

Evaluating health system interventions: A comparison of different methods

Michael 2025; 22: 26–31.

doi: 10.56175/Michael.12579

Findings from randomized controlled trials are usually considered more trustworthy than those from non-randomized studies. However, trials can be difficult or even impossible to conduct for some interventions. The interrupted time-series design is one alternative approach. The aim of this project was to explore how much an interrupted time-series analysis is likely to yield results that differ from a randomized trial in health policy evaluation.

We re-analyzed several randomized trials by applying an interrupted time-series analysis on the intervention arm only (single arm design) and then compared the results with the conventional trial analysis. The comparisons showed that excluding control group data can lead to erroneous conclusions about intervention effects, but the findings from the single-arm interrupted time series analyses were mostly consistent with the randomized trial analysis.

The randomized controlled trial (RCT) is widely regarded as the gold standard method for measuring the impacts of interventions. In clinical medicine, RCTs dominate effectiveness research. Pharmaceutical products, for example, are rarely approved for marketing without prior evaluation in RCTs.

The field of health policy evaluation is different: Randomized experiments are seldom carried out before, during, or after a policy has been implemented, even when it is highly uncertain whether the policy will have the anticipated effects.

Several factors contribute to the lack of RCTs in health policy evaluation (1), including:

- Practical barriers to limit the intervention to only one part of the population, e.g., when evaluating a mass-media campaign or a new taxation scheme.

- Practical barriers to randomly assign people or geographical areas to intervention and control groups, e.g., when implementation has already begun before an evaluation was planned.
- Political resistance to allow a “lottery” to decide who receives the intervention and who does not.
- Lack of interest in learning about the impact of the policy intervention.
- Legal and ethical barriers, e.g., requiring informed consent — a requirement that rarely is feasible when groups of people (e.g., communities) must be assigned to study arms.

Partly due to these practical challenges, other approaches are being proposed and promoted as alternatives. The simplest method is to observe change from before to after policy implementation. However, since other factors beyond the policy may influence observed change (or lack thereof), this is generally considered a weak method (2).

An improvement of this approach is the interrupted time-series (ITS) design, where several measurements before and after policy implementation help identify trends that might be missed with a simple pre-post comparison. ITS is widely recommended for impact evaluation of policies and is promoted as “the strongest quasi-experimental approach for evaluating longitudinal effects of interventions” (3).

It is well known that different study designs provide different levels of evidence. Thus, evaluation methods can be placed within a hierarchy of evidence strength. Multiple frameworks categorize study designs by reliability. In most, evidence from RCTs is ranked above findings from non-randomized studies. Similarly, many systematic reviews prepared through the Cochrane Collaboration only include RCTs. However, some reviews justify including non-randomized studies, especially for interventions that “cannot be randomized, or that are extremely unlikely to be studied in randomized trials” (4).

Assuming that ITS results are less reliable than those from RCTs, how confident can we be in their validity? Are ITS results likely to align with RCT findings, or is there a significant risk of bias?

While randomization effectively reduces allocation bias, most research on this topic has focused on clinical trials rather than health system and policy interventions (5). Studies comparing RCTs with non-randomized designs often group many different study types together, which may be misleading, as not all non-randomized methods introduce the same level of bias.

Due to a lack of empirical data, debates about study design reliability are largely theoretical, especially regarding ITS.

Thus, more empirical research is needed to inform the debate on evaluation designs. Our aim was to explore how the choice of evaluation design influences findings, specifically whether an interrupted time-series analysis is likely to yield results that differ from an RCT.

Methods

This study was inspired by a cluster-randomized trial of a quality improvement intervention in Norway, where groups of primary care physicians (working in the same practice) were randomized to receive the intervention (6). The trial took longer than anticipated, and by the time results were published, their usefulness was reduced. Also, the study could have been conducted faster using a non-randomized approach, raising the question: Would the results have differed with a less rigorous but more efficient method?

