

# “What’s the evidence?”—Towards more empirical evaluations of the impact of OR interventions in healthcare

Guillaume Lamé <sup>a,b</sup>, Sonya Crowe <sup>c</sup> and Matthew Barclay <sup>a</sup>

<sup>a</sup>The Healthcare Improvement Studies Institute (THIS Institute), University of Cambridge, Cambridge, UK; <sup>b</sup>Laboratoire Génie Industriel, Université Paris-Saclay, CentraleSupélec, Gif-sur-Yvette, France; <sup>c</sup>Clinical Operational Research Unit, University College London, London, UK

## ABSTRACT

Despite an increasing number of papers reporting applications of operational research (OR) to problems in healthcare, there remains little empirical evidence of OR improving healthcare delivery in practice. Without such evidence it is harder both to justify the usefulness of OR to a healthcare audience and to learn and continuously improve our approaches. To progress, we need to build the evidence-base on whether and how OR improves healthcare delivery through careful empirical evaluation. This position paper reviews evaluation standards in healthcare improvement research and dispels some common myths about evaluation. It highlights the current lack of robust evaluation of healthcare OR and makes the case for addressing this. It then proposes possible ways for building better empirical evaluations of OR interventions in healthcare.

## ARTICLE HISTORY

Received 8 October 2019  
Accepted 24 November 2020

## KEYWORDS

Health Quality and Evaluation; healthcare Improvement Research; impact

## 1. Introduction

Many operational research (OR) practitioners and academics involved in healthcare will have been confronted with the comment “I like the idea of OR, but what’s the evidence for it?” when talking to clinicians, funders, policy-makers or editors and reviewers in medical journals. With this question, they are asking for evidence of OR improving healthcare organisations and often expect this evidence to have a certain form and be generated through specific evaluation processes.

Research applying OR techniques (including Soft OR and Problem Structuring Methods) to improve healthcare delivery is developing rapidly (e.g., Brailsford et al., 2009; Zhang et al., 2018), but falls short in demonstrating impact through sound empirical evaluations. Most healthcare OR papers do not discuss implementation (Brailsford et al., 2009; Brailsford & Vissers, 2011), let alone assess the *impact* of the OR intervention. Therefore, we have little to support the potential of OR in bringing about desirable change to the quality, safety, and efficiency of healthcare delivery, and lack key information to learn and iteratively refine our approaches.



Evaluation helps decision-makers understand what works in a given context, as well as how and why it works, before choosing a course of action (The Health Foundation, 2015). In this viewpoint, we argue that we need more, and better, evaluations of the impact of OR


interventions in healthcare. We take a broad view of OR, including but not limited to Problem Structuring Methods, mathematical modelling and simulation (see (Pitt et al., 2016) for examples). First, we discuss the type of evidence currently generated to support claims of OR’s potential impact. We then highlight the gap between the objectives of evidence-based healthcare improvement and the reality of current practice in healthcare OR. Finally, we identify possible ways to address this gap.

## 2. Evaluating healthcare improvement interventions

Evaluations aim “to determine merit, worth, value or significance” (Patton et al., 2014). In the context of interventions for improving healthcare, evaluation means answering the questions (Walshe, 2009):

- *Does the intervention work, and how well? How much does it cost?* The efficiency, effectiveness and efficacy of the intervention needs to be assessed to know if the resources invested are well spent. This is referred to as *outcome evaluation*.
- *Why and how does the intervention generate these outcomes in this context?* Underlying mechanisms linking an intervention to a given set of outcomes need to be analysed to understand in which other circumstances the intervention (or parts of it) may be useful. This is labelled *process evaluation*.

**CONTACT** Guillaume Lamé  [guillaume.lame@centralesupelec.fr](mailto:guillaume.lame@centralesupelec.fr)  Université Paris-Saclay, CentraleSupélec, Laboratoire Génie Industriel, 3 rue Joliot-Curie, F-91192 Gif-sur-Yvette Cedex, France.

 Supplemental data for this article can be accessed [here](#).

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

- *What is it like to use the intervention?* Learning from practical implementation by exploring the experience of those implementing and participating in the intervention, to improve the implementability and feasibility of the intervention.

The emerging consensus among healthcare improvement evaluators is to use a “programme theory” to plan the evaluation and synthesise its results. Programme theories synthesise different types of knowledge to explain how the components of an intervention will generate a certain effect through intermediate processes and subject to moderating factors (Davidoff et al., 2015; Funnell & Rogers, 2011).

The key advantage of a theory-driven approach is that it goes beyond asking whether the intervention works, to exploring where it works, for whom, and why, through unpacking the mechanisms that generate these impacts. This makes it easier to understand to what extent success is transferrable, not least by assessing what aspects of the intervention are particularly contingent on context (Davidoff et al., 2015). Programme theories also help explain unsuccessful interventions (Funnell & Rogers, 2011).

Programme theories can be used in quantitative, qualitative or mixed-methods evaluations, and in experimental, quasi-experimental or observational designs. In general, a degree of pragmatism is considered necessary in choosing an appropriate evaluation design due to the complexity of some interventions, their dynamic and evolving character, and the cost of running multisite experimental studies (Barry et al., 2018). Nonetheless, this flexibility does not mean that all evaluation designs and methods are equivalent or provide the same level of confidence in the impact of the intervention. For example, simple uncontrolled before-and-after studies are a very weak form of evaluation since any changes observed may be caused by secular trends, changes in the environment or the phenomenon of “regression to the mean”, rather than the intervention (Eccles et al., 2003). Interrupted time-series designs, where the outcomes of interest are measured at several points before, during, and after the intervention, are a stronger design because they allow the effect of the intervention to be distinguished from secular trends, while remaining simpler to organise than controlled experiments (Fretheim & Tomic, 2015). Although the literature tends to discuss quantitative designs more extensively, qualitative studies are also part of an evaluator’s toolkit in healthcare improvement (Portela et al., 2015).

