Evidence-based policy as reflexive practice. What can we learn from evidence-based medicine?

Roland Bal, Erasmus University Rotterdam

Summary

The call for evidence-based policy (EBP) is often accompanied by rather uncritical references to the success of evidence-based medicine (EBM), leading to often-unsuccessful translation attempts. In this paper, I reflect on the practice of EBM in an attempt to sketch a more productive approach. Discussing three 'moments' of EBM—clinical trials, the production and use of clinical guidelines—I conclude that the success of EBM is based on the creation of reflexive practices in which evidence and practice can be combined productively. In the conclusion I discuss the prospects of such a practice for evidence-based policy.

Keywords: evidence-based policy, evidence-based medicine, reflexivity

Introduction

During one of the first weeks at my then new workplace, the Department of Health Policy and Management in Rotterdam, the then Minister of Health decided to exclude Viagra from the basic reimbursement package. The economists at the Department, who had just concluded from their research that Viagra was a very cost-effective drug, were shocked. How could the Minister keep a cost-effective drug away from patients, while at the same time leaving some of the most cost-ineffective therapies such as lung transplants open for reimbursement (1)? I, as a relative newcomer to the field of healthcare but experienced in studying science advice, was, in turn, shocked by the naivety of my colleagues. How could the Minister be expected to reimburse a medication that might be used for a medical condition but is mostly considered to be a lifestyle drug? And yet, the phenomenon of politicians not eager to follow research results is no longer strange to me either. Let me give you two examples from my own research.

Recently, Dutch healthcare policy has placed an increasing pressure on the concentration of hospital care fuelled by the assumption that doing more of the same diagnostic and therapeutic processes improves the quality of care. The Ministry of Health and health insurers especially have stimulated hospitals and medical specialists to concentrate all kinds of care processes—complex oncology, emergency departments, intensive care, stroke services. I have been involved in different projects evaluating such concentrations. These evaluations have illustrated that the evidence for the relation between volume and quality is restricted to just a few care processes, mainly involving highly complex care, but that even for those causality is not clear. Activities of hospitals and medical specialists to concentrate care are mainly driven by organisational and financial motives (while the evidence in this regard is also poor). Meanwhile, concentrating care also creates lots of challenges, mainly in the coordination of

care for multi-morbid patients and in the levels of expertise in departments where care is 'concentrated away', i.e. neurology departments that loose their stroke specialists (2, 3). Within health policy, these issues barely receive attention however and concentration policies are increasingly pursued. Second example: in recent years I have been involved in research on 'transparency' policies and especially the quantification of the quality of care through performance indicators, benchmarks and rankings (4, 5). These projects invariably showed that while there are some positive effects of transparency policies, overall they have led to lots of perverse effects. Rather than 'representing' the level of quality of care, quantitative measures 'constitute' such qualities. Again, policy makers have largely ignored this research and have continued in their quest for transparency. The Dutch Ministry of Health has even called 2015 the 'year of transparency' without paying any attention to demonstrated negative effects of such policies (6).

Now, I'm not so naive as to think that my research would change policy and I do not foster the fantasy that influencing policy through research is a clear-cut and linear process. Students of the relation between science and policy have convincingly shown the complexity of that relation (7-9), also indicating that this is not a typical Dutch problem. But still, systematically ignoring all unwelcome research results is yet another extreme A bit of evidence in policy might not be a bad thing, one would think. But how do we get it there?

The example of Evidence-Based Medicine

In health care, one quickly refers to evidence-based medicine (EBM) in this situation. The EBM movement, as generally understood, has succeeded in bringing about a radical change in the way medicine in taught and practiced. Just as with evidence-based policy (EBP), this has brought both praise and criticism. Critics argue that EBM has lead to standardized care in which the individual patient has suffered under the violence of the evidence produced in randomized controlled trials, and in which the autonomy of individual doctors has been denied (10). Advocates of EBM on the other hand have argued that EBM has lead to reducing bad quality of care and that standardization has strengthened rather than weakened the profession. Both groups however tend to plead from a rather formalistic version of EBM that does little justice to the ways in which EBM is practiced (11).

