Use of the Case-Control Approach in Vaccine Evaluation: Efficacy and Adverse Effects

L. C. Rodrigues and P. G. Smith

INTRODUCTION

The least ambiguous method with which to estimate the protective effect of a vaccine and to assess possible adverse effects associated with its use is a double-blind randomized controlled trial (RCT). Double-blindness, ensuring that neither trial participants nor those assessing vaccine effects know who has received the vaccine under evaluation and who has received the comparison product, should eliminate bias. Randomization should ensure, if the study is sufficiently large, that the vaccinated and unvaccinated groups are similar with respect to risk factors, other than vaccination, for the disease and adverse reactions under study. That is, randomization controls for both known and unknown factors that might otherwise confound the measurement of vaccine effects. If the vaccine is designed to protect from a disease for which there is currently no effective vaccine, the comparison product may be a placebo or, if a placebo injection is considered unacceptable, another vaccine which is not expected to have an impact on the endpoints the trial is designed to assess. If, however, a partially effective vaccine already exists, ethical considerations may dictate that this be used as the comparison product. In such circumstances, the assessments of efficacy and adverse effects are made relative to the existing product, and absolute estimation of effects is precluded.

Once an RCT of a vaccine has been conducted and the vaccine has been shown to be efficacious, further such trials may be considered unethical, especially after the vaccine has been introduced into public health use. However, it is often of interest to know not only whether a vaccine is effective in the context of a trial but also what its efficacy is under routine conditions, which are often less favorable than those in a trial. Estimates may also be required of efficacy in subgroups of the population and of the frequency of any adverse effects too rare to have been detected in controlled trials. In addition, sometimes vaccines are reformulated, and although the changes might be considered unlikely to diminish efficacy (e.g., as judged by induced immunologic markers of protection), reassurance is required that this is the case, and assessments using approaches that are quicker and less costly than RCTs are required.

The use of observational approaches to evaluate the impact of interventions was initially discouraged by epidemiologists (1) as being too great a departure from the “gold standard” of RCTs. While this was reasonable at a time when RCTs were not widely used (in circumstances where they should have been used), it also had the effect of discouraging alternative approaches, even in situations where RCTs were not applicable. For example, the case-control approach has been widely used to assess disease risks associated with “noninterventional” exposures, but with rare early exceptions (2), only relatively recently have case-control studies been applied in the context of vaccine evaluation.

One factor that prompted increased use of the case-control approach for vaccine evaluation was controversy regarding the efficacy of Bacillus Calmette-Guérin (BCG) vaccine against tuberculosis. By the early 1980s, BCG vaccine was recommended for routine use as a protective measure against tuberculosis in many countries, but large, well-designed RCTs had found efficacies ranging from nil to >80 percent. It became clear that undertaking another large RCT would not resolve the conundrum created by these conflicting results. Indeed, the very large trial that was set up in southern India in an attempt to answer many of the outstanding questions regarding the efficacy of BCG vaccine against tuberculosis raised more questions than it answered (3). Following the release of the early results of that trial, the World Health Organization recommended that further studies be conducted to evaluate the efficacy of BCG vaccine using the case-control approach in settings where BCG
was given routinely (4). In 1982, Smith (5) promoted the use of case-control studies for estimating the efficacy of BCG vaccine against tuberculosis and discussed key methodological issues. Clemens and Shapiro (6) also suggested that case-control and other observational designs could be used to help resolve controversies about the efficacy of pneumococcal vaccines. Since then, over 100 papers have been published using or discussing case-control studies in the context of the evaluation of vaccines.

In a case-control study, a measure of the influence of the exposure of interest on the disease of interest is obtained by comparing the proportions of exposed cases and controls. In studies of vaccine efficacy, cases have the disease the vaccine is designed to prevent, and controls are chosen to be representative of the population from which the cases were derived. Histories of vaccination are established for cases and controls, ensuring that the collection of information is done in the same way for both groups. In RCTs, randomization ensures that the intervention and comparison groups are similar, but in case-control studies the investigators have no influence over who receives the vaccine under study (since the evaluation is retrospective), and attempts must be made to control for differences between cases and controls that might influence the estimate of vaccine efficacy. Control for confounding variables, such as age or socioeconomic status, may be taken into account in both the design of a study and the analysis.

The odds ratio for vaccinated persons versus non-vaccinated persons among cases and controls gives a measure of the association between vaccination and the disease. This may be based on the McNemar formula (7), if matched controls are selected for each case, or it may be derived following conditional or unconditional logistic regression analysis if other variables need to be controlled for. For "rare" diseases, the odds ratio approximates closely the risk ratio and the rate ratio. With appropriate selection of controls, the rate ratio or the risk ratio may be estimated using case-control methods even if the disease is common (8, 9).

An example of the use of the case-control approach for vaccine evaluation is a Brazilian study of serogroup B meningococcal vaccine (10). This vaccine was licensed, but doubts were raised about its efficacy following its apparent failure to control an epidemic of meningitis. For the case-control investigation, the study population was defined as children aged 3–84 months who lived in the city of Sao Paulo, Brazil, during the mass vaccination campaigns with serogroup B meningococcal vaccine. Cases in the study were hospitalized cases of meningococcal disease identified from a hospital-based surveillance system. For each case, four neighborhood controls were selected, matched by age (for cases under age 24 months, controls were matched within 6 months of age; for cases aged ≥24 months, controls were matched within 1 year). The researchers identified controls by visiting houses, following a systematic procedure starting from the case child’s place of residence. They interviewed parents or guardians to obtain data on potentially confounding variables and inspected vaccination cards to determine vaccination status. A child was deemed to have been vaccinated if his or her vaccination card showed a meningococcal vaccine stamp for each of two doses, with an interval between the doses of 40 days–12 weeks, and more than 21 days had elapsed between the last dose and the onset of disease (in the matched case for controls). Children were considered unvaccinated if they had a card with no record of meningococcal B vaccination and the parent or guardian reported no vaccination. Children without cards, children with cards that showed an incomplete vaccination history, and children for whom there was conflicting information between the card and the information given by the person interviewed were excluded from the analysis. Conditional logistic regression methods were used to adjust for confounding variables and to estimate the vaccine efficacy and confidence intervals.

The case-control design has also been used, though much less widely, to study adverse effects of vaccination. The size of an RCT of a vaccine is usually determined on the basis of the statistical power needed to detect a protective effect of a given size. An important aspect of such trials is their ability to enable an unbiased assessment to be made of adverse effects that may be associated with vaccination. However, such trials may not have sufficient power to identify rare but important adverse effects of vaccination that may become apparent, or be suspected, only when the vaccine is applied on a widespread basis. When a potential adverse effect has been thus identified, usually initially by case reports, the association between the effect and vaccination can be studied using case-control methodology.

