Clinical Trials and Intervention Studies

The Intervention Study

In a controlled observational **cohort study**, two groups of subjects are selected from two populations that (hopefully) differ in only one characteristic at the start. The groups of subjects are studied for a specific period and contrasted at the end of the study period. For instance, smokers and nonsmokers are studied for a period of 10 years, and at the end the proportions of smokers and nonsmokers that died in that period are compared. On the other hand, in an intervention study, the subjects are selected from one population with a particular characteristic present; then, immediately after baseline, the total study group is split up into a group that receives the intervention and a group that does not receive that intervention (control group). The comparison of the outcomes of the two groups at the end of the study period is an evaluation of the intervention. For instance, smokers can be divided into those who will be subject to a smoking-cessation program and those who will not be motivated to stop smoking.

Interventions have the intention to improve the condition of an individual or a group of individuals. Some examples of intervention studies in public health research are studies that evaluate the impact of a program: (a) to promote a healthier lifestyle (avoiding smoking, reducing alcohol drinking, increasing physical activity, etc.), (b) to prevent HIV-transmission, (c) to start brushing teeth early in babies, and so on. Ample intervention studies can also be found in other disciplines; two examples illustrate this. First, Palmer, Brown, and Barrera [4] report on an intervention study that tests a short-term group program for abusive husbands against a control program. The two groups are compared with respect to the recidivism rates of the men regarding abuse of their female partners. Second, Moens et al. [1] evaluated in a controlled intervention study the effect of teaching of how to lift and transfer patients to nursing students in a nursing school. After two years of follow-up, the incidence risk of one or more episodes of back pain was compared between the two groups of nursing students.

Controlled clinical trials constitute a separate but important class of intervention studies. There, the aim is to compare the effectiveness and safety of two (or more) medical treatments or surgical operations or combinations thereof. Clearly, now the target population constitutes patients with a specific disease or symptom. More aspects of clinical trials will be highlighted in section 'Typical Aspects of Clinical Trials'.

Intervention studies are often applied on an individual level but they can also be applied on a group level. For instance, promoting better brushing habits for children could be done on an individual basis, for example, by means of a personal advice to the parents of the child, or on a group basis, for example, by introducing special courses on good brushing habits in school. Intervention studies operating on a group level need dedicated statistical methods. We will start with the intervention studies on individual level but come back to intervention studies on group level in section 'Intervention Studies on Group Level'.

Basic Aspects of an Intervention Study

The first step in any intervention study is to specify the target population, which is the population to which the findings should be extrapolated. This requires a specific definition of the subjects in the study prior to selection. In a clinical trial, this is achieved by specifying inclusion and exclusion criteria. In general, the inclusion criteria specify the type of patients who need the treatment under examination and the exclusion criteria exclude patients for which there will be most likely safety concerns or for which the treatment effect might not be clear, for example, because they are already on another, competing, treatment.

To obtain a clear idea about the effect of the intervention, the two groups (intervention and control) should be comparable at the start. More specifically, at baseline, the two groups should be selected from the same population – only in that case a difference between the two groups at the end of the study is a sign of an effect of the intervention. Comparability or balance at baseline is achieved by randomly allocating subjects to the two groups; this is known as

Reproduced from the *Encyclopedia of Statistics in Behavioral Science*. © John Wiley & Sons, Ltd.
ISBN: 0-470-86080-4.

randomization. Simple **randomization** corresponds to tossing a coin and when (say) heads, the subject will receive the intervention and in the other case (s)he will be in the control group. But other randomization schemes exist, like block- and stratified randomization (*see* **Block Random Assignment** and **Stratification**). It is important to realize that randomization can only guarantee balance for large studies and that random imbalance can often occur in small studies.

For several types of intervention studies, balance at baseline is a sufficient condition for an interpretable result at the end. However, in a clinical trial we need to be more careful. Indeed, while most interventions aim to achieve a change in attitude (a psychological effect), medical treatments need to show their effectiveness apart from their psychological impact, which is also called the *placebo effect*. The **placebo effect** is the pure psychological effect that a medical treatment can have on a patient. This effect can be measured by administering placebo (inactive medication with the same taste, texture, etc. as the active medication) to patients who are blinded for the fact that they haven't received active treatment. Placebo-controlled trials, that is, trials with a placebo group as control, are quite common. When only the patient is unaware of the administered treatment, the study is called *single-blinded*. Sometimes, also the treating physician needs to be blinded, if possible, in order to avoid bias in scoring the effect and safety of the medication. When patients as well as physician(s) are blinded, we call it a double-blinded clinical trial. Such a trial allows distinguishing the biological effect of a drug from its psychological effect.

