Chapter 2
Introduction to Clinical Research Concepts, Essential Characteristics of Clinical Research, Overview of Clinical Research Study Designs

Stephen P. Glasser

To educate is to guide students on an inner journey toward more truthful ways of seeing and being in the world.

(Parker J. Palmer) [1]

Abstract This chapter addresses some of the central concepts related to clinical research such as sampling, hypothesis generation, and what is meant by the strength of scientific evidence. We also begin to discuss the different clinical research designs along with their respective strengths and weaknesses.

Keywords Sampling • Hypothesis • Prospective and retrospective cohort design • Case-control design • Case cohort design • Cross-sectional design • Type I and type II error

Principles for the conduct of research are set forth in internationally recognized documents such as the Declaration of Helsinki and the Guideline for Good Clinical Practice (GCP) of the International Conference on Harmonization (ICH-see Chap. 6). The principles of these and other standards are translated into legal requirements through laws and regulations that are enforced by national authorities such as the US FDA (see Chap. 6). The issues addressed by GCP include such things as protecting research subjects, ensuring objectivity in research, communication information about clinical trials, informed consent, and the very conduct of clinical trials including independent review and safety monitoring. In recent years clinical research has been discussed in the lay media, and this has (mostly) negatively impacted recruitment (also see Chap. 8).

S.P. Glasser, M.D. (✉)
Division of Preventive Medicine, University of Alabama at Birmingham,
1717 11th Ave S MT638, Birmingham, AL 35205, USA
e-mail: sglasser@uabmc.edu

S.P. Glasser (ed.), *Essentials of Clinical Research*, DOI 10.1007/978-3-319-05470-4_2,

Sampling

An essential characteristic and the goal of any clinical research is to make inferences from the population under study (the sample or study population) and apply those inferences to a broader population (the target population i.e. the population about which we want to draw conclusions). Imagine if the investigator could only learn about and apply the results in the sample population? Rather we must be able to extrapolate the results of the findings in the sample population to a broader group of patients-otherwise the results would have little utility (Fig. 2.1). Thus, one of the most important weaknesses of any study is that inferences drawn from a study are based on a limited sample (again, a sample is a select subset of a population that the investigator hopes represents the general population perfectly, but which is unlikely to do so). This aforementioned limitation is further compounded by the fact that disease is not distributed randomly, whereas samples tend to be, and that the causes of disease are multifactorial. Thus, ideally, when performing clinical research, we would like to include everyone in our study who has the disease of interest. Because this is impossible we settle for a sample of the diseased population, however, the researcher now has to deal with a degree of uncertainty (see Chap. 18). Because different samples contain different

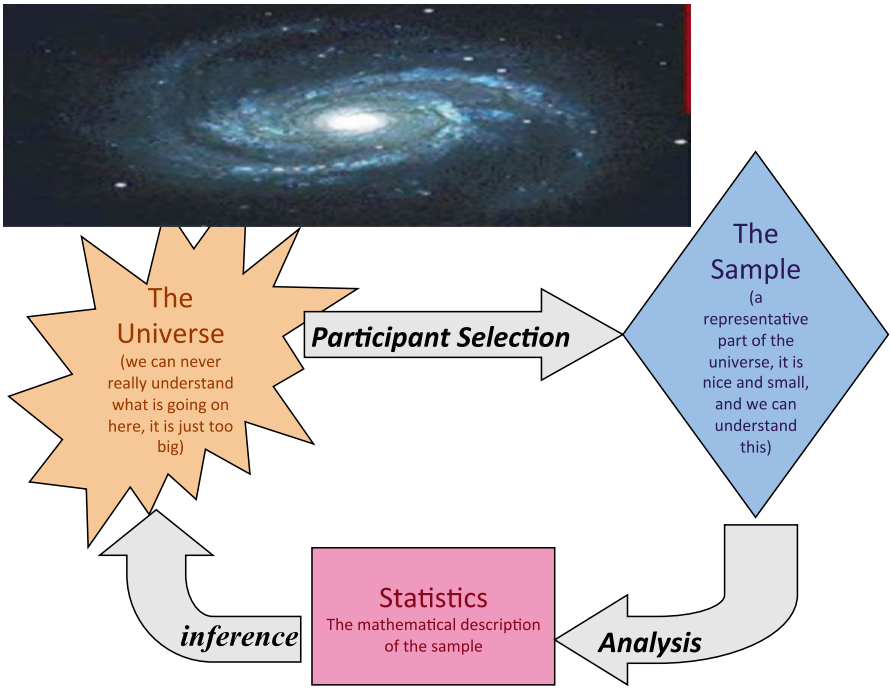


Fig. 2.1 The sample and how it relates to the universe

Table 2.1 Potential sampling errors

Selecting study participants
Selection bias
Non-respondent bias:
Volunteer or referral bias
External validity
Sampling bias
Ascertainment bias
Prevalence-incidence bias
Berkson bias
Healthy worker effect
Detection bias: the risk factor investigated itself may lead to increased
Diagnostic
Overmatching bias

people with different co-morbidities, and differing experiences, we end up with different data. The question now facing the researcher is which data from which sample is most representative of the entire population? Sampling errors commonly result in type I and type II errors. For example, if the researcher finds a certain effect of an interventional therapy, the question to be asked is 'how likely is it that this therapy observation that was made from this sample is falsely representing the total population (that is the intervention in the sample population shows no effect, but if the total population had been exposed to the intervention there would have been an effect)? This potential false result is a type II error. The reverse situation is a total population would in fact have a therapy effect, but the sample studied shows no such effect. This is a type I error and is reflected by the p value.

Sampling bias is also a major problem (Table 2.1). For example, considering who responds to certain types of advertisement to recruit subjects can bias the sample. If random digit telephone dialing is used, subjects who do not have a phone cannot be recruited, if newspaper advertisement is utilized people who do not read newspapers cannot respond, etc.

A suggested solution to the sampling issue is to use random sampling; but, random sampling does not guarantee 'good' sampling. As an example, consider If you draw repeated random samples of size 100 and 1,000 from a population with 50 % women the largest and smallest number of women in a sample of 100 can range from 33 to 68 and in a sample of 1,000 from 450–550.

