Relation between experimental and non-experimental study designs.
HB vaccines: a case study

Tom Jefferson, Vittorio Demicheli

Abstract

Study objective—To examine the relation between experimental and non-experimental study design in vaccinology.

Design—Assessment of each study design’s capability of testing four aspects of vaccine performance, namely immunogenicity (the capacity to stimulate the immune system), duration of immunity conferred, incidence and seriousness of side effects, and number of infections prevented by vaccination.

Setting—Experimental and non-experimental studies on hepatitis B (HB) vaccines in the Cochrane Vaccines Field Database.

Results—Experimental and non-experimental vaccine study designs are frequently complementary but some aspects of vaccine quality can only be assessed by one of the types of study. More work needs to be done on the relation between study quality and its significance in terms of effect size.

Investigators have traditionally assessed aspects of vaccine quality, such as effectiveness and safety, using analytical studies. Such studies can be classified as either experimental or non-experimental (see the appendix for study definitions). Although the volume of vaccine analytical literature is notable (the journal Vaccine alone having published 231 trials on different vaccines between 1983 and 1994), so far little attention has been paid to the relation between the two broad types of study and its significance to the assessment of vaccine quality.

There are several reasons why an assessment of this relation is important in vaccinology. Firstly, vaccines are one of the most widely used preventive technologies available to protect all members of society, including the most vulnerable, from the effects of disease. For example, an estimated 1479 million doses were used in humans worldwide in 1990.

Secondly, vaccination programmes are mostly publicly funded and, as such, their quality should be carefully scrutinised and decisions to immunise whole populations made on the best available evidence, especially if law regulates such practice. Thirdly, the current array of vaccines is likely to be augmented considerably in scope and variety by new vaccines (currently at an experimental stage) that will impose a resource burden on society. Such a burden, set against the backdrop of increased healthcare costs, will necessitate discrimination of funding of preventive programmes. Discrimination should be partly based on the best available evidence of vaccine effectiveness and safety and on its economic efficiency, acceptability, and importance of the target disease (or “quality” of the vaccine in question).

Evidence of quality is therefore likely to be a most important deciding factor on whether to adopt a particular vaccination policy or not. As evidence of quality of healthcare interventions is scarce (especially that which is methodologically sound), decision makers are not in a position to afford to rely solely on experimental data and disregard information coming from non-experimental designs. Additionally, as pointed out by Black, both experimental and non-experimental designs should be considered complementary and not mutually exclusive.

In this paper we assess how this statement can be applied to vaccine evaluation designs.

Experimental and non-experimental designs: two different approaches

An experimental study is one in which the participants are exposed to the agent or putative cause because the investigator has assigned the exposure to the subject(s) on the basis of the study design. The aim of experimental studies is the creation of duplicate sets of circumstances in which only one factor that is relevant to the outcome varies, making it possible to observe the effect of variation in the factor.

Among experimental studies are the clinical trials, which are experiments with patients as subjects with the aim of evaluating one or more treatments for a disease or condition. There are four types of clinical trials and their description is in the appendix.

Non-experimental studies are those in which participants are exposed to the agent or putative cause in a natural way, as the investigator cannot control the circumstances of the exposure to the subject(s).

As in the future many new vaccines produced using different technologies are likely to be made available to humanity to prevent both infectious and non-infectious diseases, there will be an increased requirement for methods of evaluation of the effectiveness, safety, and efficiency of vaccines. Traditionally studies of vaccines have tested four aspects of vaccine performance: immunogenicity—the capacity to stimulate the immune system; duration of immunity conferred; incidence and seriousness of side effects; number of infections prevented by vaccination.

In general each of these aspects can be tested only through studies that assess quality of the
vaccine in a limited population, such as the study cohort, and generalise from the study cohort to the rest of the population and although different study designs can be expected to contribute to provide answers in different degrees, this will also be a function of the capacity to generalise from study to general population. This function in turn will depend on the capacity of the study to minimise systematic bias and detect and minimise random error.

