Social Science & Medicine 71 (2010) 685–692

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Pragmatic evidence and textual arrangements: A case study of French clinical cancer guidelines

Loes Knaapen [a,*], Hervé Cazeneuve [b], Alberto Cambrosio [a], Patrick Castel [c], Beatrice Fervers [d]

[a] Social Studies of Medicine, McGill University, 3647 Peel Street, Montreal, Quebec, H3A 1X1, Canada
[b] Mondes et Dynamiques des Sociétés, Université Lumière Lyon2, France
[c] Centre de Sociologie des Organisations/CNRS, Sciences Po, Paris, France
[d] Centre Léon-Bérard and Santé Individu Société, Université de Lyon, France

ARTICLE INFO

Article history:
Available online 4 June 2010

Keywords:
France
Evidence-based medicine
Clinical practice guidelines
Oncology
Pragmatic sociology
Dispositif
Regulation

ABSTRACT

Both critics and supporters of evidence-based medicine view clinical practice guidelines as an important component of this self-defined "new paradigm" whose goal is to rationalize medicine by grounding clinical decision-making in a careful assessment of the medical literature. We present an analysis of the debates within a guideline development group (GDG) that led to the drafting, revision and publication of a French cancer guideline. Our ethnographic approach focuses on the various aspects of the *dispositif* (or apparatus) that defines the nature and roles of participants, procedures, topics and resources within the GDG. Debates between GDG members are framed (but not dictated) by procedural and methodological rules as well as by the reflexive critical contributions of the GDG members themselves, who justify their (tentative) recommendations by relating to its (possible or intended) audiences. Guideline production work cannot be reduced to an exchange of arguments and to consensus-seeking between pre-defined professional interests. It is about the production of a text in the material sense of the term, i.e. as a set of sentences, paragraphs, statements and formulations that GDG members constantly readjust and rearrange until closure is achieved. As such, guidelines partake in the emergence and stabilization of a new configuration of biomedical knowledge and practices grounded in the establishment of mutually constitutive links between two processes: on the one hand, the re-formatting of clinical trials into a device for producing carefully monitored evidence statements targeting specific populations and clinical indications and, on the other hand, the increasingly pervasive role of regulatory processes.

Introduction

Clinical practice guidelines (CPGs) are key components of evidence-based medicine (EBM), the self-styled "new paradigm" which "de-emphasizes intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision-making", and argues for "using the medical literature [especially results from randomized clinical trials] more effectively in guiding medical practice" (EBM Working Group, 1992, p. 2420). Medical reformers and administrators consider CPGs "the tool of choice to weed out unwarranted variation in diagnostic or therapeutic practice and to enhance the scientific nature of medical care delivered" (Berg, Horstman, Plass, & van Heusden, 2000, p. 766). Thousands of CPGs have been produced in the past decades by a great variety of institutions and associations in many different countries (Weisz et al., 2007).

Unsurprisingly, CPGs have attracted the attention of many commentators. A substantial part of the social science literature on this topic (e.g. Castel & Merle, 2002; Timmermans & Berg, 2003) focuses on the *use* of guidelines as distinct from their *production*. This distinction, however, may be challenged for, as we will see, guideline producers often openly discuss its potential use and users; as noted, more in general, by science studies scholars (Akrich, 1992), technical devices contain built-in *scripts* of their expected deployment: examining those scripts can deepen our understanding of future uses. Most articles on the production of guidelines have been published in medical journals and usually consist of methodological recommendations and suggestions on how to improve the process (e.g. Eccles et al., 1996; Eddy, 1990). Among the social science studies that investigate guideline production, a number have resorted to an experimental or a retrospective design to correlate the professional characteristics of guideline group members with their decisions (e.g. Hutchings & Raine, 2006). Yet, as analysts of procedural rationality would argue (e.g. Reynaud & Richebé, 2007, p. 8), guideline development

* Corresponding author. Tel.: +1 514 398 4400; fax: +1 514 398 1498.
E-mail address: Loes.Knaapen@mail.mcgill.ca (L. Knaapen).

0277-9536/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.socscimed.2010.05.019

cannot be equated simply to a decision about a preset number of choices, but often leads to novel, unexpected solutions. Only ethnographic investigations of guideline development can account for the dynamics and peculiarities of processes that take place in time.

Pagliari and Grimshaw (2002) observed interactions among group members, focusing on the effect of professional role and status on group discussions. However, by implying that decisions were constrained or even pre-determined by pre-existing social variables such as professional status, their study foreclosed any consideration of their emergent nature as predicated upon interactions between group members. In contrast, Moreira's (2005) ethnographic study of guideline development more subtly portrays the debates taking place during group meetings. Borrowing from Boltanski and Thévenot's (1999) "pragmatic sociology", Moreira focused on the actors' own critical capacity, identifying five types of "repertoires" used by the participants to justify the guideline's content by reference to the *actions* to which the guideline would presumably lead in the "external world". While participants from different professional groups made preferential use of specific repertoires, Moreira attributed this fact less to the presence of *a priori* interests than to the observation that group members envisioned different (future) practices and users.

