University of Glasgow

Fischbacher, C. M., Lewsey, J., Muirie, J. and McCartney, G. (2021) A critical reflection on the use of improvement science approaches in public health. Scandinavian Journal of Public Health, (doi: 10.1177/1403494821990245).

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/234739/

Deposited on: 23 February 2021

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk

A critical reflection on the use of Improvement Science approaches in Public Health

Word count: 2639 (main text excluding abstract, references and table)

Colin M Fischbacher, Consultant in Public Health, Public Health Scotland, Edinburgh, Scotland, UK, EH12 9EB (Colin.Fischbacher@phs.scot).

Jim Lewsey, Professor of Medical Statistics, Institute of Health and Wellbeing, University of Glasgow, 1 Lilybank Gardens, Glasgow G12 8RZ. (Jim.Lewsey@glasgow.ac.uk)

Jill Muirie, Public Health Programme Manager, Glasgow Centre for Population Health, Olympia Building, 2-16 Orr Street, Bridgeton, Glasgow G40 2QH. (Jill.Muirie@glasgow.ac.uk)

Gerry McCartney, Consultant in Public Health, Public Health Scotland, Edinburgh, Scotland, UK, EH12 9EB. (Gerard.Mccartney@phs.scot)

Abstract

Objective

'Improvement science' is used to describe specific quality improvement methods (including tests of change and statistical process control). The approach is spreading from clinical settings to population wide interventions and is being extended from supporting the adoption of proven interventions to making generalisable claims about new interventions. The objective of this narrative review is to evaluate the strengths and risks of current improvement science practice, particularly in relation to how they might be used in population health.

Study design

Narrative review.

Methods

A purposive sampling of published studies to identify how improvement science methods are being used and for what purpose. The setting was Scotland and studies that focussed on health and wellbeing outcomes.

Results

We have identified a range of improvement science approaches which provide practitioners with accessible tools to assess small scale changes in policy and practice. The strengths of such approaches are that they facilitate consistent implementation of interventions already known to be effective and motivate and empower staff to make local improvements. However, we also identified a number of potential risks. In particular, their use to assess the effectiveness of new interventions often seems to pay insufficient attention to random variation, measurement bias, confounding and ethical issues.

Conclusion

The use of current improvement science methods to generate evidence of effectiveness for population wide interventions is problematic and risks unjustified claims of effectiveness, inefficient resource use and harm to those not offered alternative effective interventions. Newer methodological approaches offer alternatives and should be more widely considered.

[abstract word count: 250]

Key words

Improvement Science, Quality Improvement, Evaluation, Causality

Introduction

Public health professionals have a responsibility to use robust, high quality evidence to inform decisions relating to health and health care. This ensures that intended and unintended consequences are better understood, potential risks are mitigated and the most effective, resource efficient and safe policies and practices are adopted.

'Improvement science' (IS) is a term used to describe various approaches to quality improvement. Understanding and applying IS approaches in public health is made difficult by the variety of ways in which this term is used. Traditionally it has been described as a method of studying approaches to improve outcomes by making effective changes to practice. For example, Eccles[1] defines improvement science as:

> "the scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practice, and, hence, to improve the quality and effectiveness of health services."

Other definitions contrast this approach with a focus on the question of the effectiveness of a particular initiative:

> "Improvement science ..[looks] ..explicitly and critically at which factors and techniques are most useful for facilitating improvement and roll out and by encouraging more rigorous use of scientific methods for planning, implementing, analysing and disseminating research findings. This is different to solely focusing on whether a particular initiative is effective or not"[2]

Others suggest a more specific approach to generating change. The Institute for Healthcare Improvement (IHI) define IS as:

> "an applied science that emphasizes innovation, rapid-cycle testing in the field, and spread in order to generate learning about what changes, in which contexts, produce improvements."[3]

Building on this definition is the suggestion, by a leading proponent of improvement science in Scotland, that IS is a way of stimulating innovation, particularly amongst staff involved in service delivery:

