Lecture 6: Intervention studies

Learning objectives

By the end of this session, participants should be able to:

i. Understand the basic design features of an intervention study.
ii. Decide when it would and would not be appropriate to use an intervention study to provide the epidemiological information needed to take important public health decisions.
iii. Describe the main features of randomised controlled trials.
iv. Describe the different types of randomised studies.
v. Understand the ethical issues to be considered.
vi. Understand the principal analytical approaches.
vii. Appreciate how to interpret findings and assess their policy implications.
viii. Describe non-randomised intervention studies and when they may be used.

NB: Before the lecture, please look over Bailey et al. "Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomized controlled trial." *Lancet*. 2007; 369: 643-56. DOI https://doi.org/10.1016/S0140-6736(07)60312-2. We will use this trial as an example in the lecture to illustrate the principles of randomised controlled trials.

1. What is an intervention study?

Interventions are preventive or therapeutic measures, such as sex education, insecticide-treated bed nets, antiretroviral regimens, or drugs to reduce pain. In an intervention study the investigator starts with two (or more) groups of individuals (**participants**) who are similar in all respects other than the fact that one group (the **intervention group**) is deliberately assigned to receive the preventive or therapeutic measure under investigation (the **intervention**), while the other group (the **control or comparison group**) does not receive this intervention. All the participants are followed up to assess the **outcome** of interest (e.g. diarrhoea episodes, cervical cancer, death). The **incidence** of this outcome is then compared in the two groups.

The gold-standard intervention study is the **randomised controlled trial (RCT)**, and this lecture will focus on this design. Other designs of intervention study include quasi-experimental designs, in which researchers are not able to control the allocation of an intervention, and before-after studies in which data are not collected on a comparison group.

2. Randomised controlled trials

When conducted properly, RCTs generate the highest quality evidence of all the epidemiological studies. With appropriate additional safeguards, such as **blinding** (masking), RCTs minimise **biases** and **confounding** (see Lecture 4 on Confounding and Causality). The ≥