An early case-control investigation of possible adverse effects following vaccination in the United Kingdom was the National Childhood Encephalopathy Study (NCES), a study of neurologic illnesses in early childhood and their relation to immunization with pertussis vaccine (11). The initial hypothesis with respect to the forms of neurologic illness pertussis vaccination might induce was not well defined, and thus a study encompassing a range of such illnesses had to be set up. National surveillance for the defined conditions was instituted through contacts with consultant pediatricians, infectious disease specialists, and neurosur-

*Epidemiol Rev* Vol. 21, No. 1, 1999
geons, who were asked to notify the investigators of all children aged 2–24 months who were admitted to hospital with these conditions during a 3-year period. Over 1,000 such cases were reported. Children with a neurologic abnormality upon discharge were examined at home by a study physician. Two controls were selected for each case, matched for age, sex, and area of residence, using either birth registers or lists of children eligible for immunization. Vaccination histories of cases and controls were obtained from general practitioner and community physician records. Controls were also visited at home for collection of information on risk factors that might have predisposed them to neurologic illnesses.

Analysis focused on the comparison between cases and controls of the proportion vaccinated and the time interval between vaccination and the estimated date of clinical onset of the neurologic illness (using a “pseudo-” date of onset for controls corresponding to that of the matched case). The investigators reported a significantly increased risk within 7 days of diphtheria-pertussis-tetanus (DPT) vaccination (3.5 percent of cases and 1.7 percent of controls) but not for diphtheria-tetanus vaccination. The risk associated with DPT was greatest during the 72 hours following vaccination. When analyses were confined to children who had been “normal” prior to vaccination (e.g., excluding children with a history of fits, which was a contraindication for DPT vaccination), the relative risks were larger. The authors concluded that serious neurologic reactions may occur very rarely after DPT vaccination but most of the cases studied did not appear to be associated with vaccination (11). On the assumption that the surveillance system set up had identified all of the potential cases, the NCES estimated the attributable risk of serious neurologic disorders occurring within 7 days of vaccination in previously normal children to be 1 in 110,000 (95 percent confidence interval: 1 in 360,000, 1 in 44,000).

The conclusions of the NCES have been the subject of considerable controversy. The parents of a child with permanent neurologic injury sued for damages, and this led to a detailed legal review of the findings from the NCES (12, 13). The court report judged that “the evidence supports the conclusion that DPT sometimes causes febrile convulsions; it does not provide evidence that such convulsions following DPT cause brain damage... the results of the NCES do not support... that DPT can cause permanent neurological damage” (13, p. 76). The judge highlighted a number of issues in making his ruling, including the study team’s declared bias of trying to maximize the risk rather than minimize it and the suspicion that vaccinated cases were more likely to be reported to the study team than unvaccinated cases. In addition, the results of a review of the relevant cases and their controls indicated that although many case children were recruited into the study, very few cases were of permanent neurologic damage in previously normal children, and a careful case-by-case review reclassified about half of these. Debate on the findings of the NCES has continued (14, 15). The American Academy of Pediatrics concluded that pertussis vaccine has not been proven to be a cause of brain damage (16).

The controversy caused by the NCES raised many of the methodological issues related to the use of the case-control approach in the study of adverse reactions. These included the possibility that the presence of contraindications may result in lower vaccination rates among those likely to suffer the event presumed to be caused by the vaccine (confounding), a greater likelihood of diagnosis or ascertainment of presumed adverse reactions among vaccinated cases than among unvaccinated cases (diagnosis bias), and faulty recall of the date of onset of symptoms or the date of vaccination (recall bias). These issues are discussed below. The controversy and methodological difficulties surrounding the NCES may be at least part of the reason why the case-control approach has been used much less widely for the study of adverse reactions than for the study of vaccine efficacy.

As was noted above, case-control studies have been used to assess vaccines mainly with respect to two issues: vaccine efficacy and adverse effects. These issues are discussed separately below. We give formulae for the measure of interest, contrast advantages and disadvantages of the case-control approach in relation to other available designs, and discuss relevant methodological issues. We refer to case-control studies in the literature to illustrate particular points.

**VACCINE EFFICACY**

Usually, a vaccine is not licensed for general use without undergoing evaluation in at least one RCT (Phase III trial) (17). However, for any of several reasons, it may be necessary to measure the performance of a vaccine after it has been introduced into routine use (18). Firstly, the results of RCTs may not have been consistent; this was the case for BCG vaccine and tuberculosis (19, 20). Secondly, the efficacy measured in the Phase III trials may be greater than that applicable when the vaccine is in public health use. Phase III trials are generally designed to evaluate the efficacy of a vaccine under carefully controlled conditions, and may exclude persons who would not be excluded from a routine vaccination program (e.g., those with concurrent illnesses). In addition, the conditions of storage of the vaccine, the interval between doses, and the population selected for vaccination can be controlled more...
carefully in a trial than in a routine vaccination program. Case-control studies can be used after the implementation of routine vaccination to estimate the protection given by the vaccine under "normal" conditions (17, 20, 21). Thirdly, case-control investigations may be used to determine whether an outbreak of a vaccine-preventable disease was due to poor vaccine efficacy (22, 23) and to identify causes for this (24, 25). Fourthly, there may be reasons to study the protection conferred by a vaccine in populations or against disease types that were not included in Phase III investigations. There may be interest in the efficacy of a vaccine in a subgroup of the population, e.g., pneumococcal vaccine in the elderly (26) or BCG vaccine in children of Indian-subcontinent ethnic origin living in England (27), in a specific setting, e.g., influenza vaccine among institutionalized persons (28), against a specific form of disease, e.g., BCG vaccine against miliary or meningeal tuberculosis (20) or level of severity, e.g., influenza vaccine against influenza mortality (29), or against a different disease, e.g., BCG vaccine against leprosy (30). Case-control studies have also been used to assess the protection of a vaccine not against a specific disease but against a group of conditions presumed to consist in part of the disease of interest, e.g., studies of the efficacy of measles vaccine against all-cause mortality (31, 32) and of pneumococcal vaccine against hospitalization for pneumonia in the elderly (33) and to study effects not originally measured, e.g., the protective effect in infants of cholera vaccine given to women who breastfeed (34). It is also possible to measure the efficacy of a vaccine against a long term complication of a disease, e.g., the efficacy of measles vaccine against subacute panencephalitis (35).

Methodological issues associated with the use of case-control and other observational approaches to the evaluation of vaccine efficacy have been discussed in a number of papers. Comstock (21, 36) provided two overviews; Joseph and Joseph (37) discussed possible biases; Smith et al. (8) and others (9, 38, 39) reviewed the methodological implications of the presumed biologic mechanism underlying vaccine protection; Wunsch et al. (40) discussed different sources of controls; and Clarkson and Fine (41) and Noah (42) reviewed public health applications. A number of papers discussing the field evaluation of vaccine efficacy (17, 43, 44) or reviewing efficacy studies for specific diseases (e.g., pertussis (45, 46) and tuberculosis (47, 48)) contain relevant sections on case-control studies.