A similar emphasis on the formalistic, external characteristics of EBM can be found in arguments to translate EBM to the world of policy and management (12, 13). These authors thereby sidestep the situated and emergent character of EBM as described especially by scholars from Science & Technology Studies. Despite all claims, EBM is most certainly not a universalistic operating machine but is displaying much local variation (14). Moreover, translating the formalistic character of EBM to policy contexts has practically proven to be of little use. Even Chalmers, in the just cited paper, has to admit that despite all attempts to do just this, an *evidence-based* policy has not come about. For example, the practice of the Dutch Centre for Healthy Living to create a list of evidence-based interventions has not lead to interventions that can be used, let alone be successful in public health practices (15). Reasons for this are that the

complexity and dynamics of policy contexts are usually not taken into account in these translations and that political realities are ignored.

In this essay, I want to explore another, pragmatic, picture of EBM to add to the discussion on the (im)possibilities of using EBM as an example for evidence-based policy. Doing this, I take an *insiders* perspective—without trying to suggest that an outsiders critique of EBM or EBP would not be interesting or valuable—in which I am especially interested in the question how EBM, and in its wake EBP, *can* be successful. I will do this by analysing three 'episodes' in the practice of EBM. Those three episodes can also be seen as three consecutive phases in the practice of EBM: the *production of evidence* through medical research (especially the clinical trial); the *gathering of evidence* in the production of clinical guidelines; and the *use of evidence* in healthcare practices through the implementation of guidelines. The analysis of those episodes is based on research I have been involved in over the past years.

The clinical trial as a reflexive space between research and practice The randomized clinical trial (RCT) is probably the greatest good of the evidence-based medicine movement and referred to by Timmermans and Berg as the 'gold standard' (14). The RCT has been developed during the 20th Century in reaction to the disadvantages of a medicine based on clinical practice, translating all kinds of methodological prescriptions from the natural sciences to healthcare, and alongside developing an ethics of medical research (e.g. the practice of *informed consent*). Historians and sociologists of science have thoroughly documented this development, which I am not going to repeat here (16, 17).

Like EBM in general, the RCT also has its critics. And again, mainly the standardization central to RCT is the main target. To perform a clinical trial, one has to standardize the group that is being researched, the treatment (or diagnostic procedure) under investigation and the context in which the research is taking place. That necessity produces frictions. Standardizing the group being researched leads to the exclusion of all kinds of groups—children, the ageing population, people with multiple diseases. Interventions are kept simple, that is 'evaluable', because too complex interventions are harder to research in the RCT framework. And the dynamics and complexity of practical contexts in usually ignored. RCTs thus lead to knowledge production about 'ideal' patients in 'ideal' circumstances that usually has little value for practice. In the wake of the RCT, as a result of these criticisms, other forms of experimental research have been developed, like the 'pragmatic trial', but these usually have the same problems (18).

Now the criticism on the RCT on the one hand is justified but is on the other also missing out on an important point: the practice of medical research is different from what is prescribed in the methods textbooks, or described in the methods sections of scientific papers. Trials in practice are much messier than being pictured in the methods literature, if only because of problems with the inclusion of patients, the impossibilities of completely standardizing treatments or the keeping under control of contexts. Resultantly, Ethnographies of trials often paint a different picture.

An example is research that we did on the 'PreCare' trial, a randomized trial that was done in youth care, in which an intervention that was originally developed and tested in the United States on 'problematic teen mothers' was translated to the Dutch context (19). The 'intervention' that was researched in this trial entailed that youth workers would pay a total of 18 visits to mothers in the research group on the basis of a standardized protocol. In practice, this proved to be near to impossible, due to *no show* of the mothers, or due to their home situation, e.g. the presence of aggressive fathers, etc. Also, the home workers were dissatisfied with the trial because they didn't want to withhold an intervention they thought to be effective from the control arm of the trial.

To deal with these and similar problems, within the context of the trial 'practice groups' were organized, bringing together the researchers, the youth workers and their managers. These groups mediated between the research and practices and devised situated, practical solutions that would enable resolutions for the frictions found. Not all conflicts could be remedied through this procedure; the youth workers remained dissatisfied with not being able to give the intervention to the control group, although they better understood the necessity of this for the research context—but many practical problems were resolved. In many cases this meant that the trial had to compromise on the standardized protocol, and the researchers had to find solutions to legitimate this in the context of the trial methodology.

