A survey of French general practitioners and a qualitative study on their use and assessment of predictive clinical scores

Background: Predictive clinical scores, diagnostic as well as prognostic, are considered to be useful tools for making decisions under conditions of uncertainty. They are not intended to replace clinical judgment or medical experience, but to help physicians in the interpretation of clinical information. The general practitioner (GP), the gateway to care in the French health system, should be the main beneficiary of their utilization. However, there is no information on the prevalence of their use in general practice in France.

Methods: A national, transversal epidemiological survey was conducted by electronic mail among GPs belonging to the French Sentinelles network. GPs were asked about their use of scores, the context of their utilization and the expected benefit. A qualitative study (focus groups) was also carried out with three groups of GPs within the context of continuous medical education.

Results: The study consisted of 358 GPs. They were questioned on their use of seven predictive clinical scores (six diagnostic and one prognostic). Clinical scores were used by 75% of GPs, with no statistical difference with regard to their age or sex. The most often used were: the Mini Mental Status Examination (MMSE) (95%), Fagerström test (90%), Hamilton scale (65%), McIsaac scores (61%), DETA/CAGE (45%), Simple Calculated Osteoporosis Risk Estimation (SCORE) for osteoporosis (33%), and the only prognostic score CHADS2 (28%). Clinical scores were especially used when elderly people were involved (77%) and when the diagnosis was uncertain (63%). The qualitative study gave additional information on the barriers and obstacles to the use of predictive clinical scores.

Conclusion: This study, the first one in France, gives information on the perception of clinical scores and on the rationale for their use by GPs. Suggestions to improve the situation (availability and rate of utilization of clinical scores) are provided.

Keywords: predictive scores, general practice, epidemiological survey

Introduction

Medical decision making is a mental act of choice among several possibilities, based on the knowledge acquired during medical education and professional experience. The decision also depends on two general modes of thinking: evidence-based medicine, which promotes the use of the highest possible level of scientific evidence, and case-based medicine, which emphasizes the importance, among several components, of empirical evidence, experiential evidence, as well as patients’ goals and values.1

The coexistence of these two apparently contradictory viewpoints is the likely consequence of the complex nature of medical care. A similar opposition appears to exist between the use of predictive scores and intuitive, implicit, clinical judgment. Predictive scores are intended to reduce the uncertainty related to diagnosis by the
explicit use of simple clinical and nonclinical data. They represent a valuable tool in decision making, provided that their impact has been assessed and their limits understood. In general practice, predictive prognostic scores could be useful as they can predict, for one particular patient, his risk of developing a certain disease, especially cardiovascular, and thereby stimulate the prescription of well-established preventive measures. In this respect, predictive scores satisfy the requirements of evidence-based medicine. On the other hand, conflicting results have been reported when clinical diagnosis has been compared to score-based diagnosis. Comparable problems exist for prognostic scores, where an index as simple as age yields information similar to that of more complex cardiovascular risk scores. These considerations make it important to obtain better information on the use of predictive scores by general practitioners (GPs) in their daily practice and their appropriateness in such a setting, as well as the reasons for their non-utilization.

In France, the prevalence of the use of clinical scores in general practice is currently unknown. There are no data to assess the use of predictive scores and their impact on the daily care for patients by GPs. Because GPs are the gatekeepers in the French health care system, it seemed important to conduct a quantitative and qualitative study to analyze the use of predictive scores in GPs’ practices in France and to identify potential barriers to their integration into the primary care process.

Methods
Quantitative study
Type of survey
Observational and transversal epidemiological study conducted among GPs in France.

Participants
GPs of the French Sentinelles Network (FSN), a national system for the electronic surveillance of diseases, invited to participate in June 2010 (n = 1087 GPs) via electronic mail. They could connect via a single password to a system for the electronic surveillance of diseases, and thereby stimulate the prescription of well-established preventive measures. The FSN was given formal ethical approval for this study by the national ethics committee [Commission nationale de l’informatique et des libertés (CNIL) N° 471393]. All data were handled confidentially and the results were anonymous.

