The value of pragmatic and observational studies in health care and public health

Maxwell S Barnish  
Steve Turner  
Child Health, University of Aberdeen, Aberdeen, UK

Abstract: Evidence-based practice is an important component of health care service delivery. However, there is a tendency, embodied in tools such as Grades of Recommendation, Assessment, Development, and Evaluation, to focus principally on the classification of study design, at the expense of a detailed assessment of the strengths and limitations of the individual study. Randomized controlled trials (RCTs), and in particular the classical “explanatory” RCT, have a privileged place in the hierarchy of evidence. However, classical RCTs have substantial limitations, most notably a lack of generalizability, which limit their direct applicability to clinical practice implementation. Pragmatic and observational studies can provide an invaluable perspective into real-world applicability. This evidence could be used more widely to complement ideal-condition results from classical RCTs, following the principle of triangulation. In this review article, we discuss several types of pragmatic and observational studies that could be used in this capacity. We discuss their particular strengths and how their limitations may be overcome and provide real-life examples by means of illustration.

Keywords: research methods, randomized controlled trials, pragmatic trials, observational studies, disease registries, evidence-based medicine

Introduction
Evidence-based practice is an important component in health care service delivery, although its development into a formal paradigm is a relatively recent phenomenon. One characteristic feature of common conceptualizations of evidence-based medicine (EBM) is a focus on quality of evidence. It is of course important to consider the quality of evidence, since clinical practice recommendations based on low-quality evidence may not lead to high-quality clinical outcomes. However, one potential danger of this approach has been the emergence of an established hierarchy of evidence, which is based principally on the classification of study design selected, rather than the methodological quality of the individual study. Randomized controlled trials (RCTs) hold a privileged place in this hierarchy. For example, the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) system, which is one common framework for EBM clinical guideline development, classifies randomized studies as “high” evidence and nonrandomized studies as “low” evidence. Therefore, in this approach, a well-designed robust nonrandomized study would be considered poorer evidence than a poorly designed randomized study with critical methodological flaws that make its conclusions highly unlikely to be valid. As peer review has its limitations, it is important not to discount the possibility of the existence of poor-quality RCTs in
the published scientific literature. As such, it is concerning that systems such as GRADE place so much emphasis on study design classification, such as randomized trial or non-randomized trial, at the expense of a more holistic evaluation of the methodological rigor of the individual study.

There are a variety of trial designs that can be classified as RCTs. Nevertheless, the classical (sometimes called “explanatory”) RCT remains at the pinnacle of the hierarchy of evidence, apart from systematic reviews and meta-analyses of RCTs. Nevertheless, the privileged place of the classical RCT, even if well designed, has come into question, especially in recent years. Such studies may be just too far removed from the realities of clinical practice to be able to reliably inform it. Classical RCTs seek to achieve optimal internal validity, but this comes at the cost of markedly reduced external validity (or generalizability). Sociodemographic biases frequently occur, often as a combination of barriers to recruitment of certain types of patients and very stringent inclusion criteria. For example, the classical RCT routinely excludes potential participants with multiple comorbidities, and this may exclude at least 89% of patients encountered in routine care. Older people are also markedly underrepresented in trials of pharmaceutical therapies they are likely to receive. Moreover, a study of recruitment to asthma and chronic obstructive pulmonary disease trials showed that only 1% and 7% of the routine common patient populations, respectively, met common clinical trial inclusion criteria. For example, the classical RCT routinely excludes potential participants with multiple comorbidities, and this may exclude at least 89% of patients encountered in routine care. Therefore, these methodological concerns about classical RCTs are longstanding history. Indeed, the distinction between classical RCTs and a more pragmatic RCT design was discussed back in the 1960s. The concept of the pragmatic RCT arose out of a recognition of some of the issues with classical RCTs that we have discussed earlier, most notably a lack of generalizability and applicability to routine clinical practice. Therefore, these methodological concerns about classical RCTs are longstanding rather than a recent occurrence. Nevertheless, pragmatic trials long remained as a minority approach on the edge of the RCT scene and undervalued by regulators. Starting in the 1990s, there has been an increased interest in pragmatic trials within the academic medical community, leading to the publication of a number of commentaries. For example, Roland and Torgerson published an article in the BMJ outlining the key features and utility of pragmatic RCTs. Nevertheless, it was not until 2009 when the first formal tool to characterize the extent to which a given trial was pragmatic or explanatory was published: the Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) tool.

