

Method

Cite this article: Formoso G, Jeronic A, Bonvicini L, Djuric O, Erdos J, Pezzarossi A, Ballini L (2022). Synthesizing quantitative and qualitative information on multiple comparisons of health interventions to facilitate knowledge transfer: an example from an EUnetHTA multi-HTA. *International Journal of Technology Assessment in Health Care* **38**, e20, 1–6. <https://doi.org/10.1017/S0266462322000046>

Received: 22 September 2021

Revised: 7 December 2021

Accepted: 12 January 2022

Key words:



Information synthesis; Systematic reviews; Knowledge transfer; Multiple interventions; Clinical relevance

Author for correspondence:

Giulio Formoso,

E-mail: giulio.formoso@ausl.re.it

Synthesizing quantitative and qualitative information on multiple comparisons of health interventions to facilitate knowledge transfer: an example from an EUnetHTA multi-HTA

Giulio Formoso¹ , Ana Jeronic² , Laura Bonvicini³, Olivera Djuric^{3,4}, Judit Erdos⁵, Annamaria Pezzarossi³ and Luciana Ballini^{6,7}

¹Clinical Governance Unit, Local Health Authority – IRCCS of Reggio Emilia. Via Amendola 2, 42122 Reggio Emilia, Italy; ²Department of Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia; ³Epidemiology Unit, Local Health Authority – IRCCS of Reggio Emilia, Reggio Emilia, Italy; ⁴Center for Environmental, Nutritional and Genetic Epidemiology (CREAGEN), Section of Public Health, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy; ⁵HTA Austria – Austrian Institute for Health Technology Assessment GmbH, Vienna, Austria; ⁶Local Health Authority, Modena, Italy and ⁷Regional Health and Social Care Authority, Emilia-Romagna, Italy

Abstract

Objective. Presenting an approach to synthesize quantitative and qualitative information from systematic reviews of multiple health interventions.

Methods. Within the context of an EUnetHTA multi-health technology assessment of twenty-three surgical techniques, we developed synthetic single tables, using color gradients and abbreviations, with information on which technologies had been compared, estimates of the size of differences for available comparisons, their clinical relevance, and certainty of the related evidence.

Results. The proposed methodology provided, through a single depiction, information normally included in multiple figures/tables such as network plots, league tables, and summary of findings tables.

Conclusion. Transferring information on benefits, risks, and certainty of the available evidence on health interventions may be challenging, especially when assessing multiple treatments: more pieces of information need to be integrated in order to show an overall picture for each of the chosen outcomes, and usual reporting tools may be targeted to researchers more than to different kinds of decision makers. While more in-depth layers of information can always be added to satisfy needs of different audiences, the proposed tools could favor a quick interpretation of articulated scientific data by both decision makers and researchers.

Introduction

Systematic reviews can provide a wealth of information, through meta-analyses of data and qualitative judgments about the certainty of available evidence. They aim at providing information on relative effects of treatments that are directly compared, including estimates of differences among treatments or between each treatment and the common comparator, and information related to the weight of each comparison (and to the related certainty of evidence). When the effects of multiple treatments need to be assessed, network meta-analysis (NMA) may be used—besides multiple direct comparisons—to allow a quantitative estimation of the effect of different interventions using both direct and indirect comparisons against a common comparator [1].

Several NMA-specific tools are available to present graphical and numerical summaries of the analyses. Network plots (Figure 1) can show which direct comparisons are available within the network and how many trials or patients are available for each direct comparison [2]; sometimes a judgment on the risk of bias associated with each direct comparison is also shown by applying coloring of edges.

League tables (Figure 2) report the meta-analytic estimate with confidence interval (CI) of the difference in effect sizes between each pair of treatments, derived from both direct and indirect comparisons [1].

Summary of findings (SoF) tables (Supplementary Figure 1) report the meta-analytic estimate described above referred to single comparisons, generally between each treatment and a common reference treatment, adding information on the overall judgment of the certainty of evidence which is derived from risk of bias, indirectness, inconsistency, and publication bias [3;4].