Selection of scores
Scores to be taken into consideration were selected from a literature review in the area of general medicine using either PubMed or the recommendations of the French National Authority for Health. All French reviews and articles about predictive scores were selected. A ranking of the predictive scores most often cited in the last 10 years was established. The six most used scores (according to percentage of citation) were chosen to be presented to the participating GPs with the exception of the SCORE risk of osteoporosis (Simple Calculated Osteoporosis Risk Estimation) which was selected to allow comparison with another study, already published. An open question allowed GPs to list other scores.

The selected scores were:
- Diagnostic scores: Mini Mental Status Examination (MMSE) score (cognitive functions); Fagerström score (risk of tobacco dependence); DETA/CAGE score (excessive consumption and risk of alcohol dependence); SCORE for the diagnosis of osteoporosis; Melisa score (diagnosis of bacterial pharyngitis); and Hamilton score (depression scale).
- Prognostic score: CHADS2 score (thromboembolic risk in patients with atrial fibrillation).

A scale of usefulness was created from 1 (useless) to 5 (very useful) to quantify the value of the scores as assessed by GPs. To evaluate the context of score use, the following items were considered: patient age (>65 years); uncertain diagnosis; uncertain treatment; part of medical recommendations; and no particular context.

Qualitative study: focus groups (FGs)
A second study was carried out, consisting of three FGs (held from September to November 2010) with GPs enrolled in continuous medical education groups in the Rhône-Alpes area. The groups consisted of volunteers who did not belong to the FSN. The FGs were led by an experienced FG moderator. GPs were asked to discuss predictive clinical scores: general attitude and professional approach to predictive scores, context of use, usefulness of scores, barriers to predictive score use, suggestions to improve their use, and knowledge about the construction of predictive scores.

The FG discussions were recorded, transcribed, and entered into the analysis software NVivo 9 Categories (QSR International Pty Ltd, Doncaster, VIC, Australia) and themes were then identified and refined in a continuous and iterative process. GPs were classified according to their place of practice: rural (<2000 inhabitants), semi-rural (2000 to 5000 inhabitants), urban area (>5000 inhabitants).
Analysis of results

All variables in this study were analyzed using the R software program (http://www.r-project.org/). Categorical variables were compared using the Chi-square test or Fisher’s exact test, and continuous variables were compared using a Student’s t-test. Logistic regression analysis was used to represent the GP sample and find the determinants of users and non-users of predictive clinical scores. The focus group analysis was performed using the grounded theory.18

Results

Quantitative survey

Participation and characteristics of GPs

A total of 358 GPs completed the quantitative study; ie, a participation rate of 33% of the GPs invited to the survey. All metropolitan French regions were represented. Participating GPs were mostly men (sex ratio 4.3 [3.4–5.2]), had an average age of 53 years (43–63 years), women being younger than men (50 years versus 54 years, P < 0.05). They had an average of 22 years (12–32 years). Demographic characteristics: there was no significant difference between responders, non-responders, and French Sentinelles Network (FSN) in terms of average age and practice location (data not shown).

Assessment of prevalence of predictive scores use

Seventy-five percent (n = 268) of the GPs reported that they use predictive scores in their everyday practice without significant differences in terms of age, sex, and location (urban or rural) between users and nonusers (Table 1).

Among the proposed scores, MMSE and Fagerström were the most commonly used scores (95% and 90% of GPs, respectively) Hamilton, McIsaac, DETA/CAGE, and SCORE (for osteoporosis) were used by 65% to 33% of the GPs, respectively) Hamilton, McIsaac, DETA/CAGE, and SCORE (for osteoporosis) were used by 65% to 33% of the GPs (Table 2). Again, no difference could be accounted for by the demographic characteristics of GPs. The CHADS2 test for embolic risk in atrial fibrillation was used by only 28% of GPs. Other scores mentioned by the GPs in the open question were: one diagnostic score, the Framingham score for cardiovascular risk without any other precision (n = 9);19 and six prognostic scores, the Epworth scale for somnolence (n = 12),20 the Clock Drawing Test for dementia (n = 14),21 the instrumental activities of daily living scale for dementia (n = 8),22 the Get Up and Go Test for mobility (n = 6),23 and the Ottawa rule for the radiological diagnosis of ankle injuries (n = 7).24 Among the GPs who used scores, over one third considered that scores were not useful.

