Abstract—The paper discusses the potentially disruptive role of ICT in consolidating an approach to medical knowledge that is as much rationalistic as it is far from actual practice. We also discuss the role that lightweight web-based tools like on-line survey systems can play in making doctors of highly distributed and loosely integrated medical associations aware of treatments of choice of their colleagues and of the potential agreement on these preferences within their community. We present a case study that involved a medical association counting more than one thousand surgeons; we sent to these surgeons an invitation to a web-based survey and gathered responses from 409 of them (response rate: 40%). The findings are discussed as implications for design of tools that could promote discussion and collaboration within large medical communities to overcome the above mentioned drawback.

I. ICT AND MEDICAL KNOWLEDGE

From the very beginning, the adoption of ICT in the healthcare domain was aimed at rationalizing healthcare processes so as to induce ordered behaviors of practitioners and allow for the structured documentation of interventions for many purposes [1]; in this process, ICT has facilitated the collection and analysis of those techniques and interventions leading to positive outcomes and hence provided the basis for the construction of a reliable corpus of medical knowledge. Since then, ICTs and medical knowledge have established a profound alliance and reached a long-lasting community of interests and aims. On the one hand, in the last decades ICTs have helped doctors in both producing and getting access to a huge amount of studies and guidelines that has been progressively organized in repositories and can be accessed and queried flexibly from almost everywhere. Moreover, computer systems have been used for decades to analyze the large amounts of data generated from electronic medical records and clinical research, with the aim to suggest doctors unexpected associations, facilitate the validation of existing theories or assist the design of new drugs [2].

These technological and cultural trends are at the basis of the raise of the so called Evidence-Based Medicine (EBM); this is a quite recent cultural phenomenon that drives doctors to apply the best available evidences gained through scientific methods to their daily decision making and interpretation of human cases. Here the role of communication technologies is deemed so central that some authors have even claimed that without them, EBM would not exist [3, p.17] [4], [5]. In fact, computer-based technologies enable the coordination of complex clinical trials, move large quantities of data in almost no time and make the resulting evidences available in any corner of the world in almost no cost. The spread of EBM, backed by ICT specifically supporting medical practice [6], is however flanked by some emergent perplexities. A number of scholars are witnessing a problematic relationship between doctors and this body of "official" medical knowledge (see e.g. [7]) and between doctors and the technologies that should neutrally help them leverage it (see e.g., [8]). There are many reasons why doctors do not (want to) follow practice guidelines and evidence-based recommendations, ranging from barriers that affect physician knowledge (i.e., lack of awareness or lack of familiarity), attitudes (lack of agreement, lack of self-efficacy, lack of outcome expectancy) or behaviors (e.g., the inertia of previous practice) [9]. On the other hand, when evidence-based recommendations are extensively and systematically reviewed in meta-analyses, the number of interventions that actually show either no effect or insufficient evidence is unexpectedly high [10]. These and similar contributions confirm the widespread opinion among doctors that scientifically grounded certainty is a sort of myth in their profession. In fact, these studies highlight how doctors prefer to intervene on the body of patients on the basis of personal habits, available resources and equipment, familiarity and proficiency with a technique, personal experience of outcomes, school tradition as well as for “political” reasons and economic advantages, rather than on the basis of evidences coming from clinical trials performed at a thousand miles away or just by some – possibly competing – colleagues. For example, Greenhalg [11] challenges the basic tenets of EBM and confronts them with the outcomes of studies that report on the “dissonance between the “science” of objective measurement and the “art ” of clinical proficiency and judgement”[p.323], and Tonelli [12] advocates a more case-based approach that “balances empirical evidence and experiential evidence […] before coming to a presumptive conclusion regarding the appropriate course of action” [p.255]. These studies support the promotion of more individualistic approaches aimed at maintaining the autonomy of individual doctors towards the peculiar needs of their particular patients (cf. the recent rhetoric of patient-centredness) [13].