The Linear-Semilinear Relationship of Biological Variables

Another important concept of clinical research is the fact that most, if not all biological variables have a linear–semilinear relationship in terms of exposure and outcomes, whereas clinical medicine is replete with the use of 'cut-points' to separate

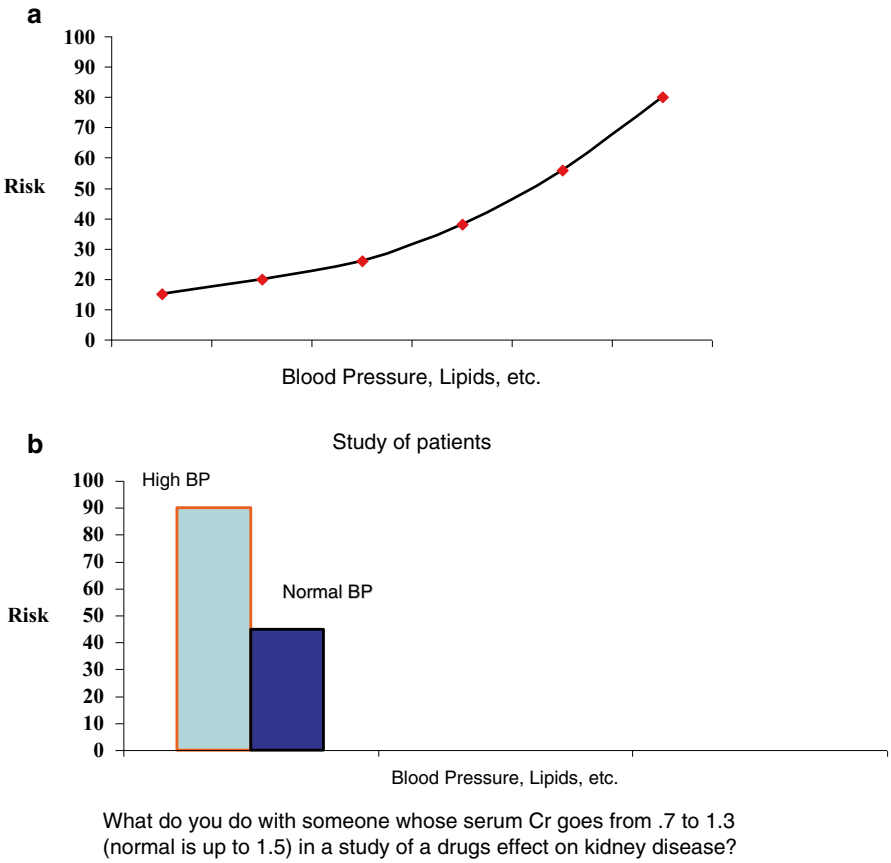


Fig. 2.2 A comparison of the epidemiologic way of determining risk (**a**) vs. the clinical way (**b**)

normal and abnormal or effect and no effect (Fig. 2.2a, b). A cut-point presumes that there is some value or range of values that separates normal form abnormal rather than considering that the relationships tend to be linear.

Strength of Relationships

Additionally, clinical research relates to what we mean when we talk about 'the strength of evidence.' The greatest strength of evidence is often attributed to the randomized clinical trial (RCT). In fact, in response to the question of what is the best clinical research design, the answer generally given is 'the RCT,' when in fact the correct answer should be 'it depends,' an answer which will be further discussed later in this book. What is actually meant by 'the highest level of

evidence' is how certain we are that an exposure and outcome are causally related, that is, how certain we are that an effect is the result of a given cause, and that the observations do not just reflect that an association exists; but, that they are not causally related.

The Hypothesis

Let's return to the question: 'What is the best study design?' This is a different question from 'What is the best study design for a given question, and given the specific question, which study design leads to the highest level of evidence?'; which may finally be different from asking 'What is the study design for a given question that will result in the greatest certainty that the results reflect cause and effect?' This latter question is really the one that is most often sought, and is the most difficult to come by (see Chap. 16). Other important factors in considering the most appropriate study design, besides the most important factor—ethics—include the natural history of the disease being studied, the prevalence of the exposure, disease frequency, the characteristics and availability of the study population, measurement issues, and cost.

Let us now return to our quest for 'universal truth.' What are the steps we need to take in order to achieve 'truth'? The fact is that truth is at best elusive and is not actually achievable since truth is more a function of our interpretation of data, which is in part dictated by our past experiences, than any finite observation that is absolute. The steps needed to achieve this uncertain quest for truth begins with a research question, perhaps the result of a question asked during teaching rounds, or stimulated by contact with a patient, or provoked during the reading of a book or journal, and so on. The research question is usually some general statement such as 'Is there an association between coffee drinking and myocardial infarction (MI)?' or 'Is passive smoke harmful to a fetus?' Let us examine this last research question and consider its limitations in terms of a testable hypothesis. In addressing a question such as 'Is passive smoke harmful to a fetus?' one needs first to ask a few questions such as: 'what is the definition of 'harmful'; how will passive smoke be measured and what do we mean by the term i.e. how is it to be defined in the study to be proposed?' Answering these questions comes nearer to something that is testable and begins to define the clinical research design that would have the greatest level of evidence with that specific question in mind. For the question proposed above, for example, it would be best, from a research design perspective, to randomize exposure of pregnant women to both passive smoke and 'placebo passive smoke.' But considering the ethics issue alone, this would not be acceptable; thus, an RCT would not be the 'best study design' for this research question, even if it would lead to the 'highest level of evidence'.

The hypothesis is generally (for the traditional approach of superiority testing) stated in the null (Ho). The alternative hypothesis (H_A) i.e. the one you are really interested in is, for example, that a new drug is better than placebo. That is, if one

wants to compare a new investigational drug to placebo, the hypothesis would be constructed in the null, i.e. that there is no difference between the two interventions. If one rejects the null, one can then say that the new drug is either better (or worse-depending on the results of the study) than placebo. By the way, if the null is not rejected one cannot say that the new drug is the same as placebo, one can only claim that no difference between the two is evident from these data (this is more than a nuance as will be discussed later).