Box 1. Key steps to conduct an RCT

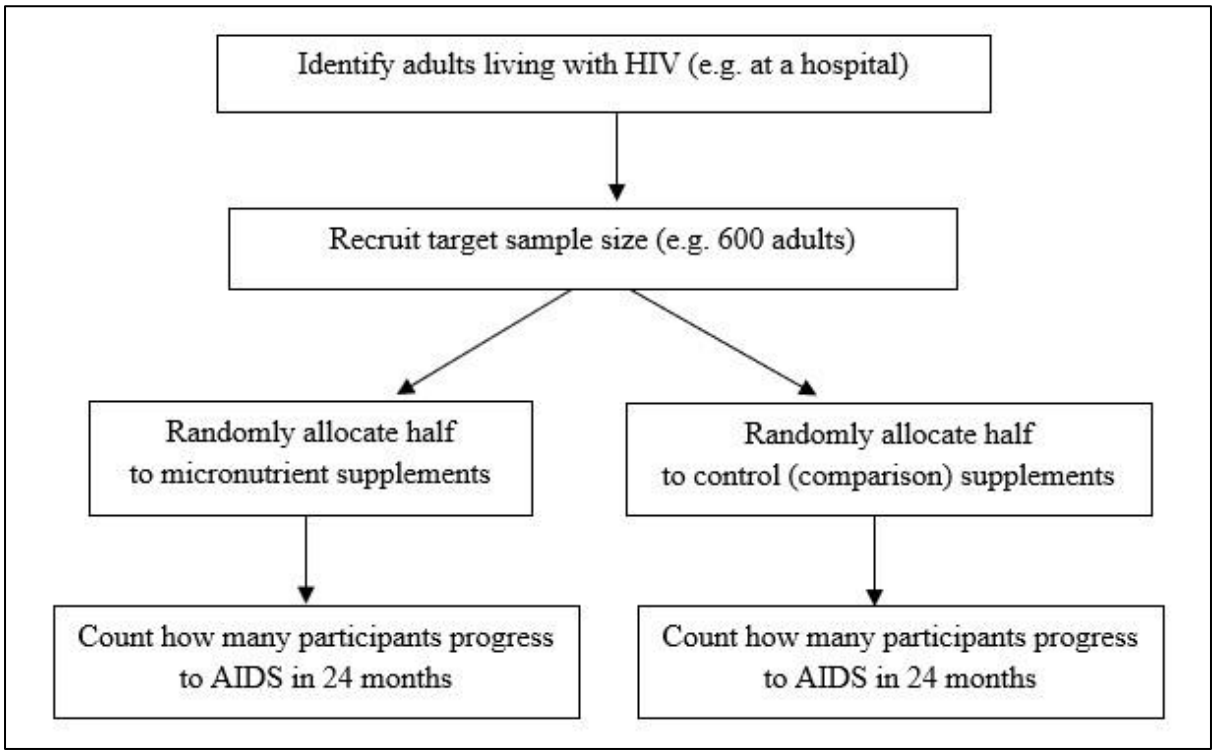
1. Clear formulation of hypothesis
2. Specify objectives of trial
3. Define the reference population
4. Select the study population
5. Select eligible participants
6. Obtain informed consent
7. Collect baseline data
8. Randomly allocate the participants to the (new) intervention or to the standard or placebo treatment (controls)
9. Follow up all participants in both groups (minimizing and monitoring defaulters and participants lost to follow up).
10. Assess defined outcome(s) continuously, intermittently, or at the end of trial ("blind" the assessment if possible)
11. Analysis - comparison of outcomes between intervention and control groups
12. Interpret the magnitude of effect
13. Consider alternative explanations of effect e.g. bias in composition or follow-up of groups
14. Policy implications
15. Feedback results to the participants
16. Communicate key results to the relevant official bodies (e.g. Ministry of Health, nongovernmental organizations), other researchers, trial participants and the general public.

2.1. Clear formulation of the hypothesis to be tested

Simple though this may sound, it is surprising how often investigators spend insufficient time formulating and clarifying the **hypothesis** they wish to test in an intervention study. For example, in an evaluation of the effectiveness of a package of health education messages, it is essential to specify exactly what the messages will be and what the direct and indirect effects of these messages are expected to be, in order to select the outcome(s) to be studied.

In the trial described in Box 2, the hypothesis was that micronutrient supplementation provided to adults living with HIV would reduce the risk of progression to disease over 24 months. This hypothesis had arisen from several well-designed observational studies which indicated an association between micronutrient intake, serum concentrations of micronutrients, and/or reported dietary intakes of micronutrients and the rate of progression of HIV. The question was of considerable public health importance, because antiretroviral drugs were, at that time, not widely available in low-income countries and were usually started relatively late in the course of HIV disease. Also, micronutrient supplements are relatively inexpensive.

Box 2: Trial of micronutrient supplements for adults living with HIV in Thailand



2.2. Definition of the reference population and study sample

The reference population is the population to which the results of the trial will be generalised. To enable such a generalisation, the study participants should be a random sample of the reference population. For example, it would be inappropriate to extrapolate the results of a trial of the effect of a new drug on the duration and severity of hospitalised cases of pneumonia among infants aged 3-11 months in Bangladesh to all children hospitalised with pneumonia aged 0-59 months in the world, or even to all 3- to 11-months old infants with pneumonia in Bangladesh.

Both the reference population and the study sample should be well defined, and you should be able to describe the important characteristics of the study sample with baseline data.

In the micronutrient supplementation trial, it was important that the results could be generalised to all adults living with HIV in Thailand. Patients were recruited from a hospital clinic which served a large, well-defined, urban population. Data on various characteristics of the participants were collected and described. Because the researchers wanted the results of the trial to apply to a wide spectrum of adults living with HIV, they kept the inclusion and exclusion criteria to a minimum. These criteria are described in Box 3.

