

Chapter 4

Study Population

Defining the study population in the protocol is an integral part of posing the primary question. Additionally, in claiming an intervention is or is not effective it is essential to describe the type of participants on which the intervention was tested. Thus, the description requires two elements: specification of criteria for eligibility and description of who was actually enrolled. This chapter focuses on how to define the study population. In addition, it considers two questions. First, what impact does selection of eligibility criteria have on participant recruitment, or, more generally, study feasibility? Second, to what extent will the results of the trial be generalizable to a broader population? This issue is also discussed in Chap. 10.

In reporting the study, the investigator needs to say what population was studied and how they were selected. The reasons for this are several. First, if an intervention is shown to be successful or unsuccessful, the medical and scientific communities must know to what population the findings apply [1].

Second, knowledge of the study population helps other investigators assess the study's merit and appropriateness. Unfortunately, despite guidelines for reporting trial results [2], many publications contain inadequate characterization of the study participants [3]. Therefore, readers may be unable to assess fully the merit or applicability of the studies.

Third, in order for other investigators to be able to replicate the study, they need data descriptive of those enrolled. Before most research findings are widely accepted, they need to be confirmed by independent scientists. Although it is small trials that are more likely to be repeated, these are the ones, in general, that most need confirmation.

Fundamental Point

The study population should be defined in advance, stating unambiguous inclusion (eligibility) criteria. The impact that these criteria will have on study design, ability to generalize, and participant recruitment must be taken into account.

Definition of Study Population

The study population is the subset of the population with the condition or characteristics of interest defined by the eligibility criteria. The group of participants actually studied in the trial, which constitutes the trial participants, is selected from the study population. (See Fig. 4.1). There are two main types of exclusions. First, patients who have absolute or relative contraindications to the study intervention. Second, trial design issues that may interfere with the optimal conduct of the trial and factors that could interfere with participant adherence (see below).

The extent to which the obtained trial results can be generalized depends on its *external validity* [1]. External validity refers to the questions whether the trial findings are valid for participants other than those meeting the protocol definition of the study populations, but from a comparable clinical setting. Rothwell identified