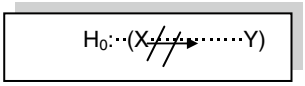
In order to understand why the hypothesis is stated in the null and why one cannot accept the null but only reject it, consider the following three examples (taking a trip with your family, shooting baskets with Michael Jordon, and contemplating the US legal system). Consider the scenario outlined by Vickers [2] where you have just finished packing up your SUV (a hybrid SUV no doubt) with all of your luggage, the two kids, and your dog, and just as you are ready to depart; your wife says 'honey, did you pack the camera?' At least two answers present themselves; one that the camera is in the automobile, or two that the camera is in the house. Given the prospect of unpacking the entire SUV, you decide to approach the question with, 'the camera is not in the house (Ho) i.e. it is in the car'. If you in fact do not find the camera in the house you have rejected your null and your assumption is that it is in the car. Of course, one can easily see that the camera could be in the house (you just did not find it), and even if you did such a thorough job of searching the house that you can be almost certain that it is not there, it still may not be in the car (you might have left it elsewhere (the office, a prior vacation, etc.)) Another way to look at this issue is to envision that you are out on the basketball court when Michael Jordon comes in. You challenge him to a free throw shooting contest and he makes 7 of 7 while you make 3 of 7. It turns out the p value for this difference is 0.07 i.e. there is no "statistically significant difference between the shooting skills of MJ and your shooting skills" you can draw your own conclusions about this likelihood [2]. In the Woman's Health Initiative (WHI), women eating a low fat diet had a 10 % reduction in breast cancer compared to controls P = .07. This was widely interpreted, as low fat diets don't work. In fact, the NY Times trumpeted that 'low fat diets flub a test' and that the study provided 'strong evidence that the war against all fats was mostly in vain'. This is what we call accepting the null hypothesis (i.e. it was not rejected so it was accepted) and is to be avoided i.e. failure to reject it does not mean you accept it, rather it means that these data do not provide enough evidence to reject it. By the way, guess what happens when the next study does reject the null-'but they said it did not work!'.

Finally, consider our Anglo-American legal system. *It is no mere coincidence that the logic of hypotheses testing in scientific inquiry is identical to that which evolved in the Anglo-American legal system and most of the following descriptions are taken from The Null Logic of Hypothesis Testing found on the World Wide Web* [3]. *Much of the pioneering work in the logic of hypothesis testing and inferential statistics was done by English mathematicians and refined by their American counterparts. For instance consider the contributions made by W.S. Gossett, R.A. Fisher, and Karl Pearson to the logic of hypothesis testing and statistical inference. The concept of the null hypothesis can be compared to the legal concept of guilty*

vs. non guilty, the latter of which does not mean innocence. What is interesting is that the guilt vs. innocent scenario involves two diametrically opposed logics, one affirmative and the other null. From the time a crime is reported to the police an affirmative, accusatory, and inductive logic is followed. Detective X gathers the evidence, follows the evidentiary trail, and based upon the standard of probable cause, hypothesizes that the accused is guilty and charges him accordingly. The District Attorney reviews the case for probable cause and quality of evidence and affirms the accusation. The case is argued affirmatively before the grand jury, and they concur. But relative to the jury, at the point the trial begins, the logic is reversed, it is no longer affirmative, it becomes null. The jury, the trier of the facts, is required to assume that the defendant is not guilty unless the facts established otherwise. Let's abstract this two part logical process and represent it symbolically. The police, the prosecutor, and the grand jury hypothesized (H_A) that the accused (X) committed the crime (Y).

$$H_A : (X \rightarrow Y)$$

The jury on the other hand hypothesizes (H_0) that the accused (X) was not guilty of the crime (Y) unless the evidence reached the standard of "beyond a reasonable doubt".



Formulating the logic in this manner, one can be certain of three things. Either:

H_0 is true, the accused is not guilty, or
H_A is true, accused is guilty,
and
H_0 and H_A cannot both be true.

The logic of establishing someone's guilt is not the simple converse of the logic of establishing his/her innocence. For instance, accusing someone of a crime and requiring them to prove their innocence requires proving a negative, something that is not logically tenable. However, assuming that someone is not guilty and then assessing the evidence to the contrary is logically tenable (Fig. 2.3).

The decision matrix in Table 2.1 shows the possible outcomes and consequences of this legal logic as applied to the case of the accused, our hypothetical defendant. Assume H_0: the accused is not guilty unless the evidence is convincing beyond a reasonable doubt. Notice that in terms of verdicts and outcomes, there are two kinds of errors the jury might have made, identified as (I) and (II).

Type I Error The jury finds the accused guilty when in fact he is not guilty.
Type II Error The jury finds the accused not guilty when in fact he is guilty.

Compare this with the Table 18.2

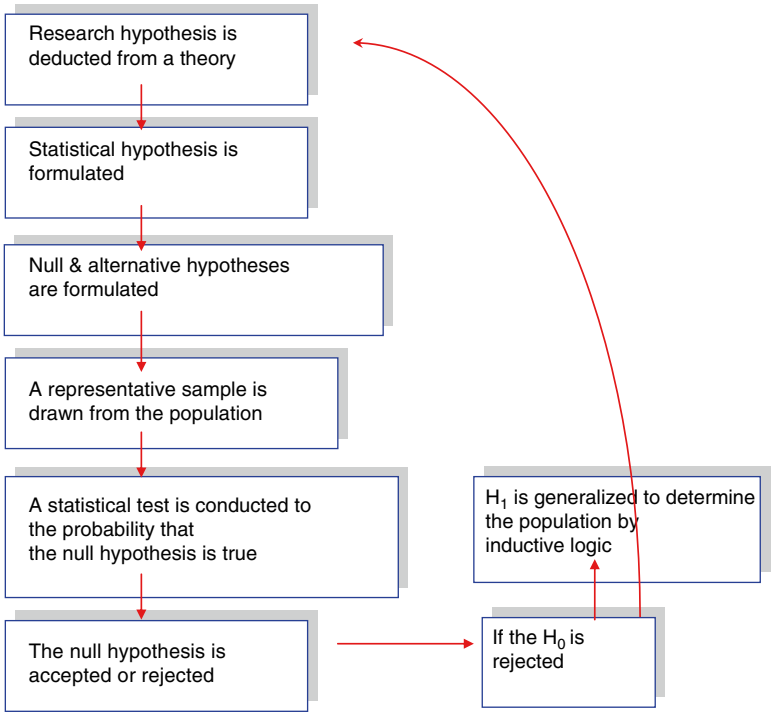


Fig. 2.3 Deductive and inductive logic of hypothesis testing

In the Anglo-American legal tradition, the consequences of these two possible errors are not considered equivalent. On the contrary, considerable safeguards have been incorporated into the criminal law to minimize the probability (α) of making a Type I error (convicting an innocent person), even at the risk of increasing the probability (β) of making a Type II error (releasing a guilty person). Indeed, this is where the concept of innocent until proven guilty comes from, and the quote: Finally, this logic also assumes that justice is better served if, as the noted 18th Century English jurist Sir William Blackstone stated, "…ten guilty persons escape than that one innocent suffer" [4, p. 358] (Fig. 2.4).

