The use of quasi-experimental designs for vaccine evaluation

James A Lopez Bernal¹,³
Nick Andrews²
Gayatri Amirthalingam¹

1. Immunisation, Hepatitis and Blood Safety Department, Public Health England, London, UK
3. Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

Corresponding author: James Lopez-Bernal, Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK, james.lopez-bernal@lshtm.ac.uk

Summary: This article reviews and critically appraises the use of quasi-experimental designs for the evaluation of vaccines. We highlight where quasi-experimental designs can be applied, provide examples and discuss the relative strengths and limitations of each design.
Abstract
Randomised control trials are not always possible to evaluate interventions targeting infectious disease. This is frequently the case when evaluating the population level impact of vaccines or when evaluating interventions aiming to increase vaccine uptake. Under such circumstances an array of quasi-experimental designs is increasingly being used to evaluate the effect of vaccines on a wide range of morbidity and health service outcomes. These studies can provide valuable information on the impact of vaccination programmes and other related interventions in real world settings. Nevertheless, not all quasi-experimental designs are equal and it is important that authors and readers are aware of their relative strengths and potential sources of bias. In this paper we discuss what a quasi-experimental design is, when they might be used for vaccine evaluation, their strengths and limitations, and examples of their application.

Keywords:
Quasi-experimental designs
Immunisations
Evaluation
Interrupted time series
Regression discontinuity
Introduction
During their development, licensure and surveillance, vaccines undergo a series of phase I to IV randomised controlled trials (RCTs) in order to assess their safety and efficacy against the targeted infectious diseases.[1-4] While RCTs have strong internal validity and provide the most robust evidence on the direct biological effects of the vaccine at the individual level, they cannot examine overall population level effects of a vaccination programme.[5] Cluster RCTs (CRCTs) can be used to measure overall effects nevertheless both RCTs and CRCTs are frequently not feasible or not appropriate. Examples include evaluations of vaccines that have already been implemented, when there are cost or time constraints or when the outcome of interest is rare and large efficacy trials are not possible.[5]. Under such circumstances an array of quasi-experimental designs are increasingly being used to evaluate the effect of vaccines on a wide range of outcomes.[6-8] These studies can provide valuable information on the overall effect (or impact) of vaccination programmes in real world settings, vaccine safety, the effectiveness of interventions to increase vaccine uptake, and the associations between vaccination and both morbidity and health service indicators.[6, 7, 9-11] Nevertheless, not all quasi-experimental designs are equal and it is important that authors and readers are aware of their relative strengths and potential sources of bias. In this paper we discuss what a quasi-experimental design is, when they might be used for vaccine evaluation, their strengths and limitations, and examples of their application. Although we focus on vaccination, the methods described can equally be applied to evaluations of other infectious disease interventions that act at a population level.[12-14]

What is a quasi-experimental design?
All evaluations require a comparison between what happened and what would have happened in the absence of the intervention. The latter is known as the counterfactual. Evidently, it is not possible to observe the intervention both being implemented and not being implemented in the same population at the same time, therefore the true counterfactual is unknown. Evaluation design centres on creating the best approximation of the true counterfactual.[15] RCTs and CRCTs achieve this by randomly assigning individuals or population to receive the intervention or to a control group, with the control group representing the counterfactual. Random assignment has strong internal validity as both known and unknown confounders will, on average, be evenly distributed between the two populations, therefore they would be expected to act in the same way under the same treatment option.[16]

Shadish et al define quasi-experiments as “experiments that lack random assignment of units to conditions but that otherwise have similar purposes and structural attributes to randomised experiments”. [15] Whereas in experimental designs, the participants are actively assigned to either the intervention or control group, quasi-experimental methods take advantage of exogenous sources of assignment to the intervention, this is known as a ‘natural experiment’. [17, 18] For example, a vaccine may be introduced in one place but not another, creating a natural control group which can approximate the counterfactual. Alternatively, the vaccine may be introduced at a known time point allowing a pre-post comparison in which pre-intervention observations are used to approximate the counterfactual. Finally, the vaccine may be introduced above or below a given threshold, for example an age threshold. Evaluators can, again, take advantage of this threshold assignment to approximate a counterfactual. Importantly, in each of these situations the participant
does not choose whether they are eligible for the vaccine. Therefore volunteer selection bias is not an issue.