**Measurement of vaccine efficacy**

Vaccine efficacy is defined as the percentage reduction in the disease rate among vaccinated subjects that is attributed to vaccination. In a situation where there are no differences between vaccinated and unvaccinated individuals with respect to confounding variables, as in an RCT, vaccine efficacy (VE) is defined as

$$VE = 100 \left( \frac{I_u - I_v}{I_u} \right) = 100 \left( 1 - \frac{RR_{v/u}}{RR_{u/v}} \right),$$

where $I_u =$ disease incidence in the unvaccinated group, $I_v =$ disease incidence in the vaccinated group, and $RR_{v/u} = I_u/I_v$ (the relative risk of disease in vaccinated persons compared with unvaccinated persons).

Relative risk may be estimated in case-control (or cohort) studies, and thus the formula may also be used in such studies to estimate vaccine efficacy (usually after the relative risk estimate has been adjusted for potentially confounding variables).

In some circumstances—for example, in a disease surveillance program—the information most readily available may be the proportion of cases who have been vaccinated ($P_v$) and the proportion of the target population who have been vaccinated ($P_p$). A rapid estimate of vaccine efficacy (though possibly biased, since there is no adjustment for confounding variables) is given by

$$\left\{ 1 - \left( P_v (1 - P_p) \right) / \left( P_p (1 - P_v) \right) \right\} \times 100.$$

Application of this formula illustrates that in vaccination programs with high vaccine coverage and high vaccine efficacy, it is to be expected that a high proportion of persons with disease will have a history of vaccination. For example, in a vaccine program which achieves 95 percent coverage with a vaccine that has 90 percent efficacy, approximately two thirds of all cases of disease will be in vaccinated individuals.

**Mode of vaccine protection**

Protection to the individual. Vaccine efficacy is a measure which summarizes the impact that a vaccine has on disease incidence among those vaccinated. The level of protection given to different members of the vaccinated population may vary, and different models of vaccine action have been discussed. Under one model, the vaccine is assumed to provide full protection to some individuals and leave the others as susceptible as they were before vaccination. In another model, the vaccine is assumed to confer the same level of partial immunity to all persons vaccinated (8). In the former “all-or-nothing” model, a vaccine efficacy of 80 percent corresponds to a situation in which 80 percent of those vaccinated are completely immune to disease subsequent to vaccination and the other 20 percent remain as susceptible as they were prior to
vaccination. That is, the number of subjects at risk of disease is reduced by 80 percent and the incidence of disease remains the same in the 20 percent who are vaccinated but not protected. In the “partial immunity” model, all persons vaccinated remain at risk of developing disease but each of their risks of disease is reduced by 80 percent. Both of these models are idealized, and for most vaccines the true mode of action probably fits neither of them, but for a specific vaccine the mode of action may be closer to one model than to the other.

The “all-or-nothing” model may be most relevant for vaccines for which protection depends upon inducing a sufficiently high immunologic response, where persons beyond some critical level of a specific antibody after vaccination are immune from disease and persons under this level have a rate of disease similar to that of those who were not vaccinated. Often, however, those with lower induced antibody levels may have some protection, and the degree of protection may be correlated with antibody level. The implications of this, as well as variations in other assumptions under these models (including homogeneity of protection among persons not made fully immune by vaccination, homogeneity of exposure to infection, and duration of immunity), have been discussed elsewhere (49–55) and are also addressed in another paper in this issue (119).

If the disease against which the vaccine is directed is common, the underlying model of vaccine action is relevant to the design and interpretation of studies of vaccine efficacy. If the immunity of vaccinated subjects does not wane with time (which is perhaps unlikely), under the “all-or-nothing” model, the proportion of fully immune individuals among those who have not yet suffered from disease will increase with time since vaccination (because an increasing proportion of the nonprotected vaccinated subjects will develop disease). Under the “partial immunity” model, the risk of disease among those who have not yet developed disease will remain constant. The implications of the model of vaccine action for the design of case-control studies have been discussed in detail elsewhere (8). In brief, under the “all-or-nothing” model, controls should be selected from the population initially at risk after vaccination. That is, individuals should be eligible for inclusion as controls even if they develop the disease under study. If such individuals are excluded from the control group, the efficacy of the vaccine may be overestimated. For example, in a mumps outbreak at a school, the estimated efficacy of mumps vaccine was 71 percent, but it increased to 78 percent when subjects with a prior history of mumps were excluded (56). Under the “partial immunity” assumption, controls should be selected to represent the person-years at risk in the population that produced the cases during the period covered by the study. That is, potential controls should be excluded if they develop disease before the date of disease onset in the case to whom they are matched, but not if they develop it afterwards (9). The underlying model is of little consequence, however, unless the disease under study is common (e.g., the attack rate in those vaccinated from the time of vaccination to the end of the study period is ~20 percent or more).

Direct and indirect protection. The protection against disease conferred by a vaccine consists of both a direct effect and an indirect effect. The direct effect results from the immunity conferred by the vaccine and vaccinated individuals; the indirect effect results from the fall in the number of sources of infection caused by the direct effect, leading to a reduction in the risk of infection in the community (57, 58). A number of RCT designs have been suggested for measurement of direct and indirect effects, based on group randomization (59). The relevance for observational studies is small. In most situations, vaccinated and unvaccinated people mix, and the impact of the indirect effect is the same in both groups and will not be measurable in either cohort or case-control studies.

Advantages and disadvantages of using the case-control approach to assess efficacy. The advantages and disadvantages of the case-control approach derive from its features: being observational, having subjects selected on the basis of disease status, and using controls to represent the population from which the cases were derived.

Case-control studies compared with randomized controlled trials. In general, a double-blind RCT is the best way to evaluate a vaccine, because through this approach issues of confounding and bias have minimal influence on the estimation of vaccine efficacy. In an RCT, participants are allocated at random to a vaccine group or a placebo group (if there is no existing efficacious vaccine), with both trial participants and researchers blinded as to which group received placebo and which received vaccine. Both trial groups are followed prospectively, and the incidence of disease is compared. In some circumstances, groups rather than individuals may be randomized (60). In contrast, case-control studies, in common with cohort and case-population studies, are observational. The individuals studied may or may not have been vaccinated, but the allocation to either of these groups has not been random. Thus, the investigators must attempt to control for differences between the groups that may confound estimation of the vaccine effect.

The advantages of the case-control approach in comparison with RCTs are several. Firstly, case-control
studies avoid ethical issues in situations where there is already evidence that the vaccine being evaluated is better than the comparison product, be that another vaccine or no vaccine. Secondly, because of their retrospective nature, case-control studies can generally be conducted much more rapidly than RCTs. Thirdly, case-control studies require smaller numbers of subjects, and thus they are substantially cheaper to conduct than RCTs.