It is logical and critical to distinguish between the concepts of not guilty and innocent in the decision paradigm used in criminal law, i.e.:

If H_A = guilty, then does …
H_0 = not guilty, or does …
H_0 = innocent?

Here, "not guilty" does not mean the same thing as innocent. A not guilty verdict means that the evidence failed to convince the jury of the defendant's guilt beyond a reasonable doubt (i.e. "The scientific corollary is that data in this study was

Finally, this logic also assumes that justice is better served if, as the noted 18th Century English jurist Sir William Blackstone stated, "...ten guilty persons escape than that one innocent suffer." (Blackstone 1753-65)

http://en.wikipedia.org/wiki/William_Blackstone

Fig. 2.4 Sir William Blackstone quote regarding guilt and innocence

insufficient to determine if a difference exists, rather than there is no difference"). By this logic it is quite conceivable that a defendant can be found legally not guilty and yet not be innocent of having committed the crime in question.

The evaluation of a hypothesis involves both deductive and inductive logic. The process both begins and ends with the research hypothesis.

Step 1 Beginning with a theory about the phenomenon of interest, a research hypothesis is deduced.

This hypothesis is then refined into a statistical hypothesis about the parameters in the population.

The statistical hypothesis may concern population means, variances, medians, correlations, proportions, or other statistical measures.

The statistical hypothesis is then reduced to two mutually exclusive and collectively exhaustive hypotheses that are called the null (H_0) and alternative hypothesis (H_A).

Step 2 If the population is too large to study in its entirety (the usual case), a representative sample is drawn from the population with the expectation that the sample statistics will be representative of the population parameters of interest.

Step 3 The data gathered on the sample are subjected to an appropriate statistical test to determine if the sample with its statistical characteristics could have come from the associated population if the null hypothesis is true.

Step 4 Assuming that the null hypothesis (H_0) is true in the population, and that the probability that the sample came from such a population is very small ($p \leq 0.05$), the null hypothesis is rejected.

Step 5 Having rejected the null hypothesis, the alternative hypothesis (H_A) is accepted, and, by inductive inference is generalized to the population from whence the sample came.

These five steps are illustrated in Fig. 2.3, that is, the conduct of research involves a progressive generation of four kinds of hypotheses: Research hypothesis, Statistical hypothesis, Null hypothesis; and, Alternative hypothesis.

A research hypothesis is an affirmative statement about the relationship between two variables. For instance, consider the following example of a research hypothesis: "there is a positive correlation between the level of educational achievement of citizens and their support of rehabilitation programs for criminal offenders". From the research hypotheses three other kinds of hypotheses can be formulated:

A statistical hypothesis
A null hypothesis
An alternative hypothesis

Again, a statistical hypothesis is a statement about the parameters of a population. The null hypothesis, which is symbolized H_0, is the negative statement of the statistical hypothesis; and, the alternative hypothesis, usually symbolized H_A, is the obverse of the null hypothesis and by custom, is stated to correspond to the research hypothesis being tested. Statements that are mutually exclusive are such that one or the other statement must be true. They cannot both be true at the same time. For instance:

Something is either "A" or "not A". It cannot be both "A" and "not A" at the same time.
Or, the object on the kitchen table is either an apple or a non-apple.
Saying the object on the kitchen table is either an "apple" or a "non-apple" covers every possible thing that the object could be.

It is critical to understand that it is the null hypothesis (H_0) that is actually tested when the data are statistically analyzed, not the alternative hypothesis (H_A). Since H_0 and H_A are mutually exclusive, if the analysis of the data leads to the rejection of the null hypothesis (H_0), the only tenable alternative is to accept the alternative hypothesis (H_A). But, this does not mean that the alternative hypothesis is true, it may or may not be true. When we reject the null hypothesis it is because there is only a remote possibility that the sample could have come from a population in which the null hypothesis is true. Could we be wrong? Yes, and that probability is called alpha (α), and the error associated with alpha is called a Type I error (Table 2.2).

What about the converse situation, accepting the null hypothesis? If the null hypothesis is accepted, the alternative hypothesis may or may not be false. For example, the null hypothesis may be accepted because the sample size was too small to achieve the required degrees of freedom for statistical significance; or, an uncontrolled

Table 2.2 Compares the US legal system determination of guilt and innocence, to the Ho and Ha

	The verdict	**The verdict**
The truth	Accused in not guilty	Accused is guilty
Accused is not guilty (Ho true)	Justice is served	Innocent person is convicted probability = α
Accused is guilty (Ho false)	Guilty person is set free probability = β	Justice is served

extraneous variable or spurious variable has masked the true relationship between the variables; or, that the measures of the variables involved are grossly unreliable, etc. The issue is the same as a "not guilty" verdict in a criminal trial. That is, a verdict of not guilty does not necessarily mean that the defendant is innocent, it only means that the evidence was not sufficient enough to establish guilt beyond a reasonable doubt. There is a further discussion about the null hypothesis in Chap. 18.

An Overview of the Common Clinical Research Designs (Tables 2.3 and 2.4)

The common clinical research designs are listed in Tables 2.3 and 2.4 and summarizes some of their characteristics. There are many ways to classify study designs but two general ways are to separate them into descriptive and analytic studies and observational and experimental studies. These designations are fairly straightforward. In descriptive studies one characterizes (describes) a group of subjects; for example 'we describe the characteristics of 100 subjects taking prophylactic aspirin in the stroke belt.' In contrast, with analytic studies where there is a comparator group, for example, 'we compared the characteristics of 100 subjects in the stroke belt taking aspirin to 100 subjects not taking aspirin'. In experimental studies the investigator is "controlling" the intervention in contrast to observational studies where the exposure (intervention) of interest is occurring in nature and as the investigator you are observing the subjects with and without the exposure. Basically, experimental trials are clinical trials, and if subjects are randomized into the intervention and control (comparator) group it is a RCT.