### Pragmatic trials of clinical interventions

Pragmatic trials of clinical interventions have a relatively longstanding history. Indeed, the distinction between classical RCTs and a more pragmatic RCT design was discussed back in the 1960s. The concept of the pragmatic RCT arose out of a recognition of some of the issues with classical RCTs that we have discussed earlier, most notably a lack of generalizability and applicability to routine clinical practice. Therefore, these methodological concerns about classical RCTs are longstanding rather than a recent occurrence. Nevertheless, pragmatic trials long remained as a minority approach on the edge of the RCT scene and undervalued by regulators. Starting in the 1990s, there has been an increased interest in pragmatic trials within the academic medical community, leading to the publication of a number of commentaries. For example, Roland and Torgerson published an article in the BMJ outlining the key features and utility of pragmatic RCTs. Nevertheless, it was not until 2009 when the first formal tool to characterize the extent to which a given trial was pragmatic or explanatory was published: the Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) tool.

In summary, according to PRECIS, pragmatic trials tend not to exclude based on comorbidities or compliance, they offer practitioners flexibility in how to apply the intervention, use a wide range of clinical practitioners for both the
intervention and the control group, use standard care as the comparator intervention, rely on administrative databases rather than formal follow-up visits, use a clinically meaningful outcome measure, do not seek to measure or improve participant or practitioner compliance to the protocol, and use intention-to-treat analysis to assess how the treatment works under usual clinical conditions. An updated version called PRECIS-2\(^2\) removed four of the original 10 domains and added three new domains: recruitment, setting, and organization. PRECIS-2 is also a well-validated tool and benefits from the increased granularity of responses offered by a Likert scale rather than binary answers for each component domain. Investigators considering designing a pragmatic trial could use PRECIS-2 to consider which combination of design features is likely to provide the most appropriate fit for a given situation.

A trade-off as we move along the continuum from explanatory to pragmatic comes in terms of increased generalizability at the cost of reduced internal validity. As a challenger to the classical RCT paradigm, it is unsurprising that pragmatic trials have faced criticism in some quarters, as reviewed, for example, by Patsopoulos.\(^2\) Potentially related to the traditional focus by regulators on explanatory RCTs, funding was reported in 2003 to be a barrier to the wider adaptation of pragmatic RCTs, because they were seen as poorly aligned to the priorities of major funding bodies.\(^2\) However, in light of increasing recognition of the value of pragmatic RCTs for informing clinical practice, this situation may have changed or at least would be expected to do so soon. No more recent reports of such issues could be identified.

Increased funding for pragmatic trials may be expected to follow an increased focus by governments on the social value.\(^2\) For example, a landmark series of pragmatic RCTs funded by the US National Institute of Mental Health\(^2\) provided invaluable evidence that currently available psychotropic medications, licensed on the basis of the results from classical explanatory RCTs, are ineffective in a substantial proportion of real-life clinical practice patients within their licensed clinical indication.

Nonrandomized and observational studies of clinical interventions

Traditionally, observational research has been seen as a way of determining risk factors and mechanisms of actions that could serve as potential targets for therapeutic intervention, to be assessed in future RCTs. Nonrandomized studies have been seen as simply too prone to methodological issues such as confounding to be useful in the assessment of potential health care interventions, barring an overriding ethical or practical reason why randomized studies could not be conducted. Clearly, observational research is useful for identifying potential intervention targets. However, the results of observational research can be used to inform our clinical understanding of the nature of particular clinical conditions, which can in turn inform clinical decisions regarding which of a range of approved and available treatments to offer a particular patient, as well as guiding decisions regarding which treatments to fund. For example, the Aberdeen Schools Asthma Survey (ASAS), one of the longest-running asthma epidemiology studies worldwide, which is not explicitly a study of treatment options, has provided information that
eczema and asthma, despite frequent co-occurrence, are likely to have separable mechanisms of action and that clinicians should be aware that these are distinct conditions and use this to inform their diagnostic and treatment choices. Moreover, ASAS has demonstrated how asthma risk factors are not static, but instead involve over time and are responsive to social and environmental changes, and that, therefore, if clinicians encourage modification of risk factors such as protecting children from second-hand smoke, asthma risk may be reduced. Moreover, a review by Turner shows how observational evidence about childhood asthma and asthma exacerbations can inform the potential of greater personalization of care in the future. Stratified medicine is an important concept, and one that should expand in terms of adaptation in routine practice in future. An important principle of stratified medicine is the realization that not all patients are the same. In contrast, RCTs, especially classical RCTs, assume that all patients are the same, and this is not a valid assumption.