As ICT designers we should then consider the risk that embedding standards and clear-cut evidences in supportive “rationalistic” technologies, like e.g., treatment algorithms and guidelines enacted by computer-based systems, could end up by weakening either the “healing power” or the decisional
autonomy of the social institution of medicine. This risk is not a concern of secondary importance and brings to profound implications with respect to design. In fact, it brings to taking the social nature of preferences and attitudes in the medical community more seriously, and to considering how information on preferred treatments and relevant indications is shared, and how consensus on this information can be achieved or, at least, agreement increased. Including these concerns in the agenda of the ICT design for the medical profession should be consequently prioritized.

Starting from this basic tenet, this paper takes its motivations from two complementary considerations: On the one hand, the medical traditional and institutional modes of ratification and dissemination of medical knowledge seem to be to a large extent substantially blind to the “silent evidences” of daily practices: this uncovered “dissonance” shows the need of an adequate technology that could support the sharing of other forms of empirically based knowledge. On the other hand, ICT designers can not side with an approach to medical decision or the other, not only because they do not possess the necessary competence to do it, but also because it is quite obvious that no approach (e.g., evidence-based vs. consensus-based vs. case-based and the like) is better than the other per se. Yet, we notice that, nowadays, the technological support is unevenly oriented towards the management, diffusion and self-reflecting ratification of institutional medical knowledge as it is “produced” by scientific communities with clear-cut boundaries and precise rules of the (publishing) game (e.g., by embedding recommended pathways and care programs into clinical workflow management systems and institutional guidelines into clinical decision support systems).

The present research moves from a previous contribution [14] and presents the development undergone by the research activities we presented therein. More specifically, the common objective of both the previous and current work is to contribute in the identification and characterization of a technology that could adequately support the emergence, sharing and discussion of contrasting voices, idiosyncratic preferences and local practices coming from medical milieus and traditions that, otherwise, could hardly come in contact and take reciprocal advantage from their mutual acquaintance.

II. Supporting a Bottom-up Approach to Knowledge

Several technologies can be used to support knowledge internalization and sharing within a highly distributed and heterogeneous group of professionals: e.g., mailing lists, knowledge markets (e.g., Yahoo Answers, Wikipedia Reference desk) and traditional forums where people can look for solutions and discuss problems, as well as document on-line publication systems and collaborative editing systems and wikis, where case reports, glossaries, guidelines, and procedures can be more or less collaboratively drafted and discussed. These contents are collected in different kinds of repositories where techniques based on agreed keywords, social tagging and the like can be put under the control of end-users and can support the management and retrieval of this “collective memory” in a lightweight way.

That notwithstanding, all these technologies tend to fall short with respect to two important aspects. First of all, efficiency in participation: that is the ratio between participation rates and the costs and incentives spent (or compromises made) to get those rates. In other words, it is always the same old ones who participate in public discussions or wiki drafting on a particular subject, while the others – either because of reserved temperament, because not interested in the topic or willing to share what they know – do not easily externalize their opinion/knowledge (even if they possess one). Second, efficiency in knowledge extraction. Even by assuming that people like to share what they know about a particular subject in the middle of their conversations or narratives, it can be difficult to “distill” usable and generalizable knowledge from those ad-hoc interactions, either because positions are not clearly stated nor problems, or because there is no way to reconcile diverging statements. In other words, these technologies interpret the knowledge management support as a repository where to record all what has been said and where the only structure provided to support fruition is often based on more or less predefined sets of keywords or categories supporting information retrieval and aggregation.

With respect to these two intertwined aspects, effective technologies to be deployed in communities of experts should be conceived with the twofold aim of fostering the participation of their members and of letting hidden/tacit knowledge emerge from the ground floor, respectively. To address these requirements, we started focusing on collaborative technologies that could i) foster and support a better comprehension and awareness of the relationship between formal, institutionalized knowledge (what in the medical domain is called ‘evidence’) and the actual practices of the members of a community (e.g., the degree of inclusion and compliance); ii) enable the reliable detection of attitudes and preferences in daily practice: for instance, attitudes of “trust” towards the truthfulness of specific statements and their common interpretation; preferences towards some particular way of performing a task, typically a specific medical treatment; iii) to highlight the collective (dis)agreement on these preferences and notions and, indeed, support the progressive and incremental building of consensus about relevant topics within a community.