As this example shows, trials are not standardized evaluation machines for the uniform application of interventions in practice, but need flexibility and compromise for their execution. The PreCare trial operates as a reflexive space in which situated, pragmatic solutions for frictions between research and healthcare practices can be formulated and applied. It is this reflexive space that makes the trial—and by extension, EBM—possible and workable.

Guidelines and dealing with uncertainty

Clinical guidelines have emerged as a response to reports from clinical epidemiologists in the United States in the 1980s that that were huge variations in medical practice and healthcare outcomes between medical doctors and healthcare organisations. Guidelines at first were mainly intended to support medical practitioners in making decisions in diagnostics and treatment. Over time, guidelines have developed as being increasingly based on *evidence* from RCTs and meta-analyses. By now, a guideline industry has emerged that is governed itself by numerous guidelines. In the Netherlands, the method of *evidence-based guideline development* ('evidence based richtlijn ontwikkeling' or EBRO) is popular, offering many methodological handles, for example on the weighting of studies from the literature in which systematic, experimentally gathered evidence is preferred over other types of studies. The newest kid on the block, *Grading of Assessment, Development and Evaluation* (GRADE) allows for more leeway and among others, considers the relevance of the evidence found in the literature in weighing studies to be taken on board in guidelines (20).

Guidelines have met similar criticism as the RCT, in that they promote the standardization of care and thus surpass the complexity of individual patients. Moreover, Guidelines tend to become ever more complete and as a result give less guidance to practice. In most countries guidelines have been taken up to *control* medical practice by third parties (governments, payers) and have thus neglected their original purpose of supporting practical decision making. Critics have also pointed out that patients have too little influence on the development of guidelines and that in guideline development the voice of healthcare practitioners is bypassed (21).

Here again, critics (and advocates) have largely ignored that the practice of developing guidelines differs from the methodological guides such as EBRO and GRADE. Sociologist of science Tiago Moreira has shown for example that in the development of guidelines by the English *National Institute for Health and Care Excellence* (NICE), at least four types of evaluation are used, of which the methodological is only one (22). Additionally, Guidelines show large international differences in which differing practices of dealing with uncertainty or conflicting evidence is just one possible explanation (23).

To gain insight in the practice of guideline development, we researched the ways in which guideline developers try to deal with uncertainty. While guidelines were originally a means to give medical practitioners support in dealing with uncertainties, guideline developers themselves are also confronted with all kinds of uncertainties, for example because of a lack of evidence. From interviews with guideline developers in different sectors of health care (e.g. medical specialist care, public health, general practice), it appeared that guideline developers have developed all kinds of mechanisms to deal with such uncertainties (24). The methodological guidelines surely play a role in this, but intertwined with a practice-based strategy in which the knowledge and experience of medical practitioners is used for guideline development. Guideline development for infectious diseases proved to be the most advanced in this respect, probably driven by a structural lack of research results with the spread of new infectious diseases combined with a need to act immediately. Guideline developers in this area had developed a network with practitioners from Public health services across the Netherlands in order to be able to respond to new developments and monitor the effects of the proposed guidelines.

Also, with the development of clinical guidelines we thus see a much more reflexive practice that could have been expected on the basis of the guidelines for guideline development and which allows the gathering and use of all kinds of knowledge, next to that produced in experimental research. Following sociologists May and Finch we could speak of a practice of *reflexive monitoring* (25), allowing for and stimulating the mediation between evidence and practice and stimulating productive practices of guideline development; that is: stimulating the development of guidelines that both reflect the state of the art in research and can actually be used in practice.

Using guidelines in healthcare practice

Within the EBM movement that have long been concerns about the use of clinical guidelines in practice. While guidelines were meant to reduce practice variation and quickly disseminate the results of experimental research into healthcare practices, the results of this endeavour seem to be disappointing. Practice variation has in fact hardly been reduced over the last decades, as is shown by the unending stream of publications on the topic (26), and the take-up of guidelines in medical practices has also not been met with much enthusiasm. Research on the implementation of guidelines consistently shows that guidelines are used in about 50% of cases (27).