"the approach of an improvement collaborative is to let the front-line staff come up with the ideas, to test them and see what works, and then to implement what works across the whole system. ... Some things that they tested did not work. When that happens, we should leave those things to one side, move on to things that do work and then share the learning."[4]

Definitions vary in the extent to which IS is seen as a source of new generalisable knowledge about "what works". For example, Marshall says that "improvement science aims to generate . . generalisable or transferable knowledge".[5] Geng includes "large-scale implementation and effectiveness trials of complex interventions" within the scope of 'implementation science'.[6] Oliver argues that these approaches may be better at identifying effective interventions than the methods advocated by the evidence based medicine movement.[7] Interest in measuring the impact of public

health interventions and in discovering what works best, for whom, and in what contexts,[8] has continued to grow in recent years.The objective of this narrative review is to evaluate how IS methods are currently being used. In particular, we use examples from the application of IS in Scotland on health and wellbeing outcomes to illustrate the strengths and weaknesses of the approach in terms of its potential relevance and utility to improve public health. We used purposive sampling to identify these examples from published and grey literature. Our results and reflections on the findings of this review are structured into the following sections: rationale, features and current uses of IS in public health; potential strengths and weaknesses of the use of IS in public health.

Rationale for improvement science in public health

The widespread promotion of IS in healthcare reflects a concern that traditional approaches to designing, planning and evaluating interventions and policies are slow, inflexible, resource intensive and cannot easily be adapted during an intervention. In contrast, repeated small 'tests of change' and frequent measurement is central to the IS approach, allowing interventions to be adapted before they are implemented widely. The IS approach is argued to be cheaper because new interventions have been extensively tested before they are implemented widely. Many proponents of IS are more concerned to see improvements in a specific local context than to answer causal questions such as 'why did this improvement occur?'. However, we argue that rigorous and valid answers to causal questions are essential before local change can be attributed to an intervention or the intervention can be promoted for use in other settings.

Features of improvement science in public health

Within the IS approach, 'rolling-out' an intervention focuses on the integrity of the intervention (i.e. ensuring that the key components are implemented) rather than intervention fidelity (i.e. ensuring that the intervention remains true to the original design). This allows adaptation to the local context, in contrast to traditional evaluation approaches[9] and is an attractive feature for public health interventions, which may be highly context specific.

The IS approach accepts 'consistent bias' (i.e. the pre-existing differences in outcomes between settings) and seeks to identify 'special cause' variation (such as that due to the intervention under consideration) and distinguish it from 'common cause' variation in outcomes, which is assumed to be random. Statistical Process Control (SPC) charts[10] such as 'run charts' or Shewhart charts are used to identify shifts in time series data. These use simple heuristic methods that can be used by practitioners without specialist statistical training and which are designed to allow users to identify changes that are thought unlikely to be due to chance.

Current uses of Improvement Science in public health

As noted above, IS can take an 'implementation science' approach to ensure that a proven 'bundle' of interventions is implemented consistently within a particular setting, achieving maximal impact.[11] The intervention is selected based on published evidence and the effectiveness and safety of the intervention is assumed rather than investigated.[11,12] Examples of this approach include the promotion of hand-washing in clinical settings,[12] or the implementation of clinical guidelines within a general practice setting.[13,14] It is worth noting that in this approach, ethical

approval is not required because the intervention is recognised to be effective on the basis of an existing body of evidence and the focus is simply on ensuring effective implementation.[15]

The second use we have observed is to check the extent to which an intervention known to be effective in one context is effective in a new setting.[16] This moves further towards assessing the causal relationship between the intervention and outcomes in the new setting. The limitations of IS for this purpose need to be balanced against the feasibility of collecting rigorous evidence of efficacy in every context in which an intervention is to be used.