In the Appendix to this article, we dispel some common myths about the evaluation of healthcare improvement interventions along with illustrative examples.

### 3. The current state of evaluation in healthcare OR

Much of the healthcare improvement research literature focuses on publishing evaluations but the healthcare OR literature has largely evolved separately. Evaluations of OR interventions are rarely reported. OR case studies typically describe the process of building models and improving their performance with few papers mentioning implementation. Literature reviews have highlighted this in healthcare OR in general (Brailsford & Vissers, 2011; Mahdavi et al., 2013; Van Sambeek et al., 2010), in modelling and simulation studies (Brailsford et al., 2009; Fone et al., 2003; Jahangirian et al., 2012; Long & Meadows, 2018; Mohiuddin et al., 2017; Van Lent et al., 2012; Wilson, 1981), multi-criteria decision analysis (Marsh et al., 2014), the application of Soft Systems Methodology (Augustsson et al., 2019), optimisation (Ahmadi-Javid et al., 2017), scheduling (Marynissen & Demeulemeester, 2019; Samudra et al., 2016) or in specific areas of healthcare, such as outpatient chemotherapy (Lamé et al., 2016), global health (Bradley et al., 2017) or community healthcare (Palmer et al., 2018). The results of Brailsford et al. (2009) that only 5 to 8% of modelling and simulation papers in healthcare mention the implementation of results in practice seem to hold for healthcare OR more broadly.

OR researchers and practitioners sometimes argue that the learning generated through an OR project is more important than the “answers”, with stakeholders gaining understanding through the process about how their organisations work and what affects their performance (Serman, 1994). However, we still need to assess who learns what, and how (Lamé & Simmons, 2020). Some researchers have started to do so through lab experiments (Monks et al., 2014), or retrospective interviews with experts (Thompson et al., 2016). Empirical evaluations in real improvement projects have also been published in a corporate context (Cavaleri & Serman, 1997; Read et al., 2012), but remain rare. In many cases, there is no measurement of changes to participants’ behaviour that could be linked to modelling projects, nor of participants’ reactions to and opinions of the modelling effort (Kunc et al., 2018). There also does not appear to be a consensus on what constitutes learning in OR interventions or on how to assess it.

### 4. The gap between standards of evaluation in healthcare improvement and current practice in healthcare OR

The lack of robust evaluation of the impact of OR interventions is particularly problematic in healthcare, where the evidence-based paradigm is spreading from clinical practice to management and policy,

leading to increasing pressure for better evidence on what works to improve healthcare delivery (Auerbach et al., 2007; Bevan et al., 2005; Grady et al., 2018; The Health Foundation, 2015). In Section 2, we described the standards for evaluation of healthcare improvement interventions. The gap between these standards and current practice in healthcare OR raises pragmatic and scientific arguments for empirical evaluations of the impact of OR interventions in healthcare organisations:

- Research on OR interventions in healthcare must produce evidence that is acceptable to those who will use the results, providing empirical evidence to meet the needs and expectations of healthcare improvement funders, practitioners and policy-makers.
- Expectations for sound evidence that OR brings about improvement are healthy and should drive us to better evaluate interventions not only as an end in itself, but as a means to enhance the effectiveness of the work we do in bringing about positive change in healthcare.

Empirical data on observed changes (or absence of such changes) linked to OR interventions is needed so that we can understand their impact. Recent examples show this is not out-of-reach. For instance, Monks et al. (2015) present a case study of the impact of a simulation study in emergency stroke care combining different types of quantitative evidence (before-and-after analysis on different process duration metrics, time series analysis on the implementation of certain good practices) with an analysis of the modeller's field notes during the intervention. Outcomes of interest improved after the intervention and the time-series analysis gives confidence that this was attributable to the intervention, whilst the qualitative evidence helped identify aspects of the intervention that supported stakeholder engagement and the credibility of the results. In another example, Crowe et al. (2017) use ethnography to provide insights into the role and specific contribution of OR in multidisciplinary projects in healthcare. Other studies have also started to look at the cost-effectiveness of building a model to tackle issues in healthcare (Soorapanth & Young, 2019; Young et al., 2018). Impact evaluations of Soft OR approaches are also appearing (Emes et al., 2018).

Another salient aspect of current healthcare OR publications is that the evaluator is often *embedded* in the intervention team (Barry et al., 2018). Many papers are self-reports of interventions carried out by the authors, with very few *external* evaluations of OR projects. The stroke project mentioned above (Heaton et al., 2016) is an exception, albeit aimed at evaluating the funding initiative that supported the project rather