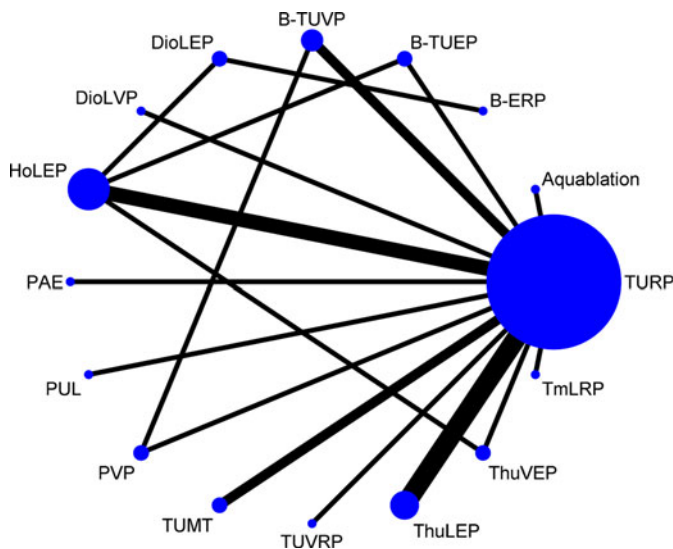


Figure 1. Example of a network plot.

These reporting tools can provide information on the number of available comparisons, on estimates of their relative effects, and on the certainty of evidence associated with these estimates. However, such information is usually split among different figures and tables and cannot be easily displayed as a single synoptic graphic tool, especially in presence of multiple comparisons among different health treatments and technologies.

Synthesizing such information is then challenging. When NMA cannot be used (e.g., when the anchor treatment differs systematically between trials and when transitivity cannot be assumed), multiple direct comparisons would make the interpretation of results particularly difficult. Whether NMAs or just direct comparisons are performed, they should help provide readers with quantitative and qualitative information to facilitate the assessment of relative effects of treatments, as well as of their clinical relevance and the level of certainty of the related evidence.

A specific approach to facilitate knowledge transfer, that can be particularly useful in case of multiple comparisons, is proposed here with examples from a complex multi-health technology assessment (HTA) on surgical techniques for the treatment of benign prostatic hyperplasia, carried out by a team of European HTA Agencies as part of the outputs of the EUnetHTA Joint Action 3 [5].

Our goal was to provide an easily interpretable overview of the available research and specifically about: (1) which technologies had been compared, (2) estimates of the size of the differences for available comparisons, (3) their clinical relevance, and (4) the certainty of the related evidence.

Methods and Results

When assessing twenty-one minimally invasive technologies alternative to the standard transurethral resection of the prostate (TURP) or to open prostatectomy (OP) (twenty-three technologies in total), NMA could not be performed on each of the twenty-seven selected outcomes since the transitivity assumption could not be warranted. We needed to find an alternative way to comprehensively visualize: all the available research, the quality of included studies, and the effect of each technology in the context of all the included alternatives; so that decisions related to the likely effect (including its uncertainty) of using each technology in clinical practice and about the research which is still needed could be facilitated. From this example, we developed two different (although conceptually similar) approaches that could be applied in any situation where synthesis of information from pairwise comparisons is at stake, providing information about different pairwise treatments on a single outcome (examples 1A and 1B), or about each treatment versus a common standard treatment on multiple outcomes (example 2).

Example 1A

Table 1 shows all the comparisons, available from the studies selected within our HTA, assessing International Prostate Symptom Score (IPSS) at 6 months. The rows display all