A third use of IS is to assess the effectiveness of new interventions, or of interventions where the impact was previously unknown or uncertain in clinical and non-clinical settings. This use of IS inevitably involves more causal interpretations. For example, the impact of the Ventilator Acquired Pneumonia (VAP) 'bundle' of interventions[17] has been tested in intensive care units.[18,19] The rationale is that a small number of 'best practice' interventions, previously evaluated individually, could work together reduce the incidence of VAP. Compliance and outcome measures are plotted on a 'run chart' and changes in the trends in outcomes are related to changes in compliance and used as evidence of effectiveness.

Many (though not all) of those using IS approaches in this way draw causal conclusions in support of recommendations about the wider use of the specific intervention beyond their own local setting.[18] For example, Al-Tawfiq concludes that "implementing the IHI VAP bundle significantly resulted in the reduction of the VAP rate".[18] A Scottish study concluded that a similar intervention was "associated with a significant reduction in VAP acquisition" and that "this work adds to our knowledge surrounding effective guideline and bundle implementation".[19] The authors of this latter study may have been drawing conclusions about a particular improvement approach rather than the specific intervention, but this is not always clear.

Other authors are less guarded about making causal claims and recommendations for 'scaling up' interventions based on IS projects. For example, in relation to the Scottish 'Early Years Collaborative' (EYC) (emphases added):

> "In the EYC, frontline practitioners, using their professional expertise, identify where they consider **changes could be made that could lead to improved outcomes for children and families**. Using the Model for Improvement and starting by making small tests of change, **they are able to measure whether their theory is correct** before scaling it up".[20]

This wording suggests both causal conclusions and generalisable results. Again there are claims that appear to relate to intervention efficacy:

> "In 2014, two South Lanarkshire nurseries agreed to run a literacy project that encouraged their children to learn about literacy, their areas and meeting people from different areas for the second year. They knew that the approach had helped the children's literacy skills – **but they wanted to see if their hunch that it also had wider benefits for the children was correct.** To do this they used the improvement approach promoted by the Early Years Collaborative to test **how they might measure how being involved in the 'our literacy journeys' programme benefited the children**".[20]

In other Scottish improvement work there seems to be a clear intention to draw conclusions about effectiveness and make wider recommendations:

> "QI is a way to test, measure, evaluate and implement **new and more effective ways of working** with the resources that teams have available . . . In many areas, there are examples of **the most effective improvements being spread** to more local services and schools."[21]

There are other similar examples.[22]

Potential strengths of the use of Improvement Science in public health

One of the strengths of the IS approach is the enthusiasm it engenders among practitioners (as evidenced, for example, by the widespread participation in the Scottish Early Years Collaborative, now the Children and Young People Improvement Collaborative[23]). It provides accessible tools for small teams to implement and assess improvements without the need for statistical expertise. It removes the requirement for authorisation to improve quality, de-mystifies the use of scientific techniques for a broader range of staff and encourages them to incorporate improvement in their routine work. It can increase job satisfaction and create a more questioning and engaged workforce, patient or service user group. It may promote a greater focus on outcomes,[24] which if combined with appropriate use of effectiveness evidence could make substantial positive impacts.

The approach may prompt practitioners to assess the effectiveness of an intervention within the local context and to make improvements.[16] Adapting interventions to the local context has frequently been a missing step in approaches taken to Evidence-Based Medicine and is recognised as a difficulty in the use of evidence to guide decision-making.[25,26]

Potential risks of the use of Improvement Science in public health

Alongside these strengths there are difficulties that are not widely recognised, specifically when IS is used to draw conclusions about the impact of specific interventions or to make recommendations about their wider use.

Confounding and inappropriate attribution of effects arises from the absence of a comparison group to provide an estimate of the counterfactual outcome – what would have happened to the intervention group had they not received the intervention. Without a comparison group (ideally obtained from a randomised design), attributing changes in outcomes to the intervention is problematic. The implied counterfactual may be provided by participants prior to the intervention, but observed changes may be due to other trends over the period of the intervention, a point that is not always explicitly considered. Portela notes that "caution is needed in treating the outputs of quality improvement projects as generalisable new knowledge".[27]

Regression to the mean is an important source of bias when the intervention group is selected on the basis of increased risk of an outcome. Extreme values tend to move towards the population mean, so that changes are wrongly attributed to the intervention. Again, the lack of a comparison group in IS projects prevents this problem from being assessed.