The advantage of randomization (plus blinding in a clinical trial) is that the analysis of the results can often be done with simple statistical techniques such as an unpaired t Test for continuous measurements or a chi-squared test for categorical variables. This is in contrast to the analysis of controlled observational cohort studies where **regression models** are needed to take care of the imbalance at baseline since subjects are often self-selected in the two groups.

To evaluate the effect of the intervention, a specific outcome needs to be chosen. In the context of clinical trials, this outcome is called the *endpoint*. It is advisable to choose one endpoint, the primary endpoint, to avoid multiple-testing issues. If this is not possible, then a correction for multiple testing such as a Bonferroni adjustment (*see* **Multiple Comparison Procedures**) is needed. The choice of the primary endpoint has a large impact on the design of the study, as will be exemplified in the section 'Typical Aspects of Clinical Trials'. Further, it is important that the intervention study is able to detect the anticipated effect of the intervention with a high probability. To this end, the necessary sample size needs to be determined such that the **Power** is high enough (in clinical trials, the minimal value nowadays equals 0.80).

Although not a statistical issue, it is clear that any intervention study should be ethically sound. For instance, an intervention study is being set up in South Africa where on the one hand adolescents are given guidelines of how to avoid HIV-transmission and on the other hand, for ethical reasons, adolescents are given general guidelines to live a healthier life (like no smoking, etc.). In clinical trials, ethical considerations are even more of an issue. Therefore, patients are supposed to sign an informed consent document.

Typical Aspects of Clinical Trials

The majority of clinical trials are drug trials. It is important to realize that it takes many years of clinical research and often billions of dollars to develop and register a new drug. In this context, clinical trials are essential, partly because regulatory bodies like the Food and Drug Administration (FDA) in the United States and the European Medicine Agency (EMEA) in Europe have imposed stringent criteria on the pharmaceutical industry before a new drug can be registered. Further, the development of a new drug involves different steps such that drug trials are typically subdivided into phases. Four phases are often distinguished. Phase I trials are small, often involve volunteers, and are designed to learn about the drug, like establishing a safe dose of the drug, establishing the schedule of administration, and so on. Phase II trials build on the results of phase I trials and study the characteristics of the medication with the purpose to examine if the treatment should be used in large-scale randomized studies. Phase II designs usually involve patients, are sometimes double blind and randomized, but most often not placebo-controlled. When a drug shows a reasonable effect, it is time to compare it to a placebo or standard treatment; this is done in a phase III trial.

This phase is the most rigorous and extensive part of the investigation of the drug. Most often, phase III studies are double-blind, controlled, randomized, and involve many centers (often hospitals); it is the typical controlled clinical trial as introduced above. The size of a phase III trial will depend on the anticipated effect of the drug. Such studies are the basis for registration of the medication. After approval of the drug, large-scale studies are needed to monitor for (rare) adverse effects; they belong to the phase IV development stage.

The typical clinical trial design varies with the phase of the drug development. For instance, in phase I studies, an **Analysis of variance** design comparing the different doses is often encountered. In phase II studies, **crossover designs**, whereby patients are randomly assigned to treatment sequences, are common. In phase III studies, the most common design is the simple parallel-group design where two groups of patients are studied over time after drug administration. Occasionally, three or more groups are compared; when two (or more) types of treatments are combined, a factorial design is popular allowing the estimation of the effects of each type of treatment.

Many phase III trials need a lot of patients and take a long time to give a definite answer about the efficacy of the new drug. For economic as well as ethical reasons, one might be interested in having an idea of the effect of the new drug before the planned number of patients is recruited and/or is studied over time. For this reason, one might want to have interim looks at the data, called *interim analyses*. A clinical trial with planned interim analyses has a so-called group-sequential design indicating that specific statistical (correction for multiple testing) and practical (interim meetings and reports) actions are planned. Usually, this is taken care of by an independent committee, called the *Data and Safety Monitoring Board* (DSMB). The DSMB consists of clinicians and statisticians overlooking the efficacy but especially the safety of the new drug.

Most of the clinical trials are superiority trials with the aim to show a better performance of the new drug compared to the control drug. When the control drug is not placebo but a standard active drug, and it is conceived to be difficult to improve upon the efficacy of that standard drug, one might consider showing that the new drug has comparable efficacy. When the new drug is believed to have comparable efficacy and has other advantages, for example, a much cheaper cost, a noninferiority trial is an option. For a noninferiority trial, the aim is to show that the new medication is not (much) worse than the standard treatment (*see* **Equivalence Trials**). Currently, noninferiority trials are becoming quite frequent due to the difficulty to improve upon existing therapies.