Compiling those scores by utility category, scores on the severity of disease (MMSE and Hamilton) are the most commonly used (59% of GPs), scores on the diagnosis of addiction (Fagerström, DETA/CAGE) are used on average by 50% of GPs, while scores conditioning a specific treatment (McIsaac, SCORE for osteoporosis, CHADS 2) are used on average by only 30% of GPs.

However, except for DETA/CAGE and SCORE for osteoporosis (as more than 60% of GPs declare using these two scores) only 20% of GPs use predictive clinical scores systematically (Table 2). These differences between score usages do not depend on age, sex and location (P > 0.05).

Utility scale

All of the GPs were asked to assess the usefulness of the proposed predictive scores. The tests most commonly used were not always considered to be the most useful ones. The MMSE, widely used, was considered to be very useful by 82% of respondents (n = 221). Scores such as Fagerström (n = 148), McIsaac (n = 137) and CHADS2 (n = 135) were considered to be valuable by only about half of the GPs (55%, 51%, and 50%, respectively). Hamilton (n = 126) and DETA/CAGE (n = 113) were considered to be even less useful, while SCORE for osteoporosis (n = 73) was the lowest rated of the group (47%, 42%, and 27%, respectively) (Figure 1).

Context of use

Only the GPs who used predictive scores answered the question on the context of use (Table 3). Age of patients and uncertain diagnosis were the two most commonly quoted contexts, which suggests a somewhat targeted use.

Qualitative study: FGs

Questionnaire

The qualitative questionnaire considered the same themes as for the quantitative study: most commonly used scores, utility, context of use, and the addition of a new item, integration of scores in daily general practice.
Table 2 Use of predictive scores by GPs

<table>
<thead>
<tr>
<th>Prognostic scores</th>
<th>Used n (%)</th>
<th>Systematically</th>
<th>On a case by case basis</th>
<th>Only if I have time</th>
<th>Rarely</th>
<th>Not used n (%)</th>
<th>I don’t use it</th>
<th>I don’t know that score</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHADS2</td>
<td>74 (28)</td>
<td>13 (5)</td>
<td>27 (10)</td>
<td>6 (2)</td>
<td>28 (10)</td>
<td>194 (72)</td>
<td>61 (22)</td>
<td>133 (50)</td>
</tr>
<tr>
<td>Diagnostic scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>251 (95)</td>
<td>39 (14)</td>
<td>149 (55)</td>
<td>30 (11)</td>
<td>33 (12)</td>
<td>17 (6)</td>
<td>17 (6)</td>
<td>0</td>
</tr>
<tr>
<td>Fagerström</td>
<td>240 (90)</td>
<td>50 (19)</td>
<td>121 (45)</td>
<td>28 (10)</td>
<td>41 (15)</td>
<td>28 (10)</td>
<td>23 (9)</td>
<td>5 (2)</td>
</tr>
<tr>
<td>Hamilton</td>
<td>174 (65)</td>
<td>15 (6)</td>
<td>84 (31)</td>
<td>23 (9)</td>
<td>52 (19)</td>
<td>94 (35)</td>
<td>88 (32)</td>
<td>6 (2)</td>
</tr>
<tr>
<td>McIsaac</td>
<td>163 (61)</td>
<td>58 (21)</td>
<td>71 (26)</td>
<td>8 (3)</td>
<td>26 (10)</td>
<td>105 (39)</td>
<td>53 (20)</td>
<td>52 (19)</td>
</tr>
<tr>
<td>DETA/CAGE</td>
<td>121 (45)</td>
<td>4 (1)</td>
<td>56 (21)</td>
<td>12 (5)</td>
<td>49 (18)</td>
<td>147 (55)</td>
<td>80 (30)</td>
<td>67 (25)</td>
</tr>
<tr>
<td>SCORE for osteoporosis</td>
<td>89 (33)</td>
<td>6 (2)</td>
<td>34 (13)</td>
<td>13 (5)</td>
<td>36 (13)</td>
<td>179 (67)</td>
<td>99 (36)</td>
<td>80 (30)</td>
</tr>
</tbody>
</table>

Abbreviations: DETA/CAGE, Diminuer Entourage Trop Alcool; GP, general practitioner; MMSE, Mini Mental Status Examination; SCORE, Simple Characteristic Osteoporosis Risk Estimation.