Case-control studies compared with cohort studies. Cohort studies are also observational studies in which subjects have already been “self-allocated” to vaccinated or unvaccinated groups. Thus, they share with case-control studies the potential for bias and confounding in the estimation of vaccine effects. Except for this important aspect, the logic of a cohort study follows that of an RCT. Persons in the population at risk (those with no previous history of disease) are classified as vaccinated or unvaccinated according to their status at the beginning of the study period, and disease incidence rates between the two groups are compared. Cohort studies may be conducted in the context of a routine vaccination program, in which there is surveillance for new cases of disease and records are kept centrally on both the population eligible for vaccination and the persons who have actually been vaccinated (e.g., Sutherland et al.’s (61) studies of BCG vaccine efficacy in the United Kingdom). Two settings are particularly efficient for application of the cohort approach: outbreaks and studies of contacts. Methodological issues relevant to the design and conduct of cohort studies have been discussed elsewhere (43, 44).

RCTs are prospective. Disease events are ascertained following randomization of participants to vaccinated and unvaccinated groups. The groups must be followed until a sufficient number of disease events have occurred to enable a good estimate of vaccine efficacy to be made. Cohort studies may be prospective or retrospective, but in both situations, as in RCTs, disease events may occur over a considerable time period. This may make it difficult to ensure uniform standards of diagnosis, particularly for retrospective studies. In contrast, in case-control studies, cases are typically recruited over a relatively short time period, and, particularly if this is done prospectively, it is easier to ensure that strict diagnostic criteria are adhered to.

In case-control studies, information is collected on a small representative sample of persons without the disease (the control group), whereas in cohort studies (and RCTs) it is necessary to collect information on and follow up all persons in the population from which the cases were derived. This gives case-control studies a very important advantage in terms of speed, cost, and logistics—an advantage that increases with the rarity of the disease. Because information needs to be collected only for a small number of subjects, it is possible to collect information on a much larger number of variables than in cohort studies, enabling better control of confounding. A major disadvantage of the case-control approach is the possibility of selection bias due to selection of a group of controls who are not representative of the population whence the cases came.

Case-control studies compared with case-population studies. The case-population approach is similar to the case-control approach, but instead of data on a control group, data on the whole population are used for contrast with vaccine coverage in the cases. The approach has also been called rapid screening (for vaccine efficacy) or case coverage (43, 44, 62, 63). Vaccine efficacy is estimated by comparing vaccine coverage among cases with the vaccine coverage in the population. The only advantage the method has over the case-control approach is that it is cheaper. It is a crude method by which to assess vaccine efficacy, but it has some uses. For example, the method may be used to obtain a preliminary and rapid estimate of vaccine efficacy, or to obtain reassurance that the proportion of cases who have a history of vaccination is compatible with vaccine coverage in the population and presumed vaccine efficacy. The main limitation of this method is that it allows for no control of confounding; therefore, it is particularly important to make sure that the “cases” and the “population” relate to the same age group and geographic area.

Methodological issues

Study population. The study population will often be selected on the basis of logistic considerations, but the choice is likely to be influenced by two issues: the generalizability of results and the proportion of the population that may have been infected prior to vaccination.

Generalizability. Vaccine efficacy may vary in different subpopulations and settings. The study must be designed to estimate efficacy in the settings or subpopulations of interest. Thus, if the main public health target for a vaccine is the elderly, the study population and thus the cases and controls must be drawn from elderly persons. Examples of the use of case-control studies to estimate vaccine efficacy in particular subgroups of the population include studies of pneumococcal vaccine in 1) the immunocompromised elderly (26), 2) the immunocompetent elderly (64), and 3) persons with human immunodeficiency virus infection (65).

Infection or disease prior to vaccination. Efficacy studies of vaccines are difficult to interpret when they...
are undertaken in settings where a substantial number of persons may have been infected (with the disease agent under study) prior to vaccination. For most vaccines, vaccination would not be expected to influence disease incidence among those who have been infected prior to vaccination. For diseases with short incubation periods, cases can be excluded from the study or can be considered unvaccinated if the interval between the last dose of vaccine and the onset of symptoms is shorter than the incubation period (66, 67). In matched case-control studies, controls should be considered unvaccinated if they received vaccine after the presumed date of infection of the matched case. This is an argument for matching, since each control can be given a "pseudo-" date of disease onset which is the same as that of the matched case.

For some diseases, the time between infection and onset of disease may be both long and variable. With tuberculosis, for example, clinical disease may develop years after the initial infection, either because of reinfection or because of reactivation of the original infection. In most of the RCTs of BCG vaccine, previously infected subjects were excluded on the basis of a tuberculin skin test. In many of the BCG mass vaccination campaigns, however, an attempt was made to administer BCG vaccine to all children under a certain age, without prior tuberculin testing. Many of these children, especially the older ones, will have been infected with the tubercle bacillus prior to vaccination. In a case-control study it would be impossible to exclude those who were infected before vaccination, and the efficacy measured would be expected to be lower than that measured in RCTs restricted to persons with no evidence of prior tuberculosis infection. Smith (5) presented tables to illustrate the magnitude of the impact of this on the measure of vaccine effect in a case-control study, taking account of the efficacy of the vaccine in persons who were not infected at the time of vaccination, the proportion infected at the time of vaccination, and the incidence of tuberculosis in those infected and not infected in the absence of vaccination (table 1). The estimated vaccine efficacy may be the measure of public health interest, of course, in populations where BCG vaccine is given without a previous skin test. The issues would be similar for a disease such as malaria, where neither previous episodes of disease nor current vaccines confer good protection (68, 69).

**Exposure. Validity of information on vaccination.** Reliable information on vaccination history is essential for estimating vaccine efficacy in case-control studies, but it is often difficult to obtain, particularly

---

**TABLE 1. Overall efficacy of Bacillus Calmette-Guérin (BCG) vaccine against tuberculosis in populations in which some members are tuberculin-positive**,†

<table>
<thead>
<tr>
<th>Proportion of population tuberculin-positive</th>
<th>Incidence in the tuberculin-positive/incidence in the tuberculin-negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td>80% protection in the tuberculin-negative</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>80</td>
</tr>
<tr>
<td>0.1</td>
<td>76</td>
</tr>
<tr>
<td>0.2</td>
<td>71</td>
</tr>
<tr>
<td>0.3</td>
<td>66</td>
</tr>
<tr>
<td>0.4</td>
<td>60</td>
</tr>
<tr>
<td>0.5</td>
<td>53</td>
</tr>
<tr>
<td>50% protection in the tuberculin-negative</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>50</td>
</tr>
<tr>
<td>0.1</td>
<td>47</td>
</tr>
<tr>
<td>0.2</td>
<td>44</td>
</tr>
<tr>
<td>0.3</td>
<td>41</td>
</tr>
<tr>
<td>0.4</td>
<td>38</td>
</tr>
<tr>
<td>0.5</td>
<td>35</td>
</tr>
<tr>
<td>20% protection in the tuberculin-negative</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>20</td>
</tr>
<tr>
<td>0.1</td>
<td>19</td>
</tr>
<tr>
<td>0.2</td>
<td>18</td>
</tr>
<tr>
<td>0.3</td>
<td>16</td>
</tr>
<tr>
<td>0.4</td>
<td>15</td>
</tr>
<tr>
<td>0.5</td>
<td>13</td>
</tr>
</tbody>
</table>

