Applicability of Textual Clinical Practice Guidelines: Impact of Physician Interpretation

Jean-Charles Dufour\textsuperscript{a}, Roch Giorgi\textsuperscript{a}, Jean-Robert Harlé\textsuperscript{b}, Dominique Fieschi\textsuperscript{a}, Françoise Volot\textsuperscript{a}, Marius Fieschi\textsuperscript{a}

\textsuperscript{a} Laboratoire d'Enseignement et de Recherche sur le Traitement de l'Information Médicale, Faculté de Médecine, Université de la Méditerranée, Marseille, France.
\textsuperscript{b} Service de Médecine Interne, Hôpital de la Conception, Assistance Publique des Hôpitaux de Marseille, France

Abstract

OBJECTIVES: 1) Determine whether textual Clinical Practice Guidelines (CPGs) are interpreted accurately and unequivocally by targeted physicians. 2) Specify audience and perception of the CPGs.

METHOD: Comparative analysis of answers given by a panel of general practitioners to a series of questions and clinical case studies related to three textual CPGs produced and published by the French National Agency for Accreditation and Evaluation in Health (ANAES).

RESULTS: 68 to 96% of physicians are aware of the existence of the CPGs studied. Less than 50% state having read them. On average, 38% of physician interpretations of CPGs are incorrect (i.e., not in agreement with expert interpretation). Furthermore, there is disagreement among physicians responses.

CONCLUSION: This study credits the argument of disparities in practice which derive from inaccurate and discordant CPGs’ interpretations. The results should prompt those responsible for producing such decision-making support to design documents that are better structured, less ambiguous, and more precise. In a model which facilitates their computerisation the expression of CPGs provides a solution that should be included upstream in the publication process.

Keywords
Guideline Adherence; Practice Guidelines; Physician’s Practice Patterns; Decision Support Systems, Clinical

1. Introduction

Clinical Practice Guidelines (CPGs) are designed to compile the best recognised medical knowledge in order to provide physicians with a practical decisional aid. Their main objective is to improve the quality and effectiveness of medical care. Also they aim at reducing unjustified disparities in medical practice by providing physicians with consensual and reliable referential documents on which physicians can base their practice. To achieve such objectives, CPGs must be accessible, correctly interpreted and properly applied by the physicians for whom they are designed.

Currently, CPGs are expressed in essentially text format. They are published in various
medical journals and/or published on the Internet to guarantee a wide dissemination and a relative accessibility [1]. This mode of dissemination cannot in itself impact medical practice sufficiently and modify habits in an appropriate and lasting way. For greater effectiveness, physicians should be regularly reminded of CPG content during their everyday clinical practice [2]. A way to achieve this objective is to provide computerized CPGs which can be readily and easily consulted in day-to-day practice and integrated within the care process [3]. We have been developing such computerized CPGs [4]. The first stage in computerizing CPGs is based on the analysis and interpretation of textual documents published by medical professional societies or specialized organisations like ANAES (the French National Agency for Accreditation and Evaluation in Health). The aim is to identify, within the textual document, the criteria and conditions that should influence a diagnosis or a therapeutic approach during medical consultation [6]. This analysis allows us to structure a CPG in the form of a decision tree and to implement it. As a result, we produce a computerized version which provides quick and direct access to the recommendations adapted to the patient profile. Structuring CPGs from an initial textual document often is difficult because of text ambiguities and incompleteness. For example, concepts contained within the initial CPG can be imprecise; certain notions implicitly rely on specialized expert knowledge; the narrative structure of a document does not always indicate an unequivocal care pathway; there are “knowledge gaps” which are not explicitly mentioned; etc [7]. Consequently, there is a real risk of partial or even erroneous interpretation in the development of the computerized version.

These considerations lead us to further question the editorial content of CPGs. We observe from experience that the latter is ill-adapted to the computerisation of CPGs. Further, do they properly reflect their original objectives (recommend an optimal care pathway, reduce disparities in practice) when used by physicians in appropriate conditions? In other words, are the difficulties of computerising textual CPGs specifically due to the change in formalisation, or rather do they reveal that the CPGs should be improved before distribution to physicians?