Participation and characteristics of GPs

Twenty-three GPs attended three different FGs. GPs in FGs were similar to GPs of the quantitative study in terms of sex and age distribution, but not of location, since their practice was essentially in urban or suburban areas (data not shown).

FGs confirmed the ranking of scores as found in the quantitative study: MMSE, Fagerström, Hamilton, DETA/CAGE, and McIsaac were the most used scores.

Main conclusions from the FGs

Personal attitude and professional approach to predictive scores: needs for tools in general practice

The multidisciplinary nature of general practice, the complex relationship between practitioners and patients, as well as the diversity of personal experiences make practitioners especially sensitive to a subjective view of their practice. However, the need for communication between professionals requires a common language with explicit content, a condition fulfilled by predictive scores. GPs, when dealing with uncertain situations, must use all available and relevant information, such as predictive scores, to give their patients appropriate care.

General practice is the opposite of treating a patient with scores. We cannot operate in that way systematically, but it would be foolish to do the opposite, which is to deny the scores that are practical and have an impact.

Context of use

In France and most countries, GPs represent the first line in patient health care. Therefore, predictive scores are considered to be useful when GPs send their patients to a specialist. GPs will provide the specialist with the appropriate tool: for example, the CHADS2 for the cardiologist or the MMSE for the geriatrician:

When I refer a patient to a neurologist or neuro-geriatric consultation, I don’t want to look stupid, so I use MMSE.

Predictive scores are also used in therapeutic decision making, especially to justify preventive treatment and possibly to provide reassurance to the patient:

MMSE results reassure the oldest patients about their memory.

On the other hand, there is sometimes an intuitive use of score items which may reinforce the a priori clinical judgment of the GP.

Barriers to routine risk score use

GPs explained that scores can be experienced as an intrusion into the consultation:

Discussing with your patients could be more useful than the use of a score, which is very impersonal.
Table 3 Distribution of responses depending on the context of use among physicians users

<table>
<thead>
<tr>
<th></th>
<th>Used in that context</th>
<th>On a case by case basis</th>
<th>Systematically</th>
<th>Rarely/if I have time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient age (&gt;65 years)</td>
<td>264 (98)</td>
<td>207 (77)</td>
<td>16 (6)</td>
<td>41 (15)</td>
</tr>
<tr>
<td>Uncertain diagnosis</td>
<td>258 (96)</td>
<td>168 (63)</td>
<td>61 (23)</td>
<td>29 (11)</td>
</tr>
<tr>
<td>Uncertain treatment</td>
<td>247 (92)</td>
<td>139 (52)</td>
<td>43 (16)</td>
<td>57 (21)</td>
</tr>
<tr>
<td>Part of medical recommendations</td>
<td>221 (83)</td>
<td>90 (34)</td>
<td>16 (6)</td>
<td>115 (43)</td>
</tr>
</tbody>
</table>

Note: n = 268.

When treating a depression, the practitioner/patient relationship is very important, so we cannot pollute the consultation with a scale, you lose in effectiveness efficacy relevance?

Scores can be integrated into the clinical inquiry, but they make it somewhat artificial and lengthen the consultation time:

Is there a real advantage to use them considering time lost in doing it?

As a result, a new consultation dedicated to the completion of the score may be required. In addition, the technical problems related to their availability can be a barrier to use. Other comments are related to the potential risk of overtreatment. The problem of a lack of accuracy of risk scores, especially if derived from a different population was also mentioned:

In France there is a lack of surveys about scores that correspond to the population.

Suggestions to improve predictive scores

GPs need to have a real assessment of the impact of scores on health management:

It must be for the interest of the patient. All that is in the interest of the patient is to be validated.

The purpose of the score must be easy to understand.

It must be short. Not too many items.