To validate our approach to these general requirements, we started a research collaboration with a team of doctors. The common goal was to understand how current evidences from their reference literature had been internalized and included in medical daily practice. These doctors were looking for innovative ways that could help groups of experts discuss recent findings and evidences without requiring the frequent organization of expensive co-located workshops and conferences. We collaboratively identified the need of some easy-to-use and scalable technology that could improve the participation of members of collective organizations, associations and societies (with which doctors and surgeons interested in a particular sub-speciality, disease or surgery technique often affiliate) especially when they are highly fragmented in several countries.

The lack of ad-hoc technologies that could specifically meet this requirement made us orientate on a simple general-purpose
technology that could be configured to meet the specific needs exhibited by a specific community regarding the objectives mentioned above. We focused on on-line survey systems, for at least four reasons. First, surveys, as a general technology and science of data analysis, are wide-spread tools and traditional methods to collect information from or about a collection of people to describe, compare or explain their knowledge, attitude and behavior [15]. Population-based and Delphi-based surveys have also been used in the medical domain to detect treatment preferences and prescribing trends in both general and specialized practice (e.g., [16]), although the corresponding analyses have always been kept at the simple level of comparison of response percentages and agreement has been usually characterized in very qualitative ways (e.g., by merely confronting percentages of responses). Second, surveys, especially if conveyed by electronic means, are technologies that can reach every single member of a community and ask them clear-cut questions on specific topics: in so doing, they enable researchers and stakeholders to get an aggregate picture while being sure to give voice to every “chorister”. Third, several online survey systems are currently available on the web and some of them are also open-source and currently under a lively programme of testing and improvements. Last, but not least, in the last years these technologies have reached a sufficient degree of technological maturity [17]: for instance, current platforms allow to minimize (or at least reduce) question- and answer-order bias (by implementing random sequencing for each respondent), fatigue bias (e.g., by the dynamic routing of questions according to run-time answers or by funnel sequencing implementing reactive skip rules), non-response bias (e.g., by making a limited set of answers mandatory), acquiescence bias (e.g., by guaranteeing the perception of anonymity) and other non-sampling errors (like the Hawthorne effect, by allowing respondents to participate in multiple asynchronous sessions).

III. THE DESIGN OF AN AWARENESS PROMOTING SYSTEM

The system we designed with the collaboration of those doctors leverages a freely available and open-source platform that allows for the development, publishing and collection of responses to online questionnaires, self-interviews and surveys1. Upon this common technology, we developed a system that can be considered something in between an electronic voting system, a consensus building system and a survey system that enables collective consultations and informal census of practices. Since we used the system to collect individual preferences on treatments of choice and promote the awareness of each doctor on the preferences of her colleagues as a whole, we refer to our system as to an Awareness Promoting System (APS) on medical preferences: the name clarifies the goal we want to achieve. The role of this kind of awareness in managing medical knowledge is part of our research approach: in [18], we identified a range of different Awareness Promoting Information (API) that computer-based system could employ for an unobtrusive cognitive support to sense- and decision-making activities in collaborative settings. An APS can be employed in several heterogeneous domains, but we deployed and experimented this technology in the medical domain; in so doing, we leveraged the expertise and sensibility of the medical team we closely collaborated with and, at the same time, could get a picture of a domain that is highly representative of socio-technical milieus where extremely specialized and competent practitioners apply articulated and uncertain bodies of knowledge to cases that only seemingly are directly traceable back to recurring patterns and categories.

In the case study, we configured the system to have expert surgeons from all around the world consider a series of short and not trivial clinical cases and answer a couple of questions for each case about what their “treatments of choice” would be if they were to cope with that case and they had to assess the appropriateness of alternative options in the light of what has been disclosed in the concise case description provided. More specifically, the system we developed was aimed at allowing doctors to either: i) express their preferences toward alternative options regarding difficult clinical cases that were described in terms of short summaries; ii) rank alternatives by means of ordinal scales; or iii) select those alternatives that they deem more appropriate/correct for the case in hand.