Ecologic Studies

An ecological study is an epidemiological study in which the unit of analysis is a population rather than an individual. Ecologic studies are usually regarded as inferior to non-ecological designs such as cohort and case-control studies because of ecological fallacy (ecological fallacy refers to when inferences about the nature of individuals are deduced from inference for the population to which those individuals belong).

Table 2.3 General overview of study types

Observational
Ecologic studies
Case reports
Case series
Cross-sectional
Case-control
Cohort
Experimental
Clinical trials
Group clinical trials

Table 2.4 Types and descriptions of observational trials

Descriptive	Definition	Best used	Limitations
Case-series	Describes clinical course of one or more patients	Identify pathological, disease or treatment patterns	No comparison group
Ecologic	Associations of exposures and outcomes over time extracted from large databases	Trends over time	Impossible to adjust for confounding; Ecologic fallacy
Cross-sectional	Associations in a population at a single point in time	Generate data for further study	No temporality

Ecological studies can be easily confused with cohort studies, especially if different cohorts are located in different places. The difference is that in the case of ecological studies there is no information available about the individual members of the populations compared; whereas in a cohort study the data pair exposure/health is known for each individual. Ecologic studies use available population data to determine associations. For example, to determine an association between coronary heart disease (CHD) and the intake of saturated fat, one could access public records of beef sales in different states (or counties or regions of the country) and determine if an association existed between sales and the prevalence of CHD. Another example is that one might look for geographical correlations between disease incidence or mortality and the prevalence of risk factors. For example, mortality from coronary heart disease in local authority areas of England and Wales has been correlated with neonatal mortality in the same places 70 and more years earlier. This observation generated the hypothesis that coronary heart disease may result from the impaired development of blood vessels and other tissues in fetal life and infancy.

Case Reports and Case Series

Case reports and case series are potential ways to suggest an association, but, although limited in this regard, should not be deemed unimportant. For example, the recognition of the association of the diet drug combination of Fen-phen was the

result of a case series [5]. These are, for the most part, descriptive observations about a single patient or a group of patients relative to some outcome of interest. It can be retrospective or prospective and usually involves a smaller number of patients than more powerful case-control studies or randomized controlled trials. Case series may be *consecutive* or *non-consecutive,* depending on whether all cases presenting to the reporting authors over a period were included, or only a selection. Case series may be confounded by selection bias, which limits statements on the causality of correlations observed; for example, physicians who look at patients with a certain illness and a suspected linked exposure will have a selection bias in that they have drawn their patients from a narrow selection (namely their hospital or clinic).

Cross-Sectional Studies

In cross-sectional studies, one measures and/or describes disease status (or outcome), exposure(s), and other characteristics at a point in time (point in time is the operative phrase), in order to evaluate associations between them. Cross-sectional studies are different from cohort studies in that cohort studies observe the association between a naturally occurring exposure and outcome (e.g., between health and a disease or between disease and an event) over a period of time rather than at a point in time. With cross-sectional studies, the exposure and outcome are evaluated at a point in time-i.e. there is no follow-up period where a subsequent evaluation of exposure/outcome is observed. Indeed, this measure "at a point in time" is both the strength and weakness of the cross-sectional (X-sectional) study design. Lack of a follow-up period means the study can be performed more rapidly and less expensively than a cohort study, but one sacrifices temporality (an important component for determining causality). In addition, because cross-sectional studies are evaluating cases (disease, outcomes) at a point in time, one is dealing with prevalent cases (not incident cases as is true of a cohort study). Confusing to some is that a X-Sectional study may take years to complete, so it is not the duration of the study that determines whether it is a x-sectional or a cohort design, it is the time between the exposure and outcome that makes that determination. In other words, if the exposure and outcome are measured at a single point in time, it is x-sectional. If the outcome is ascertained at some time point distant from the exposure it is a cohort study. There are a number of factors that must be considered when using prevalence (rather than incidence) and these are summarized in Fig. 2.5.

An example of a cross sectional study might be the assessment of arterial stiffness and hormone replacement therapy (HRT). Let's say a study is designed where age matched women receiving HRT are compared to women not taking HRT, and arterial stiffness is measured in each to determine if differences in arterial stiffness differ between the two groups. Some have likened this to taking a snapshot of the association at that point in time.

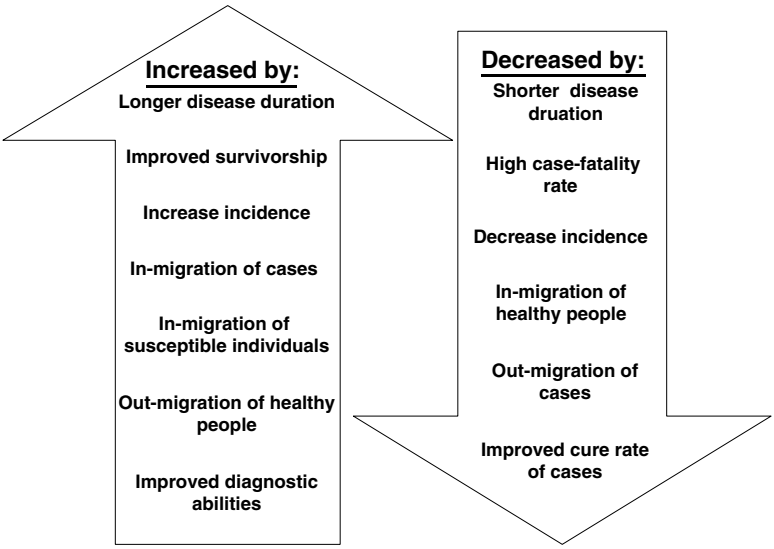


Fig. 2.5 The balance of factors that affect prevalence

Case-Control Study

In a case-control study (CCS), the investigator identifies a certain outcome in the population, then matches the 'diseased group' to a 'healthy group,' and finally identifies differences in exposures between the two groups.

With a CCS one approaches the study design the opposite of a cohort design (in fact some have suggested the use of the term 'trohoc design' – cohort spelled backwards). The term case-control study was coined by Sartwell to overcome the implication that the retrospective nature of the design was an essential feature [6]. That is, patients with the outcome of interest are identified, a control group is selected, and one then looks back for exposures that differ between the two. Two major biases exist with the CCS; first the selection of the control group is problematic, and second, one is usually looking back in time (i.e. it is a retrospective study in that sense). Selecting the control group for a CCS is problematic because if one selects too many matching criteria it becomes difficult to find an adequate control group, while if one has too few matching criteria, the two groups can differ in important variables. For CCS designs, recall bias is also an issue (this is even a greater issue if death is an outcome, in which case one not only has to deal with recall bias, but the recall is obtained from family members, caregivers, etc. rather than the subject).