Studying the effects of vaccines:
Halloran et al 2010 define four different types of effects which may be studied when evaluating the population-level effects of a vaccine.[5] Under the assumption that there exists one intervention population which is eligible for the vaccine and another control population which is ineligible and that the ineligible population is unaffected by changes to infection dynamics within the eligible group: 1) The direct effect of the vaccine is the incidence of disease (or other outcome measure) among those vaccinated in the eligible population compared to those unvaccinated in the eligible population. 2) The indirect effect of the vaccine is the incidence among those who are unvaccinated in the eligible population compared to the incidence in the ineligible population. 3) The total effect is the incidence of disease among those vaccinated within the eligible population compared to the incidence in the ineligible population. 4) The overall effect is the incidence within the eligible population as a whole compared to the incidence in the ineligible population.[5, 19] The direct effect can be used to measure vaccine efficacy, however, it does not take into account levels of uptake or herd immunity effects. The indirect effect includes herd immunity whereby there is protection of unvaccinated individuals due to a reduced probability that they will come into contact with an infected individual.[20, 21] The total effect takes into account both direct and indirect effects, however and assesses the total benefit of vaccination to vaccine recipients, it does not incorporate real world levels of vaccine uptake. Finally, the overall effect examines the population level impact taking into account both levels of vaccine uptake and indirect effects due to herd immunity, this is a measure of the public health benefit of vaccination.[22]

RCTs, cohort studies and case-control studies can be used to examine the direct effects of a vaccine. Cluster RCTs, conversely, are used to evaluate the overall, total or indirect effects.[5] Quasi-experimental designs are typically compare eligible to ineligible populations and are thus used to evaluate the overall effects of vaccination programmes. They may also be used to evaluate indirect and total effects if there is suitable data on the vaccination status of individuals within the eligible cohort.[22]

When might you use a quasi-experimental design to evaluate a vaccine?
There are a broad range of scenarios in which quasi-experimental designs may be used for vaccine evaluation. Most obviously, if a vaccine has already been implemented a RCT or CRCT will not be possible. Frequently, we are interested in outcomes within a specific population or outcomes that may not have been evaluated during the initial trials. For example, the impact of vaccination on health service utilisation, on possible adverse events or on rates of diseases potentially associated with the targeted infection.[9, 23, 24] The latter can provide evidence on possible infectious aetiologies of chronic disease. Furthermore, there are circumstances in which it would be unethical to withhold a vaccine from one group or where a RCT will not be possible due to time constraints, for example, in the evaluation of seasonal influenza vaccinations or in epidemic situations.[25] Other observational studies which compare individuals who have been vaccinated to those who have not
can be used to assess the direct effect of vaccines in these situations (such as cohort or case-control studies). Nevertheless, individuals who are willing and able to get vaccinated may differ systematically from those who are not, for example healthy vaccinee effects and frailty bias, and it is often difficult or impossible to disaggregate the effects of these differences from those of the vaccine.[26] As quasi-experimental designs rely on some exogenous assignment to vaccine eligibility, they avoid these sources of bias.

Quasi-experimental designs can also be used to examine the effect of population level interventions on rates of vaccine uptake (for example, promotional campaigns or media scares). Finally, individual level observational studies require good data on an individual’s vaccination status, conversely, in quasi-experimental designs because exposure is based on eligibility for vaccination rather than vaccination status, they can be used in situations where data on vaccination status is poor. This may be the case when using routine data sources, for example, in UK primary care data from the Clinical Practice Research Datalink (CPRD), only around 40% of girls had a record of HPV vaccination, whereas in reality coverage is known to be much higher than this.[23]

Quasi-experimental designs
There are a broad range of quasi-experimental designs from relatively weak cross-sectional or pre-post studies to strong designs such as controlled interrupted time series and regression discontinuity designs that have consistently been found to closely approximate the findings of RCTs.[27-30] We discuss the features of the main categories of quasi-experimental designs and classify them into a hierarchy. The designs are represented graphically in Figure 1.

Cross-sectional controlled designs:
In a cross-sectional design (also known as a post-test only controlled design), one group receives the intervention, another does not (the latter thereby acts as the counterfactual). Here the exogenous assignment variable depends on how the vaccine was implemented, typically one age group is eligible but another is not, it may also be that the vaccine was introduced in one area but not another. Outcomes after the intervention are compared between the two groups (Figure 1a).[14, 15] For example, the childhood live attenuated influenza vaccine was initially introduced in England as a pilot programme to selected geographically discrete areas. As part of their analysis Pebody et al (2015) compared rates of consultations for influenza like illness and cases of laboratory confirmed influenza in the pilot areas to non-pilot areas with reductions demonstrated in primary school vaccinated pilot areas.[31] While cross-sectional studies can give an early indication of associations, they are generally weak designs for evaluating the effectiveness of interventions. Because there are no baseline observations, it is not possible to know whether one group had higher or lower rates of the outcome even prior to the intervention. Furthermore, unknown or unmeasured confounding factors may explain population differences in the outcome. Although cross-sectional designs can be strengthened by including multiple control and intervention populations, they should generally only be used if no pre-intervention data is available and the intervention cannot be randomised. Fortunately, routine data on vaccine uptake as well as many morbidity and health service outcomes are increasingly available, therefore including pre-intervention data is normally possible.
Uncontrolled pre-post designs

