similarly found that 48% were "based on expert opinion, case studies, or standards of care." [8]; we show comparable results for current SIGN guidelines. Where patients are involved in clinical decisions, honestly declaring uncertainty has merit. In the absence of good scientific evidence, recommending a course of action without understanding the circumstances of the individual to whom it is applied seems both risky and, assuming a right to patient choice, unwarranted. Other guidelines that use high levels of poor evidence should evaluate the proportion of poorly evidenced recommendations and seek explanations for such trends.

This study did not examine why longer guidelines use poorer evidence. Groups of experts, indulging in "group think" may view their own opinion as more authoritative

than science can support[9]. It has been postulated that there is security in "just doing what everyone else is doing – even if what everyone else is doing isn't very good."[3] Reliance on expert opinion has a poor track record. Blinded by certainty, expert groups defining established practice have, in the past, perpetuated radical mastectomy instead of conservative surgery, Class 1C antiarrhythmics[10], pulmonary artery catheters in heart failure[11], electronic foetal monitoring in low risk pregnancies: even then practice can take a decade to reverse[12].

Even good evidence is subject to the phenomenon of reversal where new evidence contradicts current practice. Reversal can affect around 13-16% of publications [5,6]. This may partly explain why the implementation of even the most soundly evidence based national guidelines fails to improve outcome [13-15] . There is potential harm[16,17] from guidelines in real clinical settings, for example increasing radiation dose without benefit[18] or increased risks of anticoagulation[19].

SIGN 116 (diabetes), is a notable exception. It is more than 50% larger than the next largest, 2.5 times longer than the average and yet uses the 4th lowest level D recommendations. There are a number of hypotheses why this group reports differently. SIGN guidelines inform Quality Outcomes Framework (QOF) policy. Diabetes is the largest clinical UK QOF indicator and is associated with substantial payment incentives. The need for objective evaluation of performance drives a use of surrogate outcomes without appropriate clinical endpoints. [20] Diabetes guidelines have suffered several noteworthy reversals. Examples include the recommendation of glycosylated haemoglobin reduction resulting in increased use of rosiglitazone (still mentioned in the current document) both associated with harm including mortality. [21,22] Aspirin recommendations have also been changed from previous guidelines. Is it possible that the repeated use of surrogate outcomes arises from group dynamics driven by a powerful external agenda?

Many doctors whose expertise cross several guidelines[23,24] express concerns about guideline development groups. The inappropriate exclusion of disease groups from general population data is common. Smoking cessation advice for testicular

cancer survivors is level C, although studies in the general population (without excepting specific disease groups) advises everyone to stop smoking. Overall smoking cessation was level D and C once each and B on three occasions. Using evidence in this way may imply group dysfunction. Differently constituted groups, or greater oversight might avoid problems.

In 1993, SIGN guidelines stated intention was to be evidence based, brief and succinct. Brevity increases value as a quick reference guide. Removing or reducing poorly evidenced recommendations would reduce size by more than a third overall and in some up to two thirds. The two volumes Oxford Textbook of Primary Medical Care (2005) is a relatively brief 1420 pages, more than a thousand less than the 2559 pages of guidelines. Evidence based medicine is described as "the use of mathematical estimates of the risk of benefit and harm, derived from high-quality research on population samples, to inform clinical decision-making in the diagnosis, investigation or management of individual patients"[25]. Guidelines relevance to daily practice, the reliability of evidence and whether the application of evidence will improve outcomes are important questions.

These results may reflect how professional groups deal with uncertainty. If so, this is not good for individual patients faced with the same uncertainties (whether aware of it or not), nor is it good for scientists who actively seek unanswered questions by challenging established practice, an area in which medicine has a poor record from Semmelweis to the present day.

The finding of a significant increase of level D recommendations in larger guidelines has not happened by chance. A wider debate about how guideline groups can create greater clarity about the reliability of evidence used is needed.[26] Reducing the use of poorly evidenced recommendations has potential to create a shorter, more reliable and usable clinical support. The GRADE working group was formed in 2000.[27] SIGN moved to a new grading system in 2001[28] and from 2013 a new system based on GRADE principles. Whether these changes will resolve the challenges that underpin the inconsistencies we have outlined remains to be seen.

The authors would like to acknowledge the help of Heather Barrington, Dumfries & Galloway Royal Infirmary, Research & Development Support Unit
Dumfries, Dumfries & Galloway, United Kingdom who gave statistical advice and support, and Anne B Baird, The white House, Sandhead, Wigtownshire DG9 9JA who checked the data for inconsistencies.

Table 1

GRADES OF RECOMMENDATION

A	At least one meta-analysis, systematic review, or RCT rated as 1++, and directly applicable to the target population; or A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results
B	A body of evidence including studies rated as 2++, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 1++ or 1+
C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 2++
D	Evidence level 3 or 4; or Extrapolated evidence from studies rated as 2+

Table 2

1++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1 -	-Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort studies
2+	High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2 -	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, eg case reports, case series