The potential for success of such a system and its impact either on the community of practitioners (e.g., discussion, workshops, panels) or on the behavior of individual doctors is determined by three socio-technical factors: i) how questions are expressed; ii) how questions and alternative options are presented (in case of closed-question items); iii) how responses are aggregated, statistically analysed and eventually presented in a second-turn survey on the same matters2. As trivial as these factors can seem, in all of them state-of-the-art indications must be applied to improve user-experience, response rate and minimize any source of bias [19]. To this aim, after a comprehensive review of the fragmented specialist literature on survey methods (also in the medical domain, e.g. [20], [21]), we conceived our system as a multi-page survey to keep attention focused; we implemented automatic routing (or skipping) of pages according to the responses collected at run-time to make the survey shorter and to relieve respondents from the error-prone effort of skipping steps; we used only radio buttons or check-boxes for multiple options instead of drop-down lists to minimize visibility effects [22], [23], and inserted a progress indicator to minimize the drop-out rate due to poor estimation of how close respondents are to the end of the questionnaire [22].

IV. THE ESSKA CASE STUDY

In order to see how a web-based technology would impact a real community of experts, we contacted an international association of surgeons specialized in sports traumatology and knee surgery, ESSKA3. This association counts more than 1,000 members from 66 countries from all five continents (including Australia, Brazil, Japan, China, Russia, Iran, just to mention a

---

1(www.limesurvey.org)

2Not to mention a fourth factor: the involvement of the respondents, in that the final findings can be deeply influenced by their sincerity, willingness to participate, attitude towards web-based tools, et cetera.

3The name of the association is blinded for the blind review process.
few), with the 67% of members coming from European countries. In order to address concerns of privacy and transparency, we configured the platform to keep responses anonymous, prevent external users from getting access to the survey, and prepared two specific web pages, mentioned in the invitation email, where we stated both the aims and scopes of the research and our privacy statement on data management.

The survey consisted of eight cases, displayed each in a different page (see Fig. 1); for each case, respondents were asked to say if they would opt for either a surgical or conservative treatment on the basis of the available information; then, according to this first choice, each respondent was asked to either rank alternative treatments or select her treatment of choice. Time to complete the questionnaire was estimated with informal tests under the conventional threshold for fatigue bias [24]. The case summaries had been prepared by the three doctors involved in the research team, which were all expert orthopaedic surgeons holding an academic position, after a comprehensive literature survey for the detection of recent evidences in sport traumatology to confront with the alternatives proposed for each case.

We contacted the potential respondents by personalized email and decided to proceed with a single reminder. After four weeks from the first invitation we closed the survey and gathered 412 questionnaires: of these, the complete questionnaires were 374 (the 36% of the target population); 38 partial questionnaires were returned where at least one case had been considered, accounting for a final response rate of 40% of the target population; of these, the 70% of the respondents were from a European country. This was in line with expectations, considering the tight schedule of surgeons and in particular orthopaedic ones (57% of respondents declared to perform more than 200 operations yearly), the nature of their work (which does not imply long stays at a computer) and other surveys involving orthopaedic surgeons [25] as well as populations of similarly great size (i.e., >1000 participants) characterized by loose social control and high distributedness. To this respect, then, the initiative was considered a success: in fact, to our knowledge, such a collective consultation of clinical cases was never attempted before and, anyway, the participation in the survey closely matched the average attendance of ESSKA members in the last three international conferences organized by their association to gather together practitioners in the field of sport traumatology (m=376, sd=34). This result can be interpreted in terms of efficiency in participation, which we discussed in Section II.

In regard to the cases, the findings of clinical relevance are currently under process for publication in the medical literature. To our aims and to assess the concept of efficiency in knowledge externalization (cf. Section II), here we report that respondents exhibited a significant agreement only with respect to four cases out of eight and in two cases (the 6th and 7th) the distribution of their preferences was almost perfectly evenly balanced. For those cases where an agreement was detected at significant level of statistical confidence (p<.001), findings either confirmed (as in case 8 in regards to the kind of implant to use for the total knee prosthesis in case of diffuse arthritis) or opposed (as in case 1 with respect to the appropriateness of shoulder immobilization with brace in external rotation) institutional evidences reported in the literature.