In uncontrolled pre-post designs observations following an intervention are compared to baseline observations from before the intervention. The intervention effect is calculated as the difference between the pre- and post-intervention observations (figure 1b). Here the exogenous assignment is time whereby the population prior to a certain time point does not receive the intervention, whereas the population after a certain time point does. For example, Leitmeyer et al (2006) examined influenza vaccine uptake among healthcare workers before and after a nationwide awareness campaign.[11] Unlike in cross-sectional studies, pre-post designs include a baseline measure, therefore the change can be measured in the intervention group. However, they do not include a control group. The main limitation is that it is assumed that the outcome would remain static in the absence of an intervention. In reality, data may be highly variable or may follow a pre-existing trend. Leitmeyer et al found that vaccine uptake increased from 21% to 26% following the campaign, however, while this increase could have been due to the intervention, it could also have been due to a pre-existing increasing trend in vaccine uptake, random variability in uptake from one year to the next or regression to the mean following a year of particularly low uptake.[11] The pre-post design is also unable to control for other events or interventions that may influence vaccine uptake in between the pre-intervention observations and the post-intervention observations.[5] This design can be strengthened by increasing the number of pre-intervention observations which decreases the likelihood of findings being due to random variability or regression to the mean.[25] For example, Partinen et al 2012, found an association between the 2009 H1N1 pandemic vaccine and narcolepsy by looking at rates of narcolepsy in 2010 compared to the average rates from 2002-2009.[32]

Controlled pre-post designs

The controlled pre-post design combines the characteristics of the two previous designs by including both a control group and pre-intervention observations. For example, Donegan et al (2013) compared changes in rates of chronic fatigue syndrome before and after the introduction of the bivalent human papillomavirus (HPV) vaccine in girls to the change in boys (who were not eligible) with no difference seen.[23] Similarly, Parikh et al (2016) used a controlled pre-post design as part of their evaluation of the multicomponent group B meningococcal vaccine. They examined the change in the number of cases of group B meningococcal disease in the vaccination year compared to the average number in previous years in eligible versus non-eligible age cohorts and demonstrated a significant 42% reduction.[33] The effect of the intervention is estimated as the difference between the change in the outcome in the intervention group and the change in the outcome in the control group, the design is therefore also known as a difference in difference design (figure 1c). The change in the outcome in the control group acts as the counterfactual and accounts for possible changes over time unrelated to the intervention. Nevertheless, this design implicitly assumes that in the absence of an intervention both the intervention and control groups would follow parallel trends. This is not always the case, for example if one group has especially high incidence of disease in one year they may be targeted for vaccination, yet this high incidence may have been a chance outlier that would have reduced the following year simply due to regression to the mean. Conversely, the trend in an untargeted control group with no unusually high incidence of disease is more likely to
remain stable. The design can be strengthened by increasing the number of pre-intervention observations.[25]

Interrupted time series (ITS)
In an interrupted time series study multiple pre- and post-intervention observations are used to examine the change in the trend of the outcome following the intervention (Figure 1d).[34] As with the pre-post design, the exogenous assignment variable is time. The impact can be assessed by examining the change in the level of the trend immediately after the intervention or at other time points, or by examining a change in the slope of the trend after the intervention. The specific impact model should be selected a priori based on specific intervention and outcome under study.[35] For example Grijalva et al (2007) examined the effect of the introduction of routine infant immunisation with the seven-valent pneumococcal conjugate vaccine on admissions for pneumonia in the USA (Figure 2).[9] By including underlying trends: secular changes, data variability and regression to the mean can be controlled for. Furthermore, seasonal effects can be adjusted for. The design is particularly useful for examining the overall effect and in situations where no unaffected control group is available, which is again beneficial when there are herd effects. The main limitation of ITS studies is that they cannot exclude the effects of other events or interventions occurring at the same time as the intervention, known as history bias. Furthermore, as with all designs that rely on historical observations, there can be instrumentation effects if there are changes to the way the outcome is measured or recorded over time.[34]