One of the strengths of the CCS design is that if one is interested in a rare disease, one can search the area for those cases, in contrast to randomly selecting a cohort population that will develop this rare disease infrequently, even over a long follow-up time period. Also, in contrast to a cohort study in which the sample

population is followed for a time period, a CCS obviates this need so one can complete the study much more rapidly (and therefore less expensively).

There are several variations of the case-control design that overcome some of the shortcomings of a typical CCS (although they have their own limitations): a prospective CCS and a nested CCS. In the prospective CCS, one accrues the cases over time (i.e. in a prospective fashion) so that recall bias is less of an issue. However, one then has to wait until enough cases are accrued (problematic again for rare diseases); and, the selection of an appropriate control group still exists. A nested case-control study is a type of study design where outcomes that occurred during the course of a cohort study or RCT are compared to controls selected from the same cohort or clinical trial population who did not have the outcome. Compared with the typical case-control study, a nested case-control study can reduce 'recall bias' and temporal ambiguity, and compared with a cohort study, it can reduce cost and save time. One additional drawback of a nested case-control study is that the non-diseased persons from whom the controls are selected may not be fully representative of the original cohort as a result of death or failure to follow-up cases. As mentioned, the nested CCS design can be placed within a cohort study or RCT. An example is taken from the Cholesterol and Recurrent Events (CARE) Study [7]. The primary study was aimed at the prevention of recurrent MI when patients with a prior MI and 'normal' cholesterol levels were further treated with pravastatin. As part of the original study plasma was stored and after the report of the primary study was published the following study was designed: "we conducted a prospective, nested case-control study in the Cholesterol and Recurrent Events (CARE) trial. Baseline concentrations of VLDL-apolipoprotein (apo) B (the VLDL particle concentration), VLDL lipids, and apoCIII and apoE in VLDL + LDL and in HDL were compared in patients who had either a myocardial infarction or coronary death (cases, n = 418) with those in patients who did not have a cardiovascular event (control subjects, n = 370) in 5 years of follow-up. VLDL-cholesterol, VLDL-triglyceride, VLDL-apoB, apoCIII and apoE in VLDL + LDL and apoE in HDL were all interrelated, and each was a univariate predictor of subsequent coronary events. Adjustment for LDL- and HDL-cholesterol did not affect these results" [7].

Cohort Study

A cohort study is much like a RCT except that the intervention in an RCT is "investigator controlled", while in a cohort study the intervention (exposure) is a naturally occurring phenomenon. A cohort design is a study in which two or more groups of people that are "free of disease" at study onset and that differ according to the extent of exposure (e.g. exposed and unexposed) are compared with respect to disease incidence. A cohort study assembles a group of subjects and follows them over time. One follows these subjects to the development of an outcome of interest and then compares the characteristics of the subjects with and without the outcome in order to identify risk factors (exposures) for that outcome. A major assumption

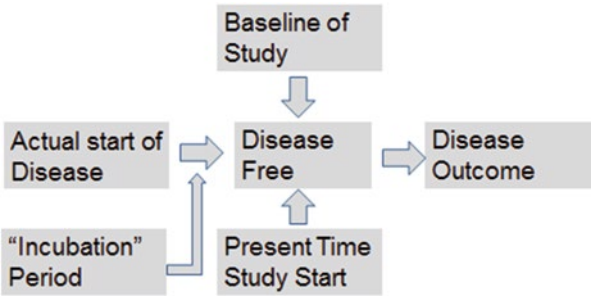


Fig. 2.6 The cohort limitation

made in cohort studies is that the subject is disease free at the beginning of the study (disease free means for the outcome- disease- of interest). For example, if the outcome of interest is a *recurrent* myocardial infarction, the subject would have had the first infarction (so in that sense he is not disease free) but in terms of the outcome of interest (a second infarction) we assume that at study onset, he is not having a second infarction. This example may seem obvious, but let us use colon cancer as another example. At study onset, one assumes that the subject is disease free (cancer-free or 'normal') at the time of enrollment, while in fact he or she may already have colon cancer that is as yet undiagnosed. This could bias the results of the study since the exposure of interest may have nothing to do with the outcome of interest (colon cancer) since the subject already has the outcome irrespective of the exposure (say a high fat diet). This also raises the issue as to what is 'normal'. One whit suggested that a normal subject is one that has been insufficiently tested! The cohort assumption mentioned above is diagrammed in Fig. 2.6. Of course, one also assumes that the incorrect assumption of no disease at onset is equally balanced in the two groups under study, and that is indeed the hope, but not always the realization. Cohort studies are considered the best way to study prognosis, but one can also do this by using a case-control design.

As an example, recall the cross-sectional study described above (the example of a cross sectional study that assessed the association of arterial stiffness and hormone replacement therapy). Let's say a study is designed where age matched women receiving HRT are compared to women not taking HRT, and arterial stiffness is measured in each to determine if differences in arterial stiffness differ between the two groups. Suppose now we follow subjects for 5 years, measure their arterial stiffness, and determine if there is a difference in that measure in women receiving HRT compared to those who are not. This would be a cohort design.

Retrospective Cohort Design

Cohort studies are generally prospective; however, retrospective cohort studies do exist. The key to the study design is identifying the exposure of interest in 'normal' subjects without disease (i.e. the outcome of interest), evaluate for that outcome