Box 3. Inclusion & Exclusion Criteria in Therapeutic Intervention Study (Clinical Trial) of Micronutrient Supplements for adults living with HIV in Thailand

Include patients if:

1. aged 18 years or over
2. CD4 count ≥ 100 cells/μL
3. residence within 20kms of the hospital
4. able to provide accurate details of their address
5. agreed to return for follow-ups
6. consented to participate in writing

Exclude patients if:

1. Already taking micronutrient supplements
2. Taking combination antiretroviral therapy

2.3. Choice of comparison 'treatment'

A key characteristic of an intervention study is the inclusion of at least one comparison group, against which the outcome(s) in the intervention group is compared. A crucial decision is selecting the most suitable comparison group: participants who receive nothing, a **placebo** (a treatment which resembles the intervention treatment in all respects except that it does not contain the active ingredient(s)), or the current best treatment, if one is available. Ethical issues often play a major role in this decision (see Box 4).

Box 4. Choice of comparison treatment in a trial of perinatal anti-retrovirals to reduce mother-to-child transmission of HIV

In Bangkok, around 1995, a cohort study showed that 25% of pregnant women living with HIV passed the infection to their offspring, a rate broadly similar to that rate reported from low income countries. Around the same time, a complex and very expensive anti-retroviral drug regimen (containing the drug AZT), was shown to be highly effective in reducing mother-to-child transmission of HIV in the USA and France. Researchers in Bangkok felt that it would not be either economically or logistically feasible to administer this regimen in Bangkok, let alone in rural areas of Thailand, or other low-income countries. They wanted to test whether a much simpler and cheaper treatment regimen of AZT would be effective in reducing mother-to-child transmission of HIV.

- Would it have been ethical to give a placebo to the comparison group?
- If there was only one comparison group in the trial and this group received the same regimen as had been given in the trials in France and the USA, what would be the policy and programmatic implications if the trial showed that the group receiving the simpler, cheaper regimen had a transmission rate of 15%, while the group receiving the much more complex and expensive regimen had a transmission rate of 5%?

Careful consideration must be given to what intervention, if any, should be provided to the control group. An important general principle is that control participants should not be deprived of effective interventions that they would receive in the absence of the trial. Thus, where existing proven interventions are available in the local context, it is generally not appropriate

to use placebo controls. For two main reasons, it is often considered unacceptable to give the comparison group nothing at all.

1. The potential for bias due to lower compliance among the comparison group who receive nothing from the trial
2. The feeling among researchers and/or the community and/or members of the ethics committee, that this is unethical, because the control participants are not getting anything from the trial

If possible, it is desirable to try to choose therapies or other regimens for controls that are extremely unlikely to interfere with the trial outcome(s). Thus, for example, in a trial to measure the impact of pneumococcal vaccination in preventing pneumococcal disease in adults, one could consider providing anti-worm drugs to the control group. In this case, these drugs could also be provided to the vaccine group. Provision of meningococcal vaccine to the controls would not be sensible as this could well influence your outcome measures, because some of the outcomes (e.g. meningitis) can be caused by both pneumococcal and meningococcal disease.

Another general principle is that intervention studies are only justified when there is genuine uncertainty as to the relative benefits of the two interventions. This is the principle of **equipoise**. The benefits considered in this decision need to be all-inclusive. For example, although researchers may want to know whether providing a piped water supply into someone's house is more effective in reducing childhood diarrhoea than building a single borehole for a village, the decision as to whether this comparison would be ethical should take into account the other potential benefits as well as those on diarrhoea, such as decreasing the time and effort expended by household members collecting water, the potential costs to the families concerned, etc.

2.4. Choice of sample size

The trial should be of a sufficient size to answer the key hypotheses. It's always a good idea to start thinking about possible sample sizes for your trial very early on in the design stage. In order to do this, you will need to calculate the expected number of events in the control group. You should then set the sample size so that you have a high chance of detecting strong evidence of a difference between the groups that are of importance clinically or in terms of the impact on public health. Large sample sizes are required to detect small differences.