Since our trial data were suitable, we conducted a retrospective ITS analysis using data only from the intervention arm and compared the results with cluster-RCT results. We also identified additional trials for similar analysis.

More specifically, we applied segmented regression analysis, estimating two key effects:

1. Change in level
2. Change in trend (slope) before and after the intervention

To compare these with the RCT estimates, we modeled the ITS effect size halfway through the post-intervention period (i.e., the difference between pre- and post-intervention regression lines at the midpoint, see Figure 1).

Findings

For the Norwegian trial, we found that the ITS result produced a somewhat higher effect estimate (12% vs. 9% in the RCT), but the ITS estimate was within the 95% confidence interval of the RCT estimate. Thus, we concluded that, in this case, the ITS analysis provided a reliable effect estimate (7).

We identified eight additional cluster-RCTs of health system interventions where the authors were willing to share their data for ITS analysis. The findings were largely—but not always—concordant with the RCT results, leading us to conclude that while “failure to use control groups can sometimes lead to erroneous conclusions about intervention effects, the

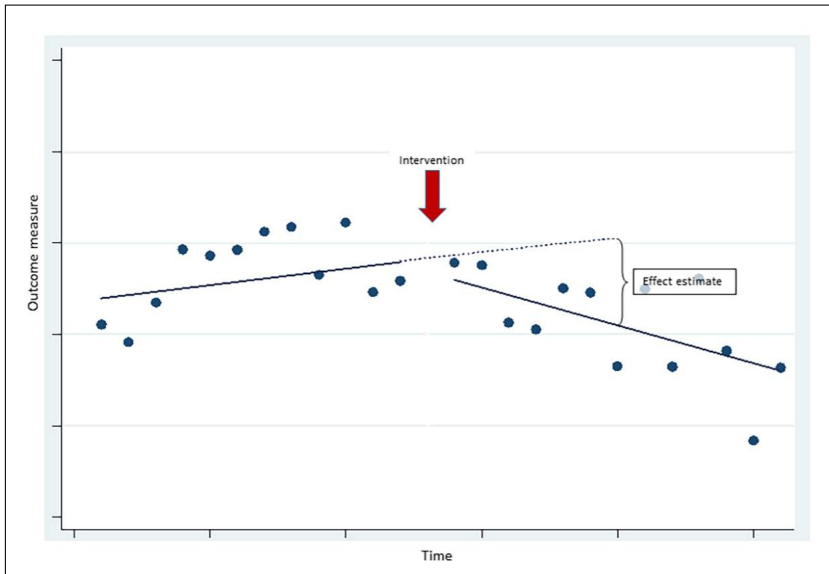


Figure 1. Illustration of how level and trend changes were combined in one effect estimate: the difference between the level of the preintervention regression line and the postintervention line halfway through the postintervention period (from reference 8).

single-arm ITS design, where the pre-intervention period serves as a control, produced findings that were mostly consistent with controlled analyses” (8).

Additionally, our access to time-series data from several RCTs enabled us to explore how results would be influenced by incorporating time trends in both intervention and control groups in RCT analyses. Again, we found that the results were mostly concordant, but not always, leading us to conclude: “If data from RCTs is analyzed without taking into account trends over time, the findings can sometimes be misleading” (8).

Impact of the Project

Both study reports (7, 8) have been widely cited, with 96 and 126 citations, respectively, according to Google Scholar (as of March 17, 2025), including in a Cochrane systematic review (9), indicating that the findings have had an impact. A review of these citations suggests that the impact is primarily methodological, influencing ITS approaches in health policy and public health evaluation (e.g., reference 10).

Impact of my Harkness Fellowship

The impact of the Harkness Fellowship extended beyond this project, fostering lasting collaborations with colleagues in various research initiatives.

The way healthcare services are organized in the United States stands in stark contrast to the universal, tax-funded healthcare system in Norway and other European countries. Experiencing U.S. healthcare services offered insight into how the Norwegian system could evolve, with an increasing share of services paid for by private health insurance. This experience also highlighted the potential consequences if Norway's current system were to weaken or fail.