Moreover, there has been renewed interest in how observational studies of health care interventions themselves can complement evidence from RCTs. However, such methods need to be used with care because such studies are more prone to limited internal validity than pragmatic RCTs. Moreover, in many cases, pragmatic RCTs can deliver comparable benefits in terms of direct applicability to routine practice settings. Sometimes, emerging fields will be dominated by a series of small observational studies. This is ill-advisable, because both negative and positive results can be difficult to interpret because of issues of confounding and power. This exact situation was encountered in a recent systematic review of potential benefit of singing for people with Parkinson’s disease, in which provisional evidence of a benefit on speech was found but many questions remained unanswered. Nevertheless, one particularly valuable use of observational studies of health care interventions that has become more common in recent years is the disease registry. Registries can provide invaluable insight into cost and patient outcomes in the context of routine health care service provision, while there may be justifiable concerns that the latter situation may not, if certain types of patient are more likely to consent to participate. In a UK context, clinical governance and ethics policies have been previously cited as barriers to the scientific validity of clinical research.

One particularly long-standing registry that has provided invaluable information about the benefits and potential safety issues associated with biological therapy in rheumatoid arthritis is the British Society for Rheumatology Biologics Register for Ankylosing Spondylitis (BSRBR-RA). Having reached 15 years and still being open to recruitment, BSRBR-RA is a good example of what a registry can achieve if sufficient financial support can be found to keep it going for long term. Such long-term registries, in addition to the benefits outlined earlier, offer insight into safety issues that are either relatively rare and only emerge in large registries or open emerge in the long term, beyond the relatively short follow-up period typically afforded by RCTs. As used in analyses from BSRBR-RA, propensity score matching techniques can be used in longitudinal analysis of registry data in an effort to overcome confounding by baseline group differences, which is ordinarily a substantial disadvantage of nonrandomized studies of health care interventions. Propensity score matching goes far beyond more traditional approaches such as covariate adjustment. This is an example of how advances in statistical techniques can help unleash the full value of otherwise limited methodologies in order to benefit from their particular strengths. As the confounding issue can now be largely addressed, the true benefits of registries in terms of observing real-life clinical care can be realized. It is important to note that while issues such as confounding and missing data can be addressed using statistical techniques, bias cannot be corrected for. Therefore, an advantage of registries, and pragmatic RCTs, over explanatory RCTs, is that they feature far less selection bias relative to everyday clinical practice – indeed the only selection bias that may arise in a registry study is if ethics board requirements necessitate individual patient consent even though the data collection occurs in the context of routine health care service provision, as discussed earlier. The rise of registries has also spawned
another new methodology that combines the strengths of registries and those of pragmatic RCTs – this is the registry RCT\(^{40,41}\) and it will be interesting to see whether and how this method takes off in the next 5–10 years.

**Evaluating public health interventions**

Health-related interventions can transcend the level of interactions between a patient and a health care provider. Public health interventions, unlike interventions in clinical practice, operate at the level of treating the population. For example, the Faculty of Public Health (UK) defines public health as “The science and art of promoting and protecting health and well-being, preventing ill-health and prolonging life through the organized efforts of society.”\(^{42}\) Public health interventions can take a variety of forms from publicity campaigns and media coverage to raise awareness of an issue related to a health behavior to changes in social policy and legislation. It is important, therefore, to define a public health intervention broadly, in terms of an intervention that is designed to benefit population health, regardless of whether or not the body implementing it is a public health authority. Indeed, there is a current trend in many localities,\(^ {43}\) including England (but hitherto not Scotland), to allocate principal responsibility to governmental local authorities rather than health authorities. Nevertheless, these local authorities employ medically qualified personnel as Directors of Public Health as well as in other consultant-level public health posts.

What many of these myriad approaches have in common is that they are not amenable to traditional evaluative methods such as the RCT as are often applied to clinical practice settings. This is because it is often either impractical to implement such a population-level initiative on a test basis or there are ethical issues that would prevent randomization, particularly in the context of regulating lifestyle and health behavior. For example, it is unethical to randomize humans to smoke cigarettes and while it may be ethical to engage individuals in a trial of smoking cessation therapy, although some ethical issues remain,\(^ {44}\) it would be unethical to randomize localities in a way that would simulate the effect of the introduction of antismoking legislation. There are indeed many areas of public health practice and policy in which RCTs are not feasible, including measures to reduce air pollution and smoking bans. Therefore, greater extrapolation and interpretation are often required when using evidence to inform the real-life implementation of such policy interventions.