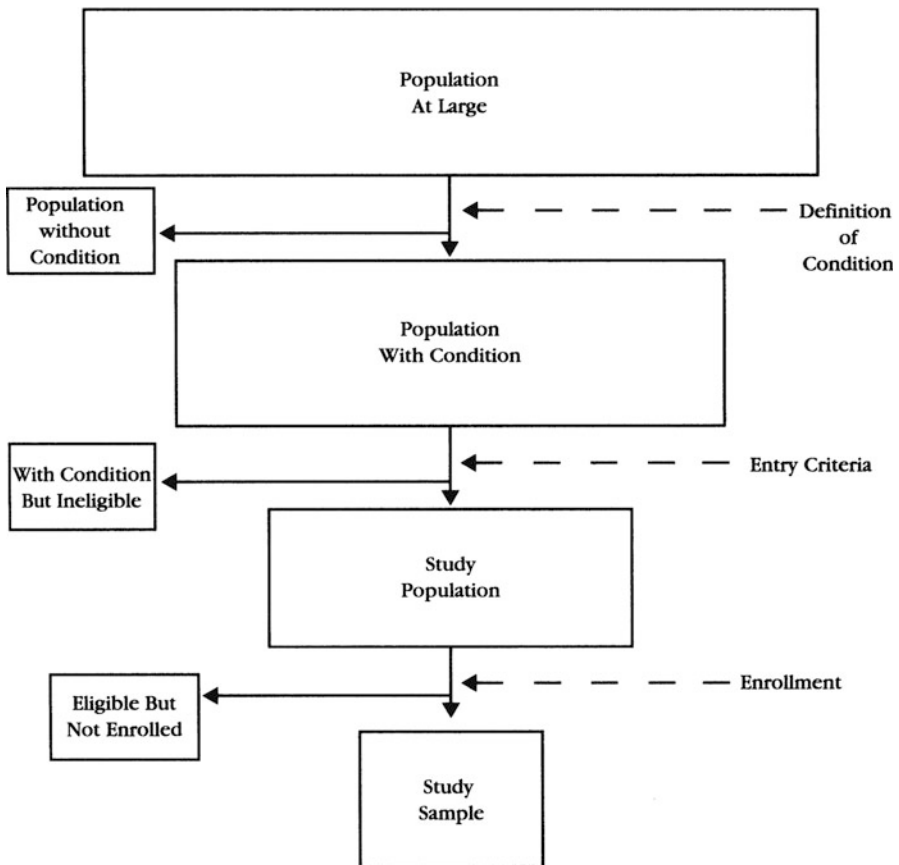


Fig. 4.1 Relationship of study sample to study population and general population (those with and those without the condition under study)

six issues that could potentially affect external validity—trial setting, selection of participants, characteristics of randomized participants, differences between the trial protocol and clinical practice, outcome measures and follow-up, and adverse effects of treatment. External validity is a measure of generalizability. The term *internal validity* refers to the question whether the trial results are valid for all participants meeting the eligibility criteria of the trial protocol, i.e., the definition of the study population.

Considerations in Defining the Study Population

Inclusion criteria and reasons for their selection should be stated in advance. Those criteria central to the study should be the most carefully defined. For example, a study of survivors of a myocardial infarction may exclude people with severe hypertension, requiring an explicit definition of myocardial infarction. However, with regard to hypertension, it may be sufficient to state that people with a systolic or diastolic blood pressure above a specified level will be excluded. Note that even here, the definition of severe hypertension, though arbitrary, is fairly specific. In a study of antihypertensive agents, however, the above definition of severe hypertension is inadequate. To include only people with diastolic blood pressure over 90 mmHg, the protocol should specify how often it is to be determined, over how many visits, when, with what instrument, by whom, and in what circumstances. It may also be important to know which, if any, antihypertensive agents participants were on before entering the trial. For any study of antihypertensive agents, the criterion of hypertension is central; a detailed definition of myocardial infarction, on the other hand, may be less important.

If age is a restriction, the investigator should ideally specify not only that a participant must be over age 41, for example, but *when* he must be over 41. If a subject is 40 at the time of a pre-baseline screening examination, but 41 at baseline, is he eligible? This should be clearly indicated. If valvular heart disease is an exclusion criterion for a trial of anticoagulation in atrial fibrillation, is this any significant valve abnormality, or is restricted to rheumatic heart disease? Does it apply to prior valve repair? Often there are no “correct” ways of defining inclusion and exclusion criteria and arbitrary decisions must be made. Regardless, they need to be as clear as possible, and, when appropriate, with complete specifications of the technique and laboratory methods.

In general, eligibility criteria relate to participant safety and anticipated effect of the intervention. It should be noted, however, that cultural or political issues, in addition to scientific, public health, or study design considerations, may affect selection of the study populations. Some have argued that too many clinical trials exclude, for example, women, the elderly, or minority groups, or that even if not excluded, insufficient attention is paid to enrolling them in adequate numbers [4–7]. Some patient groups may be underrepresented due to practical issues (the frail might not be able to attend frequent follow-up visits) and the need for informed

consent might exclude individuals with cognitive dysfunction. Policies from the U.S. National Institutes of Health now require clinical trials to include certain groups in enough numbers to allow for “valid analysis” [8]. The effect of these kinds of policies on eligibility criteria, sample size, and analysis must be considered when designing a trial.

The following five categories outline the framework upon which to develop individual criteria:

Potential for Benefit

Participants who have the potential to benefit from the intervention are obviously candidates for enrollment into the study. The investigator selects participants on the basis of his scientific knowledge and the expectation that the intervention will work in a specific way on a certain kind of participants. For example, participants with a urinary infection are appropriate to enroll in a study of a new antibiotic agent known to be effective *in vitro* against the identified microorganism and thought to penetrate to the site of the infection in sufficient concentration. It should be evident from this example that selection of the participant depends on knowledge of the presumed mechanism of action of the intervention. Knowing at least something about the mechanism of action may enable the investigator to identify a well-defined group of participants likely to respond to the intervention. Thus, people with similar characteristics with respect to the relevant variable, that is, a *homogeneous* population, can be studied. In the above example, participants are homogeneous with regard to the type and strain of bacteria, and to site of infection. If age or renal or liver function is also critical, these too might be considered, creating an even more highly selected group.