Many IS projects have small samples, relatively short time series or infrequent outcomes. The issue of multiple testing arises when multiple interventions are examined (and ones deemed ineffective are discarded) and when testing is repeated over time until significant changes are observed. Even if an intervention has no effect, the likelihood that at least one in 20 tests will be statistically

significant at the conventional 5% level simply by chance is 64%.[28] These issues arise frequently in IS projects, but there seems to be limited awareness of the resulting increased risk of false positive findings.

Typically those assessing outcomes in IS interventions are not blind to the intervention, something that is a well-recognised source of bias.[29] This is rarely considered when IS type approaches make recommendations for wider use of an intervention. This may be made worse by a lack of expertise in designing measurement tools. Portela notes that "the quality of measurement and interpretation of data in quality improvement projects is often strikingly poor".[27] In a service environment where data recording has been introduced specifically for the purposes of IS, there is also a risk of a 'Hawthorne effect' (i.e. a change in the behaviour of those being observed simply as a result of the observation process).[30] While the specific cause of positive change may not matter if the focus is local improvement, it means that the intervention may not have the same benefits in a different setting.

The lack of research governance and ethical oversight of IS projects creates the risk that existing evidence about effectiveness is not used to design service change. There is also an increased risk that unanticipated negative consequences for particular groups may not have been considered – something that is often identified through an ethical approval process.

There is also a risk of publication bias towards positive results, or even a lack of publication irrespective of the result, which may lead to a lack of generalisable learning, often the justification of experimentation within services. We suggest that the absence of IS projects that report that an intervention is ineffective should raise concerns. These risks are summarised in the Table.

Table – Risks in using IS approaches to ascertain the effectiveness of interventions

Risk of the use of IS approaches	Consequences
Lack of an appropriate comparison group (as obtained, for example, by randomisation).	Insufficient consideration of the risk of bias and confounding; observed changes may be inappropriately attributed to the intervention.
Insufficient accounting for random variation, ignoring the impact of multiple testing (multiple interventions, multiple time points), using data from small populations with short time series or infrequent outcomes.	Increased risk of false positive findings (type 1 errors). Random variation becomes confused with impact unless appropriate statistical methods are used.
Insufficient attention to measurement error (poorly valid or reliable measurement instruments) or measurement bias (increased by lack of blinding of those who assess outcomes and lack of independence of those who measure outcomes from those who implement the intervention).	Increased risk of changes being inappropriately attributed to the intervention; lack of blinding and observer independence, increasing the risk of false positive findings.
Lack of ethical approval.	Increased risk that participants will be exposed to interventions that are ineffective or have harms that have not been considered or mitigated.
Lack of research governance.	Increased risk that interventions will be planned without reference to existing evidence; that (particularly negative) results will not be published and learning will not be recorded or disseminated; that measurement and recording tools will be of insufficient reliability or validity; and that poor quality methods will be applied without sufficient support.

The dangers of treating and managing intermediate outcomes rather than the outcomes directly relevant to a population are well recognised[31] – and this is a particular risk when short-term and measurable outcomes are identified to suit the timescales required by the IS approach. More measurable intermediate outcomes are often at higher risk of measurement bias.

Conclusions

Although no doubt used with the best of intentions to improve population health, the mis-application of IS approaches could lead to substantial opportunity costs or even harms. Indeed, there are numerous examples of well-intentioned interventions producing negative outcomes.[32]

In practice, broad IS type approaches are increasingly being adopted to ascertain the impact of interventions and policies of uncertain effectiveness and as a justification for recommendations that these interventions are used more widely.