4	Expert opinion

Table 3

Main Recommendations

Number	Name	Pages	A	B	C	D	*Total*	%age D
133	Management of hepatitis C	57	20	24	7	52	***103***	50.5%
132	Long-term follow-up of survivors of childhood cancer	62	0	7	9	14	***30***	46.7%
131	Management of schizophrenia	64	10	19	3	15	***47***	31.9%
130	Brain injury rehabilitation in adults	68	0	14	7	8	***29***	27.6%
129	Antithrombotics: indication and management	68	25	11	6	19	***61***	31.1%
127	Management of perinatal mood disorders	47	0	5	6	15	***26***	57.7%
126	Diagnosis and management of colorectal cancer	56	11	19	15	29	***74***	39.2%
125	Management of atopic eczema in primary care	34	3	5	3	2	***13***	15.4%
124	Management of adult testicular germ cell tumours	63	6	6	9	21	***42***	50.0%
123	Management of early rheumatoid arthritis	27	3	7	2	0	***12***	0.0%
122	Prevention and management of venous thromboembolism	88	26	15	14	55	***110***	50.0%
121	Diagnosis and management of psoriasis and psoriatic arthritis in adults	65	11	16	6	26	***59***	44.1%
120	Management of chronic venous leg ulcers	46	5	3	4	7	***19***	36.8%
119	Management of patients with stroke: identification and management of dysphagia	42	0	6	4	20	***30***	66.7%
118	Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning	101	21	29	7	21	***78***	26.9%
117	Management of sore throat and indications for tonsillectomy	37	9	3	4	4	***20***	20.0%
116	Management of diabetes	161	57	62	23	16	***158***	10.1%
115	Management of Obesity	87	6	11	7	11	***35***	31.4%
114	Nonpharmaceutical management of depression	37	5	4	0	0	***9***	0.0%
113	Diagnosis and pharmacological management of Parkinson's disease	61	12	6	6	4	***28***	14.3%
112	Management of attention deficit and hyperkinetic disorders in children and young people	45	6	4	3	4	***17***	23.5%
111	Management of hip fracture in old people	49	10	9	8	14	***41***	34.1%
110	Early management of patients with a head injury	76	1	7	6	17	***31***	54.8%
109	Management of genital Chlamydia trachomatis infection	40	3	6	9	29	***47***	61.7%
108	Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention	100	42	27	18	14	***101***	13.9%
107	Diagnosis and management of headache in adults	81	17	16	9	34	***76***	44.7%
106	Control of pain in adults with cancer	71	5	7	3	19	***34***	55.9%
105	Management of acute upper and lower gastrointestinal bleeding	57	14	5	2	15	***36***	41.7%
103	Diagnosis and management of chronic kidney disease	50	9	6	4	3	***22***	13.6%
102	Management of invasive meningococcal disease in children and young people	46	1	4	6	26	***37***	70.3%
99	Management of cervical cancer	73	1	13	19	29	***62***	46.8%
97	Risk estimation and the prevention of cardiovascular disease	72	16	12	2	4	***34***	11.8%
96	Management of stable angina	59	13	10	3	11	***37***	29.7%
95	Management of chronic heart failure	55	9	12	1	1	***23***	4.3%
94	Cardiac arrhythmias and coronary heart disease	42	22	11	13	23	***69***	33.3%
93	Acute coronary syndromes	60	11	14	9	8	***42***	19.0%
91	Bronchiolitis in children	42	4	3	6	14	***27***	51.9%
90	Diagnosis and management of head and neck cancer	92	42	8	26	60	***136***	44.1%
89	Diagnosis and management of peripheral arterial disease	37	11	2	0	4	***17***	23.5%
88	Management of suspected bacterial urinary tract infection in adults	45	8	10	2	10	***30***	33.3%
87	Management of oesophageal and gastric cancer	70	3	26	23	28	***80***	35.0%
61	Investigation of postmenopausal bleeding	26	2	7	4	4	***17***	23.5%
	TOTAL	2559	480	491	318	710	1999	

Acknowledgements.

Heather Barrington, Statistical Adviser; Bridget Bird, Administrative Assistant; Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries DG1 4AP.
Dr A Gordon Baird affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; any discrepancies from the study as planned have been explained.

Contributorship Statement

Both authors discussed the hypothesis that there was a disproportionate use of poor evidence in longer guidelines; both checked the raw data, and agreed on a statistical approach to discover whether the trend was significant or not. Both have been involved in writing and researching the evidence.

Funding.
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data Sharing.
Technical appendix, statistical code, and dataset available from the corresponding author, who will provide a permanent, citable and open access home for the dataset.

Competing interests
No competing interests

MeSH Headings

- Practice Guideline N04.761.700.350.650
- General Practice H02.403.340
- Evidence-Based Medicine H02.249.750

How this fits in.
Guidelines should encourage an evidence-based approach to clinical practice. Longer guidelines used significantly higher levels of poor evidence. WHO has proposed a different system of grading evidence. The effect of group behavior altering guideline development has been hypothesised. New research often challenges established clinical practice. Improving the quality of evidence, acknowledging uncertainty and shortening guideline length would make guidelines more clinically relevant and effective.