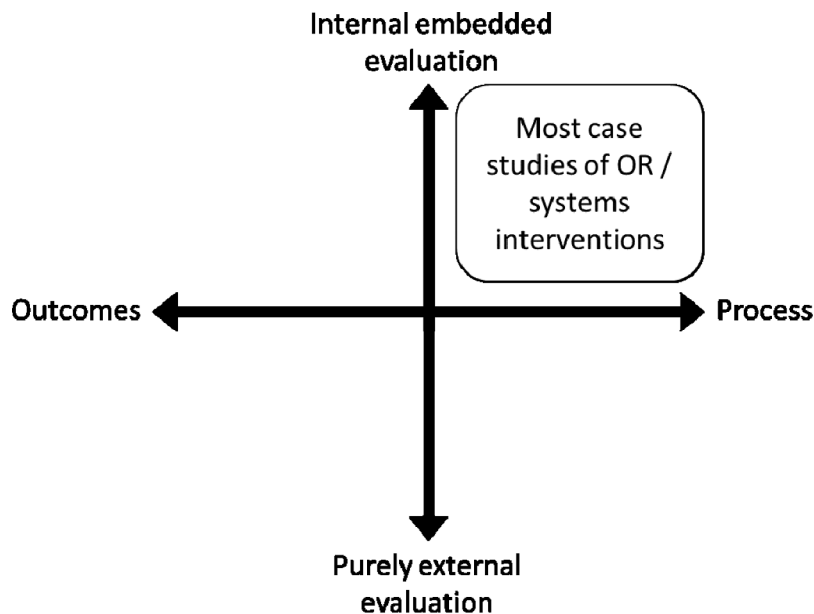
than the OR intervention. In another case, the design of the UK NHS Direct national 24-hour telephone helpline service, an external evaluation was reported independently from the OR intervention itself (Munro et al., 2000). However, the external evaluation focused on the solution proposed by the OR project (the telephone helpline) and does not mention the OR intervention. Therefore, it is not possible to understand from this evaluation how the OR intervention affected decisions that led to the implementation of the chosen solution. Elements on the role played by operational researchers in this project have been reported by the OR team (Royston et al., 2003), but fall into the category of self-reports rather than external evaluations.

Existing systematic reviews of the literature, detailed in Section 3, show most evaluations of healthcare OR are reported by the operational researchers who led the intervention, and focus on the modelling stages rather than the implementation of the findings. Current OR evaluations published in the academic literature primarily cover only one of the four quadrants of possible evaluation modalities (Figure 1):

- We rarely look at the *outcomes* of our interventions,
- We rarely design *external* evaluations.

There is nothing wrong with internal process evaluations of OR interventions, which can address many interesting questions. The problem comes when we avoid other types of evaluation. By not measuring what happens after our interventions, or relying on self-reported data and internal evaluations, we risk introducing biases that make interventions appear useful when they might do little to solve issues in practice. In reality, many interventions do not work as well as planned or equally well in every context (Øvretveit, 2011), yet few unsuccessful studies are reported in OR journals (for examples of OR interventions in healthcare reported as partially unsuccessful by their authors, see Bennett & Worthington, 1998; Connell, 2001). This suggests either some form of publication bias (where only positive studies are reported), or outcome reporting bias (where reporting is biased towards the more positive aspects of interventions, overlooking less successful dimensions) (Fanelli, 2012).

External evaluation and outcome evaluations are not panaceas. They are not always needed by healthcare stakeholders, nor are they always appropriate. Yet including more of them in our research portfolio could help us to understand how OR interventions work, specify which work best in different settings, and anticipate their likely impact. Systematic evaluation and reporting would also allow us to critique and improve the interventions we develop and our practice as operational researchers. Ultimately, these types of



**Figure 1.** Positioning current evaluations of OR approaches in a landscape of evaluation approaches.

evaluation would provide stronger arguments for using OR methods to bring about improvement in healthcare organisations, and enable continuous learning and improvement.

## 5. Towards empirical evaluations of OR in healthcare

### 5.1. What do we need to evaluate?

How should we explore the other quadrants of [Figure 1](#) and what exactly do we need to evaluate in OR interventions? When using OR methods to structure, model and better understand a problem situation in order to take informed action, the intervention and potential outcomes are harder to define upfront than for, say, an annual training module on patient safety. For instance, an OR intervention may use simulation to understand patient flows in an emergency department and evaluate the impact of different ways of organising resources and processes. The proposed changes from this OR intervention (e.g., shifting resources across different parts of the pathway) will then be debated and either implemented or not, and if implemented may or may not bring improvement.

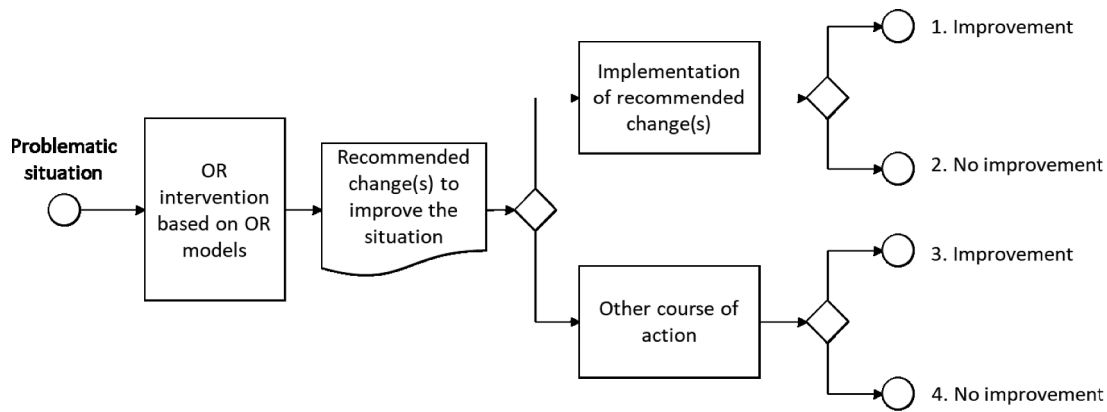
Determining whether improvement has occurred is not always straightforward and notions of improvement may be contested by different stakeholders. Therefore specifying appropriate evaluation outcomes for OR interventions can be challenging, particularly when the situations being tackled are multi-faceted and messy ([Williams, 2008](#)). However, some OR interventions in healthcare aim to address relatively bounded issues, such as patient flow ([Mohiuddin et al., 2017](#); [Palmer et al., 2018](#)), for which indicators

can be defined and used to measure impact. In more complex situations, qualitative methods may be better suited to assessing whether and in which ways things improve or worsen.