* Reproduced with permission from Smith (5).
† The table shows the overall protective effect (%) that would be estimated in various circumstances according to 1) the protective effect of BCG in those who were tuberculin-negative at vaccination; 2) the relative incidence of tuberculosis in those who were tuberculin-positive compared with the incidence in those who were tuberculin-negative and not vaccinated; 3) the proportion of the population who were tuberculin-positive before vaccination.
when there is interest in the protective effect several years after vaccination. The ideal vaccination history consists of a record written at the time of vaccination. Such records may be available in some situations, particularly in developed countries with good central recording of vaccines administered. In many circumstances, however, vaccination records will not be available or will be incomplete, and cases and controls (or their parents or guardians) must be asked to recall the vaccination history. Such recall may be imperfect.

The consequences of misclassification of vaccine status vary. If the information about vaccination status is subject to error but to a similar extent in cases and controls (nondifferential misclassification), the estimate of vaccine efficacy will be biased toward zero. Some case-control studies of the efficacy of childhood vaccines have presented results based on information extracted from vaccination cards and also based on maternal recall. In general, estimates of efficacy based on recall have tended to be lower than those based on vaccination cards (70, 71). Sometimes the sources of vaccine history may vary between individuals, and a decision must be made as to whether to restrict the analysis to subjects with information based on written records or to include those with information based on parental recall only. In such circumstances, it is probably desirable to present both sets of results to indicate the robustness (or lack thereof) of the estimate.

BCG and smallpox vaccine are among the few vaccines to leave a long-lasting physical mark after vaccination. BCG scars have been used as an indication of vaccination in case-control studies of the efficacy of BCG vaccine. The sensitivity and specificity of this measure of vaccination are less than 100 percent. In one trial in Malawi, the validity of scar reading, estimated where vaccine history was known, was low (72), but further studies are necessary to establish whether this would be true in other settings. Smith (5) gives a formula and a table (table 2) and Fine et al. (72) give tables showing the effect on the estimated vaccine efficacy caused by misclassification of BCG vaccination history based on the sensitivity and specificity of scar reading.

When the validity of vaccine history is different in cases and controls (differential misclassification), the efficacy may be biased towards or away from zero. This is more likely to occur if information on vaccination is based solely on recall (recall bias), since cases may be likely to recall their vaccination history with different reliability than controls.

Vaccination of controls after disease onset in the matched case. For most diseases, cases are unlikely to be vaccinated after they have developed the disease under study, and any such vaccination would be ignored in a case-control evaluation. Care must be taken that comparable vaccinations in controls are treated in the same way in the analysis of the results. This is straightforward in matched case-control studies, where the vaccination history of controls is considered only up to the age at which the matched case developed disease. This strategy cannot be adopted when frequency matching is done, or when matching is not used at all. This is an additional reason for using a matched design in such studies.

Variation in efficacy. Features associated with low efficacy. It is possible to examine how the efficacy of a vaccine varies according to factors such as age at vaccination, vaccination schedule, place of vaccination, and time since vaccination. Two methods of examining such variations have been employed, both using the case-control approach. In the first approach, there are no restrictions on inclusion in the study on the basis of vaccination. Studies using this approach have examined vaccine efficacy according to number of doses, interval between doses, and schedule (73–79). The effect of age at vaccination on the efficacy of measles vaccine was established in early case-control studies using this approach (80–83). Other studies have examined other specific features suspected of lowering vaccine efficacy (84).

In the second approach, the analysis is restricted to cases and controls who give a history of vaccination. The exposures studied are features of the vaccine that are suspected of being associated with low efficacy—for example, age at vaccination or number of doses. In the simplest case, the analyses compare the proportions of vaccinated cases and vaccinated controls with and without the feature of interest, and the estimated odds ratio provides a measure of how much vaccine efficacy varies according to the presence or absence of the feature (23, 71, 85).

Time since vaccination. It is surprising how rarely the case-control approach has been used to examine changes in vaccine efficacy with time since vaccina-
tion, especially given our ignorance of the long term protective effect of many vaccines. For example, the duration of protection given by BCG vaccine against tuberculosis is unclear, despite the long history of use of this vaccine (86). Estimation of efficacy according to time since vaccination is straightforward in case-control studies, if information on date of vaccination and date of disease is collected (table 3).

One of the few examples of estimating change in vaccine efficacy with time since vaccination using a case-control approach is that of Reingold et al. (87). These authors conducted successive case-control studies in the years following a mass vaccination campaign with meningococcal polysaccharide A vaccine, and they presented vaccine efficacy according to the amount of time elapsed between the campaign and the case-control studies.

Fine and Clarkson (46) have reviewed studies of the efficacy of pertussis vaccine according to age at disease onset. In this instance, age at disease onset is closely correlated with time since vaccination, as the vaccine is given in early infancy.

Definition and selection of cases. Defining the outcome of interest. The efficacy of a vaccine may be different for different forms of a disease. For example, efficacy may vary in persons with any clinical disease, persons with severe disease (e.g., severe influenza (88), mortality associated with influenza (29)), or persons with specific forms of clinical disease (e.g., miliary and meningeval tuberculosis (20), invasive Haemophilus influenzae type b (89), pertussis carriage (46)).

A strength of the case-control approach is that it can be employed to study the efficacy of a vaccine against endpoints which cannot be studied in RCTs. Under the special circumstances of an RCT, there is an ethical obligation to treat cases of disease when they occur in the study population. The surveillance systems set up for the trial may have the effect of ensuring that most cases of disease are treated early in their natural history, often preempting the occurrence of more severe disease. Thus, it may be difficult or impossible to measure the impact of vaccination on severe disease, which is likely to be the measure of most public health importance. Because of their retrospective nature, case-control studies do not suffer from this restriction. Examples in which case-control studies have been used to assess the efficacy of vaccines against serious disease endpoints include assessments of the efficacy of measles vaccine against all-cause mortality (31, 32) and against hospitalization for dehydration (90).

Case definition. In case-control studies, as in RCTs, it is important that the case definition be as specific as possible, since the inclusion of "false positive" cases will decrease the estimated vaccine efficacy. It is not necessary to have a highly sensitive case definition, as long as the case definition is applied without bias (as to vaccine status) and it is possible to recruit enough cases for the study to obtain the desired statistical power. Cases can also be classified according to diagnostic certainty, and efficacy estimated for each level of certainty (91).