Note that in the control group, the number of outcome events might lower than expected due to the better care that is usually available to all participants during the trial. When calculating the sample size, investigators must consider the number of outcome events in the intervention and control groups, along with likely number of losses to follow-up and the expected adherence to the intervention.

2.5. Allocation to intervention and control groups

When allocating the intervention, the aim is to obtain two groups which are similar in all respects other than the introduction of the preventive or therapeutic measure under study. Unless this is done, the underlying differences between the study groups may be responsible for any differences in outcome. For example, when two drugs are being compared for their effectiveness in treating pneumonia, if one group of participants includes more young children than the other group, a fair comparison of the effectiveness of the two drugs will not be possible since young children are more likely to get more pneumonia than older children and adults.

In order to check the similarity of the groups, the collection of baseline data on the participants is useful. These **baseline data** should include all the variables which are known or thought to affect the outcome(s) of interest.

However, there will always be risk/protective factors which have not been measured either because they are not known to be risk/protective factors, or because they are difficult or expensive to measure. And there may be many more risk/protective factors which individually affect the outcome only slightly, but collectively could affect it substantially. Furthermore, in many intervention studies the participants are not all recruited at the same time. Trying to ensure that the groups will eventually resemble each other for a large number of potential risk factors by conscious selection processes would be extremely difficult. No matter how conscientiously the investigators try to eliminate biases related to the allocation, such biases may still creep in due to either conscious or subconscious choices on the part of the people who do the allocation.

The best way to try to get around these **selection biases** is to entirely remove the element of human choice from the whole allocation process. This can be done by the allocation being determined in advance, for example by allocating alternate participants to each treatment group. However, such **systematic** allocation can still result in substantial allocation biases if either the participants or the person doing the enrolling favour one group over the other and can influence the order of enrolment, or if there is some inherent structure within the order in which the participants are enrolled (e.g. alternate participants are male and female).

The best solution to this is to use **random allocation** of the participants to the treatment groups. This is where the `random' comes from in the title `**randomised controlled trial'**. The allocation should not be known in advance either by the participants or by the person doing the enrolment. This is known as **allocation concealment**. Depending on the trial the actual method of random allocation may be done in a number of ways, including paper-based (e.g. the ‘sealed envelopes’ described in the trial in the lecture), electronically (e.g. a telephone number or internet portal/app where the investigator can request the next participant to be randomised) or a combination (e.g. clusters have already been randomised to ‘A or B’ by an independent statistician and a coin is flipped at a public ceremony to decide which is control or intervention).

Sometimes **restricted randomisation** is used to generate a sequence to ensure particular allocation ratios to the intervention groups (e.g. 1:1). Blocked randomization (random permuted blocks) is a common form of restricted randomization. Blocking ensures that the numbers of participants to be assigned to each of the comparison groups will be balanced within blocks of, for example, five in one group and five in the other for every 10 consecutively entered participants. The block size may be randomly varied to reduce the likelihood of foreknowledge of intervention assignment. Also common is **stratified randomisation**, in which restricted randomisation is performed separately within strata. This generates separate randomisation schedules for subsets of participants defined by potentially important prognostic factors, such as disease severity and study centres.

Note that random allocation is not a completely fool-proof method of achieving similar groups, particularly when the number of units being allocated (e.g. individuals, families, communities) is relatively small. The methods for improving group similarity for small RCTs is beyond the scope of this module.

Lastly, when reporting the results of an RCT it is useful to present the baseline data by study arm, in order for the reader to assess whether the study arms were balanced (i.e. that the

profile of participants in either study arm was similar). However, it is not appropriate to show a p-value when comparing baseline characteristics by arm. This is because a p-value shows the probability that a difference has occurred due to chance – but if randomisation has been conducted correctly, by definition any differences between the study arms **must** be due to chance.