Even if the mechanism of action of the intervention is known, however, it may not be feasible to identify a homogeneous population because the technology to do so may not be available. For instance, the causes of headache are numerous and, with few exceptions, not easily or objectively determined. If a potential therapy were developed for one kind of headache, it would be difficult to identify precisely the people who might benefit.

If the mechanism of action of the intervention is unclear, or if there is uncertainty at which stage of a disease a treatment might be most beneficial, a specific group of participants likely to respond cannot easily be selected. The Diabetic Retinopathy Study [9] evaluated the effects of photocoagulation on progression of retinopathy. In this trial, each person had one eye treated while the other eye served as the control. Participants were subgrouped on the basis of existence, location and severity of vessel proliferation. Before the trial was scheduled to end, it became apparent that treatment was dramatically effective in the four most severe of the ten subgroups. To have initially selected for study only those four subgroups who benefited was not possible given existing knowledge. This is an example, of which there are many, of the challenge in predicting differential intervention effects

based on defined subgroups. For most interventions, there is uncertainty about the benefits and harms that makes enrolling a broader group of participants with the condition prudent.

Some interventions may have more than one potentially beneficial mechanism of action. For example, if exercise reduces mortality or morbidity, is it because of its effect on cardiac performance, its weight-reducing effect, its effect on the person's sense of well-being, some combination of these effects, or some as yet unknown effect? The investigator could select study participants who have poor cardiac performance, or who are obese or who, in general, do not feel well. If he chose incorrectly, his study would not yield a positive result. If he chose participants with all three characteristics and then showed benefit from exercise, he would never know which of the three aspects was important.

One could, of course, choose a study population, the members of which differ in one or more identifiable aspects of the condition being evaluated; i.e., a *heterogeneous* group. These differences could include stage or severity of a disease, etiology, or demographic factors. In the above exercise example, studying a heterogeneous population may be preferable. By comparing outcome with presence or absence of initial obesity or sense of well-being, the investigator may discover the relevant characteristics and gain insight into the mechanism of action. Also, when the study group is too restricted, there is no opportunity to discover whether an intervention is effective in a subgroup not initially considered. The broadness of the Diabetic Retinopathy Study was responsible for showing, after longer follow-up, that the remaining six subgroups also benefited from therapy [10]. If knowledge had been more advanced, only the four subgroups with the most dramatic improvement might have been studied. Obviously, after publication of the results of these four subgroups, another trial might have been initiated. However, valuable time would have been wasted. Extrapolation of conclusions to milder retinopathy might even have made a second study difficult. Of course, the effect of the intervention on a heterogeneous group may be diluted and the ability to detect a benefit may be reduced. That is the price to be paid for incomplete knowledge about mechanism of action.

Large, simple trials are, by nature, more heterogeneous in their study populations, than other sorts of trials. There is a greater chance that the participants will more closely resemble the mix of patients in many clinical practices. It is assumed, in the design, that the intervention affects a diverse group, and that despite such diversity, the effect of the intervention is more similar among the various kinds of participants than not. In such trials, not only are the interventions relatively easy to implement, and the baseline and outcome variables limited, so too are the eligibility criteria. Definitions of eligibility criteria may not require repeated visits or special procedures. They may rely on previously measured variables that are part of a diagnostic evaluation, or on variables that are measured using any of several techniques, or on investigator judgment. For example, a detailed definition of myocardial infarction or hypertension may be replaced with, "Does the investigator believe a myocardial infarction has occurred?" or "Is hypertension present?" The advantage of this kind of criteria is their simplicity and greater generalizability. The disadvantage is the possible difficulty that a clinician reading the results of the

trial will have in deciding if the results are applicable to specific patients under his care. It should be noted, however, that even with the large simple trial model, the criteria are selected and specified in advance.