There should be a clear definition of the target population from which cases are to be ascertained, and an effort must be made to ascertain all cases, or a representative sample of all cases, occurring in that population over the study period. Controls should then be selected to represent that population. Clarity regarding the definition of the study population whence the cases were derived is essential in order to avoid selection bias when choosing controls. It is generally inefficient to include cases and controls who are under the minimum age recommended for vaccination, since they will probably be unvaccinated and will not contribute information to the study analysis.

Selection bias will also be introduced in a case-control study if the chance of a case's or control's being recruited into the study is influenced by his or her vaccination history. Disease diagnosis or designation by the study as a case must be as independent as possible from markers of vaccination. Thus, criteria for case definition should not be based on immunologic/serologic tests that are affected by vaccination. For example, tuberculin sensitivity is induced by infection with Mycobacterium tuberculosis and by vaccination with BCG. If tuberculin testing were included in the criteria for a diagnosis of tuberculosis, more BCG-vaccinated children than unvaccinated children would be diagnosed as cases (because of their past history of vaccination). This would lead to a reduction in the estimated vaccine efficacy.

Definition and selection of controls. Controls should be representative of the population that produced the cases. Selection bias is liable to be introduced when they are not.

Selection bias. The easiest situation in which to rule out selection bias is the one where the case-control

Table 3. Estimation of vaccine efficacy (VE) in case-control studies, by time since vaccination

<table>
<thead>
<tr>
<th>Status</th>
<th>Cases</th>
<th>Controls</th>
<th>Odds ratio</th>
<th>VE/100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unvaccinated</td>
<td>a</td>
<td>b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time since vaccination (years)</td>
<td>c</td>
<td>d</td>
<td>cb/adb</td>
<td>1 - (cb/adb)</td>
</tr>
<tr>
<td>&lt;5</td>
<td>c</td>
<td>d</td>
<td>cb/adb</td>
<td>1 - (cb/adb)</td>
</tr>
<tr>
<td>5-10</td>
<td>e</td>
<td>f</td>
<td>eb/afa</td>
<td>1 - (eb/afa)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>g</td>
<td>h</td>
<td>gb/ah</td>
<td>1 - (gb/ah)</td>
</tr>
</tbody>
</table>
study is nested within a cohort study (or in a well-defined population). In such circumstances, all cases arising in that cohort or population can be included in the study, and controls can be selected at random from the cohort or population (possibly matched with the cases according to characteristics such as age and sex). In other situations, there should be a clear definition of the population from which the cases were derived, and the controls should be ascertained from, and be representative of, that population. A good general rule for deciding whether or not an individual should be eligible for being a control is whether, if that person developed the disease of interest, he or she would become a case in the study.

Sources of controls. The selection of an appropriate group of controls will be influenced largely by the method of case selection, taking account of the need to avoid selection bias and of logistic considerations. Control selection is easiest in the context of a case-control study set within a cohort study, as discussed above, where the cases are derived from a closed population and the controls can be selected from the same population. Such circumstances arise in a variety of situations: For example, the study population may consist of children in particular schools (80, 92), patients registered with a general physician (93), persons visiting an outpatient clinic (65), children listed in the child health computer database in England (46), persons enrolled in medical care programs in the United States (33, 94), or infants on a list of births compiled by local midwives (95).

Other potential sources of controls include hospitals (if cases are ascertained through hospital admissions), the neighbors of cases, friends of cases, and random digit telephone dialing. When controls are selected from hospitalized subjects, it is advisable to avoid including controls with diseases that can be influenced by the disease of interest (e.g., pneumonia if the disease of interest is measles) or by the vaccine (e.g., leprosy if the study is designed to assess the protective-ness of BCG vaccine against tuberculosis). Children who have received one vaccine are more likely to have received others, so often children with other vaccine-preventable diseases will also be excluded from the control group. When controls are selected through random digit telephone dialing, cases should be excluded from the study if they live in a home without a telephone (96).

In studies conducted in developing countries, selection of controls from neighbors is often a reasonable and efficient method. Controls are selected using a systematic approach starting at the house of the case. Neighborhood controls are matched (by neighborhood), and thus matched analysis must be used.

It is possible to use more than one control group from different sources (97, 98). This is a good strategy if there is no strong reason for preferring one type of control group to another. If, in the analysis of the data, similar results are obtained using either control group, this gives some reassurance that the results are not biased (or that they are consistently biased!). If, however, the analysis gives results which differ according to the control group utilized, then bias is likely to be present (with respect to at least one of the control groups). If it is clear at the outset of the study that one type of control is better than another on theoretical grounds, it will only potentially complicate and confuse the interpretation of the results to include both control groups.

Confounding. Case-control studies are observational, and therefore there must be appropriate control of confounding variables to ensure that the estimate of vaccine efficacy is not biased by differences between the characteristics of cases and controls. Confounding variables are those which are independently associated with the outcome and the exposure of interest. Thus, in an assessment of a vaccine, a potentially confounding variable will be one that influences the risk of the disease under study and is also associated with the likelihood of vaccination. Factors such as age, sex, social class, and, for some diseases, birth order may be associated with rate of disease and with risk of vaccination, and therefore are potentially confounding factors.

An issue that is sometimes discussed in the context of using case-control studies to evaluate vaccine efficacy is the extent to which the controls must have been exposed to the same risk of infection as the cases (37). With infections for which exposure is highly likely to result in disease in the absence of vaccination, all of the cases must clearly have been exposed, and the unvaccinated controls are unlikely to have been exposed (or else they would have been cases!). Thus, ensuring equal chances of exposure to infection among cases and controls cannot be necessary. The issue is relevant only if the likelihood of vaccination is influenced by the chance of exposure to the infection under study. For example, if a vaccination program is targeted at the poorer segment of a population, who may be more exposed to infection, controls should be matched to the cases with respect to socioeconomic status (or area, if there is marked geographic variation in vaccine coverage, due, for example, to variable access to health services). If, however, poor and rich are equally likely to be vaccinated, socioeconomic status will not be a confounding factor, even if it is related to the risk of infection or to disease, given infection.

Some vaccines are not universally recommended but have defined target groups (e.g., pneumococcal vac-
ADVERSE EFFECTS

When high population coverage is achieved with an efficacious vaccine, the occurrence of cases of the disease that the vaccine prevents may be rare. In these circumstances, morbidity associated with the vaccine itself or presumed to be associated with vaccination may cause more public concern than the prevented disease.