Although there is considerable methodological overlap between Moreira's approach and ours, there are also several differences. Firstly, we have chosen to focus on a different empirical domain, oncology. All medical professionals attending our guideline group meetings were specialists, albeit from different disciplines. This probably accounts, in part, for the absence of a structuring effect of professional parameters on group dynamics. Oncology, moreover, has a long multidisciplinary tradition, which, in the French case we studied, is entrenched both in the institutional nature of comprehensive cancer centres and in state regulations. Secondly, the role of material and textual artifacts in the shaping of judgments and actions is a key element of pragmatic sociology, but this is notably absent in Moreira's analysis. In line with science & technology studies' longstanding focus on textual inscriptions and translations (e.g. Latour, 1990), and following up on Mykhalovskiy and Weir's (2004) programmatic suggestion to investigate the textual dimension of EBM, we pay special attention to textual practices. Guideline group meetings cannot be reduced to an exchange of arguments to select a winning position, after which the actual writing of the guideline would amount to a mere formality. Textual activities do not happen *after* consensus has been reached, they are *part* of the debate. Closure of debate does not necessarily imply that participants share the same opinion or interpretation. The collective production of a text — i.e. of a specific sequence of sentences and paragraphs that group members constantly readjust and rearrange until a final version is agreed upon — signals the end of debate. Thirdly, we have borrowed from the sociology of organizations with respect to procedures and organizational routines. One of the most striking features of the dynamics of a guideline development group lies in the role of (local) procedures, rules and distinctions, as set by the guideline developing institution and flexibly enforced, interpreted, adapted and modified by group members. Our analysis will focus precisely on this "apparatus" (or *dispositif*, to use Foucault's notion (1994)) and, in particular, on the organizational and methodological routines that are deployed in the course of group activities.

Our focus on *dispositif* and texts has led us to an additional point. Clinical trials do not test substances nor do agencies such as the FDA approve them; both institutions test and process specific *claims* about substances. Amounting to carefully crafted textual statements about the scope and results of a clinical trial (e.g. substance X works against condition Y affecting patient population Z), claims are excerpted from publications, submitted for drug market approval and embedded in guidelines. As we will see, this is far from a mechanical transposition, but this process presupposes and depends on the upstream production of specifically formatted textual claims. As a result, guidelines no longer appear as self-contained evidence-based tools targeting individual clinician's behavior; they are elements of a chain of textual *translations* linking knowledge production about therapeutic substances and pathological processes, drug marketing and the regulation of medical practices. In other words, they partake in the emergence and stabilization of a new biomedical configuration grounded in the establishment of mutually constitutive links between two processes: on the one hand, the re-formatting of clinical trials into a device for producing carefully monitored evidence statements targeting specific markets (Greene, 2007) and, on the other hand, the increasingly pervasive role of regulatory processes within biomedicine (Cambrosio, Keating, Schlich, & Weisz, 2006).

Material and methods

Our ethnographic analysis centres on an oncology guideline development group convened by a French program called "Standards, Options, Recommandations" (SOR). Established in 1993 by the National Federation of the French comprehensive cancer centres (FNCLCC), with additional financial support from a national charity, the French National League against Cancer, and the government's Health General Directorate (HAS), the SOR program was given the mandate to develop and update oncology guidelines in order to harmonize "clinical practices between cancer centres concerning diagnostic, classification, treatment and follow-up procedures" (Fédération Nationale des Centres de Lutte Contre le Cancer, 1994, p. 50; our translation). The FNCLCC is the umbrella organization of the 20 regional comprehensive cancer centres, whose origin goes back to the 1930s (Pinell, 2002) and which combine clinical research and treatments within a multidisciplinary framework. In 2008, the French National Cancer Institute (INCa, established in 2004) took legal responsibility for the SOR program. The program relies on a distinctive framework for the production of guidelines that emphasizes the need to follow the tenets of evidence-based medicine by producing recommendations resting on the best available scientific evidence or on expert consensus when adequate evidence appears to be lacking (Fervers, Hardy, & Philip, 2001). Between 1993 and 2006, SOR published 81 guidelines, i.e. 54% of the 148 published French clinical practice guidelines (Castel, 2009). Professional and public bodies have formally endorsed the SOR programme and medical audits conducted by Social Insurance use them as reference. The SOR guidelines have been diffused outside France, namely in the *British Journal of Cancer*, and the SOR program was one of the founding members of international initiatives such as the Guidelines International Network.

Our research strategy was to follow the development of a particular guideline from the initial stages to the final drafting and circulation of the guideline document, a process that in our case took place in 2007–2008 over a period of 21 months. We selected a therapeutic guideline centered on a trans-disciplinary medical condition that affects patients with different forms of cancer. The study was approved by McGill University's Research Ethics Board I. For confidentiality reasons, we have removed any information that could allow readers to identify the specific topic of the guideline and thus individual group members. The group included clinicians from the two core medical specialties treating that particular condition, as well as medical oncologists, pathology/laboratory specialists, academic researchers investigating the condition, anesthesiologists, and two SOR methodologists, for a total of 22

participants. LK and HC attended and recorded guideline development meetings that were transcribed in full and jointly analyzed by all the co-authors. In addition to field notes, we collected the documents used in the development of the guideline, including the original and modified versions of PowerPoint slides, successive drafts of the guideline, comments of external reviewers and the group's responses to those comments. We interviewed the methodologist who was responsible for selecting and reviewing the literature, drafting the recommendations, and moderating the guideline development group meetings. (Since all our original material is in French, all translations are ours.) Last but not least, we could count on a considerable amount of background information and "experiential data" given that the three French co-authors of this paper have many years of acquaintance with the SOR program: PC has written a doctoral thesis on French cancer guidelines (Castel, 2002), BF has acted as director of SOR for several years and HC was employed as a sociologist by SOR.

Clinical oncology is, in a sense, pre-formatted for guidelines, given its reliance on protocols derived from large, multicenter clinical trials — a large proportion of which are carried out by non-commercial cooperative groups — and the aforementioned multidisciplinarity. The SOR procedures analyzed in this paper, however, are by no means unique to this program nor are they specific to oncology. We therefore expect our analysis to apply to other guideline development programs, in particular those sponsored by public or professional organizations.