It is our view that IS methods are best suited to supporting the wider or more consistent implementation of interventions of known efficacy and to implementing local quality improvement rather than generating generalisable evidence of effectiveness. Assessing whether expected change from interventions known to be effective in other contexts has occurred locally works best: with frequent intervention cycles over short periods; with sufficient events for robust estimation; with clearly defined outcomes and attention to the risk of measurement bias; where there is a short causal chain between the intervention and the outcome, and where there is minimal scope for confounding.

We suggest that greater use should be made of newer methods which can provide more robust conclusions when these conditions cannot be met. Interrupted time series methods,[33–36] instrumental variable approaches,[37] differences in differences analyses[38] and regression discontinuity designs[39] are all well established methods which provide alternatives to traditional improvement science approaches when the focus is on examining the effectiveness of population-wide interventions. Arguably these methods are under-used, largely due to lack of awareness or difficulties in accessing statistical expertise. Despite this there are clearly opportunities to use more robust methods that provide stronger causal evidence and that quantify the extent of change rather than offering a simple binary conclusion about whether change has occurred.[35]

List of abbreviations

EYC Early Years Collaborative

IS Improvement Science

SPC Statistical Process Control

VAP Ventilator Acquired Pneumonia

Declarations

Ethical approval

No ethical approval was sought for this study as no patients were involved and no patient data used. The issue of consent to participate does not apply.

Consent for publication

Not applicable

Availability of data and material

Not applicable

Competing interests

None declared

Funding

None

Acknowledgements

None

References

1. Eccles MP, Mittman BS. Welcome to Implementation Science. Implement Sci [Internet]. 2006 [cited 2018 Oct 1];1. Available from: http://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-1-1

2. The Health Foundation. Evidence scan: Improvement science [Internet]. 2011. Available from: http://www.health.org.uk/sites/health/files/ImprovementScience.pdf

3. Institute for Healthcare Improvement. Science of Improvement [Internet]. [cited 2018 Oct 1]. Available from: http://www.ihi.org/about/Pages/ScienceofImprovement.aspx

4. Scottish Parliament. COVID-19 Committee [Internet]. 2020. Available from: http://www.parliament.scot/parliamentarybusiness/report.aspx?r=12631&i=114294

5. Marshall M, Pronovost P, Dixon-Woods M. Promotion of improvement as a science. The Lancet. 2013;381:419–21.

6. Geng EH, Peiris D, Kruk ME. Implementation science: Relevance in the real world without sacrificing rigor. PLOS Med. 2017;14:e1002288.

7. Oliver D. Should practical quality improvement have parity of esteem with evidence based medicine? BMJ. 2017;357:j2582.

8. Pawson R, Tilley N. Realistic evaluation. London: SAGE; 1997.

9. Deming WE. The New Economics for Industry, Government, Education. 2nd ed. London: MIT Press; 2000.

10. Flowers J. Statistical process control methods in public health intelligence. London: Association of Public Health Observatories; 2007.

11. Pearson A. Improvement science: getting the evidence into practice. Int J Evid Based Healthc. 2010;8:109.

12. White CM, Statile AM, Conway PH, Schoettker PJ, Solan LG, Unaka NI, et al. Utilizing improvement science methods to improve physician compliance with proper hand hygiene. Pediatrics. 2012;129:e1042–50.

13. Hull S, Chowdhury TA, Mathur R, Robson J. Improving outcomes for patients with type 2 diabetes using general practice networks: a quality improvement project in east London. BMJ Qual Saf. 2014;23:171–6.

14. Vander Schaaf EB, Seashore CJ, Randolph GD. Translating Clinical Guidelines Into Practice: Challenges and Opportunities in a Dynamic Health Care Environment. N C Med J. 2015;76:230–4.

15. Lynn J, Baily MA, Bottrell M, Jennings B, Levine RJ, Davidoff F, et al. The Ethics of Using Quality Improvement Methods in Health Care. Ann Intern Med. 2007;146:666–73.