The potential biases in experimental and non-experimental study designs can be classified as follows: selection bias, when subjects included in a study are not a random sample of all relevant people that fit the inclusion criteria. There are systematic differences between participants and non-participants in a study, allocation bias, when there is a systematic difference between experimental and control groups in a clinical trial. Performance bias, when there are differences in exposure to other factors that could affect outcomes. Attrition bias, when there is a systematic difference between those accounted for and those not accounted for at the end of the study. Detection bias, when case definition used to enter subjects into a study is related to the known cases of, for example, vaccines side effects.

Moreover, the various dimensions of the quality of vaccines must be assessed using two different perspectives, each with its strengths and weaknesses: an individual perspective—in which the protection of a vaccinated individual is taken as a measure of the vaccine efficacy; a population perspective—in which the protection of a whole population is regarded as a measure of the vaccine effectiveness.

The first perspective is similar to that of clinical studies and the evaluation can be better performed through experimental designs. Only susceptible subjects can be enrolled, random allocation and double blinding can minimise bias, the outcome measure can be chosen in a way to be independent from the setting (for instance antibodies) and size and time of the study can be calculated on a sound statistical basis.

A number of problems arise from this perspective: (1) the possibility of generalising will be low if the patients and the setting are specific (this typically happens when the efficacy of the vaccine is tested in a high risk group); (2) the measure of outcome will probably be a surrogate measure (such as a rise in antibody titres) with an undemonstrated relation to the vaccine ability to avoid the disease.

The population perspective requires both study setting and design to be within the context of the population in which the vaccine effectiveness is to be tested. This condition is more likely to be fulfilled with less rigorous experimental designs or with non-experimental studies, especially: (1) when both susceptible and non-susceptible patients are included; (2) when outcome measures are context dependent (for instance, avoided cases that are dependent on the initial incidence, level of natural immunity, vaccination coverage, individual protection, duration of immunity, level of herd immunity, etc).

Thus large population studies such as community intervention studies will be required. In these, the application of rigorous criteria of randomisation and blinding will be difficult or even impossible (for instance because allocation is made on a village cluster basis), leading to the use of observational designs where the probability of biases will obviously be higher.

**HB vaccines**

As the two design types are fundamentally different, it is necessary to analyse the nature of their complementarity to describe the strengths and weaknesses of each type. It is our intention to do this through the example of hepatitis B (HB) vaccines. The choice of HB vaccines was based on their importance (as HB vaccines target a global infectious disease of great importance to the international community both in its acute and chronic forms) and on the generalisability of the results of the study to other vaccines targeting chronic diseases of global importance.

At present HB vaccines broadly fall into two types. The first type was licensed in 1981 and is made of purified viral particles from the serum of infected persons (Plasma-Derived Vaccine or PDV); the second type of vaccine is derived from recombinant technology—that is, antigenic components produced by a yeast whose genome has been manipulated (Yeast-derived vaccines or YDV). HB vaccines are currently being used in 95 countries and the WHO is actively pursuing their worldwide adoption through their integration in the EPI schedule.\(^6\)

Tables 1 and 2 summarise the performance of the various study designs and their relation to the potential biases in relation to the four dimension of quality of vaccines.

The strongest design for immunogenicity testing is traditionally held to be the double blind, randomised controlled trial, in which participants are allocated to either receiving one or more HB vaccines or placebo, or another vaccine. In the less robust CCT, quasi-randomisation (using

---

**Table 1** Strengths (indicated by the + symbol) and weaknesses (indicated by the − symbol) of experimental (RCT, CCT, and community intervention trials) and non-experimental (cohort and case-control) study designs in relation to their objectives