The standards, options, recommandations (SOR) *dispositif*

During guideline development meetings, the methodologist (see below) and the coordinator (or chair) constantly reminded group members of the procedural and methodological routines guiding and structuring the guideline development process. As mentioned in our introduction, observers have overlooked the extent to which social interactions and arguments are embedded in procedural rules and methodological devices that act simultaneously as flexible constraints and as resources. We can borrow (and slightly distort) Foucault's (1994, p. 299) notion of a *dispositif* — defined as a system of relations that can be established between heterogeneous elements such as discourses, institutions, regulatory decisions, laws, administrative measures and scientific statements — to portray SOR as an apparatus that associates and aligns participants, procedures, texts, topics and resources. In contrast to procedural rationality analyses that tend to grant a determining role to procedures and blur the contribution of participants into the background, we conceive of both methodological procedures and participants as constitutive elements of the *dispositif*; one does not take precedence over the other.

Participants include professionals such as methodologists and research librarians who are part of the SOR staff, as well as outside clinicians and experts from the relevant domains. Methodologists derive statements from the published articles and use technologies such as PowerPoint slides and Excel tables to present participants with summaries and overviews of the existing literature. These tools provide an interactional infrastructure: debates take place around the projected slides, on which textual amendments are entered in real time. Two central notions, discussed below, are used to discuss, modify and rearrange statements: *Levels of Evidence* and *Standard/Options*. They allow group members to fine tune their recommendations so as to translate the degree of (un)certainty and (dis)agreement that surrounds certain procedures. A number of explicit and implicit procedural rules and categories are used to operationalize these two notions.

Participating physicians come not only from the network of cancer centres to which SOR was affiliated, but also from public and private hospitals. By the turn of the century, about 2000 practitioners had been involved with the SOR guidelines, half of them from the Federation's cancer centres. SOR can thus claim to staff its guideline development groups with the best experts, regardless of institutional membership. Methodologists, in contrast, are project managers hired by SOR to prepare a synthesis and critical appraisal of the evidence and to organize the guideline group meetings. Originally trained as biostatisticians, information specialists or health professionals, methodologists are usually not MDs; some of them are, but they no longer practice and are thus unable to claim clinical expertise. Still, they are full-fledged participants in group discussions, officially entrusted with the task of ensuring that the resulting guidelines remain congruent with the available evidence. In 2000, because of the growing amount of clinical publications and the complexity of managing the resulting databases, SOR recruited research librarians to support the methodologists' work.

Systematic reviews of the scientific literature often show that evidence on a given procedure is either absent, of limited quality, contradictory or derived from a different patient population, clinical setting or procedure. The SOR *dispositif* thus also elicits and encompasses expert opinions. As the methodologist explained during an interview, "when many data are available, it's a no-brainer, data speak by themselves … when we have to deal with less data, it's up to the expert to speak up". To counter the potential liability caused by this situation, the formal SOR methodology suggests that the members of the development group ought to be chosen "by the most objective criteria" (Dosquet, Goldberg, & Matillon, 1995, p. 759). "Objectivity", in this context, is defined as a "civic worth" (Boltanski & Thévenot, 1999) insofar as it refers to the inclusion of members from a wide variety of specialties and groups, public as well as private institutions, at least one rank-and-file clinician and sometimes nurses and patient representatives (in our case the latter two groups were absent). While professional organizations can and, in our case, did formally endorse the procedure by jointly supporting the guideline development project and by approving the resulting guideline, they did not elect or select experts from their midst to act as their official representatives on the guideline committee. Clinical coordinators recruited them from amongst their professional and personal contacts on the basis of their expertise and national visibility. As a methodologist explained in an interview, "they are colleagues that they meet at meetings and workshops, with whom they have exchanged or co-authored articles. … One takes the two best French [name of specialty] and the two best [name of specialty], and that's it".

As can be easily seen, the SOR *dispositif* follows the tenets of EBM by relying on literature reviews and published evidence, but has also adopted a pragmatic approach by supplementing formal methods with a (necessary) reliance on expert opinion for the assessment of that information. The official endorsement by professional organizations, and the presence of experts from the relevant specialties within the development group, increases the legitimacy of the procedure. It also facilitates the practical requirement of promoting the guideline within each of the experts' distinct professional and geographical milieus (their "parishes"). SOR expects members of a guideline group to act as the guideline's advocates and promoters: "we have here representatives of specialties from very different horizons, we have to think how to communicate [the recommendations] to the colleagues in our own parishes because, first of all, there is a time issue, the issue of being recognized [as experts in the field] and, more importantly, there is the issue of the legitimacy of this group".[1]

[1] When not otherwise indicated, all quotations are from meeting transcripts.

External review of draft recommendations is another important part of the SOR *dispositif*. They are generally mailed to 150–250 external reviewers (230 in the present case) from different specialties and domains. Presented as a form of external quality control testing, this "national review" process acts as a way of internalizing the external world and thus of fostering the acceptance and implementation of the final recommendations by potential users. In practice, only 25–30% of the reviewers (28% in our case) actually provide comments and suggest modifications. These feedbacks often lead to changes of the actual content of the guideline. Equally as important, and in spite of the relative small percentage of respondents, the review is used to demonstrate the approval of the medical community at large. The final guideline document includes tables and pie charts stating, in our case, that 93% of consulted physicians approved the guideline and 94% said they would use it.

The *dispositif* is what makes the process accountable, in the dual sense of providing participants with elements and rationales for defining and modifying the guideline's content, and for establishing the legitimacy of those activities. Not only should procedures be followed, but this should be done in a visible and transparent way: search keywords, evidence tables, external review results, the names and specialties of participants are all listed in the guideline or its appendices. The *dispositif* itself gains its legitimacy from this transparency, its institutional entrenchment in the FNCLCC and its interfaces with international initiatives. It is too early to know whether and how SOR's takeover by a government institution (INCa) will change the situation.