Score results must be reproducible, reliable, easy to use and helpful for therapeutic decisions.

Discussion

This study describes, for the first time in France, how predictive scores are used in general practice. When compared to similar studies in the literature, it also introduces several new features that need to be commented upon. Our approach was score-based and not disease-based as in other studies. This allowed a more global analysis on the concept of predictive scores than would have been the case if focused on a particular disease or group of diseases.

The choice of the predictive clinical scores to be tested by GPs was not based on an individual, potentially biased selection, but was made as a function of their rate of quotations in journals of general medicine and national clinical guidelines; this criterion makes it likely that these scores are the ones most commonly known by GPs. It turned out that the vast majority of them were diagnostic scores. Furthermore, nearly half of the scores were related to the diagnosis of diseases commonly found in elderly patients: cognitive disorders (MMSE), osteoporosis (SCORE for the risk of osteoporosis), or the risk of thromboembolic complications in atrial fibrillation (CHADS2). These two elements, predominance of diagnostic scores and strong implication of geriatric problems, could have been the consequence of our mode of selection of scores. Therefore, there might be a need for further studies with different protocols to better comprehend the complex relationship between clinical scores and their utilization by physicians. Moreover, both parts of this study revealed what worried GPs: age as a determining factor for GPs’ use of a score and the diagnostic decision as probably a very emotional situation. On the other hand, the selected scores covered a rather wide range of situations: addiction (CAGE/DETA for alcohol and Fagerström for cigarettes), cognitive disease (MMSE for cognitive disorders, Hamilton scale for depression), and diagnostic procedure (McIsaac score for bacterial tonsillitis and SCORE for the risk of osteoporosis for bone densitometry). In this respect, the selected scores were representative of the wide range of medical activities encountered in primary care. GPs need to constantly adapt according to very different patients’ situations, that explain such different scores. Additional data of the typology of the patients seen by the GPs could have brought a more accurate insight into the relationship between the need and the appropriate use of scores. Obviously, some scores are limited to specific patients
(CHADS2: patients with non-rheumatismal chronic atrial fibrillation; McIsaac: more likely to be used in children and adolescents). It follows that the need for certain scores is directly related to the type of patients recruited and scores must reflect this activity.

Our study combined a questionnaire sent to a large number of GPs with the use of FGs of 23 GPs from a separate population. Thus, we had the opportunity to have both a quantitative evaluation of the use of predictive scores and a qualitative assessment of physicians’ behavior and feelings in response to the existence of scores to provide help in decision making. Findings from the qualitative analysis provided explanations about the discrepancy between theoretical and practical use revealed during the quantitative study. Although GPs need tools in medical practice, many limitations to the acceptance and use of predictive scores are still present.

Wide variations exist in the rate of use of predictive scores. The MMSE and Hamilton scale, two commonly used scores, are considered as a somewhat compulsory prerequisite prior to referring a patient to a specialist, as if the GP needed them to validate his request for professional advice. The fact that the Hamilton scale was frequently tested could also be related to a particular incentive since it is the only predictive score on this list to be financially rewarded. As completion of a score can be very time consuming, the use of a predictive score depends on the number, quality, and ease of production of the score items. The MMSE is easy to perform and has been demonstrated to be a useful tool to identify cognitive impairment.\textsuperscript{25} Other explanations can be offered for the less frequent use of scores, such as their limitation to specific patient groups (SCORE to women only, CHADS2 to patients with atrial fibrillation). Surprisingly, the DETA/CAGE score was used two times less frequently than the Fagerström score, perhaps because GPs may be reluctant to tackle alcohol problems with their patients and the score does not help in those consultations.\textsuperscript{26}

The analysis of GPs’ judgment on the usefulness of predictive scores gave a somewhat disturbing picture. Apart from the MMSE, which was very frequently applied and considered to be helpful, the rating for the usefulness of the other scores was a meager 27%–50%. These results could also suggest that most GPs did not really trust the information provided by the scores, relying perhaps more heavily on their own clinical judgment.\textsuperscript{27,28} Lack of knowledge about the possible weaknesses of clinical reasoning could explain this attitude. Alternatively, GPs may not be knowledgeable enough about the scores available for their use and their theoretical background, creating, as a consequence, an additional cause for uncertainty.