Controlled interrupted time series (CITS)
The controlled interrupted time series adds a control group, that did not receive the intervention, to the uncontrolled ITS design (Figure 1e). Many different types of controls are possible, including locations where the vaccination was not introduced or non-eligible age cohorts.[36] For example Amirthalingam et al (2017) examined the impact of the herpes zoster vaccination programme in adults aged 70-79 years by comparing time series in eligible and non-eligible age cohorts (Figure 3).[7] Parikh et al (2016) also used this design to analyse the impact of meningococcal group B vaccine in addition to their use of the controlled pre-post method (described above). They again used non-eligible age-cohorts as their control and found the impact changed from a reduction of 42% to 36%, suggesting that by incorporating the underlying trend the effect was slightly attenuated.[33] This design has the same strengths as the ITS design by controlling for underlying trends, but including a control series also helps to exclude other events or interventions occurring around the time of the intervention that could have affected both groups. Confounding factors in a CITS design would have to occur at the same time as the intervention and only affect the intervention group which is unlikely in many situations. Control groups should be unaffected by the intervention. It is therefore important that indirect effects do not apply to the control groups as this would lead to underestimation of the intervention effect (because an effect will be seen in both the intervention group and the control group). The design can be strengthened by including multiple controls, especially if different groups receive the intervention at different times (known as a multiple baseline design) as any confounding events would have to occur in multiple groups at multiple different times.[37] Amirthalingam et al (2017) in fact used multiple different age cohort
controls and took advantage of the phased implementation of the programme for the eligible cohorts between 70-79 years in a multiple baseline design.[7]

Regression discontinuity design (RDD)
Whereas cross-sectional designs define eligible and non-eligible as two discrete groups, the regression discontinuity design takes advantage of the way an intervention is allocated according to some continuous variable (the ‘assignment variable’) and examines the change in the outcome at the point where individuals become eligible (Figure 1f).[38] This could be applied for vaccines given above or below a certain age or before or after a certain date of birth. For example, there were concerns in Canada that low HPV vaccine uptake may be due to a perception among parents that the vaccination increased risky sexual behaviours. Smith et al (2015) used a RDD based on the date of birth that girls became eligible for HPV vaccination to look at the impact on indicators of risky sexual behaviour.[8] In another example, Dykstra et al (2015) looked at the impact of funding by the Gavi Vaccine Alliance on levels of national vaccine uptake for multiple vaccines, including Haemophilus influenza type B (Figure 4). Here countries with a gross national income (GNI) below $1,000 per capita in 1998 were eligible for funding, whereas those with a GNI above $1,000 were not. The effect of funding can therefore be estimated by the change in uptake at the point at which countries become eligible for funding.[39] This is a powerful design as confounding is very unlikely and at the cut-off point it is essentially random whether an individual (or in the latter example, a country) is eligible or not eligible for the intervention. The only confounding factors are other interventions or events that are assigned according to the same threshold as the intervention under study. As an example, another possible application of RDD would be to examine the effect of the seasonal influenza vaccine on respiratory admissions, however, in the UK both the influenza vaccine and the pneumococcal vaccine are given from age 65 therefore pneumococcal vaccination could be a confounder. Ideally, the ineligible group should be unaffected by the treatment of the eligible group. This may not be the case in with vaccines that may have indirect effects, Aronow et al describe approaches to using RDD when such ‘interference’ exists.[40]

Discussion
As we have highlighted quasi-experimental designs have a broad range of applications in the evaluation of vaccines. These include evaluating the overall effect of the vaccine and evaluating the impact of different interventions on vaccine uptake. These designs also be applied for the evaluation of other infectious disease interventions that act at a population level.[12-14] We have discussed the strengths and limitations of each design and in table 1 we present a hierarchy.[41] It is important to note that different designs may be most appropriate for different situations depending on the nature of the outcome, the data available, the availability of controls and the way the intervention is implemented. For example having pre-intervention trends generally strengthens the design, however, it is less useful for outcomes that are highly variable. Furthermore, it is important to consider how far back trends can be extrapolated from, for example if there are substantial secular changes in disease incidence.[35] Nevertheless, in general RDD and CITS can be regarded as the most powerful designs, ITS and controlled pre-post designs as intermediate designs and uncontrolled pre-post and cross-sectional designs as relatively weak designs.
Researchers should also be aware that while we have classified these designs as discrete, different designs can frequently be combined to provide greater strength. For example, while we classified the herpes zoster study by Amirthalingam et al 2017 as a controlled interrupted time series or multiple baseline design, the authors also included age as an assignment variable so it can therefore also be considered a regression discontinuity design.[7]