Homogeneity and heterogeneity are matters of degree and knowledge. As scientific knowledge advances, ability to classify is improved. Today's homogeneous group may be considered heterogeneous tomorrow. Patients with mutations in BRCA1 and BRCA2 genes discovered in the 1990s have different susceptibility and course of breast and ovarian cancer. Patients with breast cancer tissue with HER2 and/or estrogen receptors respond differently to chemotherapy treatments [11]. Thus, breast cancer is now defined and treated based on genomically defined subsets.

High Likelihood of Showing Benefit

In selecting participants to be studied, not only does the investigator require people in whom the intervention might work, but he also wants to choose people in whom there is a high likelihood of detecting the hypothesized effects of the intervention. Careful choice will enable investigators to detect results in a reasonable period of time, given a reasonable number of participants and a finite amount of funding.

For example, in a trial of an antianginal agent, an investigator would not wish to enroll a person who, in the past 2 years, has had only one brief angina pectoris episode (assuming such a person could be identified). The likelihood of finding an effect of the drug on this person is limited, since his likelihood of having many angina episodes during the expected duration of the trial is small. Persons with frequent episodes would be more appropriate. One option is to enrich the population with high risk patients, as was done in the ROCKET-AF trial of rivaroxaban versus warfarin for stroke prevention in atrial fibrillation [12]. Patients were required to have three risk factors for stroke that resulted in a population with higher risk and higher stroke rate than the general population with indication for oral anticoagulation. This allowed for a smaller sample size, since the calculation of sample size (Chap. 8) takes into account the expected incidence of the primary outcome. The results were consistent across the risk levels of patients enrolled, and the FDA provided approval for the drug across the spectrum of risk, including even lower risk patients who were not included in the trial. Although one might have somewhat less confidence that the treatment is safe and effective in lower risk patients, trials of related drugs have subsequently shown consistency across risk and thus it seems reasonable to extrapolate to the lower risk population.

Another approach is to begin with a higher risk population and if the results from a first trial are positive, the investigator can then enroll groups with lower risk levels. The initial Veterans Administration study of the treatment of hypertension [13] involved people with diastolic blood pressure from 115 through 129 mmHg. After therapy was shown to be beneficial in that group, a second trial was undertaken using people with diastolic blood pressures from 90 to 114 mmHg [14]. The latter study suggested that treatment should be instituted for people with

diastolic blood pressure over 104 mmHg. Results were less clear for people with lower blood pressure. Subsequently, the Hypertension Detection and Follow-up Program [15] demonstrated benefit from treatment for people with diastolic blood pressure of 90 mmHg or above. The first trial of angiotensin converting enzyme inhibitors in heart failure, the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS) [16], enrolled 253 patients with advanced heart failure. There was a 40% relative risk reduction in mortality at 6 months with enalapril versus placebo. Subsequent larger trials defined the treatment effects in patients with less severe heart failure with lower event rates. Studies Of Left Ventricular Dysfunction (SOLVD) consisted of two individual trials. One involved symptomatic participants [17] and the other asymptomatic participants with reduced ejection fraction [18].

Medical conditions with low event rates represent a challenge. One example is the relapse-remitting disease, multiple sclerosis. Its attack or relapse rate is reported to average 0.54 episodes annually with a slightly higher rate in the first year [19]. Properly designed clinical trials in this population would have to be very large and/or have a long duration. Similarly, many people accept the hypothesis that LDL-cholesterol is a continuous variable in its impact on the risk of developing cardiovascular disease. Theoretically, an investigator could take almost any population with moderate or even relatively low LDL-cholesterol, attempt to lower it, and see if occurrence of cardiovascular disease is reduced. However, this would require studying an impossibly large number of people. From a sample size point of view it is, therefore, desirable to begin by studying people with greater levels of risk factors and a consequent high expected event rate.

Generally, if the primary response is continuous (e.g., blood pressure, blood sugar, body weight), change is easier to detect when the initial level is extreme. In a study to determine whether a new drug is antihypertensive, there might be a more pronounced drop of blood pressure in a participant with diastolic pressure of 100 mmHg than in one with diastolic pressure of 90 mmHg or less. There are exceptions to this rule, especially if a condition has multiple causes. The relative frequency of each cause might be different across the spectrum of values. For example, familial hypercholesterolemia is heavily represented among people with extremely high LDL-cholesterol. These lipid disorders may require alternative therapies or may even be resistant to usual methods of reducing LDL-cholesterol. In addition, use of participants with lower levels of a variable such as cholesterol might be less costly due to lower screening costs [20]. Therefore, while in general, use of higher risk participants is preferable, other considerations can modify this.