2.6. Collection of outcome data

Misclassification

The investigator's main aim in the assessment of the outcome(s) of an intervention study is to minimise **misclassification**. **Non-differential misclassification** (i.e. misclassification of outcome to a similar degree in all arms of the trial) will lead to an underestimate of the effect of the intervention. Outcome measures used in trials should ideally be **repeatable**, or reliable, and **valid**, or sensitive and specific.

The method of assessment should be standardised, so that all the participants, irrespective of their group assignment, are monitored for the outcome(s) of interest in an identical manner. The importance of this is far greater when the outcome(s) involve subjective judgments (such as the mental state of the participant).

Blinding

Differential misclassification is misclassification of outcome which occurs to a different degree in different arms of the trial. This bias can arise both from the participant and from the observer, but can be minimised by **blinding** or ensuring that neither the participant nor the observer know which group the participant is in. When reporting the results of the trial, the investigator will state who was blinded/masked, i.e. the participants, intervention deliverers, outcome assessors or other study staff.

Keeping both the participant and observer ignorant of the treatment group allocation is not always possible. For example, the participant cannot be masked if the study is comparing the effect on the incidence of malaria parasitaemia of sleeping under a mosquito net versus the incidence among individuals who are given weekly antimalarial tablets. However, it should be possible to keep the microscopists (observers) masked. Conversely, in a study comparing the recurrence of inguinal hernia among patients receiving two different surgical procedures which involve slightly different incisions, it might be possible to keep the participants masked, but not the surgeons who make the assessment of the outcome.

Biases related to participants

Bias can be induced from the following scenarios:

- Recruited persons who decline participation
- Participants in the intervention arm who do not fully receive the intervention (“nonadherence”)
- Participants in the control arm who are exposed to the intervention (“contamination”)
- Participants who are lost to follow-up during the study.

Almost invariably, people who decline participation differ from people who consent to participate in terms of their risk of outcomes. Wherever possible, baseline data on at least some of the key risk factors should be collected even on people who decline participation, to allow the potential bias to be explored.

Similarly, participants who are lost to follow-up are usually different from those who have complete follow-up in terms of their underlying risk of outcomes. It is essential to make every effort to minimise the number of non-adherers and participants lost to follow-up, and to monitor both of them as much as possible. Also, for studies with long duration, it is beneficial to measure outcomes periodically, rather than to wait until the very end when the numbers who are lost to follow-up might have reached unacceptably high levels.

3. Classes of randomised controlled trials

One way of classifying trials is as an **efficacy** or an **effectiveness** trial. Efficacy trials are designed to estimate the maximum potential benefit to be derived from the intervention in ideal circumstances, including as close as possible to 100% **adherence**. Such ideal interventions are very rarely achieved in a real health programme. The benefits derived from a real-life programme are known as the programme's **effectiveness**. It is possible to design intervention studies, even RCTs, to measure effectiveness by randomising the intervention(s) to take place within routine (health) programmes.

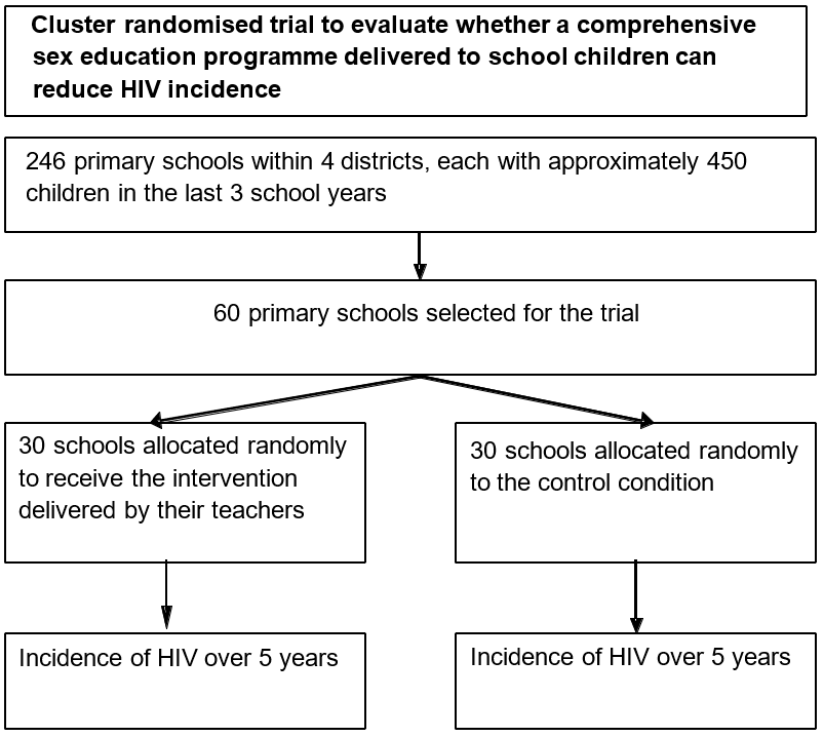
There are many different types of randomised controlled trial. Some are described further below.