The choice of the primary endpoint can have a large impact on the design of the study. For instance, changing from a binary outcome evaluating short-term survival (say at 30 days) to survival time as endpoint not only changes the statistical test from a chi-square test to, say, a logrank test but can also have a major practical impact on the trial. For instance, with long-term survival as endpoint, a group-sequential design might become a necessity.

Despite the fact that most clinical trials are carefully planned, many problems can occur during the conduct of the study. Some examples are as follows: (a) patients who do not satisfy the inclusion and/or exclusion criteria are included in the trial; (b) a patient is randomized to treatment A but has been treated with B; (c) some patients drop out from the study; (d) some patients are not compliant, that is, do not take their medication as instructed, and so on. Because of these problems, one might be tempted to restrict the comparison of the treatments to the ideal patients, that is, those who adhered perfectly to the clinical trial instructions as stipulated in the protocol. This population is classically called the *per-protocol population* and the analysis is called the *per-protocol analysis*. A per-protocol analysis envisages determining the biological effect of the new drug. However, by restricting the analysis to a selected patient population, it does not show the practical value of the new drug. Therefore, regulatory bodies push the **intention-to-treat** (ITT) analysis forward. In the ITT population, none of the patients is excluded and patients are analyzed according to the randomization scheme. Although medical investigators have often difficulties in accepting the ITT analysis, it is the pivotal analysis for FDA and EMEA.

Although the statistical techniques employed in clinical trials are often quite simple, recent statistical research tackled specific and difficult clinical trial issues, like dropouts, compliance, noninferiority studies, and so on. Probably the most important problem is the occurrence of dropout in a clinical trial. For instance, when patients drop out before a

response can be obtained, they cannot be included in the analysis, even not in an ITT analysis. When patients are examined on a regular basis, a series of measurements is obtained. In that case, the measurements obtained before the patient dropped out can be used to establish the unknown measurement at the end of the study. FDA has been recommending for a long time the Last-Observation-Carried-Forward (LOCF) method. Recent research shows that this method gives a biased estimate of the treatment effect and underestimates the variability of the estimated result [6]. More sophisticated methods are reviewed in [7] (*see* **Missing Data**).

Intervention Studies on Group Level

Many intervention studies act on the group level; they are called *group-randomized studies*. For instance, Murray et al. [2] describe an intervention study to evaluate four interventions to reduce the tobacco use among adolescents. Forty-eight schools were randomized to the four interventions. After two years of follow-up, the proportion of children using 'smokeless tobacco' was compared. The proportions found in two of the four treatment groups were $58/1341 = 0.043$ and $91/1479 = 0.062$. A simple chi-square test gives a P value of 0.03. However, this test assumes independence of the subjects. When adolescents are motivated in a school context, there will be a high interaction among adolescents of the same class/school, that is, the outcome of one adolescent will depend on the outcome of another adolescent. Hence, the chi-square test is not appropriate. An adjusted chi-square test taking the correlation among the adolescents into account (see [2]) gives a P value of 0.18.

In general, the appropriate statistical techniques for group-randomized studies need to take the correlation among subjects in the same group into account (*see* **Intraclass Correlation**). This implies the use of techniques like **Generalized Estimating Equations** (GEE) and random effects models; see, for example, [7] and **Linear Multilevel Models** and **Generalized Linear Models (GLM)**.

Further Reading

An excellent source for clinical trial methodology can be found in [5]. Intervention studies operating on group level gained in importance the last decade; for an overview of these designs, we refer to [3].

References

[1] Moens, G., Johannik, K., Dohogne, T. & Vandepoele, G. (2002). The effectiveness of teaching appropriate lifting and transfer techniques to nursing students: results after two years of follow-up, *Archives of Public Health* **60**, 115–223.

[2] Murray, D.M., Hannan, P.J. & Zucker, D. (1998). Analysis issues in school-based health promotion studies, *Health Education Quarterly* **16**(2),, 315–330.

[3] Murray, D.M. (1998). *Design and Analysis of Group-randomized Trials*, Monographs in Epidemiology and Biostatistics, 27, Oxford University Press, New York.

[4] Palmer, S.E., Brown, R.A. & Barrera, M.E. (1992). Group treatment program for abusive husbands: long-term evaluation, *American Journal of Orthopsychiatry* **62**(2), 276–283.

[5] Piantadosi, S. (1997). *Clinical Trials: A Methodological Perspective*, John Wiley & Sons, New York.

[6] Verbeke, G. & Molenberghs, G. *Linear Mixed Models in Practice: A SAS-oriented Approach*, Lecture Notes in Statistics 126, Springer-Verlag, New York, 1997.

[7] Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*, Springer Series in Statistics, Springer-Verlag, New York.

EMMANUEL LESAFFRE and GEERT VERBEKE