Procedural rules and categorizations

At the beginning of every guideline development meeting, the methodologist shows two PowerPoint slides to remind group members of the categories and procedures that are supposed to guide the process. By accepting to become members of the group, participants are expected to abide by these "rules of the game". Confronted with a participant who broke this implicit contract by questioning the basic procedures, a methodologist reacted rather swiftly by stating that "the approach we take is the classical SOR approach, to which we all subscribed … the fact of being a group member, and we thank all of you for that, means that you accept to play by the rules of the game, we all accepted it at the outset, and it's a clear methodological rule".

As we will see below, acceptance *in principle* of these rules does not mean that they cannot be challenged *in specific instances*. It is useful to refer, in this respect, to the distinction introduced by Feldman and Pentland (2003) between ostensive and performative aspects of routines. Borrowing a classical Wittgenstein argument, they argue that "no amount of rules is sufficient to specify a pattern of behavior fully, because the interpretation of any rule, or any part of a rule, requires more rules" (p. 101). Much of the group meetings we observed centered on how to apply rules to particular cases, i.e. how to perform ostensive routines. For instance, while participants had to agree that the literature review that provided a basis for the guideline rested on a formal assessment of the type of evidence reported by articles, they were also able to point to details of specific studies that cast doubts on their conclusions (or even mention extra-textual contingencies: "I'm fed up with [study X] … they keep trying to fob this off on us day after day, this has been going on for twenty years, I'm fed up"). They were thus able to reject or reduce the evidential status of claims that, formally speaking, ranked as high-quality evidence.

Our case study reinforces Feldman and Pentland's (2003) argument, showing that the procedural rules provided by SOR amounted to a mixture of *both* constraints *and* resources (see also Crozier, 1964): constraints because "ostensive" rules and routines had to be taken into account in group decision-making, and resources because they were flexibly "performed" to generate decisions. In other words, the SOR procedures were not a mere varnish that was used retrospectively to legitimate decisions taken on other grounds, they were mobilized in the course of discussions and made a difference. Neither did they structure these discussions in any deterministic way or led to pre-established conclusions.

So, what were the rules and procedures displayed on the two introductory slides? The first slide introduced the distinction between Standards and Options, obviously two central categories insofar as they are embedded in the institution's name. The second slide defined the four main categories of Levels of Evidence that participants had to use in assessing the evidence and drafting recommendations.

Levels of evidence

The "Levels of Evidence" (henceforth LOE) slide listed four levels (A to D) ranging from evidence supported by "good quality meta-analyses" or by a "coherent" set of "good quality" randomized clinical trials (level A), to no data or only case report evidence (level D). Levels contained sub-categories: the B level, for instance, included B1 evidence (randomized trials) and B2 evidence (prospective or retrospective studies). Resort to an evidential hierarchy is of course not a SOR innovation; it is a defining feature of EBM and a core methodological component of systematic literature reviews. There are, however, dozens of slightly different hierarchies, in spite of the fact that an international collaboration has attempted for several years to devise a single standardized version (GRADE working group, 2004).

As previously mentioned, one of the participants launched a wholesale attack against the hierarchy used by SOR and, one could argue, against the very idea of establishing formal rules for determining LOE, since his target was the failure to provide a detailed assessment of each study: "the methodological quality, the type of randomization, the non-existence of treatment differences between the two arms of the trial, the analysis that led to the reported results, the criteria that have been taken into account, and so on … all these elements are totally absent [from the LOE scheme]". The methodologist and the clinical coordinator, both acting as spokespersons for the SOR *dispositif*, rejected this global challenge on pragmatic grounds: the need for consistency across SOR guideline groups ("this particular hierarchy has been used for all the guidelines in the past 10 years", "we cannot change LOE tables every six months for each project") and the fact that in the absence of international standards the LOE used by SOR were as good as any other:

> *Coordinator*: This being said, all these [LOE classification] tables can be criticized: I had a look at the tables used by the Italian guidelines, and then I looked at the tables used by the [name of French healthcare agency], and I said to myself, gosh, there is no coherence, and then one has to decide [which one to use], and so we might as well use the SOR tables…

When the challenger insisted with his criticism, the methodologist became more blunt: "those are the rules of the game, we don't modify them … that's how it is, you can't do anything about that". This did not mean, however, that the attribution of a LOE category would proceed automatically: to the contrary, once the general principle was accepted, its application to individual studies often raised many discussions. Did a given study amount to a "good quality" clinical trial? Did it count as a prospective or a retrospective study? If four out of five studies agreed, did that count as "a coherent set of studies"? In the case of particular studies, participants discussed criteria such as the number of patients included

("Either we stick with B2 or if we consider that too few patients were included, we stay with C") or the date of publication ("So, let's use only recent studies, because if we exclude [the older reference X], then the conclusion can be level of evidence A or B"), or even distinctively more idiosyncratic parameters such as the previously quoted statement about being "fed up with [study X]".

Quite often, discussions focused on how to categorize a given study's methodological design. For instance, a study claimed to be a retrospective analysis, but when one of the participants asked what that really meant, a long discussion followed which did not lead to consensus:

> *Methodologist*: Well, that means that they took the [study] arms a posteriori, the five studies …, they extracted data from the five retrospective studies. …
> *Participant 2*: I think that we should say: "it's an a posteriori analysis of five prospective randomized trials" and add that it's an a posteriori analysis, with cancer patients, of five studies. And in fact this is what it is, we are mixing up prospective and retrospective.

After much discussion, and in the interest of time, the coordinator put an end to the debate by deferring the decision to the methodologist: "OK my dear friends, we have to go ahead… I'll write down in red 'verify methodology' and we will check that".