3.1. Double-blind randomised controlled trial

The prototype randomised controlled trial (RCT) is the double-blind, placebo-controlled trial, in which participants are randomly allocated to either receive the intervention under test (e.g. a new drug) or a **placebo**, and neither the participants nor the investigators (including those doing the allocation and those doing the assessment of outcomes) know which each participant has received.

3.2. Cluster randomised trials

In **cluster randomised trials**, groups of individuals (known as clusters) are randomised, with all individuals within a cluster receiving the same intervention. This design is most appropriate when the intervention needs to be given at the cluster level, or there is a major risk of contamination between the trial arms. The design is sometimes used to make study logistics easier. However, cluster trials usually need to be much larger than individually randomised trials, and the analysis is more complex.



4. Ethical issues

Since the investigator is deliberately intervening rather than just observing, ethical considerations are even more important in intervention studies than in any other type of epidemiological study. Ensuring that the study is ethical is primarily the responsibility of the investigator(s). However, whether a study is judged to be ethical or unethical is a subjective judgment, based on cultural norms, which vary greatly between countries, groups within countries, and over time. However, WHO and the Council for International Organisations of Medical Sciences have published `Proposed international guidelines for biomedical research involving human participants', which provide a useful reference point (see Bibliography).

Issues which must always be considered are:

4.1. Has the question already been answered?

It is clearly unethical to conduct a trial if the answer to the question posed in the trial is already known beyond reasonable doubt. This may seem obvious, but it is not always as clear cut as may appear at first sight. Systematic reviews can be useful, but it may still be important to establish whether the results will be applicable to a specific population. For example, a trial in Indonesia in the early 1980s showed that six-monthly, high-dose vitamin A supplements reduced late infant and child mortality substantially. Yet investigators and ethics committees considered that it was still ethical for further placebo-controlled trials to be conducted in the same age group in other populations.

4.2. Informed consent

With very few exceptions, individuals must freely consent to take part, after having been told the aims, methods, anticipated benefits and potential hazards of the study. It is usual for individuals to give their consent in writing, though this may not be possible in all cases (e.g. low literacy populations). Sometimes it may be sufficient to obtain `community consent' rather

than individual consent in cluster trials, especially when the only outcomes that will be measured are aggregate/population-level outcomes. It can be acceptable for informed consent to be given by someone other than the individual participant (e.g. the parent or guardian of a child), or the guardian of a person who is unable to make an informed judgment for themselves (e.g. a mentally handicapped or an unconscious person).

4.3. Lack of coercion

Participants should be allowed to refuse to participate in the study or to withdraw their participation at any time during the study without any negative consequences to themselves, and they must be made aware of this right.

4.4. Confidentiality

The privacy of participants must be respected, and it is usual to ensure that the identity of individuals who participate in the research is not revealed to anyone who is not involved in the research team. Also, the researchers should not use any information obtained during the research to the detriment of the individuals or communities involved.