The Harkness Fellowship gave me a unique opportunity to engage with experts on the principles behind various healthcare delivery models, including policy advisors with perspectives vastly different from my own. It became very clear that ideological positions strongly impact views on health policy. For example, policies designed to ensure broad access to healthcare are seen by some as undue interference with individual rights and responsibilities, especially if tax-based funding is involved.

Exposure to ideological perspectives uncommon in Europe was both refreshing and challenging, offering early insight into the neoliberal wave at the time (2011–2012), which eventually evolved into the rise of Trumpism.

Future Research or Policy Work

In my current position as head of the Centre for Epidemic Interventions Research (CEIR), a key objective is to strengthen the evidence base for decision-makers selecting interventions in crisis situations, e.g., pandemics.

CEIR was recently designated a WHO Collaborating Centre for effectiveness research on public health and social measures in health emergencies. This entails conducting research that extends beyond the Norwegian setting. The centre also aims to improve public understanding of intervention effectiveness, e.g., regarding vaccines.

Equity considerations play a central role, ensuring that benefits from implemented interventions reach all population groups equitably.

Literature

1. Gopinathan U, Elgersma I, Dalsbø T et al. Strengthening research preparedness for crises: lessons from Norwegian government agencies in using randomized trials and quasi-experimental methods to evaluate public policy interventions. *Health Res Policy Syst* 2025; 23: 8. doi: 10.1186/s12961-024-01271-y.

2. Fretheim A, Oxman AD, Lavis JN et al. SUPPORT tools for evidence-informed policymaking in health 18: Planning monitoring and evaluation of policies. *Health Res Policy Syst* 2009; 7: Suppl 1: S18.
3. Wagner AK, Soumerai SB, Zhang F et al. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther* 2002; 27: 299–309.
4. Reeves BC, Deeks JJ, Julian Higgins JPT et al. Including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J et al., editors. *Cochrane handbook for systematic reviews of interventions*. Version 6.4. Cochrane; 2023. Available from: <https://training.cochrane.org/handbook>
5. Odgaard-Jensen J, Vist G, Timmer A et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database of Systematic Reviews* 2011; Issue 4.
6. Fretheim A, Oxman AD, Håvelsrud K et al. Rational prescribing in primary care (RaPP): a cluster randomized trial of a tailored intervention. *PLoS Med*. 2006 ; 3: e134. doi: 10.1371/journal.pmed.0030134.
7. Fretheim A, Soumerai SB, Zhang F et al. Interrupted time-series analysis yielded an effect estimate concordant with the cluster-randomized controlled trial result. *J Clin Epidemiol* 2013 66: 883–887. doi: 0.1016/j.jclinepi.2013.03.016.
8. Fretheim A, Zhang F, Ross-Degnan D et al. A reanalysis of cluster randomized trials showed interrupted time-series studies were valuable in health system evaluation. *J Clin Epidemiol* 2015; 68: 324–333. doi: 10.1016/j.jclinepi.2014.10.003.
9. Gaitonde R, Oxman AD, Okebukola PO et al. Interventions to reduce corruption in the health sector. *Cochrane Database Syst Rev* 2016 CD008856. doi: 10.1002/14651858.CD008856.pub2
10. Bernal JL, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. *International Journal of Epidemiology*, 2018; 47: 2082–2093, <https://doi.org/10.1093/ije/dyy135>

Atle Fretheim

Centre for Epidemic Interventions Research (CEIR)

Norwegian Institute of Public Health

PO Box 222 Skoyen, 0277 Oslo

Norway

Atle Fretheim is Research Director at the Norwegian Institute of Public Health and the founding director of the Centre for Epidemic Interventions Research (CEIR), established at the Norwegian Institute of Public Health in 2021. He also holds the position of Adjunct Professor at Oslo Metropolitan University.