A simple model of the possible outcomes from an OR intervention can be helpful when planning evaluations. For example, [Figure 2](#) shows a generic model appropriate for relatively simple, linear, OR interventions. It raises generic questions for each possible outcome:

1. The changes recommended by the OR intervention are implemented, and the situation improves. Can the improvement be attributed to the implementation of the recommended changes? Can the choice to implement these changes be attributed to the OR intervention (or would it have been chosen anyway)? Did the recommended changes generate side effects?
2. The changes recommended by the OR intervention are implemented, but the situation does not improve. Was the decision to implement these changes attributable to the OR intervention? Were the recommended changes implemented correctly? Was the OR model appropriate (complexity, type of model, hypotheses, factors included and excluded)? Did the recommended changes generate side effects?
3. The changes recommended by the OR intervention are not implemented, and the situation improves. Why were the recommended changes not chosen? Might the recommended changes have led to greater improvement?
4. The changes recommended by the OR intervention are not implemented, and the situation does not





**Figure 2.** Generic model of the outcomes of OR interventions (similarities can be noted with the four stages of success in simulation projects suggested by Robinson and Pidd, 1998).

improve. Why were the recommended changes not chosen?

Addressing these questions requires a range of evaluation methods. For example, matters of attribution (“did the improvement come from the intervention?”) can be approached using experimental or quasi-experimental quantitative methods (e.g., using Statistical Process Control over the course of the modelling process (Perla et al., 2011)), while qualitative methods (e.g., interviews and observations) are well suited to identifying why people did or did not accept the OR recommendations. Side effects can be investigated both qualitatively and quantitatively, either by assessing against pre-identified potential side effects or in a more exploratory manner.

Existing frameworks for evaluating “complex interventions” in healthcare (like that proposed by the UK Medical Research Council, c.f. Campbell et al., 2007) may provide a useful basis on which to develop evaluation frameworks for OR interventions. Useful insights might also be drawn from the growing field of Behavioural OR, which seeks to examine the role and impact of behaviour on the use of OR to support decision-making (Franco & Hämmäläinen, 2016), including through qualitative approaches and experimental designs that could help us to understand stakeholders’ acceptance of, and fidelity to, recommendations from OR interventions.

## 5.2. Programme theories for OR interventions

A programme theory offers a theoretical model of how an intervention is expected to generate certain outcomes in a given context. Programme theories can help in designing appropriate evaluations. Figure 2 stems from a simple generic programme theory for OR, with three elements:

- Modelling supports the establishment and sharing of a common, simplified, representation of a complex situation.
- Manipulation of this simplified representation allows assessment of the likely effect of changes, and an exploration of the importance of different factors in the overall behaviour of the system.
- This experimentation allows people to learn about the behaviour and dynamics of a system and gives a shared foundation for debate, leading to better decisions about how to make changes in their organisation.

This simple programme theory will not always be appropriate. Many OR projects are iterative or involve continuous negotiation (Williams, 2008), so the programme theory would need to reflect this. There may need to be additional steps addressing the construction and presentation of the model: for example, some would argue that building models collaboratively, in a facilitated environment, improves their acceptance by stakeholders and that effective visualisation of model outputs can affect how users perceive the model. Fine-tuning an evaluation requires all these elements be considered and integrated into a programme theory that reflects the specifics of the OR intervention and the context.

We will often be able to develop programme theories based on our understanding of OR methods and the context in which they will be applied, but theoretical approaches from other disciplines also offer a rich repertoire for building programme theories for OR interventions. Activity theory (Leroy White et al., 2016), the concept of boundary objects (Franco, 2013), the mangle of practice framework (Ormerod, 2014) and single and double-loop learning (Monks et al., 2014) have all been used for this purpose, albeit not in healthcare OR. Drawing on theory from the social sciences forms a key aspect of the Behavioural OR research agenda (Becker, 2016; Brocklesby, 2016).

**BOX 1. Recommendations.**

- Think about *interventions* rather than models or methods. The impact of OR interventions can be very diverse, with models being only one aspect (Crowe et al., 2017). The processes of interventions, the interactions with the team and the social and political dynamics at play should also be assessed.
- Think about evaluation from the beginning. Develop a monitoring and evaluation plan when designing OR interventions. What would success look like for this intervention? How could it be assessed? What could go wrong? What type of data is needed for the evaluation? Include follow-up periods in projects to assess the impact of OR interventions.
- Collaborate with specialists in qualitative and quantitative evaluation to frame and conduct evaluations. Effective models exist for such collaborations between intervention designers, implementers and evaluators (Brewster et al., 2015).
- Use a programme theory to model interventions and design their evaluation. Theory-driven evaluation is in line with past recommendations in OR (Midgley, 1998; White, 2006) and the Behavioural OR movement (Becker, 2016; Brocklesby, 2016) has explored the type of external theories that can be drawn upon.
- Adapt evaluation designs to interventions, contexts of implementation and resources (Eccles et al., 2003; Portela et al., 2015).
- Think in research programmes rather than projects, with replication of the same intervention across multiple sites. This allows comparisons of the effect of the same intervention (e.g., a group model-building approach, an optimisation strategy or a problem structuring intervention) across settings.
- Incorporate evaluation in OR training curricula, including creating professional development modules on evaluating OR.
- Support the publication of empirical evaluations in OR journals and prompt authors to mention the practical outcome of OR interventions in their articles or to write follow-up articles on the implementation (or lack thereof).