REFERENCES

1. Harbour R, Lowe G, Twaddle S. Scottish Intercollegiate Guidelines Network; the first 15 years (1993-2008). J R Coll Physicians Edinb 2011 Jun;41(2) 163-168

2. de Joncheere K, Hill S, Klazinga N, et al.The Clinical Guideline Programme of the National Institute for Health and Clinical Excellence (NICE) A review by the World Health Organization May 2006

3. Lenzer J. Why we can't trust clinical guidelines: BMJ 2013;346:f3830

4.Aylett V. Do geriatricians need guidelines? BMJ 2010 Sep;:341 :c5340

5. Prasad V, Cifu, A, Ioannidis JP. Reversals of Established Medical Practices: Evidence to Abandon Ship. JAMA. 2012 Jan4;307(1):37-38,

6. Prasad V, Gall V, Cifu A. The frequency of medical reversal. Arch Int Med. 2011 Oct 10;. 171(18): 1675-1676

7. SIGN 50: A guideline developer's handbook http://www.sign.ac.uk/pdf/sign50.pdf

8.Tricoci P, Allen JM, Kramer JM, et al. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA 2009 Feb25; 301 (8): 831-841

9. Raine, R, Sanderson C, Black N. Developing clinical guidelines: a challenge to current methods. BMJ. 331(7517):631-633, September 17, 2005

10. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo - The Cardiac Arrhythmia Suppression Trial. New Eng J Med. 1991 Mar21; 324 (12): 781-788

11. Binanay C, Califf RM, Hasselblad V, et al. Evaluation study of congestive heart failure and pulmonary artery catheterization effectiveness: the ESCAPE trial. JAMA. 2005 Oct 5;294(13):1625-33.

12. Tatsioni A, Bonitsis NG, Ioannidis JA. Persistence of Contradicted Claims in the Literature. JAMA. 2007 Dec5 ;298(21):2517-2526.

13. Anderson JE, McKenzie C, Singh N, et al. Compliance with the 62 day target does not improve long-term survival. Association of Coloproctology of Great Britain and Ireland Annual Meeting 2012 Dublin Ireland.

14. Kerr J, Smith R, Gray S, et al. An audit of clinical practice in the management of head injured patients following the introduction of the Scottish Intercollegiate Guidelines Network (SIGN) recommendations. Emerg Med J.2005 Dec; 22(12):850-854

15. Caplan LR. How well does "evidence-based" medicine help neurologists care for individual patients? Revin Neurol Dis.2007;4(2):75-84

16.Hutchison G. Guidelines can harm patients too. BMJ (Clinical research ed.). 2012 Apr18;344: e2685.

17. Woolf SH, Grol R, Hutchinson A, et al. Clinical guidelines. Potential benefits, limitations, and harms of clinical guidelines. BMJ 1999 Feb20;318(7182):527-530

18. Miller L., Kidd A. Are sign guidelines key in the decision to CT? Post-head injury CT scanning within a paediatric population. Academic Emergency Medicine. Conference: 14th International Conference on Emergency Medicine, ICEM 2012 Dublin Ireland. Conference Publication: 19 (6) (pp 776), 2012. Date of Publication: June 2012.

19. Thomson R., Eccles M., Wood R., et al. A cautionary note on data sources for evidence-based clinical decisions: Warfarin and stroke prevention. MedDecis Making. 2007;27 (4): 438-447

20. Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. BMJ. 2011 Dec28;343:d7995

21. Cohen D. Rosiglitazone: what went wrong? BMJ.2010 Sep; 341:c4848

22. Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. Action to Control Cardiovascular Risk in Diabetes Study Group. New Engl J Med. 2008 Jun12;358(24):2545-59

23. Rashidian A, Eccles MP, Russell I. Falling on stony ground? A qualitative study of implementation of clinical guidelines' prescribing recommendations in primary care. Health Policy.2008 Feb; 85 (2) 148-161

24.Guy S, Wardlaw JM. Who writes guidelines, and who should? ClinRadiol. 2002 Oct ;57 (10): 891-897.

25. Greenhalgh, Trisha. How To Read a Paper: The Basics of Evidence-Based Medicine. Wiley-Blackwell, fourth edition, 2010, p. 1

26. Zuiderent-Jerak T, Forland F, Macbeth F. Guidelines should reflect all knowledge, not just trials BMJ 2012;345:e6702

27.GRADE working group. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490

28. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ. 2001 Aug 11;323(7308):334-6

~~SIGN~~ Scottish Intercollegiate Guideline Network~~Guidelines~~ – analysis of evidence levels for their recommendations.

Alternatively

Guideline ~~s – is bigger~~ recommendations are more better? A review of Scottish Intercollegiate Guideline Network ~~SIGN~~ national guidelines.

Dr A Gordon Baird, MB ChB MRCOG FRCGP R&D General Practice Adviser, Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries, Scotland, UK DG1 4AP.

Address for correspondence; Dr A Gordon Baird, The White House, Sandhead, Wigtownshire UK. DG9 9JA. Email gordon.baird@me.com. Phone/fax 01776830281

Dr James R Lawrence, BSc MD FRCP(Glas & Edin) R&D Clinical Lead, Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries, Scotland, UK DG1 4AP.

MeSH headings
General practice / Family practice
HEALTH SERVICES ADMINISTRATION & MANAGEMENT
Protocols & guidelines
Quality in health care
Clinical governance

ABSTRACT

Background

National guidelines have become an integral part of UK general practice and aim to effectively deliver quality and consistency in clinical practice. As with any intervention, there are negative as well as positive consequences. Guideline effectiveness depends on the quality of evidence used.

Aim

To quantify and analyse the quality of evidence that is presented in national guidelines.

Design and setting

Levels of evidence used in all the current valid recommendations in the Scottish Intercollegiate Guideline Network (SIGN) guidelines were reviewed and statistically analysed.