The main objective of this work is to determine whether the textual CPGs studied are interpreted accurately and unequivocally by the physicians for whom they are intended. The secondary objective is to further define physician audience and perception.

2. Material and Method

We choose to study three textual CPGs published by ANAES and entitled:

1. « Health care management for type 2 diabetic patients without complication, march 2000 »
2. « Diagnosis, health care management and treatment of chronic lumbago patients, december 2000 »
3. « Health care management for adult patients with primary high blood pressure, april 2000 »

Our study is based on an analysis of answers given by general practitioners (GPs) or residents (Rs) in response to a series of questions and clinical cases developed by the internist co-investigator of this study and related to these three CPGs. Participants were required to adhere strictly to the recommendations given by the CPGs when responding to the clinical cases. A sample of participating physicians was pulled at random from the GPs and the Rs in contact with our research laboratory. All physicians who agreed to take part in the study were included. The physicians were remunerated for their participation.

---

1 These three CPG were mainly chosen as we had prior knowledge of their computerisation
A general questionnaire and two clinical cases were designed for each of the three CPGs:

- The general questionnaire was composed of four questions as follows: 1) Were you aware of the existence of the CPG? 2) Had you read it before? 3) How long did it take you to read it? 4) How do you rank the clarity of the CPG?

- Each clinical case was composed of 3 to 5 open-ended and focused questions such as « What treatment would you prescribe? », « When do you plan to see the patient again? » etc. Each of these questions addresses one of the following four medical tasks categories: diagnosis assessment, patient’s medical follow-up, therapeutic options, and disease check-up. Further, physicians had the option of expressing any comments or problems concerning the description of the clinical case.

The answers were recorded directly by the physicians via an on-line questionnaire. Physicians were systematically reminded to consult and strictly observe the CPG recommendations when discussing the clinical cases. The original PDF documents of the CPGs published by the ANAES were instantly available and could be downloaded from the questionnaire interface.

Each physician possessed a username and a password which allowed him to add to or modify their answers while consulting the study website. Participants could not be identified from their login. No time limit was imposed for answering the questions.

For each clinical case, we asked the internist expert to provide his answer while strictly adhering to the CPG recommendations. The expert’s answers supplied served as the “gold standard” in evaluating the physicians’ answers to the same CPG. Each physician’s answer was compared to the expert’s answer by two medical observers who worked independently. An answer-score with two values (0 = physician’s answer disagrees with expert’s answer; 1 = physician’s answer agrees with expert’s answer) was defined. In the event of scoring disagreement between the first two medical observers, a third was called upon to give a definitive answer-score. The two values answer-score enabled us to calculate the percentage of answers in agreement with the expert’s answers for each CPG.

The degree of agreement between the two medical observers was measured by the Kappa coefficient.

The degree of agreement between the participant physicians was measured by the intraclass coefficient correlation of the answers-scores. This coefficient allowed us to appreciate the degree of discordance concerning answers to the clinical cases within the physicians group.

3. Results

Twenty seven physicians took part in this study: 14 GPs (52%) and 13 Rs (48%).

The study questionnaire, primarily composed of clinical cases, contained 26 questions which the medical observers had given a score to. A total of 629 answers were given a score [26 * 27 - (73 missing data)]. The first two observers agreed for 549 answers (94.4%) and disagreed for 35 answers (6.4%). There was a good degree of agreement between the two observers (Kappa coefficient of 0.88, p < 0.0001).

3.1. CPG audience and perception

Table 1 records the percentage of physicians aware of the existence of each CPG and the percentage of physicians who had previously read each CPG. In general, we observed that whereas physicians are relatively well aware of the existence of the CPGs studied (68% to 96% according to the CPG), less than 50% of physicians stated that they had read them beforehand. The lumbago CPG attracted the lowest audience. It was also considered the
least clear: more than 66% of physicians considered it “not clear at all” or “moderately clear”. While 46.2% and 23.8% thought likewise about the CPGs on diabetes and HBP respectively.