**5.3. Practical challenges to evaluating OR**

Clarifying the scientific challenge and providing elements of a programme theory are merely first steps towards more empirical evaluations of OR interventions in healthcare, and challenges in practical attempts to evaluate healthcare OR may well occur.

For example, evaluation of OR may not fall within the remit of many traditional **funding** sources for OR projects. However, evaluations have started to appear in the UK, funded by organisations such as the National Institute for Health Research's Applied Research Collaborations (NIHR ARCs, previously CLAHRCs) and the Health Foundation (e.g., Crowe et al., 2017; Monks et al., 2015).

Most OR researchers are not trained in the **standards, methods and practices** of evaluating healthcare improvement interventions. In this case, collaborating with evaluation-driven disciplines such as health services research, implementation science or improvement research can help (Brailsford & Klein, 2015), and there may be scope to incorporate evaluation techniques within OR degrees and professional development. This resonates with one stream of the Behavioural OR movement, seeking to use theories and methods from the social and behavioural sciences to better understand what happens in and around OR interventions (Becker, 2016; Brocklesby, 2016).

Even so, **methodological challenges** arise when evaluating complex, evolving interventions (Burke & Shojania, 2018). Specific frameworks may need to be developed for OR interventions, depending on the type of problem, the modelling methods used and the scope of the project.

Finally, robust evaluation of OR methods and interventions may need multisite projects to compare the effect of an intervention in different places and contexts, requiring different **research project**

**management skills and strategies** from the one-off projects that constitute the majority of applied OR papers at present.

Although there is no "one best way" that would apply to every OR method and every situation in healthcare, some practical recommendations can be identified (Box 1).

**6. Conclusion**

Promoting the value of healthcare OR is challenging, despite steady developments since the 1950's and areas of notable success (for example, embedded OR in the UK NHS, c.f. Royston et al., 1999, 2003), because we rarely evaluate the impact of our projects. A key limitation in this analysis is our reliance on the peer-reviewed academic literature. Indeed, there could be OR practitioners generating very good evidence of the impact their OR interventions have on the organisations they are working with, enabling them to illustrate the value of the techniques within their organisation or to their next potential client. However, we are not aware of such evidence in the grey literature, and, if kept confidential, such evidence is of little use to the wider OR community.

Failing to engage in evaluation limits our impact on practices and performance in healthcare, not least because we are less able to have meaningful conversations with healthcare professionals who increasingly seek an evidence-base for change. Systematic approaches to assess the impact of our efforts are required, and there are already a range of approaches for evaluating healthcare improvement interventions that could be adapted for use with OR interventions. Evidence is not everything, and having it will not suddenly change how OR is perceived and used in healthcare, but not having it makes OR easy to dismiss. Importantly, beyond promoting OR methods, evidence from evaluations would allow us to learn

about their use in practice so that we can improve their effectiveness.

In this article, we have reviewed recommended practice in healthcare improvement research, which combines approaches prevalent in health services research, social sciences and public policy. We do not mean to place these approaches on a pedestal, nor suggest that healthcare improvement research is only of the highest methodological quality (Auerbach et al., 2007). Rather, we wish to trigger a debate on what constitutes evidence that OR is effective in bringing about desirable changes in healthcare. Many questions remain open. What outcomes should we measure? What are the key mechanisms that make OR interventions effective? We often talk about learning as a key process in OR projects, but how can we operationalise this concept for evaluation? On a more mundane level, how can this enterprise be funded? Who should we partner with? All these questions open exciting avenues for experimentation and progress for the healthcare OR community.

## Acknowledgments

The authors thank Graham Martin for valuable and constructive suggestions on earlier drafts of this paper. Any mistake remains our own.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

Guillaume Lamé  <http://orcid.org/0000-0001-9514-1890>

Sonya Crowe  <http://orcid.org/0000-0003-1882-5476>

Matthew Barclay  <http://orcid.org/0000-0003-1148-1922>

## References

Ahmadi-Javid, A., Jalali, Z., & Klassen, K. J. (2017). Outpatient appointment systems in healthcare: A review of optimization studies. *European Journal of Operational Research*, 258(1), 3–34. <http://www.sciencedirect.com/science/article/pii/S0377221716305239>

Auerbach, A. D., Landefeld, C. S., & Shojania, K. G. (2007). The tension between needing to improve care and knowing how to do it. *New England Journal of Medicine*, 357(6), 608–613. <https://www.nejm.org/doi/full/10.1056/NEJMs070738>

Augustsson, H., Churrua, K., & Braithwaite, J. (2019). Re-energising the way we manage change in healthcare: The case for soft systems methodology and its application to evidence-based practice. *BMC Health Services Research*, 19(1), 666. <https://doi.org/10.1186/s12913-019-4508-0>

Barry, D., Kimble, L. E., Nambiar, B., Parry, G., Jha, A., Chattu, V. K., ... Goldmann, D. (2018). A framework for learning about improvement: Embedded implementation and evaluation design to optimize learning. *International Journal for Quality in Health Care*, 30(suppl\_1), 10–14. <https://doi.org/http://dx.doi.org/10.1093/intqhc/mzy008>

Becker, K. H. (2016). An outlook on behavioural OR – Three tasks, three pitfalls, one definition. *European Journal of Operational Research*, 249(3), 806–815. <http://linkinghub.elsevier.com/retrieve/pii/S0377221715008978>

Bennett, J. C., & Worthington, D. J. (1998). An example of a good but partially successful or engagement: Improving outpatient clinic operations. *Interfaces*, 28(5), 56–69. <https://pubsonline.informs.org/doi/abs/10.1287/inte.28.5.56>

Bevan, H., Eglin, S., Gollop, R., Inglis, S., Laughton, J., Lee, J., ... Wilkins, J. (2005). *Evaluating improvement*. NHS Institute for Innovation and Improvement.