Method

The data was collected from published guidelines available online to the public. SIGN methodology entails a professional group selected by a national organisation to develop each of these guidelines. Statistical analysis of the relationship between the number of guideline recommendations and the quality of evidence used in its recommendations was performed.

Result

A significant correlation between the number of recommendations in a guideline and the use of level D evidence was discovered.

Conclusion

Practice guidelines should be brief and based on scientific evidence. Paradoxically the longest guidelines have the highest proportion of recommendations based on the lowest level of evidence. Guideline developers should be more aware of the need for brevity and a stricter application of evidence-based principles could achieve this. The findings support calls for a review of how evidence is used and presented in guidelines.

Article Summary

Article focus

Examines the hypothesis that larger guidelines rely disproportionately on poor evidence.

Key Messages

- One third of current national guidelines are supported only by case reports, case studies and expert opinion.
- Guidelines with large numbers of recommendations used a higher proportion of weak evidence.
- Guideline development groups appear to vary in their approach to offering recommendations.
- Guideline recommendations should be based on good evidence.
- Paucity of evidence should highlight topics for research.

Strengths and limitations of the study

Strengths.

- This is the first objective evidence of inconsistencies in approach by a national guideline developers
- This supports commentator suggestion that even without good evidence a group will prefer consensus.
- Adds to the current debate about how guidelines might be developed in the future

Limitations.

- The study is limited to only one set of national guidelines (SIGN).
- Reasons for the differences in quality of evidence preferred by the guideline development groups is unclear

Guidelines – is bigger better? A review of SIGN guidelines.

INTRODUCTION

The Scottish Intercollegiate Guidelines Network (SIGN) was founded in 1993. It is a national body, professionally led and publicly funded. SIGN's founding principles proposed direct links between evidence and recommendations, offering a brief and succinct quick-reference guide for clinicians [1]. Guidelines anticipated presenting brief, evidence based clinical advice. They have developed into long and authoritative texts used by managers and politicians to inform policy. A formal arrangement between SIGN and the National Institute of Clinical Excellence (NICE) has existed from 2003. ~~SIGN has~~Both have responsibility to consider cost-effectiveness and ~~directly~~ input~~s~~ to the Quality Outcomes Framework (QOF). ~~A formal arrangement between SIGN and the National Institute of Clinical Excellence (NICE) has existed from 2003.~~

The World Health Organisation (WHO) recognises that current grades of recommendation (Table 1) may be ambiguous[2] and encourages guideline developers to use a system which includes a category "Use only in the context of research" where doubt exists.

Guideline developers have conflict of interest policies reported as challenging to apply. Where doubt exists, groups of specialists may feel consensus more defensible than acknowledging uncertainty. [3]

Even with the best evidence, concerns are expressed about the relevance of guidelines in treating patients with multiple morbidities[4], and the emergence of the phenomenon of reversal[5,6], where established practice, sometimes evidence based, is shown to be sub-optimal or harmful. This study looks at the quality of evidence used for SIGN guidelines, and describes a significant trend for some groups to emphasise poorly evidence-backed recommendations.

METHODS

SIGN guidelines were accessed online in September 2013. SIGN guidelines were chosen because they are internationally respected, the authors were familiar with their format and they contribute to national government policy. Guidelines that were “Current” and “Current 3-7 years”. Some recommendations may be out of date.” were included. Those that had been "Withdrawn", "Recommendations being updated", "Need for update being considered" and those with no recommendations were excluded.

SIGN guideline 50 clearly describes an established process for developing guidelines.[7] It explains how the process is planned, how it is implemented and by whom. This process is independent of this study, but is stated to be an objective process. SIGN guidelines have four grades of recommendation outlined in table 1. Table 2 describes the level of evidence SIGN uses to support the recommendation grading. SIGN guideline development groups vary in size depending on the scope of the topic under consideration, but generally comprise between 15 and 25 members. SIGN states they are aware of the many psychosocial factors, including the problems of overcoming professional hierarchies that can affect small group processes.

3 investigators (JRL, AGB, ABB) independently enumerated ~~The~~the level of evidence used by each ~~examined~~ guideline ~~was independently enumerated by 3 investigators. and discrepancies resolved.~~ They discounted any duplication implicit in text-embedded key recommendations and also implementation recommendations. There were no discrepancies. A statistical analysis of the correlation between the proportion of level D evidence and the total number of recommendations was performed for the 42 guidelines.

RESULTS

The 42 guidelines consisted of 2559 pages (including references), ranging from 26 to 161 (median 59.5) pages. The longest guideline, number 116 was 61 pages longer than the next largest. The number of recommendations per page ranged from 0.2 to

1.8 (median 0.7). The number of recommendations per guideline is presented in table 3.

Of the 1999 recommendations, 480 (24.0%) were level A, 491(24.6 %) were level B, 318 (15.9%) level C, and 710 (35.5%) level D. Thus 51.4% were poorly evidenced (C&D) and over a third (D) developed almost entirely on "expert opinion". The number of level A recommendations per guideline ranged from 0-57 (median 9), level B from 2-62 (median 8.5) level C ranged from 0-26 (median 6) and D from 0-60 (median 14.5). 4 guidelines had no level A evidence.

The proportion of level D evidence increases with the number of recommendations made. This correlation is significant with Kendall's Tau=0.22 [approximate 95% confidence interval 0.008-0.45] p value =0.04, and Spearman rho=0.22 [approximate 95% CI 0.02-0.57] p value= .04.