Table 1: CPG audience and perception

<table>
<thead>
<tr>
<th></th>
<th>Diabetes</th>
<th>Lumbago</th>
<th>HBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aware of the existence</td>
<td>92%</td>
<td>68%</td>
<td>96%</td>
</tr>
<tr>
<td>Previously read</td>
<td>48%</td>
<td>36%</td>
<td>50%</td>
</tr>
<tr>
<td>Perfectly clear</td>
<td>53.8%</td>
<td>33.3%</td>
<td>76.2%</td>
</tr>
<tr>
<td>Moderately clear</td>
<td>23.1%</td>
<td>33.3%</td>
<td>19%</td>
</tr>
<tr>
<td>Not clear at all</td>
<td>23.1%</td>
<td>33.3%</td>
<td>4.8%</td>
</tr>
<tr>
<td>Average reading time (standard deviation)</td>
<td>27 mn (15)</td>
<td>42 mn (39)</td>
<td>27 mn (16)</td>
</tr>
<tr>
<td>Number of A4 pages</td>
<td>21</td>
<td>95</td>
<td>17</td>
</tr>
</tbody>
</table>

3.2. Concordance with the « gold standard »

The overall percentage of answers in concordance with the expert is only 62% (standard deviation, sd = 11%). Only 54% (sd = 18%) of answers were tallied with this reference for the lumbago CPG, 63% (sd = 14%) for HBP and 65% (sd = 15%) for diabetes (non-significant statistical difference). Table 2 shows a synthesis of the results for each physician group (GPs or Rs). We did not find any significant statistical difference in score distribution by CPG between the GPs and Rs groups, except for the CPG for HBP (p = 0.02).

Tableau 2 - Percentage of agreements with the 'gold standard' reference depending on CPG and physician group

<table>
<thead>
<tr>
<th></th>
<th>Diabetes</th>
<th>Lumbago</th>
<th>HBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPs + Rs</td>
<td>65%</td>
<td>54%</td>
<td>63%</td>
</tr>
<tr>
<td>GPs</td>
<td>65%</td>
<td>49%</td>
<td>57%</td>
</tr>
<tr>
<td>Rs</td>
<td>64%</td>
<td>61%</td>
<td>70%</td>
</tr>
</tbody>
</table>

Table 3 shows the percentage of agreements with the reference, depending on the CPG and on the category of medical task addressed by clinical case questions. Whatever the CPG, the agreement with the expert was significantly stronger for questions focused on diagnosis assessment (p<0.01 when ‘diagnosis assessment’ was compared with every other category).

Table 3 - Percentage of agreements with the 'gold standard’ reference depending on CPG and question category

<table>
<thead>
<tr>
<th></th>
<th>Diabetes</th>
<th>Lumbago</th>
<th>HBP</th>
<th>Overall CPGs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis assessment</td>
<td>98.2% (sd=9.6%)</td>
<td>69.6% (sd=36.1%)</td>
<td>87% (sd=22.4%)</td>
<td>86.4% (sd=15.4%)</td>
</tr>
<tr>
<td>Patient’s medical follow-up</td>
<td>59.3% (sd=29.7%)</td>
<td>Not Applicable</td>
<td>58.7% (sd=35.8%)</td>
<td>58.5% (sd=24.1%)</td>
</tr>
<tr>
<td>Therapeutic options</td>
<td>72.2% (sd=32%)</td>
<td>52.2% (sd=51.1%)</td>
<td>54.4% (sd=23.4%)</td>
<td>62.7% (sd=24.9%)</td>
</tr>
<tr>
<td>Disease check-up</td>
<td>29.6% (sd=28.6%)</td>
<td>44.9% (sd=25.8%)</td>
<td>Not Applicable</td>
<td>37.2% (sd=27.4%)</td>
</tr>
</tbody>
</table>

If we consider the overall CPGs, the lower agreement was for questions on ‘disease check-up’ and the percentage of agreement between each question category is statistically
significant (p<0.01), except for the comparison between ‘Patient’s medical follow-up’ and ‘Therapeutic options’.

