BMJ Open  Guidelines: is bigger better?  
A review of SIGN guidelines

A Gordon Baird, James R Lawrence

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.5 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 Care Excellence (NICE) has existed from 2005. Both have responsibility to consider cost-effectiveness and input to the Quality and Outcomes Framework (QOF).

The WHO recognises that current grades of recommendation (box 1) may be ambiguous 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.6

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 7 8 where established practice, sometimes evidence based, is shown to be suboptimal 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 evidenced recommendations.

METHODS

SIGN guidelines were accessed online in September 2013. SIGN guidelines were chosen because they are internationally

ABSTRACT

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

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

Outcome measures: The proportion of level D evidence used in each guideline and a statistical analysis.

Method: Data were 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: 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% CI 0.008 to 0.45), $p=0.04$; and Spearman $\rho=0.22$ (approximate 95% CI 0.02 to 0.57), $p=0.04$.

Conclusions: 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.

Strengths and limitations of this study

- 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.
- The study is limited to only one set of national guidelines, that is, the Scottish Intercollegiate Guideline Network (SIGN).
- Reasons for the differences in quality of evidence preferred by the guideline development groups are unclear.
SIGN states they are aware of the 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.

Three investigators ([JRL, AGB and 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 2](#).

Of the 1999 recommendations, 480 (24%) 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 and D) and over a third (D) depend almost entirely on ‘expert opinion’. The number of level A recommendations per guideline ranged 0–57 (median 9), level B 2–62 (median 8.5), level C ranged 0–26 (median 6) and level D ranged 0–60 (median 14.5). Four 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 τ=0.22 (approximate 95% CI 0.008 to 0.45), p=0.04; and Spearman ρ=0.22 (approximate 95% CI 0.02 to 0.57), p=0.04.

**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’,5 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 many psychosocial factors, including the problems of overcoming professional hierarchies that can affect small group processes.

---

**Table 1** Levels of evidence

| Level | Description |
|-------|-------------|
| 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 and case series |
| 4     | Expert opinion |

RCTs, randomised controlled trials.
Table 2 The number of recommendations per guideline

| Number | Number | Name                                                                 | Pages | A  | B  | C  | D  | Total | Percentage of 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    | Diagnosis and management of colorectal cancer                        | 34    | 3  | 5  | 3  | 2  | 13    | 15.4            |
| 124    | 124    | Diagnosis and management of colorectal cancer                        | 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    | Non-pharmaceutical 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            |
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. 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.”

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

SIGN 116 (diabetes), is a notable outlier. It is more than 50% larger than the next largest, 2.5 times longer than the average and yet uses the fourth lowest level D recommendations. There are a number of hypotheses why this group reports differently. SIGN guidelines inform 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. 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.

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 express concerns about guideline development groups. The inappropriate exclusion of disease groups from general population data is common. Smoking cessation advice is applicable to the general population almost without exception, yet the evidence to stop smoking was graded as B on 3 occasions and level C and D once each. Interpreting evidence inconsistently 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 1000 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.” 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. 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. SIGN moved to a new grading system in 2001 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.

Acknowledgements The authors would like to thank Heather Barrington, Statistical Adviser; Bridget Bird, Administrative Assistant; Research and Development Support Unit, Dumfries and Galloway, and Anne B Baird (ABB), Sandhead Surgery, Sandhead, Wigtownshire.

Contributors AGB and JRL were involved in revising the raw data and agreed on a statistical approach to discover whether the trend was significant or not, and were also 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.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Extra data can be accessed via the Dryad data repository at http://datadryad.org/ with the doi:10.5061/dryad.r2fh0.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/