DISCUSSION

This study reveals that expert groups who produce long guidelines rely on poor evidence more heavily than others. While this study only looks at SIGN, this study highlights a problem that has escaped national guideline developers, a wide range of professionals and the public to whom these guidelines are applied. National guidelines are useful and important and there is a debate about how evidence is best presented. Guidelines define standards of care, help busy clinicians and allow managers and politicians to develop governance. An American study (using 3 not 4 levels of evidence) similarly found that 48% were "based on expert opinion, case studies, or standards of care." [8]; we show comparable results for current SIGN guidelines. Where patients are involved in clinical decisions, honestly declaring uncertainty has merit. In the absence of good scientific evidence, recommending a course of action without understanding the circumstances of the individual to whom it is applied seems both risky and, assuming a right to patient choice, unwarranted. Other guidelines that use high levels of poor evidence should evaluate the proportion of poorly evidenced recommendations and seek explanations for such trends.

This study did not examine why longer guidelines use poorer evidence. ~~Cloistered~~ ~~G~~groups of experts, indulging in "group think" may view their own opinion as more authoritative than science can support[9]. It has been postulated that there is security in "just doing what everyone else is doing – even if what everyone else is doing isn't very good."[3] Reliance on expert opinion has a poor track record. Blinded by certainty, expert groups defining established practice have, in the past, perpetuated radical mastectomy instead of conservative surgery, Class 1C antiarrhythmics[10], pulmonary artery catheters in heart failure[11], electronic foetal monitoring in low risk pregnancies: even then practice can take a decade to reverse[12].

Comment [GB1]: New reference

Even good evidence is subject to the phenomenon of reversal where new evidence contradicts current practice. Reversal can affect around 13-16% of publications [5,6]. This may partly explain why the implementation of even the most soundly evidence based national guidelines fails to improve outcome [13-15] . There is potential harm[16,17] from guidelines in real clinical settings, for example increasing radiation dose without benefit[18] or increased risks of anticoagulation[19].

SIGN 116 (diabetes), is a notable exception. It is more than 50% larger than the next largest, 2.5 times longer than the average and yet uses the 4th lowest level D recommendations. There are a number of hypotheses why this group reports differently. SIGN guidelines inform Quality Outcomes Framework (QOF) policy. Diabetes is the largest clinical UK QOF indicator and is associated with substantial payment incentives. The need for objective evaluation of performance drives a use of surrogate outcomes without appropriate clinical endpoints. [20] Diabetes guidelines have suffered several noteworthy reversals. Examples include the recommendation of glycosylated haemoglobin reduction resulting in increased use of rosiglitazone (still mentioned in the current document) both associated with harm including mortality. [21,22] Aspirin recommendations have also been changed from previous guidelines. Is it possible that the repeated use of surrogate outcomes arises from group dynamics driven by a powerful external agenda?

Many doctors whose expertise cross several guidelines[23,24] express concerns about guideline development groups. The inappropriate exclusion of disease groups from general population data is common. Smoking cessation advice for testicular cancer survivors is level C, although studies in the general population (without excepting specific disease groups) advises everyone to stop smoking. Overall smoking cessation was level D and C once each and B on three occasions. Using evidence in this way may imply group dysfunction. Differently constituted groups, or greater oversight might avoid problems.

In 1993, SIGN guidelines stated intention was to be evidence based, brief and succinct. Brevity increases value as a quick reference guide. Removing or reducing poorly evidenced recommendations would reduce size by more than a third overall and in some up to two thirds. The two volumes Oxford Textbook of Primary Medical Care (2005) is a relatively brief 1420 pages, more than a thousand less than the 2559 pages of guidelines. Evidence based medicine is described as "the use of mathematical estimates of the risk of benefit and harm, derived from high-quality research on population samples, to inform clinical decision-making in the diagnosis, investigation or management of individual patients"[25]. Guidelines relevance to daily practice, the reliability of evidence and whether the application of evidence will improve outcomes are important questions.

These results may reflect how professional groups deal with uncertainty. If so, this is not good for individual patients faced with the same uncertainties (whether aware of it or not), nor is it good for scientists who actively seek unanswered questions by challenging established practice, an area in which medicine has a poor record from Semmelweis to the present day.

The finding of a significant increase of level D recommendations in larger guidelines has not happened by chance. A wider debate about how guideline groups can create greater clarity about the reliability of evidence used is needed.[26] Reducing the use of poorly evidenced recommendations has potential to create a shorter, more

reliable and usable clinical support. The GRADE working group was formed in 2000.[27][2] ~~-~~ SIGN ~~proposed a move~~moved to a new grading system in 2001[28].~~:~~[3] and from 2013 a new system based on GRADE principles. Whether these changes ~~in process at present being considered~~ will resolve the challenges that underpin the inconsistencies we have outlined remains to be seen.

The authors would like to acknowledge the help of Heather Barrington, Dumfries & Galloway Royal Infirmary, Research & Development Support Unit
Dumfries, Dumfries & Galloway, United Kingdom who gave statistical advice and support, and Anne B Baird, The white House, Sandhead, Wigtownshire DG9 9JA who checked the data for inconsistencies.