Bradley, B. D., Jung, T., Tandon-Verma, A., Khoury, B., Chan, T. C. Y., & Cheng, Y.-L. (2017). Operations research in global health: A scoping review with a focus on the themes of health equity and impact. *Health Research Policy and Systems*, 15(1), p 32. <https://doi.org/10.1186/s12961-017-0187-7>

Brailsford, S. C., Harper, P. R., Patel, B., & Pitt, M. (2009). An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation*, 3(3), 130–140. <https://doi.org/10.1057/jos.2009.10>

Brailsford, S. C., & Klein, J. H. (2015). *The value of modelling and simulation in healthcare. A review of the evidence and some possible ways forward*. B. University. [http://cumberland-initiative.org/wp-content/uploads/2015/07/the\\_value\\_report\\_web.pdf](http://cumberland-initiative.org/wp-content/uploads/2015/07/the_value_report_web.pdf)

Brailsford, S. C., & Vissers, J. (2011). OR in healthcare: A European perspective. *European Journal of Operational Research*, 212(2), 223–234. <https://doi.org/10.1016/j.ejor.2010.10.026>

Brewster, L., Aveling, E.-L., Martin, G., Tarrant, C., & Dixon-Woods, M. (2015). What to expect when you're evaluating healthcare improvement: A concordat approach to managing collaboration and uncomfortable realities. *BMJ Quality & Safety*, 24(5), 318–324. <https://qualitysafety.bmj.com/content/qhc/24/5/318.full.pdf>

Brocklesby, J. (2016). The what, the why and the how of behavioural operational research—An invitation to potential sceptics. *European Journal of Operational Research*, 249(3), 796–805. <http://linkinghub.elsevier.com/retrieve/pii/S0377221715008760>

Burke, R. E., & Shojania, K. G. (2018). Rigorous evaluations of evolving interventions: Can we have our cake and eat it too? *BMJ Quality & Safety*, 27(4), 254–257. <https://qualitysafety.bmj.com/content/qhc/early/2018/02/09/bmjqs-2017-007554.full.pdf>

Campbell, N. C., Murray, E., Darbyshire, J., Emery, J., Farmer, A., Griffiths, F., ... Kinmonth, A. L. (2007). Designing and evaluating complex interventions to improve health care. *BMJ*, 334(7591), 455–459. <http://www.bmj.com/cgi/doi/10.1136/bmj.39108.379965.BE>

Cavaleri, S., & Serman, J. D. (1997). Towards evaluation of systems-thinking interventions: A case study. *System Dynamics Review*, 13(2), 171–186. [https://doi.org/10.1002/\(SICI\)1099-1727\(199722\)13:2<171::AID-SDR123>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1099-1727(199722)13:2<171::AID-SDR123>3.0.CO;2-9)

Connell, N. A. D. (2001). Evaluating soft OR: Some reflections on an apparently 'unsuccessful' implementation using a Soft Systems Methodology (SSM) based approach. *Journal of the Operational Research Society*, 52(2), 150–160. <https://doi.org/10.1057/palgrave.jors.2601054>

Crowe, S., Turner, S., Utley, M., & Fulop, N. J. (2017). Improving the production of applied health research findings: Insights from a qualitative study of operational research. *Implementation Science*, 12(p), 112. <https://doi.org/10.1186/s13012-017-0643-3>