As it can be seen from these selected but representative examples, procedural rules and distinctions provided a frame for the debate, but they were unspecific and flexible enough to require interpretation in their application. They were in fact the starting point and a recurring element of the debates that took place within the group and were embedded in the justifications offered in support of a given statement, but they did not dictate the outcome of interactions. The methodologist and the coordinator played an important role in reminding participants of the official frame but the debate was essentially open-ended. When participants contradicted the assessment by a methodologist of the quality or LOE of a given study, they did so *thanks to*, rather than *in spite of* the existence of procedural rules. The application of the rules to specific cases was open to debate and enforcement was limited to frequent but friendly warnings that could be quite elastic: "I remind you that [in this recommendation] there is no level of evidence and, it seems, we step outside our role and our job, *but oh well!*" (Methodologist, our emphasis). Boundaries were elastic, and so, by redefining them, one could step outside them without officially doing so. But it would be wrong to infer that "anything goes", for these actions had to be justified: there were rules, routines and categories (however elastic) that *had* to be heeded. The specific justification that would eventually carry the day, however, emerged from the debate and was in no way pre-determined.

Standards and Options

In addition to LOE, a second important device contributed to the open-ended framing of the debates, namely the distinction between Standard and Options. According to the first introductory slide, a Standard refers to a clinical procedure that experts *unanimously* regard as the gold standard. Options refer to clinical procedures (notice the plural) that experts consider *appropriate*, even in cases when experts favor one of those options. It is important to note that, unlike LOE, the Standard/Options distinction is based on the *level of agreement*, and not on the *level of evidence*: standards are by unanimous decision, while options are agreed upon by a majority (Fervers et al., 2001). *Ideally*, a standard should be based on the highest LOE, but this is not a necessary condition: as noted by a methodologist in response to a commentary from an external reviewer: "A recommendation can be listed as a Standard even in the absence of data from the literature or even when supported only by a weak level of evidence".

This flexibility to recast a statement as either a Standard or an Option was a very useful device for bringing (temporary) closure to a debate by allowing recommendations to be made even in the presence of disagreements and uncertainty. After all, as mentioned by participants, it was precisely in the areas where evidence and unanimity was lacking that they were expected to provide guidance for practitioners: "those who read our recommendations, they expect from us that we give them options, even weak options based only on expert opinion, and if we don't give them options, it's not worth doing all this work".

The Standard/Options distinction was a tool for managing uncertainty in an additional, more subtle sense. It helped to define the "grey zones" in need of further investigation by redefining the boundaries of uncertainty as new arguments were introduced about the availability or absence of evidence, and by fuelling discussions about the tasks that lie ahead: how to learn what was not yet known, where to look for possible answers. Uncertain domains were not zero-knowledge domains or independent variables constraining medical activities or judgments, as Renée Fox (2003) would have it. Rather, uncertainties were the outcome of the actors' practices (Bourret & Rabeharisoa, 2008).

Some participants actually understood the distinction between Standard and Options in terms of *degrees of certainty* (and not in terms of *agreement*, as previously mentioned): Standards were certain, Options had not been proven beyond reasonable doubt. The trouble with this interpretation was that it left open the possibility of having several coexisting standards (with well-established criteria for choosing between them) or of ending up with a single (uncertain) option, contradicting what others considered the defining plurality criterion. Consider the following exchange:

> *Participant 1*: But that is a standard. It's a standard to say that we can't say anything; it's clearly not an option.
> *Participant 2*: No, a standard and an option, they should both recommend a given procedure, so it really should say "I can or cannot use [drug X]".
> *Participant 3*: If it's optional, by definition we need at least two possible procedures … there needs to be several alternatives.

Recommending several courses of action raised the question of how to define the criteria for deciding between the available choices. One way of handling this question was to leave the choice to the individual clinician facing a given clinical situation (use option X "if needed"), but other participants insisted that the recommendation should specify the conditions under which a given option became quasi mandatory ("We must list situations in which an option should be used"), thus turning it into a proto-standard.

Both in the case of LOE and of the Standard/Options categories, it is tempting to resort to the category of *negotiations*, as sociologists often do, to account for the interactions we reported. A number of situations we witnessed appeared to fit such a description, for instance in the case of long exchanges about whether the LOE of a given claim should be moved up from B2 to B1 or, to the contrary, retrograded to C. And yet, negotiations is not the right term, since it implies that the issue was well defined, that the negotiating groups had pre-existing opinions or interests about it, and that they looked for a more or less favorable compromise, possibly in exchange of other concessions. The situation in the guideline group was different: in the case of controversial statements, it did not amount to a clash between pre-established choices or opinions (underwritten by different professional interests) but to an attempt to chart uncharted territory. As discussed in the next section, the process did not imply a freewheeling discussion: framed by the SOR *dispositif*, it focused on the fine-tuning of *textual statements* to make

them compliant with two requirements: that they form a coherent sequence of interconnected textual statements, and that they include the appropriate mixture of precision and vagueness.

Textual practices

So far, we have discussed the constitutive elements (people, procedures, rules and categories) of the SOR *dispositif* and the dynamics they engendered. It is now time to focus on the *product* of these activities. Offering a forum for the exchange of arguments, or reaching an intellectual consensus on the value of a given intervention was not the final purpose of this complex procedure. Rather, it was to produce a text. The chain of textual translation began with the extraction of textual claims from published clinical trial reports: methodologists compiled extracts and created tables and summaries. Several of these statements were translated without much debate from evidence tables to conclusions and turned into recommendations, sometimes literally by copy and paste. In a significant number of cases, however, textual statements led to more or less heated debates within the group. Without discounting the presence of the former occurrences, we focus on the latter.