Treatment 1	0.77 (0.02,23.96)	0.62 (0.03,10.94)	0.29 (0.01,8.74)	0.34 (0.02,5.96)	0.23 (0.01,5.36)	0.18 (0.02,1.51)	0.14 (0.02,1.18)	0.14 (0.01,2.54)	0.15 (0.02,1.32)	0.05 (0.00,0.81)
1.30 (0.04,40.68)	Treatment 2	0.80 (0.05,12.02)	0.38 (0.01,27.04)	0.44 (0.07,3.00)	0.30 (0.01,17.54)	0.23 (0.01,6.55)	0.19 (0.00,10.57)	0.18 (0.00,8.81)	0.20 (0.01,2.94)	0.06 (0.00,2.84)
1.62 (0.09,28.85)	1.25 (0.08,18.67)	Treatment 3	0.48 (0.01,21.68)	0.55 (0.08,3.76)	0.38 (0.01,13.71)	0.29 (0.02,4.57)	0.23 (0.01,8.24)	0.23 (0.01,6.72)	0.25 (0.04,1.69)	0.08 (0.00,2.16)
3.40 (0.11,100.83)	2.61 (0.04,183.83)	2.09 (0.05,94.85)	Treatment 4	1.16 (0.03,51.86)	0.79 (0.12,5.08)	0.60 (0.04,8.20)	0.49 (0.01,26.40)	0.48 (0.06,3.88)	0.52 (0.02,14.19)	0.16 (0.02,1.01)
2.94 (0.17,51.35)	2.25 (0.33,15.23)	1.81 (0.27,12.30)	0.86 (0.02,38.73)	Treatment 5	0.69 (0.02,24.48)	0.52 (0.03,8.12)	0.42 (0.01,14.71)	0.41 (0.01,11.99)	0.45 (0.07,2.98)	0.14 (0.00,3.85)
4.28 (0.19,98.07)	3.28 (0.06,188.99)	2.63 (0.07,95.12)	1.26 (0.20,8.06)	1.46 (0.04,51.97)	Treatment 6	0.75 (0.08,7.32)	0.62 (0.01,26.75)	0.60 (0.13,2.92)	0.66 (0.03,13.70)	0.20 (0.05,0.79)
5.69 (0.66,48.94)	4.36 (0.15,124.89)	3.50 (0.22,56.01)	1.67 (0.12,22.99)	1.94 (0.12,30.49)	1.33 (0.14,12.94)	Treatment 7	0.82 (0.04,16.57)	0.80 (0.12,5.53)	0.87 (0.12,6.52)	0.26 (0.04,1.75)
6.94 (0.85,56.68)	5.33 (0.09,300.16)	4.28 (0.12,150.64)	2.04 (0.04,110.27)	2.37 (0.07,82.29)	1.62 (0.04,70.48)	1.22 (0.06,24.69)	Treatment 8	0.98 (0.03,34.92)	1.07 (0.05,21.59)	0.32 (0.01,11.23)
7.08 (0.39,127.31)	5.43 (0.11,260.03)	4.36 (0.15,127.60)	2.08 (0.26,16.84)	2.41 (0.08,69.66)	1.65 (0.34,7.99)	1.24 (0.18,8.56)	1.02 (0.03,36.28)	Treatment 9	1.09 (0.07,17.63)	0.33 (0.11,1.02)
6.50 (0.76,55.91)	4.99 (0.34,73.15)	4.00 (0.59,27.03)	1.91 (0.07,51.96)	2.21 (0.34,14.61)	1.52 (0.07,31.63)	1.14 (0.15,8.52)	0.94 (0.05,18.93)	0.92 (0.06,14.88)	Treatment10	0.30 (0.02,4.76)
21.55 (1.23,378.18)	16.54 (0.35,777.40)	13.27 (0.46,380.43)	6.35 (0.99,40.47)	7.34 (0.26,207.66)	5.04 (1.27,19.97)	3.79 (0.57,25.10)	3.10 (0.09,108.28)	3.05 (0.98,9.43)	3.32 (0.21,52.33)	Treatment11

Figure 2. Example of a league table.






Table 1. Synthesis of information on multiple comparisons (on IPSS at 6 months) [2]

Intervention	Control								
	TURP	OP	HoLEP	ThuLEP	B-TUEP	PVP	DioLEP	B-TUVP	B-TUERP
B-TUEP	VL	VL		L			L	VL	
HoLEP	L			VL			L		
DioLEP	M		L		L				L
B-TUVP	L	VL			VL	VL			
PVP								VL	
ThuLEP	L		VL		L				
DioLVP	VL								
TUVRP	VL								
Aquablation	L								
TUMT	M								
PUL	VL								
TmLRP	L								
B-TUERP	L						L		
PAE	VL								
B-VEP	L								
TURP + TUIP	L								

Key to certainty of evidence: H = high; L = low; M = moderate; VL = very low certainty of evidence [6].

Notes on reasons for downgrading the evidence and key messages were provided in Ref. [5].