This has given impetus to a whole range of flanking policies and practices that should stimulate the uptake of guidelines. As a result, guidelines play an ever-increasing role in regulation and procurement. Also, many large scale quality programs have been set up to 'roll out' guidelines—in the Netherlands, these include programs such as 'Better Faster' and 'Care for Better', see e.g. (28, 29). Also, the science and practice of implementation has been further developed and 'implementation' in healthcare can by now call itself an academic field of its own, including its own journals and academic associations (30).

Research on the use of guidelines in practice however often uses a rather problematic perspective on use, often just seen as the 'correct' application of a guideline, in which 'correct' is more often than not defined by the researchers. Questionnaires, file research or observations are then used to 'measure' the percentage of patients or medical actions in which the guideline is applied. Again, ethnographic research on medical practices paints a different picture, in which the use of guidelines and protocols receives new meaning. Care, as shown in this type of research, is not just the application of rules from guidelines, but much more a messy process in which different options are constantly weighed and tested given the often complex situation in which patients (and professionals) find themselves (31).

In a study on the use of guidelines on heart failure by cardiologists, we started off with studying the patient records to see in which and how many cases the guideline was actually followed (32). Consistent with the literature, this appeared to be in about 50% of cases. In a next step we interviewed the cardiologists—with the patient records in hand—to get a feel for their experiences in using the guideline and find out why they did (not) use the guideline. From these interviews it not only appeared that the cardiologists were well aware of the existence of the guideline and applied it in their practices, but also that often they had good reasons not to 'apply' the guideline with specific patients. Often this entailed the sensitivity of a patient to specific drugs or the fact that patients suffered from multiple conditions which made the direct application of the guideline to those patients impossible. In the end, for only 10% of patients it remained unclear why the guideline was not applied—a completely different picture than often found in the literature on guideline use.

Care workers appear to take guidelines much more into account than often assumed. Guidelines do give direction to their actions. This however does not

mean that they follow guidelines uncritically. It is precisely in the confrontation between the guideline and the patient (and possible other factors) that decisions on treatments emerge. Again, this leads to the conclusion that the use of evidence is a reflexive practice in which guidelines together with their contexts of application lead to specific diagnostic or therapeutic interventions. Guidelines are an essential part of this, but instead of dictating specific actions they rather offer a kind of 'reflexive guidance'. Like we saw in clinical research and the development of guidelines, the use of guidelines in healthcare practice can rather be described as a practice of mediation between the guideline and the specificities of the context of application.

Conclusions

I started this essay with the question what we can learn from evidence-based medicine for evidence-based policy. Rather than basing this on formalistic prescriptions of EBM often found in the literature, I took a practice-based approach, focusing on the ways in which EBM is 'done' in practice. By analysing three 'episodes' in the practice of EBM the success of EBM seemed to be based on completely different mechanisms than often assumed. Three such mechanisms seemed crucial: creating reflexive spaces for the mediation of research and practice, the reflexive weighting of evidence in the light of practical contexts, and the use of evidence as a form of reflexive guidance. What do these conclusions mean for the practice of evidence-based policy? And, what are the prospects that these conclusions—themselves a product of research—can be translated to policy contexts?

Before addressing the first question, let me start with a cautionary note. When translating evidence from one field to another one always risks not taking into account qualitative differences between those fields. Some authors have for example argued that EBP is in nature much more complex than EBM as it has to deal with a more heterogeneous context and builds on more practical epistemologies (9, 33). While I do agree that context matters and that EBP should take into account the varying complexity of the problems it tries to tackle (8), I hope to have showed in this essay that practices of EBM are complex in themselves and that the differences between EBM and EBP in this regard are not essential. Learning across domains moreover needs a more practice-based approach this gives a deeper understanding of why a specific practice is successful (or not). The formalistic translations of EBM to evidence-bases policy arrangements that I cited earlier are unproductive precisely for the reason that they lack such a practice-based approach. Too large a separation between research and policy for example is not desirable as it prevents the necessary reflexive mediation between the two—as it does in the connection between science and practice in EBM.

The analysis of EBM as presented above also means that practices of reflexive monitoring should be taken much more seriously in contexts of evidence-based policy and should lead to mutual sense-making and modification of policies. This then also entails that policy arrangements should indeed leave room for reflexive monitoring and should create platforms for reflexivity. Additionally, one may conclude that research cannot be productively organized outside of the policy

process. One the one hand it should take into account the dynamics and complexities of policies and use the insights of policy practitioners in developing research designs, while on the other hand research results can not act as the final arbiters of policies but should be seen as just one—albeit an important one—ingredient of the policy-making process.