4.5. Ethics committee

It is very difficult for the investigators to remain objective and impartial in the decision as to whether something is ethical or not. The generally accepted way to manage this is for the proposed study to be vetted by a group of individuals who have nothing to do with the study themselves (the Ethics Committee).

5. Analysis

5.1. Measures of effect

The main measures of effect obtained from an intervention study (as in a cohort study) are the **risk ratio, rate ratio, risk difference,** or **rate difference**. The choice will depend in large part on the type of data collected. In addition, the **number needed to treat (NNT)** may also be reported (refer back to the notes for Lecture 2: Epidemiologic Measures Part II for a reminder of how these measures are calculated)**.**

5.2. Methods of analysis

It is important to develop an analysis plan ***before*** the analysis is started. This plan should include precise definitions of outcomes (divided into primary and secondary outcomes) and the analytic methods that will be used.

There are two methods that you will need to consider for the primary analysis, intention-totreat analysis or per protocol analysis. These are illustrated with an example in Box 5.

Box 5. Intention to treat and per protocol analysis: A trial of a pneumococcal vaccine.

In a large, randomised trial of pneumococcal conjugate vaccine, the aim was to measure the efficacy of the pneumococcal vaccine in preventing deaths in childhood. About 50,000 Gambian infants were randomised to receive either the pneumococcal or a control vaccine (Haemophilus influenzae Type B/DTP). Three doses were given at ages 2, 3, and 4 months of age. The analysis compared the frequency of deaths in the pneumococcal vaccine group against deaths in the control group.

• **In an "intention to treat" analysis, all follow-up and outcomes would be assigned to the group as randomised. Thus, if infants got the wrong vaccine or only received one dose for some reason, they would be left in the group to which they were randomised.** • **In a "per protocol" analysis, the results would be analysed according to what the infants actually received.**

In an intention to treat analysis:

1. Intervention participants who only received 1 or 2 doses remain in the analysis as being in the intervention arm.
2. The groups are "analysed as randomised". The full advantages of the randomisation are maintained, which is that the comparison groups remain balanced for both known and unknown confounders.
3. The analysis gives an estimate of effectiveness, that is the effect you might expect in "real world" conditions.
4. The results are not easily generalisable since the measured effect is based on a mixture of partial and full vaccination. Health planners in other countries, where the rates of vaccination coverage may be different, may want to know the effect of the vaccine when given exactly as recommended (i.e. an efficacy estimate from a "per protocol" analysis).

5.3. Meta-analysis

Results from different trials can be combined in a meta-analysis, in which results from relevant randomised trials answering the same research question are combined and a summary measure of intervention effect is estimated. The summary measure is weighted, usually based on the inverse of their variance. This means that, in general, larger studies will contribute more to the summary measure than smaller studies. The results of meta-analyses are often presented in a forest plot, where each study is shown with its effect size and the corresponding 95% confidence interval.

6. Interpretation and policy implications

6.1. Magnitude and importance of observed effect

Trials should only be set up to look for effects which would be of important benefit to the individuals concerned and (in future) to others who are similar to them. Also, when interpreting the results of an intervention study, it is important to evaluate the effect both in terms of the magnitude of the effect and strength of evidence.

6.2. Consider alternative explanations for the results

- Could the results be due to biases or confounding?
- Was randomisation effective?
- Was blinding effective?
- Did the study have the power to detect an important effect?

6.3. Monitoring adverse effects and costs

Wherever possible, potential adverse effects (e.g. side effects of drugs) should be monitored as well as the main outcomes of interest.

Information on the cost of the intervention is also needed by policy makers who have to decide whether any observed benefits from the intervention are worth the costs. It is much easier to collect these intervention costs and to separate them from the research costs if this is set up prospectively.

6.4. Policy importance

In assessing the public health policy importance of the results from an intervention study, the mere finding of evidence in a research study is not sufficient. The acceptability of the intervention, its cost, the frequency and severity of the condition(s) it is targeted at, and hence the overall potential impact in the total population needs to be assessed.