Key to quantitative differences

	Intervention statistically significantly better than control (95% CI crossing MCID).
	Intervention statistically significantly better than control (but 95% CI below MCID).
	No difference.
	Intervention statistically significantly worse than control (but 95% CI below MCID).
	Intervention statistically significantly worse than control (95% CI crossing MCID).

MCID, minimal clinically important difference; CI, confidence interval.

Notes: Comparison of each technology (intervention, left column) by row to the other technologies (control, other columns). The colors denote the quantitative difference for each comparison, as shown in the key. Letters denote the quality of the evidence.

interventions considered in our scope, whereas the columns display all the possible controls considered for comparison, including the two standard comparators in our scope—TURP and OP as well as the minimally invasive technologies listed among the interventions—since they all could be compared to each other. Following the row of a specific intervention, one can see how many other technologies it had been compared to and, looking at the corresponding cells, if there are differences, in favor of which technology, whether any difference is statistically significant and clinically relevant (using different colors), and what is the certainty of the related evidence (using abbreviations).

We were particularly concerned about how to represent judgments on clinical relevance, often overlooked or missing in graphical representations of comparative results. In this regard, we used the concept of minimal clinically important difference (MCID), which was first used by Jaeschke *et al.* and defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive costs, a change in the patient’s management” [7]. When available, MCIDs are generally obtained considering patients’ or clinicians’ perspectives—e.g., asking patients to indicate, in *ad hoc* studies, how relevant observed changes are, or through expert consensus [8]. When assessing an estimate for a domain of interest (e.g., the estimated

difference in IPSS between two interventions) in light of an MCID, we should also consider how uncertain is that estimate, therefore the possibility that the true effect can be higher than the MCID threshold even if the point estimate is lower. In this regard, we used more or less intense colors so that, if the 95% CI was entirely below the MCID, a statistically significant difference was considered not clinically relevant and the blue (or red) color (depending on the direction of effect) was less intense; in case the CI crossed the MCID, we could not exclude the possibility of a clinically relevant difference and used a more intense color. We decided not to add quantitative estimates and CIs to keep the table easier to read, relying on color intensity to indicate the level of uncertainty; however, it would be possible to add CIs to provide more precise indication of it. As for judgments on the quality of evidence, notes on reasons for downgrading the evidence were also provided for each comparison. Final key messages eventually synthesized the overall results.

Since most technologies are listed among both interventions and controls, information on a specific comparison may be available in two different cells (intervention A × control B, and intervention B × control A), where colors may be reversed depending on the direction of effect (A × B, and B × A). Even if the related information is duplicated, this representation may help a reader to quickly get information on the quantity and the quality of

information available for each intervention of interest just by following the corresponding row.

Example 1B

The previous example refers to the assessment of continuous outcomes where an MCID is available. The same method can also be used with dichotomous outcomes. In this case, color gradients could also be used to highlight different strengths of relative risks, always considering the level of uncertainty (Table 2). Judgments on clinical relevance of the observed differences for binary outcomes are difficult, since no related MCIDs are available in the scientific literature. In case, frequency and perceived relevance of the specific outcomes should be considered to discuss and provide such judgments.

Example 2

When the objective is to present results of parallel comparisons for each technology versus only one comparator, a single table can display results for each technology versus that comparator, using columns to show results for each of the outcomes of interest (e.g., on IPSS and Qmax at different time periods, Table 3).

Discussion

Transferring information on benefits, risks, and certainty of the available evidence on different health interventions is challenging, especially when assessing multiple treatments. Reporting graphic tools, used in the context of NMAs or of direct comparisons, provides a set of relevant information but may be targeted to researchers more than to different kinds of decision makers. For example, SoF tables provide a wealth of information, but they refer to one comparison at a time; league tables appear as kind of mathematical artifacts providing quantitative estimates that are not easy to interpret in terms of direction of effect.

Different pieces of information need to be integrated in order to get an overall picture about what is particularly relevant for a decision maker, at least for interpreting trial results and learning about missing research. In other words, the overall picture should include information on which treatments have been (or not) compared, estimates of differences (if present), their clinical relevance, and certainty of the available evidence. This would be important considering different decision-making contexts, either at the bedside or when deciding about technologies to be acquired or implemented in clinical practice, or when deciding about which new research should get resources.