3.3. Concordance between physician answers

Whichever CPG was considered, the calculation of the intraclass coefficient correlation (ICC) revealed a low rate of answer agreement between physicians (ICC Diabetes = 0.43, confidence interval at 95% (CI95%) 0.23-0.77; ICC Lombago = 0.13, CI95% 0.03-0.53; ICC HBP = 0.39, CI95% 0.19-0.77; p<10-6 in every case). This variability in answers to the clinical cases suggested divergence in understanding and interpreting CPGs. Indeed, we can consider that problem in understanding the clinical cases themselves can be excluded since no criticism was made by the physicians concerning the description of these cases.

4. Discussion

The objective of this study was to determine if the CPGs studied helped physicians respond correctly to a given clinical problem and if they gave rise to diversions in interpretation and/or in understanding. The results obtained showed that on average, in 38% of cases, physician interpretation of the CPGs was incorrect (did not match the expert interpretation). Above all, our results revealed the disparities between the physicians’ answers themselves, a fact that minimizes the potential bias due to the choice of a “gold standard”. Although the difference was not significant, the rate of incorrect answers was higher for the CPG concerning chronic lumbago. It is worth pointing out, that this particular CPG presented the most difficulties during its computerisation stage (a stage carried out prior and independently of this study). Its editorial content did not lend itself easily to structuring as a refined decision tree. The other two CPGs (diabetes and HBP) were more easily structured in this form, but were not exempt from imprecision and ambiguity.

A frequent observation about the textual and static format of CPGs is that they are difficult to use in daily practice. Partly this might explain why physicians hardly followed the recommendations given by CPGs in their daily care activity. Thus it was implicitly felt that the observed divergence between practice and CPGs was mainly due to the fact that, for various reasons, such as inadequate format, physicians did not use CPGs and consequently did not follow their proposed recommendations. Although the physicians used and followed the CPGs, in our study we observed disparate interpretations brought to light by divergent answers to the clinical cases. Therefore our findings tend to argue that some of the disparity in practice is derived from inaccurate and unequal interpretation of CPGs, even though used by physicians. This hypothesis should be backed up by further in-depth studies involving a greater number of physicians and an increased diversity of CPGs.

In our study, there were open-ended questions in the clinical cases so as to avoid influencing physician answers (even if more easily used statistically, close-ended questions do not correspond to the reality of clinical practice). The scoring of the answers by the medical observers, indispensable for statistical result analysis, showed a certain stability (the Kappa coefficient measuring agreement between the two medical observers was 0.88).

5. Conclusion

In their classical editorial format, CPGs are not adapted to computerisation [8]. The same difficulties encountered in developing a computerised CPG from a text version are also experienced by physicians themselves when they translate a CPG into operational behaviours within a clinical context. The results of our study tend to support the argument...
that CPGs are a source of incorrect and different interpretations among physicians for whom they were specifically designed. These results should encourage those responsible for producing such decision-making supports to provide documents which are better structured, less ambiguous and more precise to ensure correct and unequivocal interpretation. The computerisation of CPGs which requires explicit knowledge clarification and rigorous content structuring is a means to detect potential stumbling-blocks upstream. The advantage of computerisation is not just to provide a support which is easier to use and more readily available than a textual document. It also imposes the need to organize and structure information in a more comprehensive, coherent and consistent way so as to guarantee a better and unequivocal understanding. The development of CPGs according to a computerisable model should precede the publication stage of classical textual CPGs. Computerisation itself gives advance warning of weaknesses that must be addressed upstream of the publication stage so as to benefit both the computer solution and the physicians who actually use the textual version.

6. Acknowledgments

We thank Dr. Omar Bouhaddou for his helpful proofreading and comments.

7. References


Address for correspondence

Dr Jean-Charles Dufour.
LERTIM, Faculté de Médecine, 27 boulevard Jean Moulin, 13385 Marseille Cedex 05.
Email : jean-charles.dufour@medecine.univ-mrs.fr / URL : http://cybertim.timone.univ-mrs.fr