Adverse effects which are relatively common and which appear shortly after vaccination should usually have been detected in Phase III studies of a vaccine, or earlier. However, such effects may be missed, for any of several reasons. For example, Phase III studies may not have been conducted (e.g., if the change to a different vaccine formulation was based on serologic studies only). The incidence of the adverse effect may be too low to be detected with reasonable reliability in Phase III studies, and the effects may only become apparent when the vaccine is introduced into general use. In addition, the adverse effect may occur after longer periods postvaccination than were studied in Phase III trials. Finally, in general use, vaccines may be administered to a population which is less homogeneous than that vaccinated in a Phase III trial, and there may be subgroups at higher risk of an adverse effect.

After Phase III trials have been completed and a vaccine is being given routinely, possible adverse effects of vaccination may first be recognized through surveillance systems set up for this purpose (e.g., the "yellow card" surveillance system in the United Kingdom (99) or the Vaccine Adverse Event Reporting System (VAERS) in the United States (100)), through a detected rise in the incidence of the putative adverse effect in other surveillance systems, or simply because doctors or other health care workers notice an apparent increase and voice their concerns. More recently, the first recognition has resulted from scanning large data sets and linking individual records of vaccination data and morbidity data (e.g., the Vaccine Safety Datalink Project in the United States (101) and linked hospital discharge data and child health and general practice records in the United Kingdom (102, 103)).

Many vaccines are given to very young children at an age when developmental abnormalities first tend to be noted and, in particular, the incidence of serious neurologic illnesses is relatively high. Thus, the temporal coincidence of vaccination and the recognition of such developmental abnormalities is often thought to indicate a causal connection (probably falsely more often than not). Such alarms can have dramatic effects on vaccination coverage rates and, indeed, may precipitate an epidemic of the disease that vaccination would prevent. For example, following extensive media coverage of a possible association between pertussis immunization and neurologic illness in the United Kingdom, the national coverage of pertussis vaccination dropped from 80 percent in 1974 to 31 percent in 1978 and coincided with the largest pertussis epidemic in 20 years (11). More recently, media coverage of a postulated association between measles-mumps-rubella vaccine and autism in the United Kingdom (104) was followed by a steep decline in measles vaccination. It is important, therefore, that appropriate epidemiologic studies be implemented rapidly to investigate whether or not unexpected adverse effects of vaccination are occurring and with what frequency.

Because adverse reactions are often first detected or suspected after a vaccine has been introduced into general use, observational designs are usually employed to try to link suspected adverse events to vaccination. The approaches adopted may include cohort studies, case-series analyses, and case-control studies. The advantages and disadvantages of each approach and the methodological issues relevant to case-control studies are discussed below.

Key papers discussing the use of the case-control approach in the study of adverse reactions to vaccines include those by Comstock (21) on the use of case-control studies in vaccine evaluation, Fine and Chen (105) on confounding in studies of adverse reactions, Fine (106) on methodological issues in the evaluation of vaccine safety, Farrington et al. (107) on different designs used in the study of adverse reactions, and much of the literature addressing the controversy over adverse effects associated with pertussis vaccines (11, 12, 14-16, 87, 108-112).

Measurement

The case-control approach may be used to produce estimates of the ratio of the risk of a specific adverse
effect among vaccinated persons to the risk among unvaccinated persons, based on a comparison of the proportion vaccinated among cases with the suspected adverse reaction and the proportion vaccinated among controls.

The main measurement of interest when assessing adverse reactions, however, is the frequency of adverse effects attributable to vaccination among those vaccinated. This is usually expressed as the number of adverse reactions induced in a given number of persons vaccinated (subtracting the baseline rate in the population). This is theoretically straightforward in cohort studies, since the rates of adverse reactions in vaccinated and unvaccinated individuals are measured directly. The rate of induced adverse reactions can be estimated in a nested case-control study (113), in a case-control study (102), and in a case series (107) if the proportion of the population that is vaccinated and the frequency of the suspected adverse reaction in the population are known or can be estimated. Thus,

$$\text{attributable risk} = \frac{r(R - 1)}{(Rp - p + 1)},$$

where \( p \) is the proportion of the population vaccinated, \( r \) is the rate of the putative adverse event in the total population, and \( R \) is the relative risk of the adverse event in vaccinated individuals compared with unvaccinated individuals.

Advantages and disadvantages of the case-control approach compared with other approaches

The "best" situation in which to investigate whether or not a suspected adverse reaction is linked to vaccination, using cohort, case-series, or case-control studies, is one where there is a clearly defined population with a record of all vaccinations, all instances of the disease event of interest (potential adverse effect), and the possibility of linking these back to individuals. Studies carried out within such a linked data set are less prone to biases, particularly if the data set contains information on dates of vaccination and dates of symptom onset and this information was collected before the link between the vaccine and the putative adverse event was suspected. Case-control and cohort studies require that vaccine coverage is not approaching 100 percent. Case series are based on examining vaccination history in a series of cases only, and this can be done even if all cases are vaccinated, as long as the occurrence of the adverse event shows a close temporal association with vaccination. Farrington et al. (107) compared the case-series method with the other two approaches.

The advantage of conducting investigations in populations in which vaccination coverage is incomplete is that it is possible, in theory at least, to study adverse effects which may occur at any time after vaccination and which may not show a close temporal relation with vaccination. The change in risk with time since vaccination may also be studied, but this is not critical in establishing vaccination as a risk factor for the condition (whereas it is in the case-series approach). If vaccination coverage is very high, there is not a contemporary unvaccinated group of sufficient size with which to compare the frequency of specific conditions, using either cohort or case-control study designs. It is in such situations that, with efficacious vaccines, there may be the most population concern about possible adverse effects (of vaccination against a disease that does not occur—because vaccine coverage is so high!). If the adverse effect does not follow vaccination after some characteristic time interval but may occur at any time following vaccination, it may be very difficult in situations of high vaccine coverage to establish an association through analytic epidemiologic investigations. Ecologic studies or before-versus-after studies, with all of their potential defects, may be the only approaches applicable.

Fortunately, many adverse effects which are attributable to vaccination tend to occur during a specific period after vaccination, and thus the critical epidemiologic measure to study is the interval between vaccination and the onset of the condition under study. If the interval between vaccination and the presumed adverse effect is very short (for example, anaphylaxis within a few minutes of vaccination), the association with vaccination may be obvious, but if the adverse effect occurs days or weeks after vaccination and is not of a nature which is clearly vaccine-associated, careful epidemiologic studies are required to elucidate whether or not a causal association is a likely explanation.