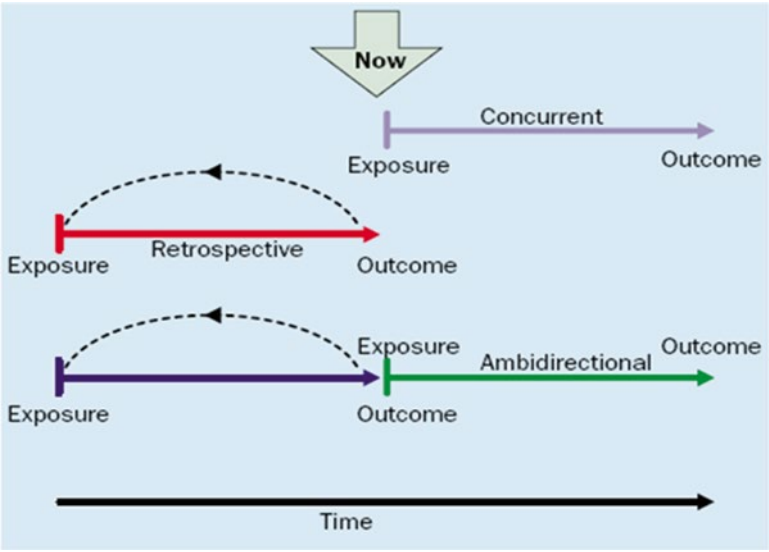


Fig. 2.7 A comparison of prospective and retrospective cohort study designs

after a period of time has elapsed, and determining the exposure as different or not in those with and without the outcome. Retrospective cohort studies are particularly well suited to the study of long-term occupational hazards. An example of a retrospective cohort study is the study of nickel refinery workers where about 1,000 nickel refinery workers were identified from company records and their outcomes identified over a prior 10 year period. Sixteen were found to have died from lung cancer (expected rate was 1 from National data), 11 died from nasal cancer (1 expected) and 67 from other causes (72 expected) [8].

Or, to continue with our HRT example from above, suppose we now identify a group of women who have and have not been taking HRT and we now measure their arterial stiffness and make comparisons of association to HRT (Fig. 2.7). It is common that this design is confused with case control studies. The differentiating factor is whether one designs the study to evaluate whether the exposure (e.g. HRT in this example) is associated with the outcome (arterial stiffness), this would be a cohort design; or, if subjects are identified by their outcome (say normal vs. abnormal arterial stiffness) and then exposure status is determined (they did or did not take HRT), this would be a case control design.

Case Cohort Design

Another modification of cohort studies is the case-cohort design. With the case-cohort design, a 'subcohort' is randomly selected from the cohort sample, a separate exposure of interest from the total cohort is identified, and cases (outcomes)

are then determined in the same manner as the primary design. An example might be a cohort study of 10,000 subjects that is assessing some outcome-let's say a CVD outcome- in relation to dietary fat. The investigator decides that she would also like to know the association of CVD with a measure of coronary artery calcium, so electron beam computed tomography (EBCT-a relatively expensive procedure to perform on the all of the original cohort) is measured in a random sample of 100 of the cohort subjects (the 'subcohort'). The association of EBCT to CVD outcome is then ultimately determined. A key feature of this design is that the cases are selected from among those with disease, while the controls are selected at the beginning of the study period, irrespective of disease status (that is some control cases may later become a case).

Randomized Control Trial (RCT)

In the randomized-controlled trial (RCT), the exposure is "controlled" by the investigator, which contrasts it to all the other study designs. A detailed discussion of the RCT will be presented in Chap. 3. However, it should be noted that RCTs cannot be used to address all important questions. For example, observational studies are more appropriate when studies are used to detect rare or late consequences of interventions, situations not best suited to the RCT.

The above discussion of study designs is not meant to be all-inclusive. For example there is a design called the "case-only design" that is somewhat unique to genetic studies. The case-only design is an efficient and valid approach to screening for gene-environment interaction under the assumption of the independence between exposure and genotype in the population. That is, if the primary purpose of the study is to estimate the effect of gene-environment interaction in disease etiology, one can do so without employing controls, thus, the case-only design requires fewer cases than the case-control design to measure gene-environment interaction, and it also requires fewer cases to measure gene-gene interactions.

One should now be able to begin to understand the key differences, and therefore limitations, of each study design; and, circumstances where one design might be preferable to another. Let's, for example, use the exposure of electromagnetic energy (EME) and cancer outcome (e.g. leukemia). With a cross-sectional study, a population is identified (target population), cancer rates determined, and exposure and lack of exposure to EME is ascertained from a sample population. One then analyzes the exposure rates in subjects with cancer and those that are cancer free. If the cancer rate is higher in those who were exposed, an association is implied. This would be a relatively inexpensive way to begin to look at the possible association of these variables, but limitations should be obvious. For example, since there is no temporality in this type of design, and since biologically, exposure to EME if it did cause cancer would likely have to occur over a long period of time, one could easily miss an association. Also reverse causation cannot be ruled out. Also remember, that even though the RCT is generally the "best"

Table 2.5 Common study designs, uses and limitations

Descriptive	Definition	Best used	Limitations
Cohort	Comparison of outcome of those with and without exposure	Rare exposure, common outcomes	Lack of randomization, bias from dropouts
Case-cohort	Exposure between cases and random sample of the original cohort	Rare exposure and outcome, long latency period	Recall bias; not suitable for chronic conditions
Case-crossover	Each case contributes one case window of time and one or more control windows	Outcome does not vary over time; exposures are brief	Recall bias; not suitable for chronic conditions
Cross-sectional	Description of associations at a single point in time	Outcome associations to generate further study	No temporality
Case-control	Odds of exposure among cases c/w non-cases	Common exposure rare outcome	Selection and recall bias; confounding
Nested case-control	Case-control nested within cohort (or clinical trial)	Rare outcome and/or long latency period	Decreases biases of case-control Design

study design, one could easily see why it would not be appropriate for this research question. Table 2.5 summarizes a few of the study designs in relation to the frequency of the exposure and outcome.

In summary, it should be evident that observational studies (e.g. cross-sectional, case-control, and cohort studies) have a major role in research. However, despite their important role, von Elm et al. discussed the lack of important information that was either missing or unclear in prior published observational studies; and why this lack of information led to a guideline document for reporting observational studies (the STROBE statement-the Strengthening and Reporting of Observational Studies in Epidemiology). The STROBE statement was designed after the CONSORT- the Consolidated Standards of Reporting Trials-; this statement outlines the guidelines for reporting RCTs. The STROBE statement is a checklist of 22 items that are to be considered essential for good reporting of observational studies (also see Chap. 19) [9].