Sometimes, it may be feasible to enroll people with low levels of a risk factor if other characteristics elevate the absolute risk. For example, the Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) [21] used C-reactive protein to select those with LDL-cholesterol levels under 130 mg/dL (3.4 mmol/L) but who were likely to be at higher risk of developing coronary heart disease. The cholesterol-lowering agent rosuvastatin was shown to significantly lower the incidence of coronary heart disease.

The concept of enrichment has received considerable attention from the FDA (Guidance for Industry: Enrichment strategies for clinical trials to support approval of human drugs and biological products) [22]. Enrichment is used in order to enroll those participants with a high likelihood of demonstrating an effect from the intervention. Participants with characteristics, including genetic features, that put them at high risk, are entered into the trial. As discussed in Chap. 5, withdrawal studies are also a way of preferentially assessing participants who are more likely to show benefit from the intervention.

The increased FDA focus on fast-track approval has already had implications for the design of randomized clinical trials and their study populations [23]. Regulatory approval without proper phase 3 trials or only based on surrogate efficacy or pharmacodynamic markers limits sample sizes and places focus on highly selected populations. These trials provide limited information about the safety of the intervention. For specific information see Chap. 22 on Regulatory Issues.

Avoiding Adverse Effects

Most interventions are likely to have adverse effects. The investigator needs to weigh these against possible benefit when he evaluates the feasibility of doing the study. However, any person for whom the intervention is known to be harmful should not, except in unusual circumstances, be admitted to the trial. Pregnant women are often excluded from drug trials (unless, of course, the primary question concerns pregnancy) particularly if there is preclinical evidence of teratogenicity. Even without preliminary evidence the amount of additional data obtained may not justify the risk. Similarly, investigators would probably exclude from a study of almost any of the anti-inflammatory drugs people with a recent history of gastric bleeding. Gastric bleeding is a fairly straightforward and absolute contraindication for enrollment. Yet, an exclusion criterion such as “history of major gastric bleed,” leaves much to the judgment of the investigator. The word “major” implies that gastric hemorrhaging is not an absolute contraindication, but a relative one that depends upon clinical judgment. The phrase also recognizes the question of anticipated risk vs. benefit, because it does not clearly prohibit people with a mild bleeding episode in the distant past from being placed on an anti-inflammatory drug. It may very well be that such people take aspirin or similar agents—possibly for a good reason—and studying such people may prove more beneficial than hazardous.

Note that these exclusions apply only before enrollment into the trial. During a trial participants may develop symptoms or conditions which would have excluded them had any of these conditions been present prior to randomization. In these circumstances, the participant may be removed from the intervention regimen if it is contraindicated, but should be kept in the trial and complete follow-up should be obtained for purposes of analysis. As described in Chap. 18, being off the intervention does not mean that a participant is out of the trial.

Competing Risk

The issue of competing risk is generally of greater interest in long-term trials. Participants at high risk of developing conditions which preclude the ascertainment of the outcome of interest should be excluded from enrollment. The intervention may or may not be efficacious in such participants, but the necessity for excluding them from enrollment relates to design considerations. In many studies of people with heart disease, those who have cancer or severe kidney or liver disorders are excluded because these diseases might cause the participant to die or withdraw from the study before the primary response variable can be observed. However, even in short-term studies, the competing risk issue needs to be considered. For example, an investigator may be studying a new intervention for a specific congenital heart defect in infants. Such infants are also likely to have other life-threatening defects. The investigator would not want to enroll infants if one of these other conditions were likely to lead to the death of the infant before the effect of the intervention could be evaluated. This matter is similar to the one raised in Chap. 3, which presented the problem of the impact of high expected total mortality on a study in which the primary response variable is morbidity or cause-specific mortality. When there is competing risk, the ability to assess the true impact of the intervention is, at best, lessened. At worst, if the intervention somehow has either a beneficial or harmful effect on the coexisting condition, biased results for the primary question can be obtained.