<table>
<thead>
<tr>
<th>Study design</th>
<th>Immunogenicity testing</th>
<th>Duration of immunity</th>
<th>Frequent and/or short term</th>
<th>Rare and/or long term</th>
<th>HB case prevention</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCT</td>
<td>+++</td>
<td>–</td>
<td>++</td>
<td>–</td>
<td>±</td>
</tr>
<tr>
<td>CCT</td>
<td>++</td>
<td>–</td>
<td>++</td>
<td>–</td>
<td>±</td>
</tr>
<tr>
<td>Community intervention trials</td>
<td>+++</td>
<td>–</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Cohort</td>
<td>–</td>
<td>+++</td>
<td>+</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Case-control</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+++</td>
<td>+</td>
</tr>
</tbody>
</table>
methods such as alternation or allocation by date of birth) is used to allocate interventions and/or participants. Notwithstanding the primacy of the RCT as a study design to test effectiveness, empirical evidence available from the Cochrane Pregnancy and Childbirth Group meta-analyses shows that robustness of concealment of allocation (that is, protection of blinding from ascertainment) of participants to trial arms is probably of crucial importance, with studies that fall below this standard having exaggerated estimates of effect. Although no similar work has been carried out on vaccines in general and HB in particular, in a recent meta-analysis of four trials of PDV versus placebo in healthcare workers we found evidence of poor reporting quality, and perhaps more importantly, of confusion on the fundamental RCT study design. All four included trials, testing the effectiveness of PDV against placebo as judged both by the number of cases of HB prevented and a rise in antibody titre, failed to report the results of the latter outcome in the placebo arms of the four trials. It is not unreasonable to assume that the results of effectiveness testing on the basis of poorly reported trials need to be accepted with great caution. Some of the possible biases present in such trials are likely to be magnified in non-experimental designs, both cohort and case-control that have not been carefully designed. In these, selection bias (see below) is likely to occur when inadequate care has been taken in selecting both cases and controls. For example in a cohort design subjects are either self selected or selected by a physician for exposure to HB and are unlikely to be similar to the reference population in terms of outcome. Such a design may be open to the action of confounders. A good example of a confounder is the presence of factors such as social class, region's expanded programme of immunisation or BMI. Presence and effects of such confounders are likely to be a function of the size of the study and the quality of its randomisation. If they are unknown to researchers, lack of control for their presence will not be carried out. This problem should be solved by a robust randomisation process in which smokers have an equal chance of being assigned to either the intervention or the placebo arms. However, if the study is large (and cohort studies tend to be larger than their experimental counterparts), smokers and high BMI subjects are likely to be homogeneously distributed in all arms of the study. Either way, inclusion of smokers and high BMI subjects in small and poorly randomised trials will lead to a mis-estimation of the effectiveness of YDV.

### Duration of immunity

Duration of immunity is probably best tested by cohort design, not because of this design's superiority in minimising bias, but simply because the majority of trials have too limited a duration to test this crucial aspect of vaccine quality. Between 1983 and 1994 Vaccine published 59 reports of RCTs and CCTs on HB vaccines. Their mean length of study duration was 270 days (median 420 days), a period that is insufficient to draw conclusions as to the long term immunity conferred by vaccination. Additionally, protection against chronic forms of HB that manifest themselves mainly from sub-clinical syndromes several years or decades after primary infection is also practically best assessed by a longitudinal design. In these, however, a variable length of follow up can be associated with the generation of attrition bias (see below).

For example the chronic nature of HB is a powerful reason to discontinue surveillance of vaccinated cohorts. It is then conceivable that the effectiveness to prevent long term consequences of the disease, such as primary hepatocellular carcinoma, may be lost to direct observation and can only be assessed by case-control studies.

### Observation of side effects

Observation of side effects would ideally be assessed through the means of large RCTs. This is because factors associated with the reporting and severity of side effects, such as social class, are likely to be continuous sources of detection bias. Such bias can be removed only by a randomised design in which participants have equal chances to be assigned any arm. Some side effects are, however, very rare. In 1995 Bell described the occurrence of post-vaccination anaphylaxis in a cohort of pre-adolescent school children in British Columbia as part of that region's expanded programme of immunisation. Bell reported an incidence of 1 in 340 000 doses of YDV. Such rare but extremely important side effects are unlikely to be identified either by trials or by meta-analyses of such trials. Trials of HB vaccines published in Vaccine had a median arm size of 110 participants. Assuming a three dose immunisation schedule, we would need a meta-analysis of over 1000 small and homogenous trials to give certainty of observation of such an event. The alternative in the case of such rare events is once again a well designed case-control study in which controls are matched to cases for the highest possible number of variables except of course exposure, in this case HB vaccination.