Formulating relevant and precise questions that can be answered can be complex and time consuming. A structured approach for framing questions that uses five components may help facilitate the process. This approach is commonly known by the acronym “PICOS” or “PECOS”, where each letter refers to a component as follows:

- P refers to the patient population or the disease being addressed,
- I (or E) refers to the intervention or exposure
- C refers to the comparator group
- O to the outcome or endpoint
- S refers to the study design chosen

Finally, the spectrum of evidence imparted by the different clinical research designs ranges from ecological studies through observational epidemiological studies to randomized control trials (RCTs). And, many people are becoming increasingly skeptical of RCTs. In fact, one researcher has claimed that 90 % of medical research is wrong [10]. Examples include: Two 1993 studies concluded that vitamin E prevents cardiovascular disease. That claim was overturned in 1996 and 2000; a 1996 study concluding that estrogen therapy reduces older women's risk of Alzheimer's was overturned in 2004; and, a major study concluded there's no evidence that statins help people with no history of heart disease –the cost of statins is more than $20 billion per year, of which half may be unnecessary. Jeffry Hyman has reviewed this subject and has published an On-Line tutorial that addresses this question [11]. Hyman points out the following in his presentation entitled "Is Most Medical Research Wrong? The Role Of Incentives And Statistical Significance: a myriad of biases are present in any type of research that includes selection bias, information bias (see Chap. 17) and a number of analytical issues (see Chap. 3). In addition, Hyman points out the power of incentives by raising the questions of whether we are looking for the truth when we do research… or are we? and is the search for truth our only reason for doing research? could we have any other incentives? In answer to this latter question he raises the financial and egocentric motivations for doing research beyond seeking the truth, such as:

- We want to get grants
- We have a financial interest in the study
- We might want to support funding for a program
- We might want to continue funding for a program
- We want tenure
- We want a promotion
- We think there is an association and we want to show it
- We want our studies to be published
- We want publicity
- We might have done work in this area before and we want to replicate it
- We have made an Investment of time and money in doing the study
- we want to show results
- We want people to think we are a good researcher
- We know about publication bias towards negative results

He asks an additional question: how does our strong desire for a $P<0.05$ affect our results? and points to the following "follies": We do extensive modeling with a range of variables. Then we only report the model with the most significant results (selective reporting) (multiple comparisons), we do extensive subgroup analyses (multiple comparisons), we compare extreme groups, such as the 1st and 5th quintiles, we use too large of a sample size for the effect we want to measure (such as national surveys), we use a 1 sided P value, we don't try to publish papers with negative results, we quickly do studies in hot fields, we change study endpoints after looking at the data,

Table 2.6 Study designs by frequency of exposure and outcomes

	Prevalence or incidence of outcome	
Drug exposure	Not rare	Rare
Not rare	Cohort or clinical trial	Case-control
Rare	Cohort	Case-control

we investigate multiple associations between exposure and outcome, and, we selectively site the literature. A hypothetical (extreme) example of these latter concepts is presented in Hymans presentation (in whom he cites www.johndcook.com) follows: Researchers test 200 completely ineffective new drugs;

- About 10 trials out of the 200 will have a "significant" result due to chance.
- Only the 10 studies with significant results will be submitted for publication.
- Five of these studies are published in major journals
- Result: The type 1 error rate of each study was 5 %, but the error rate in the literature is 100 %

Hyman concludes with the question "Can we predict which studies are more likely to be wrong?". Here is a list of his answers: small studies with significant results; studies with more flexible designs, outcome measures, and models; studies with significant results and a small effect measure (like odds = 1.1); The hotter the field and the more people doing research in it; studies where there are strong financial interests; studies with strong pre-existing beliefs by researchers. In summary: be aware of how we overstate our results in an effort to get statistically significant results; be aware of the limitations of P values and statistical significance; don't over interpret significant results, being significant does not make a results true or important; on the other hand, not being significant does not make a result false; do power calculations for each study and don't make your study too big or too small, make it just right; watch for problems like multiple comparisons, subgroup analysis, and selective reporting; be aware of the situations where study results are more likely to be wrong; remember the effects of publication bias; and, in observational studies speak about associations, not causality (Table 2.6).

Finally, it has been pointed out by some, that clinical trials are too expensive, recruit too few patients, and results in to many investigators to just give up because of the cost and complexity of clinical trials (in fact it was noted that 38 % of PIs who participated in clinical trials between 2000 and 2005, did not return to conduct another clinical trial [12]). It has also been suggested that half of RCTs never finish due to recruitment problems, and many that do finish are underpowered to answer the original research question, even as costs soar. As a solution, it has been suggested that since observational trials give results similar to RCTs, and at less expense, they can be used as a substitute [13]. While Pocock and Elbourne warn that the one critical deficiency of observational designs is the absence of the randomization that occurs with RCTs rather than each patients treatment being deliberately chosen in observational trials [14].

References

1. Parker Palmer. Accessed at http://en.wikipedia.org/wiki/Parker_Palmer
2. Vickers AJ. Michael Jordan won't accept the null hypothesis: notes on interpreting high P values. Medscape. 2006;7:3.
3. The Null Logic of Hypothesis Testing. Accessed at http://www.shsu.edu/~icc_cmf/cj_787/research6.doc
4. Cited in The Null Logic of Hypothesis Testing. 1753–65. Accessed at http://www.shsu.edu/~icc_cmf/cj_787/research6.doc
5. Connolly HM, Crary JL, McGoon MD, Hensrud DD, Edwards BS, Edwards WD, et al. Valvular heart disease associated with fenfluramine-phentermine. N Engl J Med. 1997;337:581–8.
6. Cited in Sartwell P and Nathanson N. Epidemiol Rev; 1993.
7. Sacks FM, Pfeffer MA, Moye LA, Rouleau JL, Rutherford JD, Cole TG, et al. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and recurrent events trial investigators. N Engl J Med. 1996;335:1001–9.
8. Doll R. Cohort studies: history of the method II. Retrospective cohort studies. Soz Praventivmed. 2001;46:152–60.
9. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med. 2007;147:573–7.
10. Freedman D. A researchers claim: 90% of medical research is wrong. Health Fam. 2010;2:1–2.
11. Hyman J. Why most published research findings are false. Am J Epidemiol. 2005;2:e124.
12. CardioSource World News. Nov 2013:25–35.
13. Benson K, Hartz AJ. A comparison of observational studies and randomized controlled trials. N Engl J Med. 2000;342:1878–86.
14. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations. N Engl J Med. 2000;342:1907–9.

http://www.springer.com/978-3-319-05469-8

Essentials of Clinical Research
Glasser, S.P. (Ed.)
2014, VI, 461 p. 80 illus., 42 illus. in color., Hardcover
ISBN: 978-3-319-05469-8