CRCTs are the gold standard design for evaluating overall effects of vaccines, nevertheless they are not always possible. Under such circumstances, the most powerful quasi-experimental designs can provide strong evidence on the overall effects of vaccines and related interventions. Despite their strengths, the potential for bias and confounding is greater in quasi-experimental studies than CRCTs. In quasi-experimental designs that use a control group, confounding due to population differences can be particularly problematic, in pre-post designs time-varying confounders, history bias and instrumentation effects can threaten validity. A range of approaches exist to deal with some of these threats, including inverse probability weighting, synthetic controls, propensity score matching and control outcomes.[15, 34-36, 38, 42-45]

It is therefore important that researchers have a clear understanding of the threats to validity of evaluative study designs. In this paper we have described the main quasi-experimental designs and highlighted their strengths and limitations. We would encourage researchers to focus on using the strongest of these designs possible as well as comparing the results of the designs when more than one can be used in order to provide robust evidence on the impact of vaccines and associated interventions.
Funding
Public Health England
References

21. Scarbrough Lefebvre CD, Terlinden A, Standaert B. Dissecting the indirect effects caused by vaccines into the basic elements. Human Vaccines & Immunotherapeutics 2015; 11(9): 2142-57.


## Table 1: Hierarchy of quasi-experimental designs

<table>
<thead>
<tr>
<th><strong>Strong designs</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression discontinuity design</td>
<td>Only confounded by other interventions or events if they are implemented according to the same threshold</td>
</tr>
<tr>
<td>Controlled interrupted time series</td>
<td>Control for pre-existing trends and other events/interventions affecting both groups over time. Only confounded by other interventions or events not captured by the trend that effect groups differently. Further strengthened when trends are the same in both groups allowing a longer period of extrapolation post intervention.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Intermediate designs</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Single interrupted time series</td>
<td>Control for pre-existing trends and confounding/selection bias due to population differences. Cannot control for other events occurring at the same time as the intervention or changes in trends (this limits reliability of extrapolation of trends).</td>
</tr>
<tr>
<td>Controlled pre-post design (difference in difference)</td>
<td>Control for changes over time that affect both groups. Selection bias and confounding can be a threat if baseline trends differ and this cannot be assessed without multiple time points.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Weak designs</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncontrolled pre-post</td>
<td>Do not control for pre-existing trends or other factors causing changes over time.</td>
</tr>
<tr>
<td>Cross-sectional design</td>
<td>No baseline. Confounding due to population differences is a major threat.</td>
</tr>
</tbody>
</table>

Adapted from Soumerai et al[41]
Figure 1: Graphical depiction of quasi-experimental designs

a = cross-sectional controlled design, impact = difference between study group and control group ($d_1$); b = pre-post design, impact = difference between pre and post observations ($d_1$); c = controlled pre-post design, impact = difference between pre and post observations in the study group minus the difference between the pre and post observations in the study group ($d_1-d_2$); d = interrupted time series, various models possible to assess the impact e.g. step change immediately after the intervention after taking into account the underlying trend over time ($t_1$); e = controlled interrupted time series, various models possible to assess the impact e.g. step change immediately after the intervention in the study group after taking into account the underlying trend over time and above any step change in the control group ($t_1-t_2$); f = regression discontinuity design, various models possible to assess the impact e.g. step change immediately after the intervention after taking into account the underlying trend according to the assignment variable ($r_1$).

Dotted lines = counterfactuals

Figure 2: Example of an interrupted time series examining the effect of the introduction of PCV7 on all-cause pneumonia admissions (reproduced from Grijalva et al 2007 with permission)

Blue dots = observed number of admissions. Red line = deseasonalised trend.

Figure 3: Example of a controlled interrupted time series examining effect of the introduction of the HZV vaccine for 70 year olds compared to 67 year olds who were not yet eligible for vaccination (reproduced from Amirthalingam et al 2017 with permission)

Blue dots = not eligible for HZV vaccination, red dots = eligible for HZV vaccine programme

Figure 4: Example of a regression discontinuity design looking at the impact of GAVI funding on coverage of Hib vaccine (reproduced from Dykstra et al 2015 with permission)

Countries with a GNI < $1000 were eligible for GAVI funding

Solid black lines = regression line, grey lines denote the 95% confidence interval
Figure 1
Figure 3