Table 1

GRADES OF RECOMMENDATION

A	At least one meta-analysis, systematic review, or RCT rated as 1++, and directly applicable to the target population; or A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results
B	A body of evidence including studies rated as 2++, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 1++ or 1+
C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 2++
D	Evidence level 3 or 4; or Extrapolated evidence from studies rated as 2+

Table 2

1++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1 -	-Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort studies

2+	High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2 -	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, eg case reports, case series
4	Expert opinion

Table 3

Main Recommendations

Number	Name	Pages	A	B	C	D	*Total*	%age D
133	Management of hepatitis C	57	20	24	7	52	***103***	50.5%
132	Long-term follow-up of survivors of childhood cancer	62	0	7	9	14	***30***	46.7%
131	Management of schizophrenia	64	10	19	3	15	***47***	31.9%
130	Brain injury rehabilitation in adults	68	0	14	7	8	***29***	27.6%
129	Antithrombotics: indication and management	68	25	11	6	19	***61***	31.1%
127	Management of perinatal mood disorders	47	0	5	6	15	***26***	57.7%
126	Diagnosis and management of colorectal cancer	56	11	19	15	29	***74***	39.2%
125	Management of atopic eczema in primary care	34	3	5	3	2	***13***	15.4%
124	Management of adult testicular germ cell tumours	63	6	6	9	21	***42***	50.0%
123	Management of early rheumatoid arthritis	27	3	7	2	0	***12***	0.0%
122	Prevention and management of venous thromboembolism	88	26	15	14	55	***110***	50.0%
121	Diagnosis and management of psoriasis and psoriatic arthritis in adults	65	11	16	6	26	***59***	44.1%
120	Management of chronic venous leg ulcers	46	5	3	4	7	***19***	36.8%
119	Management of patients with stroke: identification and management of dysphagia	42	0	6	4	20	***30***	66.7%
118	Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning	101	21	29	7	21	***78***	26.9%
117	Management of sore throat and indications for tonsillectomy	37	9	3	4	4	***20***	20.0%
116	Management of diabetes	161	57	62	23	16	***158***	10.1%
115	Management of Obesity	87	6	11	7	11	***35***	31.4%
114	Nonpharmaceutical management of depression	37	5	4	0	0	***9***	0.0%
113	Diagnosis and pharmacological management of Parkinson's disease	61	12	6	6	4	***28***	14.3%
112	Management of attention deficit and hyperkinetic disorders in children and young people	45	6	4	3	4	***17***	23.5%
111	Management of hip fracture in old people	49	10	9	8	14	***41***	34.1%
110	Early management of patients with a head injury	76	1	7	6	17	***31***	54.8%
109	Management of genital Chlamydia trachomatis infection	40	3	6	9	29	***47***	61.7%
108	Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention	100	42	27	18	14	***101***	13.9%
107	Diagnosis and management of headache in adults	81	17	16	9	34	***76***	44.7%
106	Control of pain in adults with cancer	71	5	7	3	19	***34***	55.9%
105	Management of acute upper and lower gastrointestinal bleeding	57	14	5	2	15	***36***	41.7%
103	Diagnosis and management of chronic kidney disease	50	9	6	4	3	***22***	13.6%
102	Management of invasive meningococcal disease in children and young people	46	1	4	6	26	***37***	70.3%
99	Management of cervical cancer	73	1	13	19	29	***62***	46.8%
97	Risk estimation and the prevention of cardiovascular disease	72	16	12	2	4	***34***	11.8%
96	Management of stable angina	59	13	10	3	11	***37***	29.7%
95	Management of chronic heart failure	55	9	12	1	1	***23***	4.3%
94	Cardiac arrhythmias and coronary heart disease	42	22	11	13	23	***69***	33.3%
93	Acute coronary syndromes	60	11	14	9	8	***42***	19.0%
91	Bronchiolitis in children	42	4	3	6	14	***27***	51.9%
90	Diagnosis and management of head and neck cancer	92	42	8	26	60	***136***	44.1%
89	Diagnosis and management of peripheral arterial disease	37	11	2	0	4	***17***	23.5%
88	Management of suspected bacterial urinary tract infection in adults	45	8	10	2	10	***30***	33.3%
87	Management of oesophageal and gastric cancer	70	3	26	23	28	***80***	35.0%
61	Investigation of postmenopausal bleeding	26	2	7	4	4	***17***	23.5%
	TOTAL	2559	480	491	318	710	1999	

Acknowledgements.

Heather Barrington, Statistical Adviser; Bridget Bird, Administrative Assistant; Research and Development Support Unit, Dumfries and Galloway Royal infirmary, Bankend Road, Dumfries DG1 4AP.

Dr A Gordon Baird affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; any discrepancies from the study as planned have been explained.

Contributorship Statement

Both authors discussed the hypothesis that there was a disproportionate use of poor evidence in longer guidelines; both checked the raw data, and agreed on a statistical approach to discover whether the trend was significant or not. Both have been involved in writing and researching the evidence.

Funding.
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data Sharing.
Technical appendix, statistical code, and dataset available from the corresponding author, who will provide a permanent, citable and open access home for the dataset.

Competing interests
No competing interests

MeSH Headings

- Practice Guideline N04.761.700.350.650
- General Practice H02.403.340
- Evidence-Based Medicine H02.249.750

How this fits in.
Guidelines should encourage an evidence-based approach to clinical practice. Longer guidelines used significantly higher levels of poor evidence. WHO has proposed a different system of grading evidence. The effect of group behavior altering guideline development has been hypothesised. New research often challenges established clinical practice. Improving the quality of evidence, acknowledging uncertainty and shortening guideline length would make guidelines more clinically relevant and effective.