### HB case prevention

HB case prevention is the last, but possibly overall the most important of the aspects of quality of HB vaccines. Although 86% of HB vaccines trials published in Vaccine between 1983 and 1994 had exclusively serological outcomes, the ultimate test of the vaccine is its ability to prevent HB morbidity burden—that is, to prevent cases in the community. Both poorly designed experimental and non-experimental...
Studies are open to performance bias when assessing this aspect of quality. For instance known factors associated with acceptance of one or more vaccine doses in healthcare workers are: social influence by other healthcare workers; knowledge of disease and vaccine. Non-acceptance of the vaccine is known to be associated with: concern about side effects; problems with access to the vaccine. Other factors related to initiating the course are: occupation (with doctors the most likely to initiate the course); increased blood exposure frequency; recent influenza vaccination.

Similar factors are also predictors of course completion. Here too possible failure to adjust design and analysis techniques for such confounders will lead to mis-estimation of vaccine effect. In addition to the presence of these possible confounders, assessment of case prevention hinges on another very important variable, that of incidence of the disease we are trying to prevent through vaccination. When the incidence of the target disease is low, large community intervention trials or cohort studies are the only studies likely to answer the question of what effect vaccination has on the incidence of the disease. An example of this can be found in the case of vaccination against anthrax, a very rare cutaneous and respiratory disease. Killed or attenuated anthrax vaccines are efficacious in preventing disease, but the bulk of the evidence comes from a community intervention trial in the former USSR Republic of Kazakh in which 157 259 persons spread in 228 sites in which anthrax was endemic took part. A total of 52 763 people were assigned to the needleless injection arm, 54 522 were assigned to the scarification arm, and 49 974 to the control arm. Such large numbers are unlikely to be practical or even necessary when assessing vaccines for less rare diseases, including HB.

Discussion

Applying the essence of Black’s discussion about the need of non-experimental studies to vaccinology, we can see a number of instances when non-experimental designs should be applied:

1. When an experiment is impossible because the aim of the evaluation is to assess the population effectiveness of a vaccine, the external validity of a trial is generally low mainly because both patients and study setting may be atypical, as in the case of very high risk healthcare workers in transplant or renal units vaccinated against HB.

2. When an experiment is not necessary: smallpox vaccination is a well known example of population effectiveness evaluation, which did not require an experimental study.

3. When an experiment is inappropriate, as in cases of a trial population not being large enough to detect the event or the outcome in question or when there is reluctance or refusal to participate, ethical objections or political and legal obstacles.

4. When the individual efficacy is to be measured in terms of infrequent adverse event (such as post-HB vaccination anaphylaxis).

5. When the interventions prevent rare events (such as primary hepatoma caused by the long term sequelae of HB infection).

We conclude that in vaccinology, experimental and non-experimental designs are frequently complementary but there are aspects of vaccine quality that can only be assessed by one of the types of study. More work needs to be done on the relation between study quality and its significance in terms of effect size.

Funding: none.
Conflicts of interest: none

Appendix

Types of experimental studies — trials

A randomised controlled trial (RCT) is any study on humans where it appears that the individuals (or other experimental units) followed up in the study were definitely or possibly assigned prospectively to one of two (or more) alternative forms of health care using random allocation.

A controlled clinical trial (CCT) is any study on humans where it appears that the individuals (or other experimental units) followed up in the study were definitely or possibly assigned prospectively to one of two (or more) alternative forms of health care using some quasi-random method of allocation (such as alternative, date of birth or case record number).

A field trial is an experiment in which subjects who have not yet got the disease and as a consequence are not patients. Typically the risk of disease in the population is small. These two characteristics make field trials expensive, as they require large numbers of people who are not administered by a central outlet such as a hospital or clinic but are in the everyday habitat (“field”).

A community intervention trial is a trial in which individual assignment or exposure is not carried out either because it is impossible or impractical. The unit of assignment is therefore a whole community, as in the case of water fluoridation to prevent dental caries.