Avoiding Poor Adherers

Investigators prefer, ordinarily, to enroll only participants who are likely to adhere to the study protocol. Participants are expected to take their assigned intervention (usually a drug) and return for scheduled follow-up appointments regardless of the intervention assignment. In unblinded studies, participants are asked to accept the random assignment, even after knowing its identity, and abide by the protocol. Moreover, participants should not receive the study intervention from sources outside the trial during the course of the study. Participants should also refrain from using other interventions that may compete with the study intervention. Nonadherence by participants reduces the opportunity to observe the true effect of intervention.

One approach of enrichment of patients who are more likely to adhere to study interventions is to use a run-in phase, either a passive run-in (in which all patients are assigned to placebo for a period of time), active run-in (in which all patients are assigned to the active treatment to assure that they tolerate and adhere to it), or a combination. The PARADIGM HF trial [24] used such an approach. In this trial, 10,521 patients were entered into run-in phase, of which 2,079 were discontinued prior to randomization during the two run-in phases that consisted of a 2 week treatment period with enalapril, followed by a 4 week treatment period with

LCZ696 (valsartan-neprololysin inhibitor) in a dose escalation. This resulted in a population more likely to tolerate and adhere to the treatments, although at the potential cost of having to apply the findings to the large number of patients excluded due to early intolerance. For a further discussion of run-in, see Chap. 14.

An exception to this effort to exclude those less likely to take their medication or otherwise comply with the protocol is what some have termed “pragmatic” clinical trials [25, 26]. These trials are meant to mimic real-world practice, with inclusion of participants who reflect general practice and who may fail to adhere consistently to the intervention. To compensate for the lower expected difference between the intervention and control groups, these trials need to be quite big, and have other characteristics of large, simple trials.

Pharmacogenetics

The field of pharmacogenetics is growing rapidly and so is also its role in clinical trials. Pharmacogenetic markers have been used to identify subgroups of patients in whom an intervention is particularly beneficial or harmful. Many of these observations have been based on post-hoc analyses of markers identified in stored samples collected at baseline. There are also situations in which the markers were known and measured in advance and used to select the study population [27, 28] or for prespecified subgroups [29].

The regulatory agencies, in particular the FDA, are paying more attention to subgroups defined by pharmacogenetic markers in their review and labeling. These markers include specific alleles, deficient gene products, inherited familial conditions and patterns of drug metabolism, such as ultra-rapid, normal, intermediate and poor metabolizer phenotypes. Thus, a very large number of drug labels in the U.S. now contain information linked to these markers. The drug labeling according to the FDA may describe:

- Drug exposure and clinical response variability
- Risk for adverse effects
- Genotype-specific dosing
- Mechanisms of drug action
- Polymorphic drug target and disposition genes

An FDA website lists approximately 140 different drugs with labeling related to genetic markers [30]. The most prevalent therapeutic areas to date are oncology, psychiatry, infectious diseases and cardiology. The experience with pharmacogenomics and psychotropic medications has been thoroughly reviewed [31].

Many large trials today collect genetic materials at baseline and store them for possible future use. We recommend investigators and sponsors to consider collection of DNA samples from participants at baseline. In doing so it is important that the informed consent specifies that permission is given for comprehensive analyses and sharing of these data in large-scale databases [32, 33].

Generalization

Study samples of participants are usually non-randomly chosen from the study population, which in turn is defined by the eligibility criteria (Fig. 4.1). As long as selection of participants into a trial occurs, and as long as enrollment is voluntary, participants must not be regarded as truly representative of the study population. Therefore, investigators have the problem of generalizing from participants actually in the trial to the study population and then to the population with the condition in a comparable clinical setting (external validity). It is often forgotten that participants must agree to enroll in a study. What sort of person volunteers for a study? Why do some agree to participate while others do not? The requirement that study participants sign informed consent or return for periodic examinations is sufficient to make certain people unwilling to participate. Sometimes the reasons are not obvious. What is known, however, is that volunteers can be different from non-volunteers [34, 35]. They are usually in better health, and they are more likely to comply with the study protocol. However, the reverse could also be true. A person might be more motivated if she has disease symptoms. In the absence of knowing what motivates the particular study participants, appropriate compensatory adjustments cannot be made in the analysis. Because specifying how volunteers differ from others is difficult, an investigator cannot confidently identify those segments of the study population or the general population that these study participants supposedly represent. (See Chap. 10 for a discussion of factors that people cite for enrolling or not enrolling in trials.)