Table 2. Synthesis of information on multiple comparisons (dichotomous outcome: urinary incontinence)

Intervention	Control								
	TURP	HoLEP	ThuLEP	B-TUEP	PVP	ThuVEP	DioLEP	B-TUVP	B-TUERP
B-TUEP	VL	L	VL				L		
HoLEP	L		L	L	VL	L	L		
DioLEP	L	L		L					L
B-TUVP	L				VL				
PVP	L	VL						VL	
ThuVEP		L							
ThuLEP	L	L		VL					
DioLVP	VL								
TUVRP	VL								
TUMT	L								
PAE	VL								
PUL	VL								
TmRLP	M								
B-TUERP	L						L		
M-TUERP	VL								
B-VEP	L								

Key to certainty of evidence: H = high; L = low; M = moderate; VL = very low certainty of evidence [6].

Notes on reasons for downgrading the evidence and key messages were provided in Ref. [5].

Key to quantitative differences

	Intervention statistically significantly worse than control (95% CI for RR > 2 and includes 3, or 95% CI for RR > 3).
	Intervention statistically significantly worse than control (95% CI for RR includes 2 but < 3).
	Intervention statistically significantly worse than control (95% CI for RR between 1 and 2).
	Intervention statistically significantly better than control (95% CI for RR < 1 and includes 0.5).
	Intervention statistically significantly better than control (95% CI for RR < 0.5 and includes 0.33).
	Intervention statistically significantly better than control (95% CI for RR < 0.33).
	No difference.

Notes: Comparison of each technology (intervention, left column) by row to the other technologies (control, other columns). The colors denote the quantitative difference for each comparison, as shown in the key. Letters denote the quality of the evidence.

Table 3. Synthesis of information on comparisons of different technologies versus a common comparator on multiple outcomes


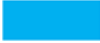



Intervention	Control: TURP			
	IPSS 6 months	IPSS 12 months	Qmax 6 months	Qmax 12 months
B-TUEP	VL	VL	VL	VL
HoLEP	L	VL	L	L
DioLEP	M		M	
B-TUVP	VL	M	L	M
PVP		VL		VL
ThuLEP	L	L	L	L
DioLVP	VL	VL	VL	VL
TUVRP	VL	L	VL	L
ThuVARP		VL		VL
Aquablation	L	L	L	L
TUMT	M	L	VL	L
TUIP		L		L
PUL	VL	VL	VL	VL
TmRLP	L	L	L	L
B-TUERP	L		L	
PAE	VL		VL	
B-VEP	L		L	
TURP+TUIP	L		VL	

Key to certainty of evidence:

H = high; L = low; M = moderate; VL = very low certainty of evidence [6].

Notes on reasons for downgrading the evidence and key messages were provided in Ref. [5].

Key to quantitative differences:

	Intervention statistically significantly better than control (95% CI crossing MCID).
	Intervention statistically significantly better than control (but 95% CI below MCID).
	No difference.
	Intervention statistically significantly worse than control (but 95% CI below MCID).
	Intervention statistically significantly worse than control (95% CI crossing MCID).

Notes: Comparison of each technology (intervention, left column) by row to the other technologies (control, other columns). The colors denote the quantitative difference for each comparison, as shown in the key. Letters denote the quality of the evidence.

Compressing data from available studies always entails the risk of overlooking “real-life” complexities, missing possibly relevant information related to the heterogeneity of patient populations, and uncertainty of “average” estimates. While some compromise between completeness of presented information and immediate comprehension may be required, further data (e.g., on the level of uncertainty through CIs and on subgroup analyses) and more in-depth layers of information can be always added through different tools so that different kinds of decision makers and researchers could find all the data they need. In other words, the ultimate goal should not just be to synthesize (thus reducing) information, but to provide with progressively enriched contents that could help to grasp the bigger picture, including all relevant data for decision making, without necessarily missing more in-depth analyses. In all regards, it should be noted that the clinical relevance of results is often overlooked in data synthesis, considering that it may represent a key element for decision makers and (of course) patients. MCIDs are especially missing for dichotomous outcomes and could be discussed in the context of HTAs looking at the frequency of these outcomes and at their perceived relevance, which may be also discussed with patients’ representatives.