By following closely the debates that took place within the group, we noticed that it was sometimes unclear which argument had won the day, or how divergent arguments added up to a given conclusion; yet, in the absence of objections to a proposed text, closure would still happen. To produce a text members of the guideline group did not need to agree on a given matter. A textual statement could be formulated in such a way as to leave room for flexible interpretations; group members would then feel comfortable with underwriting it even if its exact operational meaning remained unclear. For example, some participants argued that a given treatment should be administered for three months, while others opted for six months. Each group maintained its position and no intermediary solution (four or five months) seemed viable. Closure was reached around the following formulation: "at least three months". In this way, both options were consistent with the guideline. Having reached consensus on that formulation, group members were able to walk away with divergent opinions.

Several observers have claimed that the textual clarity and precision of a guideline have positive effects on its implementation (e.g. Michie & Lester, 2005). But, as already mentioned, lack of precision is often *needed* to produce a text providing guidance, for controlled fuzziness is a resource for generating closure on a statement (Allen, 2009). There are also other reasons why group members considered it impossible or inappropriate to provide detailed guideline statements, such as a practical impossibility to foresee the degree of detail needed by every practitioner, patient and institution operating under different circumstances. Specifications could be made later, at the local level during the implementation phase. A boundary was drawn (in principle) between the job of developing a guideline and the job of implementing it:

> *Coordinator*: … we have a [guideline] text, and its application belongs to a subsequent stage, and for that there will be [the equivalent of] pull-out recipe cards or summary cards that will become available and people will produce them in a variety of forms. Because we cannot be expected to write down bedside prescriptions. Therefore there will be, we will all be asked to be part of a six month period during which we will have to explain the [guideline] text.

The formulation of the text of the SOR guideline was a truly collective effort. Methodologists or individual group members did not write the text after the meeting according to their recollection and understanding of the debates, instead, the *precise* wording of the text was debated and worked out in real time on the PowerPoint slides. The issue was not one of editorial esthetics; rather, the choice of words was held to define the content and the (possible) consequences of the guideline recommendations. Moreover, to produce a guideline the group had to decide not only *what* to write but also *where* to locate specific statements in the document. The text of SOR guidelines follows a strict, pre-determined sequence that suggests a linear process according to which the literature provides evidence associated with LOE, which is assessed and/or supplemented by the experts, all this leading to a set of recommendations. The sequence did in part correspond to the actual processes we observed (e.g. the methodologist prepared a literature review before the first group meeting), but the group discussions were far from linear. The positioning of specific statements within the overall organization of the document did not automatically flow from the layout conventions. During the discussions, group members travelled back and forth between sections:

> *Participant*: … who says that? The literature says that? Or it's you, based on your clinical practice? … If it's you, it should go in the Expert Judgment section.

Travelling between sections also implied revisiting phases of the guideline development process. For instance, participants confronted with a particularly thorny issue agreed on the need to draft a strongly worded recommendation to prevent widespread instances of "bad practices". But a "strong" recommendation should ideally be based on "strong evidence" which, in the present case, was not available when participants reviewed the relevant publications. So, they revisited them and, by eliminating reference to a particular study whose results were not in agreement with the other studies and whose methodology seemed doubtful to some participants, were able to raise the overall LOE for the recommendation from C to B1. Although the literature review preceded the drafting of the recommendations, in this particular instance it was retrospectively modified to make it consistent with what participants felt was a crucial recommendation.

SOR guidelines must include both a set of specific recommendations and the reasons for those choices. The textual location of a statement played a critical role in establishing evidential connections between the recommendations and the different kinds of justifications supporting them. Group members were reminded of the fact that the coherence of a guideline depended on its overall organization and the need to showcase this coherence:

> *Methodologist*: The reader must get a clear sense of all this, he should not be left wondering how we ended up with this recommendation, like "there was no literature and here they go and suggest a Standard", see?

Skeptical or critically minded readers might infer, at this point, that the final layout of the guideline, with its neat argumentative sequence, was "merely" a rhetorical device for retrospectively justifying decisions and recommendations that were taken on different grounds and certainly not by following the linear process implied by the layout. Yet, this conclusion would miss the decisive fact that the linear template acted as a performative device prompting group members to move back and forth between sections of the text and to adjust its coherence little by little.

Coming to terms with the external world

The avowed objective of guidelines is to change healthcare practices by replacing substandard interventions. While maintaining a proactive stance, guideline developers are also aware of the fact that they need to adjust recommendations to prevailing external

conditions. How and to what extent remains an open question, and leads to discussions about the potential impact (or lack thereof) of specific recommendations. By evoking possible scenarios, guideline developers attempt to manage relations between recommendations, users and other relevant stakeholders.

In this respect, the aforementioned distinction between production and implementation proved to be a useful tool for framing the participant's work, but was often difficult to maintain and had to be re-specified when group members were confronted with the practical task of wording specific recommendations in such a way that they could have a reasonable chance of being of any practical import. Participants, arguing that they were "not going to recommend something that has no chance of being applied", wondered about what would make health practitioners "out there" more likely to modify established routines and replace them with the new recommendations. The obstacles they mentioned included patient preferences ("patients don't want that"), a lack of local resources ("the necessary radiological equipment must indeed be available") or legal arrangements ("If we write that down in our recommendations, there will be only 20 centres in France that will have the right to [perform the procedure]"). While these kinds of concerns were frequently raised, they were often declared off mandate or unacceptable, at least in principle:

> *Participant*: If we argue that [a procedure] should not be performed because there is no evidence, OK, I agree with that. But to argue that "this will not be accepted, this will not gather any support, it will not be done because patients don't want that" … I cannot agree with this kind of argument.