The doctors involved in the research team found those cases most interesting where the official literature provided no evidence, while respondents clearly seemed to prefer a treatment over the others (as in the case of the single bundle vs. double bundled arthroscopic reconstruction of the anterior cruciate ligament; or about the optimal length of immobilization period for external rotation in shoulder anterior dislocation cases). They also wondered if it is reasonable to rank the collective “evidences” collected by surveys like the one presented here somehow in between the “scientific” evidences drawn from case-series studies and those based on the opinions of few experts of clear and respected authority; how reliable is the attitude of the (wide) majority vs. the opinion of “he chosen few”? These were questions that were brought to the public discussion at the ESSKA conference that was held just one week after the completion of the survey, raising a debate among the community members that will probably continue on the association impacted journal.

Further details on the cases presented and the pursued methodology can be found at the following address: http://www.gl-iss.org/downloads/GLISS-TreatmentsOfChoiceESSKA-Survey-PreliminaryReport.pdf

\[ ^5 \text{Statistical hypotheses were either rejected or accepted by performing either Chi-squared tests on the frequency of selection of nominal variables or one-sided sign tests on the median of ordinal variables. Details here are omitted because out of the paper’s scope.} \]
V. Feedback for a Better Design

During the survey, twenty respondents wrote us an email to the address we mentioned for further clarifications or troubleshooting in accessing the survey. Of these messages, the 75% were short emails of either encouragement, congratulations or manifestation of interest toward the survey results. Four doctors complained that some of the cases were too specific or out of their direct experience (e.g., shoulder cases for knee-specialized surgeons) and stated that they had to quit the survey and leave it mostly incomplete for that reason. One doctor raised the point that to address properly some of the cases he would have needed more information and proposed to attach diagnostic imaging to the case description.

In addition to this informal feedback, we sent an email to all respondents one month after the conclusion of the survey to thank them for their collaboration, provide them with the preliminary results of the survey and invite them to evaluate the initiative by a very short evaluation form. Approximately one tenth of the respondents seized this opportunity. The 17% of the respondents found the survey “a little bit long”, while only the 5% did not find the cases described adequately enough to express preferences (against the 78% that found the survey of adequate length and the 89% that found case summaries adequate). The 92% stated that web-based tools like our online survey could help make associations of practitioners be more “united”. The majority of respondents also agreed that such web-based tools could promote discussion on cases that are relevant for the practice (97%) and, even more significantly, help practitioners agree on best practices and treatments that are most appropriate in their field (89%); all the respondents declared their interest in participating in a second similar survey in the next six months. In the free-text field of the evaluation form, one respondent suggested to give participants the possibility to add additional treatments of choice; one suggested us to provide each respondent with her own responses, so that a comparison between her preferences and the collective ones could be made; another respondent proposed super-specialistic surveys where cases could be limited to single articulation and only subspecialist would be asked to answer to one or another survey.

The suggestions received during the survey and from the evaluation form, although anecdotal, provided us a useful feedback for the improvement of the system design. On the basis of these suggestions and of the considerations we shared with the doctors involved in the research team, we are now designing a new survey campaign to deploy in a national association. To meet a request raised by a couple of respondents, the system will allow respondents to add additional treatments besides those proposed, to assess the appropriateness of these treatments and to provide some notes as an explanation for their choice. These new treatments, which would probably collect few preferences, could then be integrated in a second turn of collective consultation.

VI. Final Remarks and Future Work

Commentators in the medical field have discussed the fundamental non-neutrality of ICTs in their profession (e.g. [3], [13]). They notice how ICT is advocated both by those who see it as the most powerful driver towards standardization of practices and better quality of care and, on the other hand, by those who stress the importance of patient-centeredness and utmost tailoring to the patient’s idiosyncrasies and needs. Yet, we notice how currently ICT in medicine is oriented more towards the promotion of a “rational objectivity” (backed by evidences, embedded in guidelines, informing care workflows) that is “consumable” by the single liable doctor, than the promotion of the social, interpersonal and narrative dimensions of medicine that unfold between doctors and patients as well as between doctors and doctors.