Defined medical conditions and quantifiable or discrete variables such as age, sex, or elevated blood sugar can be clearly stated and measured. For these characteristics, specifying in what way the study participants and study population are different from the population with the condition is relatively easy. Judgments about the appropriateness of generalizing study results can, therefore, be made. Other factors of the study participants are less easily characterized. Obviously, an investigator studies only those participants available. If he lives in Florida, he will not be studying people living in Maine. Even within a geographical area, many investigators are based at hospitals or universities. Furthermore, many hospitals are referral centers. Only certain types of participants come to the attention of investigators at these institutions. It may be impossible to decide whether these factors are relevant when generalizing to other geographical regions or patient care settings. Multicenter trials typically enhance the ability to generalize. The growth of international trials, however, raises the important issue of relevance of results from geographical areas with very different clinical care systems.

Many trials now involve participants from community or practice-based settings. Results from these “practical” or “pragmatic” trials may more readily be translated to the broader population. Even here, however, those who choose to become investigators likely differ from other practitioners in the kinds of patients they see.

Many trials of aspirin and other anti-platelet agents in those who have had a heart attack have shown that these agents reduce recurrent myocardial infarction and

death in both men and women [36]. The Physicians' Health Study, conducted in the 1980s, concluded that aspirin reduced myocardial infarction in men over age 50 without previously documented heart disease [37]. Although it was reasonable to expect that a similar reduction would occur in women, it was unproven. Importantly, aspirin was shown in the Physicians' Health Study and elsewhere [38] to increase hemorrhagic stroke. Given the lower risk of heart disease in premenopausal women, whether the trade-off between adverse effects and benefit was favorable was far from certain. The U.S. Food and Drug Administration approved aspirin for primary prevention in men, but not women. The Women's Health Study was conducted in the 1990s and early 2000s [39]. Using a lower dose of aspirin that was used in the Physicians' Health Study, it found evidence of benefit on heart disease only in women at least 65 years old. Based on that, generalization of the Physicians' Health Study results to primary prevention in all women would not have been prudent. A subsequent meta-analysis, however, suggested that the benefits of aspirin for primary prevention were similar in women and men. A trial published in 2014 found no overall benefit of low dose aspirin in a Japanese population of men and women [40]. We must always be open to consider new information in our interpretation of study results [41].

One approach to addressing the question of representativeness is to maintain a log or registry which lists prospective participants identified, but not enrolled, and the reasons for excluding them. This log can provide an estimate of the proportion of all potentially eligible people who meet study entrance requirements and can also indicate how many otherwise eligible people refused enrollment. In an effort to further assess the issue of representativeness, response variables in those excluded have also been monitored. In a study on timolol [42], people excluded because of contraindication to the study drug or competing risks had a mortality rate twice that of those who enrolled. The Coronary Artery Surgery Study included a randomized trial that compared coronary artery bypass surgery against medical therapy and a registry of people eligible for the trial but who declined to participate [43]. The enrolled and not enrolled groups were alike in most identifiable respects. Survival in the participants randomly assigned to medical care was the same as those receiving medical care but not in the trial. The findings for those undergoing surgery were similar. Therefore, in this particular case, the trial participants appeared to be representative of the study population.

With more attention being paid to privacy issues, however, it may not be possible to assess outcomes in those not agreeing to enter a trial. Some people may consent to allow follow-up, even if they do not enroll, but many will not. Thus, comparison of trial results with results in those refusing to enter a trial, in an effort to show that the trial can be generalized, may prove difficult.