- Davidoff, F., Dixon-Woods, M., Leviton, L., & Michie, S. (2015). Demystifying theory and its use in improvement. *BMJ Quality & Safety*, 24(3), 228–238. <http://qualitysafety.bmj.com/lookup/doi/10.1136/bmjqs-2014-003627>
- Eccles, M., Grimshaw, J., Campbell, M., & Ramsay, C. (2003). Research designs for studies evaluating the effectiveness of change and improvement strategies. *Quality & Safety in Health Care*, 12(1), 47–52. <http://qualitysafety.bmj.com/content/qhc/12/1/47.full.pdf>
- Emes, M., Smith, S., Ward, S., & Smith, A. (2018). Improving the patient discharge process: Implementing actions derived from a soft systems methodology study. *Health Systems*, 8(2), 117–133. <https://doi.org/10.1080/20476965.2018.1524405>
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891–904. <http://link.springer.com/10.1007/s11192-011-0494-7>
- Fone, D., Hollinghurst, S., Temple, M., Round, A., Lester, N., Weightman, A., ... Palmer, S. (2003). Systematic review of the use and value of computer simulation modelling in population health and health care delivery. *Journal of Public Health Medicine*, 25(4), 325–335. <https://doi.org/10.1093/pubmed/fdg075>
- Franco, L. A. (2013). Rethinking Soft OR interventions: Models as boundary objects. *European Journal of Operational Research*, 231(3), 720–733. <https://doi.org/10.1016/j.ejor.2013.06.033>
- Franco, L. A., & Hämäläinen, R. P. (2016). Behavioural operational research: Returning to the roots of the OR profession. *European Journal of Operational Research*, 249(3), 791–795. <http://www.sciencedirect.com/science/article/pii/S0377221715009601>
- Fretheim, A., & Tomic, O. (2015). Statistical process control and interrupted time series: A golden opportunity for impact evaluation in quality improvement. *BMJ Quality & Safety*, 24(12), 748–752. <https://qualitysafety.bmj.com/content/qhc/24/12/748.full.pdf>
- Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models* (1st ed.). Jossey-Bass.
- Grady, D., Redberg, R. F., & O'Malley, P. G. (2018). Quality improvement for quality improvement studies. *JAMA Internal Medicine*, 178(2), 187. <https://doi.org/http://dx.doi.org/10.1001/jamainternmed.2017.6875>
- Heaton, J., Day, J., & Britten, N. (2016). Collaborative research and the co-production of knowledge for practice: An illustrative case study. *Implementation Science*, 11(1), p 20. <https://doi.org/10.1186/s13012-016-0383-9>
- Jahangirian, M., Naseer, A., Stergioulas, L., Young, T., Eldabi, T., Brailsford, S. C., ... Harper, P. (2012). Simulation in health-care: Lessons from other sectors. *Operational Research*, 12(1), 45–55. <https://doi.org/10.1007/s12351-010-0089-8>
- Kunc, M., Harper, P., & Katsikopoulos, K. (2018). A review of implementation of behavioural aspects in the application of OR in healthcare. *Journal of the Operational Research Society*, 71(7), 1055–1072. <https://doi.org/10.1080/01605682.2018.1489355>
- Lamé, G., Jouini, O., & Stal-Le Cardinal, J. (2016). Outpatient chemotherapy planning: A literature review with insights from a case study. *IIE Transactions on Healthcare Systems Engineering*, 6(3), 127–139. <https://doi.org/10.1080/19488300.2016.1189469>
- Lamé, G., & Simmons, R. K. (2020). From behavioural simulation to computer models: How simulation can be used to improve healthcare management and policy. *BMJ Simulation and Technology Enhanced Learning*, 6(2), 95–102. <https://stel.bmj.com/content/bmjstel/early/2018/10/20/bmjstel-2018-000377.full.pdf>
- Long, K. M., & Meadows, G. N. (2018). Simulation modelling in mental health: A systematic review. *Journal of Simulation*, 12(1), 76–85. <https://doi.org/10.1057/s41273-017-0062-0>
- Mahdavi, M., Malmström, T., van de Klundert, J., Elkhuizen, S., & Vissers, J. (2013). Generic operational models in health service operations management: A systematic review. *Socio-economic Planning Sciences*, 47(4), 271–280. <http://www.sciencedirect.com/science/article/pii/S0038012113000360>
- Marsh, K., Lanitis, T., Neasham, D., Orfanos, P., & Caro, J. (2014). Assessing the value of healthcare interventions using multi-criteria decision analysis: A review of the literature. *PharmacoEconomics*, 32(4), 345–365. <https://doi.org/10.1007/s40273-014-0135-0>
- Marynissen, J., & Demeulemeester, E. (2019). Literature review on multi-appointment scheduling problems in hospitals. *European Journal of Operational Research*, 272(2), 407–419. <http://www.sciencedirect.com/science/article/pii/S0377221718302108>
- Midgley, G. (1998). Theory and practice in operational research. *Journal of the Operational Research Society*, 49(11), 1219–1220. <https://doi.org/10.1057/palgrave.jors.2600653>
- Mohiuddin, S., Busby, J., Savović, J., Richards, A., Northstone, K., Hollingworth, W., ... Vasilakis, C. (2017). Patient flow within UK emergency departments: A systematic review of the use of computer simulation modelling methods. *BMJ Open*, 7(5), e015007. <http://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2016-015007>
- Monks, T., Pearson, M., Pitt, M., Stein, K., & James, M. A. (2015). Evaluating the impact of a simulation study in emergency stroke care. *Operations Research for Health Care*, 6, 40–49. <http://www.sciencedirect.com/science/article/pii/S2211692314200348>
- Monks, T., Robinson, S., & Kotiadis, K. (2014). Learning from discrete-event simulation: Exploring the high involvement hypothesis. *European Journal of Operational Research*, 235(1), 195–205. <http://www.sciencedirect.com/science/article/pii/S0377221713008047>
- Munro, J., Nicholl, J., O'Cathain, A., & Knowles, E. (2000). Impact of NHS Direct on demand for immediate care: Observational study. *BMJ*, 321(7254), 150–153. <http://www.bmj.com/content/bmj/321/7254/150.full.pdf>
- Ormerod, R. J. (2014). The mangle of OR practice: Towards more informative case studies of 'technical' projects. *Journal of the Operational Research Society*, 65(8), 1245–1260. <http://link.springer.com/10.1057/jors.2013.78>
- Øvretveit, J. (2011). Understanding the conditions for improvement: Research to discover which context influences affect improvement success. *BMJ Quality & Safety*, 20(Suppl1), i18–i23. [https://qualitysafety.bmj.com/content/qhc/20/Suppl\\_1/i18.full.pdf](https://qualitysafety.bmj.com/content/qhc/20/Suppl_1/i18.full.pdf)
- Palmer, R., Fulop, N. J., & Utley, M. (2018). A systematic literature review of operational research methods for modelling patient flow and outcomes within community healthcare and other settings. *Health Systems*, 7(1), 29–50. <https://doi.org/10.1057/s41306-017-0024-9>
- American Evaluation Association, Patton, M. Q., Asibey, E., Dean-Coffey, J., Kelley, R., Miranda, R., & Newman, G. F. (2014). *What is Evaluation?* American Evaluation Association, Washington, DC, USA. <https://www.eval.org/p/bl/et/blogid=2&blogaid=4>