In the qualitative study, the GPs assessed the positive features of the predictive scores to be their contribution to the creation of a common language and better communication between physicians by buffering often excessively subjective statements at the profit of more explicit formulations. The scores were also assessed to be potentially useful in reassuring patients. However, the negative elements were considered to be far more numerous, which confirms data from other groups dealing with prognostic cardiovascular scores:\textsuperscript{29} distrust in their validity, barrier to the doctor–patient relationship, limited availability, time requirement for the completion of scores, etc. These considerations lead to a question related to the knowledge required to make the best use of clinical scores. Because the average age of our group of GPs was around 55 years, it is unlikely that they have received systematic teaching on this particular topic. Clinical epidemiology, which encompasses the field of decision making, is a rather recent discipline and many faculties of medicine offer only a cursory coverage. Only the youngest members of the FSN may have had access to this teaching, and probably only on a limited basis.

On the other hand, many French official guidelines recognize the value of predictive scores and make strong recommendations for their use by physicians.\textsuperscript{30} Continuous medical education also supports this attitude, in line with the promotion of evidence-based medicine. It must be recognized, however, that evidence-based medicine is not always spontaneously accepted by practicing doctors, whose way of thinking is more likely to be based on the reliance of clinical experience. Furthermore, postgraduate teaching on these matters is generally done by specialists who are not exposed to the same clinical situations as GPs, hence a feeling of lack of relevance by the latter.\textsuperscript{31} Some of the answers given in the survey and in the FGs are clear indicators of a lack of understanding of the theory at the basis of clinical prediction rules. Statements issued in guidelines are certainly not sufficient to familiarize physicians with new concepts\textsuperscript{32} and better structured postgraduate teaching should be systematically undertaken. Presentation of the scores as a number can be an impediment to some GPs.\textsuperscript{31} GPs’ patient relationships can’t be reduced to such dialog. But it can be a standardized communication between physicians. A recent review\textsuperscript{33} concluded that validation and impact studies will likely improve the application of valid prediction rules in daily clinical practice.

Electronic devices are being used more frequently in the workup of patients along with the recently compulsory
use of medical electronic records of patients (by now more than 80% of GPs). As most GPs pretend to use score, most of the time they get it from their electronic patient report. This could facilitate the use of clinical scores by making them more readily available, either through specific computer programs, or via Internet resources. These opportunities exist, but they have only a minimal impact on medical practice. Our survey unfortunately did not give any information on the mode of access of GPs for the scores they use. It is obvious that the popularity of a score depends to a large extent on its availability and on the usability of the procedure. GPs appear to resent the time and barrier represented by the use of computers during a medical consultation. Because patients’ reactions were not studied, it is difficult to know if the physicians’ attitudes are mediated, at least in part, by patient comments. Because of the time required to complete them, some of the predictive scores should be presented and explained to the patients in sufficient detail to familiarize them with this new form of information gathering, a new challenge for overcharged physicians.28

In conclusion, this study raises more questions than it offers explanations, which is somewhat expected in view of the fact that predictive medicine is a rather new discipline that requires active learning to be adequately understood. It would appear that the present use of clinical scores among French GPs relies primarily on diagnostic scores, with some preference for those involved in geriatric care. Several obstacles to a more widespread utilization have been identified. We would like to suggest that a more targeted teaching in continuous medical education should be offered to practicing physicians, as some of the resistance encountered is clearly the result of insufficient knowledge.

Authors’ contributions
MS suggested the initial question. MS and AL provided the administrative and technical support for conducting the study. They supervised the data collection of the qualitative study. SGC and MK supervised the data collection of the quantitative study. MS, AL, SGC, and MK were involved in the statistical analysis and interpretation of the data. PM participated in the critical revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgments
The authors would like to thank all of the GPs of the Sentinels network who participated in the survey study, all GPs who participated in the focus group, and the scientific committee group who approved the plan and moderated the focus group. They especially thank Professor AF Junod for reviewing this paper.

Disclosure
The authors report no conflicts of interest in this work.

References