6.5. Efficacy vs. effectiveness

When interpreting the public health importance of the impact of an intervention observed in a trial, it is crucially important to know whether it was an efficacy or an effectiveness trial. If it was an efficacy trial, it must be remembered that the impact of the intervention may be lower, perhaps substantially lower, when the intervention is applied through a routine programme, if the coverage and/or the quality of the intervention is lower.

7. Non-randomised intervention studies

There are a range of intervention study designs that do not use random assignment to allocate treatment. The designs described below are more susceptible to bias and confounding than for an RCT. However, RCTs are sometimes not feasible either for technical reasons (e.g. evaluations of national mass media campaigns to promote healthier eating habits) or are not considered appropriate by the implementing agency. Non-randomised studies can still be very useful, especially if they are conducted with scientific rigour, and the investigators try to rule out (as much as is possible) alternative explanations for any differences observed (Habicht et al. 1999, Victora et al. 2004).

7.1 Before-after studies

Here the rate (or risk) of the outcome(s) is compared before an intervention has been introduced with the rate (or risk) after its introduction. This design involves the use of historical controls. Though simple, it suffers from the fact that it is impossible to completely rule out the possibility that any changes that are observed might have occurred irrespective of the intervention (e.g., due to `secular trends'). However, this is the design most commonly used in evaluations of health services.

7.2 Quasi-experimental studies

Here the participants who are given the intervention and those who act as the control group are not chosen randomly. For example, people who volunteer to receive the intervention are compared with those who do not volunteer. This design suffers from unmeasurable and potentially substantial selection bias and/or confounding and should be avoided if it is possible to do randomised allocation. However, they can also be the only practical way of having a contemporaneous comparison group in some situations (e.g. large-scale behavioural interventions, interventions using mass media approaches, etc), or when randomisation is not possible for ethical reasons. For example, a quasi-experimental study was used to assess the impact of various social distancing interventions on the spread of COVID, as a randomised trial would not be possible. The study assessed the association of the number of new COVID cases and deaths in European countries with social restrictions in place, taking into account the time when the cases, deaths and restrictions took place. They found that lifting stay-at home orders and allowing reopening of non-essential businesses would have relatively little effect on spread of COVID-19. However, the conclusions were cautious and stated that careful monitoring of how lifting each control measure affects transmissibility of COVID-19 is required and "will help to minimise the inevitably imperfect results".

References

Bailey et al. "Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomized controlled trial." *Lancet*. 2007; 369: 643-56. DOI https://doi.org/10.1016/S01406736(07)60312-2.

Bailey L, Vardulaki K, Langham J and Chandramohan D, *Introduction to Epidemiology*. Chapter 8. Open University Press, 2005 (Understanding Public Health, Series editors: Nick Black and Rosalind Raine).

Dos Santos Silva, I. Chapter 7 in *Cancer Epidemiology: Principles and Methods*. IARC, Lyon, France. 1999

Habicht, J.P., Victora, C.G. and Vaughan, J.P., 1999. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *International journal of epidemiology*, *28*(1), pp.10-18.

Hennekens CH & Buring JE. *Epidemiology in Medicine*, Chapter 8. Little, Brown and Company, 1987.

Hunter PR et al; 2021. Impact of non-pharmaceutical interventions against COVID-19 in Europe in 2020: a quasi-experimental non-equivalent group and time series design study. Euro Surveill. 26 (28): 2001401. doi: 10.2807/1560-7917.ES.2021.26.28.2001401.

Victora, C.G., Habicht, J.P. and Bryce, J., 2004. Evidence-based public health: moving beyond randomized trials. *American journal of public health*, *94*(3), pp.400-405.

http://www.consort-statement.org is a website describing a consensus statement/checklist for the reporting of randomised trials, including explanations for the inclusion of items on the checklist. This checklist is now being used by most of the top medical journals for reviewing manuscripts reporting the findings of randomised trials.