As ICT researchers, then, we are interested in how inexpensive and easy-to-run technologies, like web-based survey, can play their role in promoting the awareness of doctors that are members of large communities and associations of what their “esteemed” (and yet remote and almost unknown) colleagues do and think, of the extent these are aware of and compliant with institutional recommendations and of their level of (reported) self-confidence when facing hard-to-unravel cases. In other words, our approach is aimed at complementing those technologies that provide doctors with “nuggets” of formal and institutional knowledge (i.e., official guidelines and clinical algorithms); we aim to investigate the role of technologies that could provide doctors with the informal indications coming from the aggregative analysis of the preferences and medical positions that in other times would have hardly emerged from the single consulting rooms and doctor surgeries and, in short, from the potential of “collective intelligence” [26, p.43] that is distributed in large communities of expert practitioners at “grassroot level”.

Our research question regards then whether the web-based technologies, which have reached a sufficient degree of technological maturity and usability, can contribute, at least, in defining the contours and boundaries of collective bodies of knowledge in domains where either direct “observation” of every member of a community is unfeasible, or the opinion of the heads-of-household (whoever they are) cannot be representative of the “ground floor” of the community.

To validate our initial ideas, we conducted a case study in the domain of an international association of sport traumatology counting more than one thousand members across the world to identify treatments of choice characterized by levels of agreements statistically beyond chance. As reported in this paper, the case study gave encouraging results and hints for improving the proposed technology, which we provocatively denoted as lightweight. In particular, we use this latter term meaning that our proposal does not use any domain knowledge model or ontology that is imposed from above to help practitioners extract indications for their practices, nor it employs either frame extraction or information mining techniques to extract actual statistics on treatments of choice from the artifacts where treatments are documented, as these techniques presuppose hard models of medical knowledge and still utopian electronic medical records that are as much well structured as compliantly used.

As future work, we would like to investigate the extent case descriptions might affect preferences and the process itself of consensus building. We aim to build a system that could
display incremental details of a case summary on request of the respondents (e.g., more anamnestic details, additional laboratory test, medical imaging reports). In so doing, we aim to get useful indications on what could be the optimal level of granularity of description to express exemplificatory clinical cases for collective consultation systems. Moreover, the envisioned system should, whenever a respondent asks for further information, display and sum the risks for the patient and the expenses for the health service that each piece of additional information would entail. In doing so, we aim to get insights on what the right compromise could be between, on the one hand, representing relevant facts and context and, on the other hand, the time-, cost- and risk-based constraints that doctors are used to cope with in daily practice.

These further experimentations could help us frame an even more challenging objective. In fact, we believe that awareness-promoting and consensus-building surveys have a potential that is still overly underestimated. In medicine, for instance, qualitative consensus methods, like Delphi meetings, expert panels and diagnosis-oriented team meetings [27], have been used to support decision making in situations where there is insufficient information or an overload of (often contradictory) information; in fact these methods are capable to provide synthetic information as that provided by statistical methods but they are also liable to use a wider range of information than is common in those latter methods [28]. In these cases, as well as in the context of requirement elicitation and validation in software development processes and in the reviewing process of contributions within conference and journal boards (just to mention a few of potential application domains), web-based surveys could guarantee the necessary anonymity, agility, flexibility, inexpensiveness and scalability to probe the preferences and attitudes of either a limited or large number of stakeholders and decision-makers [29]; these usually free and easy-to-configure tools could be possibly used in multiple iterative turns, till some sound agreement and general consensus is reached on some reasonable (vs. rational) position within the usually pressing time and cost constraints of collaborative decision making.

ACKNOWLEDGMENT

The authors would like to thank Brigitte Melchior-Dolenc and Sandy Kirsch of the ESSKA Luxembourg Office for their support and collaboration, and all the surgeons that participated in this project. We also acknowledge the most valuable collaboration of Pietro Randelli, MD and Paolo Arrigoni, MD in the conception of the clinical cases and for their insightful suggestions in the survey design. Without their collaboration, this work would not have been possible.

REFERENCES