What then are the prospects for such an arrangement? I can best speak here for the Dutch context as this is the one I know best. Within the Netherlands, the prospects are both good and bad. Bad, because As a result of cuts in healthcare and in health (services) research, there is an increasing fix on 'quick wins' and on the implementation and 'roll out' of research results, without much attention for changing contexts. The space for incremental policies with forms of reflexive guidance therefore is limited. Research policy is moreover increasingly focused on 'valorisation' of research; while this means that societal relevance of research is pushed, the values that are supposed to come of that are almost exclusively economic. The quality systems of universities, still much focussed on as much as possible papers in English language journals (that are hardly read by Dutch policymakers and practitioners), offers little room for the interventionist research practices that are necessary to develop productive relations with policy practices.

Luckily there are also positive developments. The last years have seen an increasing discussion on the position of the universities, and research policies in general, amongst others fuelled by the Science in Transition movement (see http://www.scienceintransition.nl/english, accessed 8 July 2016). The Ministry of Education started an initiative to come to a national research agenda that seems to leave room for more practice-informed input (http://www.wetenschapsagenda.nl/?lang=en, assessed 8 July 2016). Interestingly, also the failure of many policy initiatives provides room for reflection. The generally felt crisis in the transparency program leads to new initiatives in different health sectors with other types of accounting for care—see e.g. the work of the association for long term care organisations on 'narrative accountability' (34). In reimbursement policies, the discussion on 'appropriate care' seems to open up space to discuss the relations between research and care in different ways than usual until recently. Many policy fields show movements that make room for the development of new ways to organize the relation between research and policy. New organisational arrangements for this relation have been developed, such as the 'Academic collaborative centres' that allow for new relations between policymakers, practitioners and researchers (35). The ACCs are partnerships between research organisations, policy departments and practitioners, stimulating the interaction across organisational and epistemological boundaries and as such furthering both policy-relevant research and evidence-informed policies. Once started in public health, these collaborative centres have now spread to all kinds of policy fields, such as youth care, supervision and reimbursement policy. Academic collaboratives, if well embedded, offer reflexive spaces for the mediation of policy, practice and research. Their success—or failure—is however dependent on the extent to which actors from these different worlds do indeed receive and use the space to experiment.

Acknowledgements

While I take full responsibility for this paper, I want to thank the two anonymous reviewers, Josje Kok and Iris Wallenburg for their careful reading of the paper. The essay is largely based on research I did with others—most of their names are listed in the reference section—and I want to thank them for this opportunity. An earlier version of this essay appeared (in Dutch) in the *Tijdschrift voor Gezondheid & Ethiek*.