16. Dixon-Woods M, Bosk CL, Aveling EL, Goeschel CA, Pronovost PJ. Explaining Michigan: Developing an Ex post theory of a quality improvement program. Milbank Q. 2011;89:167–205.

17. Institute for Healthcare Improvement. Getting started kit: prevent ventilator-associated pneumonia: how-to guide. Crit Care Nurs Q. 2006;29:157–73.

18. Al-Tawfiq JA, Abed MS. Decreasing ventilator-associated pneumonia in adult intensive care units using the Institute for Healthcare Improvement bundle. Am J Infect Control. 2010;38:552–6.

19. Hawe CS, Ellis KS, Cairns CJS, Longmate A. Reduction of ventilator-associated pneumonia: active versus passive guideline implementation. Intensive Care Med. 2009;35:1180–6.

20. Scottish Parliament. Background notes on the Early Years Collaborative [Internet]. Edinburgh: Scottish Parliament: Health and Sport Committee; Available from: http://www.parliament.scot/S4_HealthandSportCommittee/Inquiries/Note_on_Early_Years_Collaborative.pdf

21. Scottish Government. Children and Young People Improvement Collaborative [Internet]. [cited 2018 Oct 1]. Available from: https://beta.gov.scot/policies/improving-public-services/children-and-young-people-improvement-collaborative/

22. Children and Young People's Improvement Collaborative. Castlehead High School: Renfrewshire Council [Internet]. [cited 2018 Oct 1]. Available from: http://www.cypic.co.uk/wp-content/uploads/2017/12/11f19673-d69b-43c8-bbbc-4b48f5b9ccce_CYPIC_-_2017_-_Learning_Session_-_DAY_1_-_CASTLEHEAD_HS.pdf

23. Scottish Government. Early Years Collaborative [Internet]. 2014 [cited 2018 Oct 1]. Available from: https://www.webarchive.org.uk/wayback/archive/20160401140103/http://www.gov.scot/Topics/People/Young-People/early-years/early-years-collaborative

24. Wimbush E. Implementing an outcomes approach to public management and accountability in the UK:are we learning the lessons? Public Money Manag. 2011;31:211–8.

25. Fahey T. Applying the results of clinical trials to patients to general practice: perceived problems, strengths, assumptions, and challenges for the future. Br J Gen Pract. 1998;48:1173–8.

26. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" The Lancet. 2005;365:82–93.

27. Portela MC, Pronovost PJ, Woodcock T, Carter P, Dixon-Woods M. How to study improvement interventions: a brief overview of possible study types. BMJ Qual Saf. 2015;24:325.

28. Austin PC, Mamdani MM, Juurlink DN, Hux JE. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. J Clin Epidemiol. 2006;59:964–9.

29. Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. Can Med Assoc J. 2013;185:E201–2011.

30. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. J Clin Epidemiol. 2014;67:267–77.

31. Weintraub WS, Lüscher TF, Pocock S. The perils of surrogate endpoints. Eur Heart J. 2015;36:2212–8.

32. Macintyre S, Petticrew M. Good intentions and received wisdom are not enough. J Epidemiol Community Health. 2000;54:802–3.

33. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. BMJ. 2007;334:349–51.

34. Robinson M, Geue C, Lewsey J, MacKay D, McCartney G, Curnock E, et al. Evaluating the impact of the Alcohol Act on off-trade alcohol sales: a natural experiment in Scotland. Addiction. 2014;109:2035–43.

35. Fretheim A, Tomic O. Statistical process control and interrupted time series: a golden opportunity for impact evaluation in quality improvement. BMJ Qual Saf. 2015;24:748–52.

36. Ramsay G, Haynes AB, Lipsitz SR, Solsky I, Leitch J, Gawande AA, et al. Reducing surgical mortality in Scotland by use of the WHO Surgical Safety Checklist. Br J Surg. 2019;106:1005–11.

37. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000;29:722–9.

38. Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. JAMA. 2014;312:2401–2.

39. Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. BMJ. 2016;i1216.