REFERENCES

1. Harbour R, Lowe G, Twaddle S. Scottish Intercollegiate Guidelines Network; the first 15 years (1993-2008). J R Coll Physicians Edinb 2011 Jun;41(2) 163-168

2. de Joncheere K, Hill S, Klazinga N, Mäkelä M, Oxman AD.The Clinical Guideline Programme of the National Institute for Health and Clinical Excellence (NICE) A review by the World Health Organization May 2006

3. Lenzer J. Why we can't trust clinical guidelines: BMJ 2013;346:f3830

4.Aylett V. Do geriatricians need guidelines? BMJ 2010 Sep;:341 :c5340

5. Prasad V, Cifu, A, Ioannidis JP. Reversals of Established Medical Practices: Evidence to Abandon Ship. JAMA. 2012 Jan4;307(1):37-38,

6. Prasad V, Gall V, Cifu A. The frequency of medical reversal. Arch Int Med. 2011 Oct 10;. 171(18): 1675-1676

7. SIGN 50: A guideline developer's handbook http://www.sign.ac.uk/pdf/sign50.pdf

8.Tricoci P, Allen JM, Kramer JM, Califf RM, Smith Jr. SC. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA 2009 Feb25; 301 (8): 831-841

9. Raine, R, Sanderson C, Black N. Developing clinical guidelines: a challenge to current methods. BMJ. 331(7517):631-633, September 17, 2005

10. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, Arensberg D, Baker A, Friedman L, Greene HL, Huther ML, Richardson DW. Mortality and morbidity in patients receiving encainide, flecainide, or placebo - The Cardiac Arrhythmia Suppression Trial. New Eng J Med. 1991 Mar21; 324 (12): 781-788

11. Binanay C, Califf RM, Hasselblad V, O'Connor CM, Shah MR, Sopko G, Stevenson LW, Francis GS, Leier CV, Miller LW Evaluation study of congestive heart failure and pulmonary artery catheterization effectiveness: the ESCAPE trial. JAMA. 2005 Oct 5;294(13):1625-33.

12. Tatsioni A, Bonitsis NG, Ioannidis JA. Persistence of Contradicted Claims in the Literature. JAMA. 2007 Dec5 ;298(21):2517-2526.

13. Anderson JE, McKenzie C, Singh N, Gajree N, Giles L, Sharma P, McDonald A. Compliance with the 62 day target does not improve long-term survival. Association of Coloproctology of Great Britain and Ireland Annual Meeting 2012 Dublin Ireland.

14. Kerr J, Smith R, Gray S, Beard D, Robertson CE. An audit of clinical practice in the management of head injured patients following the introduction of the Scottish Intercollegiate Guidelines Network (SIGN) recommendations. Emerg Med J.2005 Dec; 22(12):850-854

15. Caplan LR. How well does "evidence-based" medicine help neurologists care for individual patients? Revin Neurol Dis.2007;4(2):75-84

16.Hutchison G. Guidelines can harm patients too. BMJ (Clinical research ed.). 2012 Apr18;344: e2685.

17. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines. Potential benefits, limitations, and harms of clinical guidelines. BMJ 1999 Feb20;318(7182):527-530

18. Miller L., Kidd A. Are sign guidelines key in the decision to CT? Post-head injury CT scanning within a paediatric population. Academic Emergency Medicine. Conference: 14th International Conference on Emergency Medicine, ICEM 2012 Dublin Ireland. Conference Publication: 19 (6) (pp 776), 2012. Date of Publication: June 2012.

19. Thomson R., Eccles M., Wood R., Chinn D.J. A cautionary note on data sources for evidence-based clinical decisions: Warfarin and stroke prevention. MedDecis Making. 2007;27 (4): 438-447

20. Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. BMJ. 2011 Dec28;343:d7995

21. Cohen D. Rosiglitazone: what went wrong? BMJ.2010 Sep; 341:c4848

22. Gerstein HC, Miller ME, Byington RP, Goff DC Jr., Bigger JT, Buse JB, Cushman WC, Genuth S, Ismail-Beigi F, Grimm RH Jr. Probstfield JL, Simons-Morton DG, Friedewald WT. Effects of intensive glucose lowering in type 2 diabetes. Action to Control Cardiovascular Risk in Diabetes Study Group. New Engl J Med. 2008 Jun12;358(24):2545-59

23. Rashidian A, Eccles MP, Russell I. Falling on stony ground? A qualitative study of implementation of clinical guidelines' prescribing recommendations in primary care. Health Policy.2008 Feb; 85 (2) 148-161

24.Guy S, Wardlaw JM. Who writes guidelines, and who should? ClinRadiol. 2002 Oct ;57 (10): 891-897.

25. Greenhalgh, Trisha. How To Read a Paper: The Basics of Evidence-Based Medicine. Wiley-Blackwell, fourth edition, 2010, p. 1

26. Zuiderent-Jerak T, Forland F, Macbeth F. Guidelines should reflect all knowledge, not just trials BMJ 2012;345:e6702

27.GRADE working group. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490

28. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ. 2001 Aug 11;323(7308):334-6

1 Zuiderent-Jerak T, Forland F, Macbeth F. Guidelines should reflect all knowledge, not just trials. BMJ 2012;345:e6702

2 GRADE working group http://www.gradeworkinggroup.org/about_us.htm

3 Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ. 2001 Aug 11;323(7308):334-6.