Group members, insofar as they were health practitioners, had a dual status: they were experts but also, and simultaneously, users, albeit "expert users" as contrasted with "rank-and-file users". They were thus able to "switch hats" during debates. Role switching did not correspond to specific sequences of the guideline production process (e.g. experts during LOE discussions and users when debating practical issues), but took place at any stage in the process. Participants, however, considered the role of state regulators or administrative decision-makers beyond their purview. To steer clear of conflict, guideline recommendations should as far as possible avoid addressing the concerns or mandates of other jurisdictions:

> *Coordinator*: It's up to them [French drug regulators] to get a move on and make [a particular version of a drug] available in France, should they judge that it is appropriate. It's not our problem.

In the discussion, group members, while agreeing on which drugs should be used, were faced with the fact that a particular drug has not yet won regulatory approval in France. They thus wondered whether they should insert a statement mentioning its "limited availability". This issue led to much soul searching: Would the recommendation have a chance, under such circumstances, to be correctly implemented? Could this be interpreted as favoring a drug over another or as meddling with approval procedures, and, if so, how would the pharmaceutical industry react? Would this affect the perception of the integrity and transparency of the guideline development process? Group members constantly reminded themselves that they would be held accountable for their work and thus of the need for a delicate balancing act so as not to antagonize a range of different parties.

The guideline development process also included a formal mechanism — the aforementioned external review — to align the guideline's content with the "external world". This procedure is part and parcel of the *dispositif*, in the dual sense of being one of its formal requirements and of supplying feedbacks, both real and apprehended, leading to text modifications. Awareness of the impending review clearly shaped the wording (and thus the content) of the recommendations. The following remark by a methodologist was typical: "[These data] should allow us to opt without major contradiction for a B2 level of evidence, and without risking a full-fledged attack when [the draft guideline] will be circulated for review".

A guideline is a text but, unlike, say, the text of a scientific article that, once published, severs its ties with the laboratory from which it originated, the text of a guideline remains attached to its producers after its publication. It will be revised at more or less regular intervals to take into account new evidence, publication of the final text is followed by a period during which guideline developers promote the results of their work to potential users, and, last but not least, recommendations will subsequently have to be rendered into summaries, decision trees and regional guidelines adapting them to local conditions, thus pursuing the cycle of textual translations.

Conclusion

Increased international collaboration and comparisons have led to the somewhat disheartening observation that guidelines on the same topic frequently contain divergent recommendations. Many guideline developers and EBM advocates argue that this is not necessarily a problem insofar as variation is warranted by different local (usually national) differences and certain procedural requirements are fulfilled (Eisenberg, 2002; Fervers et al., 2006). The most widely used instrument to evaluate guideline quality does not examine a guidelines' clinical content or its underlying evidence, but only evaluates the characteristics of its production process (AGREE Collaboration, 2003). In short, production procedures are increasingly important to legitimate guidelines. In contrast to formal, checklist-like evaluation tools of the guideline developmental process, our paper relied on ethnographic observation of how a particular guideline *dispositif* framed the production of a guideline, calling attention, in particular, to the textual dimension of these activities. While subject to limitations due to its restricted focus, this kind of detailed analysis has the potential to contribute to a better understanding of the dynamics of guideline production, especially if combined with similar studies carried out in other settings (e.g. Moreira, 2005). We hope, in particular, to promote studies of the *dispositif* of other national and international guideline development institutions, especially those mandating public or patient participation, involving commercial in addition to professional actors, or outsourcing evidence-gathering processes.

Our analysis highlights that a guideline's credibility does not rest solely on its evidential basis, as quality assessment of evidence is an intrinsic part of the process of developing guidelines. Moreover, this process is in part, self-vindicating, insofar as it is made possible by the existence of meta-regulatory documents such as other guidelines. But nor is its objectivity based on mechanical procedures for compiling facts and data. Its legitimacy rests on the articulation of heterogeneous types of expert knowledge and judgments, both within the guideline development group, and vis-à-vis an external world of textual documents (clinical trial reports, other regulatory texts, external review comments, etc.). As such, guidelines act as *mediators* at an important junction in the extended chain of textual statements produced, assessed and processed by a range of public and commercial research institutions and regulatory agencies that are part of the meta-regulatory web of contemporary biomedicine (Cambrosio et al., 2006). As mentioned in the introduction, clinical trials, the basis on which the EBM enterprise rests, are not so much devices for testing drugs in any "absolute" sense, but, rather, have become devices for producing specific claims concerning a given

molecule, test or procedure (Greene, 2007). Once excerpted from the original publications, these formatted statements can be further processed (e.g. by associating them with a LOE) and transferred to other texts, such as guidelines. This, however, is not a mechanistic account, insofar as the role attributed to expert opinion (as embedded, for instance, in the Standard/Options distinction) acts as a versatile counter-mechanism.

It would be wrong to portray this counter-mechanism as a form of resistance (for instance, by clinicians) to the reconfiguration of biomedical practices, in the same way as it would be incorrect to equate the work of methodologists with the *dispositif* and pit it against the request for an "alignment with the external world" conveyed by clinicians. "Expert opinion" and "evidence-based statements" are part and parcel of the same *dispositif* and work together to produce new knowledge and new practices that do not correspond to the somewhat paranoid picture of a corporate hijacking of medical knowledge but neither do they correspond to the positivist utopia of an EBM free from any undue interference from anecdotal, contingent or even idiosyncratic forms of justification.

Acknowledgements

Research for this paper was made possible by a grant from the "61e Commission permanente de coopération franco-québécoise" (61.405) and from the French Institut National du Cancer (INCa 07/3D1418-PL-85-04). Additional financial support was provided by Quebec's FQRSC grant SE-124896, and LK holds an FRSQ Doctoral Award (13024).