Cohort and case-control approaches in linked data sets. Developments in information technology have facilitated the setting up of both retrospective and prospective cohort studies for examining the incidence of specific outcomes in vaccinated and unvaccinated individuals, through linkage of vaccination records to databases that record episodes of morbidity (such as general practitioner and/or hospital records). The possibility of such record linkage studies exists in various forms in the United Kingdom (102, 103), in the United States (100, 113, 114), in some Scandinavian and other European countries, and among persons in some private medical insurance systems. Alternatively, or in addition, case-control investigations may be conducted in such settings in the form of "case-control within a cohort" studies. This kind of circumstance is among the best in which to conduct epidemiologic studies, because the
study population is well defined and the medical records for the population are such that all occurrences of the disease of interest may be recorded (particularly if the possible adverse effect is "severe"). The population from which to select controls for a case-control investigation is well defined, and this simplifies the conduct and interpretation of the research, since many of the common problems in control selection in case-control studies are avoided by being able to sample within the defined cohort (from which the cases are also derived). If the hypothesis being investigated is recent, it is better to restrict the study to cases diagnosed before the suspicion arose, to minimize the chance of diagnosis bias (in which cases are more likely to be diagnosed with the suspected "adverse reaction" if they were vaccinated). It is still necessary to control for confounding. If data for the control of confounding are not available in the data set and must be collected through interviews or examinations, case-control studies are cheaper and faster to carry out than cohort studies, because of the smaller sample size.

In circumstances where vaccination records are not kept in such a form that they may be directly linked to subsequent episodes of morbidity, it may still be possible to conduct case-control studies to detect associations between possible adverse effects and vaccination histories. What is necessary in these circumstances is that it be possible to detect the occurrence of all cases of the condition putatively associated with vaccination (or a "representative sample" of such cases) and that the previous vaccination records of those cases can be extracted, together with those of suitably chosen controls.

**Case-series analysis.** Ascertainment of adverse reactions based on case-series analyses among vaccinated persons was proposed by Farrington et al. (107, 115). The method is based on analysis of data on cases only (with known dates of birth, known histories of vaccination, and known dates of onset of symptoms) and examines temporal clustering of events around the date of vaccination. The approach depends upon having accurate information on dates of vaccination and the dates of onset of the putative adverse effect. Potential biases arise if a physician's likelihood of diagnosing the disease under study is influenced by knowledge of recent vaccination history. The approach is also potentially vulnerable to biased recall of dates of vaccination and dates of onset. For example, patients may tend to report dates of first onset of symptoms close to the date of vaccination if they believe that vaccination caused their disease (or that of their child). Confounding caused by a factor's being associated with vaccination and the disease event is avoided, since the method uses data on vaccinated individuals only, but factors related to timing of vaccination may still introduce bias (107). The main limitation of the case-series method is that it can only identify adverse reactions that occur after a short and defined period postvaccination. Farrington et al. present the method with an example that examines febrile reactions as an adverse effect of measles-mumps-rubella vaccine in the United Kingdom (107).

**Methodological issues**

To associate a specific adverse effect with vaccination using the case-control approach, it is usually necessary that three conditions hold. Firstly, whether or not an individual is vaccinated must be independent of the existence of conditions that may be associated with the adverse event. Secondly, diagnosis of the event must be independent of previous vaccination. Thirdly, for adverse events that are presumed to occur soon after vaccination, the date of vaccination and the date of onset of symptoms must be recalled accurately. If the first condition is not true, confounding can be introduced. If the others are not true, diagnostic and recall bias are introduced. These are discussed below.

**Confounding and selection of controls.** The confounding factors that must be taken into account when assessing adverse effects of vaccination may be different from those considered in assessing vaccine efficacy (105). Many social and medical factors are associated with avoidance of vaccination or with delaying vaccination: existing chronic diseases, acute illness, and high birth order, for example (105). Conditions which contraindicate vaccination, such as a history of fits or convulsions, may also be conditions associated with some potential adverse effects of vaccination. Unless this is taken into account, an association between vaccination and the adverse event might be obscured. Fine and Chen (105) concluded that this is why the (artifactual) significantly reduced risk of sudden infant death after pertussis vaccination has been found in so many studies, and why the estimated risk of neurologic disease after pertussis vaccination is increased when children with previous neurologic disease are excluded from the analysis. Because the contraindications for vaccination are similar for many vaccines, some investigators have examined the association between the supposed adverse effect of a vaccine and other vaccines (not thought to be associated with the adverse effect). The finding of an association which is specific for the vaccine under study strengthens the likelihood that it is a true adverse effect of the vaccine.

The choice of suitable controls is not always straightforward, but the principles are the same as those underlying the selection of controls for assess-
ment of the efficacy of a vaccine, taking into account factors which might confound any association. Matching is also useful when a reference date for the disease is needed for controls.

**Diagnostic bias.** After the hypothesis that a specific vaccine is associated with a specific adverse effect has been publicized, there is a chance that cases of the disease which occur close to the date of vaccination may be preferentially ascertained. To avoid such diagnosis bias, it is best to try to use cases that were diagnosed in an already established information system before the suspicion of a link with vaccination was raised. If the study must be concurrent, or if the hypothesis is old, cases should be sought within an established data set, so that even if diagnosis bias is not avoided, ascertainment bias is minimized.

**Information or recall bias.** In some situations, historical information on vaccination might be difficult to obtain, particularly for the study of adverse events not closely associated in time with vaccination. For example, in two case-control studies examining the association of irritable bowel disease and Crohn's disease with measles vaccination (116, 117), it proved impossible to obtain sufficient information on vaccination histories, and this component of the studies had to be abandoned. The association was finally examined in a third case-control study that was restricted to cases with a more recent date of birth, for whom it was possible to collect information on vaccinations from existing electronic records (118).

Recall bias is introduced when disease status (being a case or not) influences the likelihood of information's being collected correctly. Such bias is usually eliminated when information on exposure is based on records collected prior to diagnosis, but it may be a substantial problem when vaccination history has to be ascertained directly from the cases and controls or from their relatives. Biased recall of the date of onset of the adverse event and the date of vaccination is also a potential problem when the adverse effect is closely related in time to vaccination.

**CONCLUSION**

In summary, the case-control approach is widely used to complement RCTs in evaluating aspects of the efficacy of vaccines in routine use, and more rarely to identify adverse effects of vaccines. In addition to taking the usual precautions to prevent confounding and biases in case-control studies, when using this design to estimate efficacy one should consider the implications of the timing of infection and of vaccination, and the mechanism of vaccine protection. In the study of adverse reactions, efforts should be made to prevent a previous history of vaccination's biasing the diagnosis of the event of interest, as well as biased recall of the date of vaccination or the date of onset of symptoms. Case-control studies of the efficacy of and adverse reactions to vaccines are best done in the context of linked data sets.

**ACKNOWLEDGMENTS**

The authors thank Professor Paul E. M. Fine for helpful discussions, especially with respect to the section on adverse reactions.

**REFERENCES**

6. Clemens JD, Shapiro ED. Resolving the pneumococcal vaccine controversy: are there alternatives to randomized clinical trials? Rev Infect Dis 1984;6:589-600.
65. Gebo KA, Moore RD, Keruly JC, et al. Risk factors for pneu-

Epidemiol Rev Vol. 21, No. 1, 1999