Observational research is the most appropriate paradigm to embrace in such settings when ethical or practical considerations preclude RCTs. “Natural experiments”\(^ {45}\) can sometimes be used, for example, in comparing health outcomes in an area that has adopted a particular policy or legislative measure with a similar area that has not done so. Often, however, such an opportunity does not arise at the right time to evaluate the proposed intervention prior to implementation in a given setting, resulting in an after-the-fact evaluation. This can provide useful information about whether or not the intervention achieved its desired goals, but, due to the time sequence, cannot be used to inform that particular policy decision. A range of approaches must be considered, and the best approach will depend on the particular research question and the social context. One approach is to conduct observational assessment of the health effects of voluntary health behavior change of the same nature as would be encouraged or required by the proposed policy intervention. Another approach is to assess the magnitude of association between a particular risk factor and a given outcome to inform consideration of what the beneficial effect of introducing a policy to eliminate (or reduce) this risk factor may be. However, natural experiments can also play a useful role alongside evidence from other study designs. For example, the results of selected local implementation of a policy can be compared to other locations without the policy change. This can help inform the potential adaptation of the policy in other localities. Nevertheless, such an approach is not always possible, because it relies on the policy change being adopted in a comparable location within a reasonable temporal window. There are a range of specific designs that can be taken to a natural experiment, and while some require a long evaluation period, others such as the regression discontinuity design\(^ {46}\) do not.

Moreover, while evidence should be an important factor in guiding public health practice and policy, there are other important considerations to bear in mind. One example is the “precautionary principle,” which may lead to measures that appear to exceed the current evidence, to provide early measures to avoid and mitigate health hazards, even in the context of uncertainty.\(^ {47,48}\) Moreover, the prevailing political climate can play an important role. Perceptions of risk, which are central to public health, are culturally bound.\(^ {49}\) Moreover, political features such as social intervention and the welfare state have been shown to have a positive effect on population health,\(^ {50}\) but are associated with left wing politics,\(^ {51}\) that is not always popular in the modern world.\(^ {52}\)

Here, we shall briefly mention a couple of examples of how observational research has helped inform public health interventions in real-life policy contexts. A study led from Aberdeen by Semple et al\(^ {53}\) on observed particulate matter...
exposure in cars depending on whether or not tobacco smoking took place during the journey was cited, alongside other observational studies, in a Scottish Government consultation document for a Private Members Bill by MSP Jim Hume. In turn, this Bill ultimately led to a prohibition of smoking in a motor vehicle carrying a child under the Smoking Prohibition (Children in Motor Vehicles) (Scotland) Act 2016. Meanwhile, the evidence base has continued to become stronger and now includes a systematic review by Raoof et al. Meanwhile, following a synergy of a mass petition led by Nicola Thorp, research including a systematic review of observational studies by Barnish and Barnish and the coverage of both of these in the mass media, such as an influential article by Samuel, the UK Government has issued a clarification that forcing women to wear footwear that is detrimental to their health at work is not in accordance with the Equality Act (2010), and discussions are underway as to how to address remaining issues of this nature.

Conclusion
Although the situation may slowly be changing, classical RCTs retain a privileged place in the hierarchy of evidence for the evaluation of health care and other health-related interventions. However, these studies, although highly regarded by regulatory authorities, are limited in terms of their generalizability and, therefore, direct applicability to routine clinical practice. In this review, we have provided an overview of a range of ways in which pragmatic and observational studies can contribute to our understanding of the real-life effectiveness of interventions. We have also seen how advances, for example in statistical methodology, can help overcome some of the issues previously seen to characterize such approaches. For example, pragmatic RCTs are useful to assess whether a particular intervention is effective in the far-from-ideal conditions that characterize routine practice. Registry observational studies offer invaluable insight into the real-life effectiveness of an intervention following implementation, as opposed to in a simulated situation, especially if ethical requirements allow for an unselected population. The evaluation of public health interventions poses its own unique challenges in terms of practicalities and ethics. In some situations, natural experiments are available. However, when they are not, observational studies can be the only approach available to inform the choice of intervention. In conclusion, although classical RCTs do have a valid place in the evaluation of health care interventions, our understanding of how well these interventions will work in routine practice would likely be enhanced if there was a greater focus on triangulation to integrate evidence from different perspectives. Pragmatic and observational studies have an important, if still somewhat underrecognized, role to play in evaluating health care interventions, to ensure that they are fit for purpose in routine clinical practice settings.

Acknowledgment
There is no specific funding to declare for this review article.

Disclosure
The authors report no conflicts of interest in this work.

References
57. Samuel J. High heels aren’t glamorous, they are physically damaging and requiring women to wear them is sick. *Daily Telegraph*. 2016 12 May.