References

AGREE Collaboration. (2003). Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality & Safety in Health Care, 12*, 18–23.

Akrich, M. (1992). The de-scription of technical objects. In W. E. Bijker, & J. Law (Eds.), *Shaping technology/building society. Studies in sociotechnical change* (pp. 205–224). Cambridge, MA: MIT Press.

Allen, D. (2009). From boundary concept to boundary object: the practice and politics of care pathway development. *Social Science & Medicine, 69*, 354–361.

Berg, M., Horstman, K., Plass, S., & van Heusden, M. (2000). Guidelines, professionals and the production of objectivity. Standardisation and the professionalisation of insurance medicine. *Sociology of Health and Illness, 22*, 765–791.

Boltanski, L., & Thévenot, L. (1999). The sociology of critical capacity. *European Journal of Social Theory, 2*, 359–378.

Bourret, P., & Rabeharisoa, V. (2008). Le jugement et la décision médicale en situation de forte incertitude: l'exemple de deux pratiques cliniques à l'épreuve de la génétique. *Sciences Sociales et Santé, 26*, 33–66.

Cambrosio, A., Keating, P., Schlich, T., & Weisz, G. (2006). Regulatory objectivity and the generation and management of evidence in medicine. *Social Science & Medicine, 63*, 189–199.

Castel, P. (2002). Normaliser les pratiques, organizer les médecins: la qualité comme stratégie de changement. *Le cas des Centres de Lutte Contre le Cancer*. Doctoral Thesis. Institut d'Études politiques de Paris (Sociologie).

Castel, P. (2009). What's behind a guideline? Authority, competition and collaboration in the French oncology sector. *Social Studies of Science, 39*, 743–764.

Castel, P., & Merle, I. (2002). Quand les normes de pratiques deviennent une ressource pour les médecins. *Sociologie Du Travail, 44*, 337–355.

Crozier, M. (1964). *The bureaucratic phenomenon*. Chicago: University of Chicago Press.

Dosquet, P., Goldberg, J., & Matillon, Y. (1995). Comment élaborer des recommandations de pratique clinique? *Bulletin Du Cancer, 82*, 754–760.

EBM Working Group. (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *Journal of American Medical Association, 268*, 2420–2425.

Eccles, M., Clapp, Z., Grimshaw, J., Adams, P. C., Higgins, B., Purves, I., et al. (1996). North of England evidence based guidelines development project: methods of guideline development. *British Medical Journal, 312*, 760–762.

Eddy, D. (1990). Practice policies: guidelines for methods. *Journal of American Medical Association, 263*, 839–841.

Eisenberg, J. M. (2002). Globalize the evidence, localize the decision: evidence-based medicine and international diversity. *Health Affairs, 21*, 166–168.

Fédération Nationale des Centres de Lutte Contre le Cancer. (1994). *Rapport d'activité*. Paris: FNCLCC.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly, 48*, 94–118.

Fervers, B., Burgers, J. S., Haugh, M. C., Latreille, J., Mlika-Cabanne, N., Paquet, L., et al. (2006). Adaptation of clinical guidelines: literature review and proposition for a framework and procedure. *International Journal for Quality in Health Care, 18*, 167–176.

Fervers, B., Hardy, J., & Philip, T. (2001). Standards, Options and Recommendations. Clinical Practice Guidelines for cancer care from the French National Federation of Cancer Centres (FNCLCC). *British Journal of Cancer, 84*(Suppl. 2), 1–92.

Foucault, M. (1994). 1976–1979*Dits et écrits 1954–1988, Vol. III*. Paris: Gallimard.

Fox, R. (2003). Medical uncertainty revisited. In G. L. Albrecht, R. Fitzpatrick, & S. C. Scrimshaw (Eds.), *Handbook of social studies in health and medicine* (pp. 409–425). Thousand Oaks, CA: Sage.

GRADE working group. (2004). Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. *BMC Health Services Research, 4*, 38.

Greene, J. (2007). *Prescribing by numbers. Drugs and the definition of disease*. Baltimore: John Hopkins University Press.

Hutchings, A., & Raine, R. (2006). A systematic review of factors affecting the judgments produced by formal consensus development methods in health care. *Journal of Health Services Research & Policy, 11*, 172–179.

Latour, B. (1990). Drawing things together. In M. Lynch, & S. Woolgar (Eds.), *Representation in scientific practice* (pp. 19–68). Cambridge: MIT Press.

Michie, S., & Lester, K. (2005). Words matter: increasing the implementation of clinical guidelines. *Quality & Safety in Health Care, 14*, 367–370.

Moreira, T. (2005). Diversity in clinical guidelines: the role of repertoires of evaluation. *Social Science & Medicine, 60*, 1975–1985.

Mykhalovskiy, E., & Weir, L. (2004). The problem of evidence-based medicine: directions for social science. *Social Science & Medicine, 59*, 1059–1069.

Pagliari, C., & Grimshaw, J. (2002). Impact of group structure and process on multidisciplinary evidence-based guideline development: an observational study. *Journal of Evaluation in Clinical Practice, 8*, 145–153.

Pinell, P. (2002). *The fight against cancer: France 1890–1940*. London: Routledge.

Reynaud, J. D., & Richebé, N. (2007). Règles, conventions et valeurs. Plaidoyer pour la normativité ordinaire. *Revue Française De Sociologie, 48*(1), 3–36.

Timmermans, S., & Berg, M. (2003). *The gold standard: A challenge of evidence-based medicine and standardization in health care*. Philadelphia: Temple University Press.

Weisz, G., Cambrosio, A., Keating, P., Knaapen, L., Schlich, T., & Tournay, V. J. (2007). The emergence of clinical practice guidelines. *Milbank Quarterly, 85*, 691–727.