TABLE 1

Grades of recommendation

A	At least one meta-analysis ,systematic review, or RCT rated as 1++and directly applicable to the target population; or a body of evidence consisting principally of studies rated as 1+,directly applicable to the target population, and demonstrating overall consistency of results.
B	A body of evidence including studies rated as 2++, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 1++ or 1+.
C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or Extrapolated evidence from studies rated as 2+.
D	Evidence level 3 or 4; or Extrapolated evidence from studies rated as 2=.

Table 2

Levels of evidence

1++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1+	Well conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1 -	Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2++	High quality systematic reviews of case control or cohort studies. High quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2+	Well conducted case control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal
2 -	Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal
3	Non-analytic studies, eg case reports, case series
4	Expert opinion

Table 3 – for editing

Number	Number	Name	Pages	Main Recommendations A	B	C	D	*Total*	%age D
133	133	Management of hepatitis C	57	20	24	7	52	***103***	50.5%
132	132	Long-term follow-up of survivors of childhood cancer	62	0	7	9	14	***30***	46.7%
131	131	Management of schizophrenia	64	10	19	3	15	***47***	31.9%
130	130	Brain injury rehabilitation in adults	68	0	14	7	8	***29***	27.6%
129	129	Antithrombotics: indication and management	68	25	11	6	19	***61***	31.1%
127	127	Management of perinatal mood disorders	47	0	5	6	15	***26***	57.7%
126	126	Diagnosis and management of colorectal cancer	56	11	19	15	29	***74***	39.2%
125	125	Management of atopic eczema in primary care	34	3	5	3	2	***13***	15.4%
124	124	Management of adult testicular germ cell tumours	63	6	6	9	21	***42***	50.0%
123	123	Management of early rheumatoid arthritis	27	3	7	2	0	***12***	0.0%
122	122	Prevention and management of venous thromboembolism	88	26	15	14	55	***110***	50.0%
121	121	Diagnosis and management of psoriasis and psoriatic arthritis in adults	65	11	16	6	26	***59***	44.1%
120	120	Management of chronic venous leg ulcers	46	5	3	4	7	***19***	36.8%
119	119	Management of patients with stroke: identification and management of dysphagia	42	0	6	4	20	***30***	66.7%
118	118	Management of patients with stroke: rehabilitation, prevention and management of complications, and discharge planning	101	21	29	7	21	***78***	26.9%
117	117	Management of sore throat and indications for tonsillectomy	37	9	3	4	4	***20***	20.0%
116	116	Management of diabetes	161	57	62	23	16	***158***	10.1%
115	115	Management of Obesity	87	6	11	7	11	***35***	31.4%
114	114	Nonpharmaceutical management of depression	37	5	4	0	0	***9***	0.0%

Table 3 – for editing

113	113	Diagnosis and pharmacological management of Parkinson's disease	61	12	6	6	4	***28***	14.3%
112	112	Management of attention deficit and hyperkinetic disorders in children and young people	45	6	4	3	4	***17***	23.5%
111	111	Management of hip fracture in old people	49	10	9	8	14	***41***	34.1%
110	110	Early management of patients with a head injury	76	1	7	6	17	***31***	54.8%
109	109	Management of genital Chlamydia trachomatis infection	40	3	6	9	29	***47***	61.7%
108	108	Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention	100	42	27	18	14	***101***	13.9%
107	107	Diagnosis and management of headache in adults	81	17	16	9	34	***76***	44.7%
106	106	Control of pain in adults with cancer	71	5	7	3	19	***34***	55.9%
105	105	Management of acute upper and lower gastrointestinal bleeding	57	14	5	2	15	***36***	41.7%
103	103	Diagnosis and management of chronic kidney disease	50	9	6	4	3	***22***	13.6%
102	102	Management of invasive meningococcal disease in children and young people	46	1	4	6	26	***37***	70.3%
99	99	Management of cervical cancer	73	1	13	19	29	***62***	46.8%
97	97	Risk estimation and the prevention of cardiovascular disease	72	16	12	2	4	***34***	11.8%
96	96	Management of stable angina	59	13	10	3	11	***37***	29.7%
95	95	Management of chronic heart failure	55	9	12	1	1	***23***	4.3%
94	94	Cardiac arrhythmias and coronary heart disease	42	22	11	13	23	***69***	33.3%
93	93	Acute coronary syndromes	60	11	14	9	8	***42***	19.0%
91	91	Bronchiolitis in children	42	4	3	6	14	***27***	51.9%
90	90	Diagnosis and management of head and neck cancer	92	42	8	26	60	***136***	44.1%
89	89	Diagnosis and management of peripheral arterial disease	37	11	2	0	4	***17***	23.5%
88	88	Management of suspected bacterial urinary tract infection in adults	45	8	10	2	10	***30***	33.3%

Table 3 – for editing

87	87	Management of oesophageal and gastric cancer	70	3	26	23	28	***80***	35.0%
61	61	Investigation of postmenopausal bleeding	26	2	7	4	4	***17***	23.5%
		TOTAL	2559	480	491	318	710	1999	