As for the use of color gradients to indicate statistical significance and clinical relevance, issues of color vision deficiency, particularly affecting the vision of reds and greens, could be considered. In the examples displayed here, we used red-blue combinations, as they may be less problematic in this regard [9].

Conclusion

Knowledge translation requires easily interpretable tools, showing more attention to the information needs of different target groups [10]. Knowledge can be best promoted by paying due attention to information tools and to the characteristics of target populations [11], as successful marketing of commercial products show. The specific contents should therefore allow a reader to be able to understand, on a given topic, if there is (or there is not) evidence-based information, and can make a judgment for himself/herself of what can be expected from the use of a diagnostic, preventive, or therapeutic intervention.

We tried to implement these principles by proposing, for each of the technologies assessed in a multi-HTA, a synopsis of the information deemed relevant for both decision makers and

researchers; in just one table one can easily get information on all available comparisons, on whether the related results are statistically significant and clinically relevant and on the certainty of the related evidence. We think it may be a doable approach to ease the interpretation of articulated scientific data, which should always be one of the objectives of evidence synthesis. Such an approach could also find a relevant field of application in non-paper-based and electronic publishing, addressing information needs of those professionals seeking decision support from digital sources using related applications, as Richard Smith (former Editor of the *BMJ*) envisioned 25 years ago [12].

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0266462322000046>.

Funding. The contents of this paper arise from the project “724130/EUnetHTA JA3,” which has received funding from the EU, in the framework of the Health Programme (2014–2020). Sole responsibility for its contents lies with the author(s) and neither the EUnetHTA Coordinator, the European Commission, nor any other body of the European Union is responsible for any use that may be made of the information contained therein.

Conflict of Interest. There are no conflicts of interest.

References

1. Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analyses. *Cochrane Handbook for Systematic Reviews of Interventions*, Second Edition Available at <https://training.cochrane.org/handbook/current/chapter-11#:~:text=Network%20meta%20analysis%20is%20a,across%20a%20network%20of%20studies.&text=Domain%20specific%20assessments%20are%20combined,overall%20confidence%20in%20the%20evidence> (last accessed: Sept 20, 2021).
2. Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS One*. 2013;**8**: e76654. doi:10.1371/journal.pone.0076654. (last accessed: Sept 20, 2021).
3. Schunemann H, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Completing ‘Summary of findings’ tables and grading the certainty of the evidence. *Cochrane Handbook for Systematic Reviews of Interventions*, Second Edition. doi:10.1002/9781119536604.ch14 (last accessed: Sept 20, 2021).
4. Yepes-Nunez JJ, Li SA, Guyatt G, Jack SM, Brozek JL, Beyene J, et al. Development of the summary of findings table for network meta-analysis. *J Clin Epidemiol*. 2019;**115**:1–13.
5. EUnetHTA OTCA27 Authoring Team. Comparative effectiveness of surgical techniques and devices for the treatment of benign prostatic hyperplasia. Collaborative Assessment. Diemen (The Netherlands): EUnetHTA; 2021. Report No.: OTCA27. Available from: <https://www.eunethta.eu>
6. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE working group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017;**87**:4–13.
7. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;**10**:407–15.
8. Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, et al. Minimal clinically important differences: Review of methods. *J Rheumatol*. 2001;**28**:406–12.
9. See <https://www.nhs.uk/conditions/colour-vision-deficiency/> (accessed 2021 Nov 23).
10. Formoso G, Rizzini P, Bassi M, Bonfanti P, Rizzardini G, Campomori A, et al. Knowledge transfer: What drug information would specialist doctors need to support their clinical practice? Results of a survey and of three focus groups in Italy. *BMC Med Inform Decis Mak*. 2016;**16**:115. doi:10.1186/s12911-016-0355-7.
11. Formoso G, Marata AM, Magrini N. Social marketing: Should it be used to promote evidence-based information? *Soc Sci Med*. 2007;**64**:949–53.
12. Smith R. What clinical information do doctors need? *BMJ*. 1996;**313**:1062–8.