References

- 1. Stolk EA, Brouwer WBF, Busschbach JJV. Rationalising rationing: economic and other considerations in the debate about funding of Viagra. Health Policy. 2002;59(1):53-63.
- 2. Bal R. Met common sense komen we er niet. Zorgvisie. 2013(27 Januari 2013): http://www.zorgvisie.nl/Kwaliteit/Verdieping/2013/1/Met-common-sense-komen-we-er-niet-1157655W/.
- 3. Oldenhof L, Postma J, Bal R. Re-placing care: governing healthcare through spatial arrangements. In: Ferlie E, Montgomery K, Reff Pedersen A, editors. Handbook of Healthcare Management. Oxford: Oxford University Press; 2016. p. 415-33.
- 4. Bal R. De nieuwe zichtbaarheid. Sturing in tijden van marktwerking. Rotterdam: Erasmus MC; 2008.
- 5. Quartz J, Wallenburg I, Bal R. The performativity of rankings. On the organizational effects of hospital league tables. Rotterdam: iBMG; 2013.
- 6. Schippers E. Het jaar van de transparantie. Den Haag: Ministerie van VWS, 2015 2 maart 2015. Report No.
- 7. Jasanoff S. Designs on Nature. Science and democracy in Europe and the United States. Princeton: Princeton University Press; 2005.
- 8. Bijker WE, Bal R, Hendriks R. Paradox of Scientific Authority. The role of scientific advice in democracies. Boston: MIT Press; 2009.
- 9. Dobrow MJ, Goel V, Upshur REG. Evidence-based health policy: context and utilisation. Social Science & Medicine. 2004;58:207-17.
- 10. Armstrong D. Clinical autonomy, individual and collective: the problem of changing doctors' behaviour. Social Science & Medicine. 2002;55:1771-7.
- 11. Timmermans S, Mauck A. The promises and pittfalls of evidence-based medicine. Health Affairs. 2005;24(1):18-28.
- 12. Chalmers I. If evidence-informed policy works in practice, does it matter if it doesn't work in theory? Evidence & Policy. 2005;1(2):227-42.
- 13. Walshe K, Rundall TG. Evidence-based Management: From theory to practice in health care. Milbank Quarterly. 2001;79(3):429-57.
- 14. Timmermans S, Berg M. The Gold Standard. The challenge of evidence-based medicine and standardization in health care. Philadelphia: Temple University Press; 2003.
- 15. Kok MO, Vaandrager L, Bal R, Schuit J. Practitioner opinions on health promotion interventions that work: opening the 'black box' of a linear Evidence Based approach. Social Science & Medicine. 2012;74:715-23.
- 16. Dehue T. Testing Treatments, Managing Life: on the history of randomized clinical trials. History of the Human Sciences. 1999;12(1):115-24.
- 17. Horstman K, de Vries G. Experimenteren met mensen. Kennis & Methode. 1989;13(1):62-83.
- 18. Jansen Y. Pragmatic trials; the mutual shaping of research and primary health care practice. An ethnographic analysis of the role the pragmatic trial methodology fulfils in bridging the science-practice gap. Rotterdam: Erasmus University; 2012.
- 19. Wehrens R, Bal R. Health programs struggling with complexity. A case study of the Dutch 'precare' project. Social Science & Medicine. 2012;75:274-82.

- 20. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-6.
- 21. Zuiderent-Jerak T, Forland F, Macbeth F. Guidelines should reflect all knowledge, not just clinical trials. BMJ. 2012;345:10.1136/bmj.e6702.
- 22. Moreira T. Diversity in clinical guidelines: the role of repertoires of evaluation. Social Science & Medicine. 2005;60(9):1975-85.
- 23. Knaapen L. Being 'evidence-based' in the absence of evidence: The management of non-evidence in guideline development. Social Studies of Science. 2013;43(5):681-706.
- 24. van Loon E, Bal R. Uncertainty and the Development of Evidence-Based Guidelines. Valuation Studies. 2014;2(1):43-64.
- 25. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology. 2009;43(3):535-54.
- 26. White J. Prices, Volume, and the Perverse Effects of the Variations Crusade. Journal of Health Politics, Policy and Law. 2011;36(4):775-90.
- 27. Grol R, Buchan H. Clinical guidelines: what can we do to increase their use? Medical Journal of Australia. 2006;185(6):301-2.
- 28. Zuiderent-Jerak T. Competition in the wild. Emerging figurations of healthcare markets. Social Studies of Science. 2009;39(5):765-92.
- 29. Strating M, Nieboer A, Zuiderent-Jerak T, Bal R. Creating effective quality improvement collaboratives: a multiple case study. BMJ Quality and Safety. 2011;20(4):344-50.
- 30. Wensing M, Bal R, Friele R. Knowledge implementation in healthcare practice: a view from the Netherlands. BMJ Quality and Safety. 2012;21:439-42.
- 31. Mol A. Living with diabetes: care beyond choice and control. The Lancet. 2009;373(9677):1756-7.
- 32. Oertle M, Bal R. Understanding non-adherence in chronic heart failure: a mixed-method case study. Quality and Safety in Health Care. 2010;19:e37.
- 33. Contandriopoulos D, Lemire M, Denis J-L, Tremblay É. Knowledge exchange processes in organizations and policy arenas: a narrative systematic review of the literature. Milbank Quarterly. 2010;88(4):444-83.
- 34. Ubels GM. Narrative accountability and quality awareness: Learning about (re) presenting narrative care. Journal of Aging Studies. 2015;34:190-8.
- 35. Wehrens R. Beyond Two Communities. Rotterdam: Erasmus University; 2013.