- Perla, R. J., Provost, L. P., & Murray, S. K. (2011). The run chart: A simple analytical tool for learning from variation in healthcare processes. *BMJ Quality & Safety*, 20(1), 46–51. <https://qualitysafety.bmj.com/content/qhc/20/1/46.full.pdf>
- Pitt, M., Monks, T., Crowe, S., & Vasilakis, C. (2016). Systems modelling and simulation in health service design, delivery and decision making. *BMJ Quality & Safety*, 25(1), 38–45. <https://doi.org/10.1136/bmjqs-2015-004430>
- Portela, M. C., Pronovost, P. J., Woodcock, T., Carter, P., & Dixon-Woods, M. (2015). How to study improvement interventions: A brief overview of possible study types. *BMJ Quality & Safety*, 24(5), 325–336. <http://qualitysafety.bmj.com/lookup/doi/10.1136/bmjqs-2014-003620>
- Read, M., Gear, T., & Vince, R. (2012). Group inquiry to aid organisational learning in enterprises. *Journal of the Operational Research Society*, 63(6), 736–747. <https://doi.org/10.1057/jors.2011.80>
- Royston, G., Dost, A., Townshend, J., & Turner, H. (1999). Using system dynamics to help develop and implement policies and programmes in health care in England. *System Dynamics Review*, 15(3), 293–313. <https://onlinelibrary.wiley.com/doi/abs/10.1002/28SICI%291099-1727%28199923%2915%3A3%3C293%3A%3AAID-SDR169%3E3.0.CO%3B2-1>
- Royston, G., Halsall, J., Halsall, D., & Braithwaite, C. (2003). Operational Research for informed innovation: NHS Direct as a case study in the design, implementation and evaluation of a new public service. *Journal of the Operational Research Society*, 54(10), 1022–1028. <https://doi.org/10.1057/palgrave.jors.2601617>
- Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., & Rademakers, F. E. (2016). Scheduling operating rooms: Achievements, challenges and pitfalls. *Journal of Scheduling*, 19(5), 493–525. <https://doi.org/10.1007/s10951-016-0489-6>
- Soorapanth, S., & Young, T. (2019). Assessing the value of modelling and simulation in health care: An example based on increasing access to stroke treatment. *Journal of the Operational Research Society*, 70(2), 226–236. <https://doi.org/10.1080/01605682.2017.1421857>
- Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review*, 10(2–3), 291–330. <https://doi.org/10.1002/sdr.4260100214>
- The Health Foundation (2015). *Evaluation: What to consider. Commonly asked questions about how to approach evaluation of quality improvement in health care*. The Health Foundation, London, UK <https://www.health.org.uk/sites/default/files/EvaluationWhatToConsider.pdf>
- Thompson, J. P., Howick, S., & Belton, V. (2016). Critical Learning Incidents in system dynamics modelling engagements. *European Journal of Operational Research*, 249(3), 945–958. <http://www.sciencedirect.com/science/article/pii/S0377221715008905>
- van Lent, W. A. M., VanBerkel, P., & van Harten, W. H. (2012). A review on the relation between simulation and improvement in hospitals. *BMC Medical Informatics and Decision Making*, 12(1), p 18. <http://www.biomedcentral.com/1472-6947/12/18>
- van Sambeek, J. R. C., Cornelissen, F. A., Bakker, P. J. M., & Krabbendam, J. J. (2010). Models as instruments for optimizing hospital processes: A systematic review. *International Journal of Health Care Quality Assurance*, 23(4), 356–377. <https://doi.org/10.1108/09526861011037434>
- Walshe, K. (2009). Pseudoinnovation: The development and spread of healthcare quality improvement methodologies. *International Journal for Quality in Health Care*, 21(3), 153–159. <http://intqhc.oxfordjournals.org/cgi/doi/10.1093/intqhc/mzp012>
- White, L. (2006). Evaluating problem-structuring methods: Developing an approach to show the value and effectiveness of PSMs. *Journal of the Operational Research Society*, 57(7), 842–855. <https://doi.org/http://dx.doi.org/10.1057/palgrave.jors.2602149>
- White, L., Burger, K., & Yearworth, M. (2016). Understanding behaviour in problem structuring methods interventions with activity theory. *European Journal of Operational Research*, 249(3), 983–1004. <http://www.sciencedirect.com/science/article/pii/S0377221715006785>
- Williams, T. (2008). *Management science in practice*. Wiley.
- Wilson, J. C. T. (1981). Implementation of Computer Simulation Projects in Health Care. *Journal of the Operational Research Society*, 32(9), 825–832. <https://doi.org/10.1057/jors.1981.161>
- Young, T., Soorapanth, S., Wilkerson, J., Millburg, L., Roberts, T., & Morgareidge, D. (2018). The costs and value of modelling-based design in healthcare delivery: Five case studies from the US. *Health Systems*, 9(3), 253–262. <https://doi.org/10.1080/20476965.2018.1548255>
- Zhang, C., Grandits, T., Härenstam, K. P., Hauge, J. B., & Meijer, S. (2018). A systematic literature review of simulation models for non-technical skill training in healthcare logistics. *Advances in Simulation*, 3(1), 15. <https://doi.org/10.1186/s41077-018-0072-7>