A group of Finnish investigators conducted a retrospective chart review [44]. The typical eligibility criteria for clinical trials of patients with gastric ulcer were applied to 400 patients hospitalized with the diagnosis of gastric ulcer. Only 29% of the patients met the eligibility criteria and almost all deaths and serious complications such as gastric bleeding, perforation and stenosis during the first 5–7 years occurred among those patients who would have been ineligible. Clearly, the testing of

H₂-blockers or other compounds for the prevention of long-term complications of gastric ulcer in low-risk patients should not be generalized to the entire ulcer population, as the external validity may be low.

Since the investigator can describe only to a limited extent the kinds of participants in whom an intervention was evaluated, a leap of faith is always required when applying any study findings to the population with the condition. In taking this jump, one must always strike a balance between making unjustifiably broad generalizations and being too conservative in one's claims. Some extrapolations are reasonable and justifiable from a clinical point of view, especially in light of subsequent information.

Recruitment

The impact of eligibility criteria on recruitment of participants should be considered when deciding on these criteria. Using excessive restrictions in an effort to obtain a pure (or homogeneous) sample can lead to extreme difficulty in obtaining sufficient participants and may raise questions regarding generalization of the trial results. Age and sex are two criteria that have obvious bearing on the ease of enrolling subjects. The Coronary Primary Prevention Trial undertaken by the Lipid Research Clinics was a collaborative trial evaluating a lipid-lowering drug in men between the ages of 35 and 59 with severe hypercholesterolemia. One of the Lipid Research Clinics [45] noted that approximately 35,000 people were screened and only 257 participants enrolled. Exclusion criteria, all of which were perfectly reasonable and scientifically sound, coupled with the number of people who refused to enter the study, brought the overall trial yield down to less than 1%. As discussed in Chap. 10, this example of greater than expected numbers being screened, as well as unanticipated problems in reaching potential participants, is common to most clinical trials. We believe that exclusion criteria should include only those with clear rationale such that the negative impact on enrollment and generalizability will likely be outweighed by benefits of limiting the population.

One reason that large international trials are including a larger proportion of patients from low and middle income countries is to increase enrollment potential. This trend for globalization of trials raises a number of important issues as discussed in Chap. 22. For the results of trials to be applicable across countries and health care systems, inclusive enrollment is important. But ethical issues arise when therapies are developed in countries in which those treatments will not be used, often due to cost. And enrolled patients may be systematically different in certain countries. The TOPCAT trial enrolled patients from Russia with heart failure who, in retrospect based on B-type natriuretic peptide levels, may not have had the same degree of heart failure, and who appeared to have less treatment effect from spironolactone [46]. Careful consideration of the advantages and disadvantages of including different health care environments is needed.

If entrance criteria are properly determined in the beginning of a study, there should be no need to change them unless interim results suggest harm in a specific subgroup (see Chap. 16). The reasons for each criterion should be carefully examined during the planning phase of the study. As discussed earlier in this chapter, eligibility criteria are appropriate if they include participants with high likelihood of showing benefit and exclude those who might be harmed by the intervention, have competing risks, and conditions and are not likely to comply with the study protocol. If they do not fall into one of the above categories, they should be reassessed. Whenever an investigator considers changing criteria, he needs to look at the effect of changes on participant safety and study design. It may be that, in opening the gates to accommodate more participants, he increases the required sample size, because the participants admitted may have lower probability of developing the primary response variable. He can thus lose the benefits of added recruitment. In summary, capacity to recruit participants and to carry out the trial effectively could greatly depend on the eligibility criteria that are set. As a consequence, careful thought should go into establishing them.

References

1. Rothwell PM. External validity of randomized controlled trials: “To whom do the results of this trial apply?” *Lancet* 2005;365:82–93.
2. CONSORT. <http://www.consort-statement.org>
3. Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA* 2007;297:1233–1240.
4. Douglas PS. Gender, cardiology, and optimal medical care. *Circulation* 1986;74:917–919.
5. Bennett JC, for the Board on Health Sciences Policy of the Institute of Medicine. Inclusion of women in clinical trials – policies for population subgroups. *N Engl J Med* 1993;329:288–292.
6. Freedman LS, Simon R, Foulkes MA, et al. Inclusion of women and minorities in clinical trials and the NIH Revitalization Act of 1993 – the perspective of NIH clinical trialists. *Control Clin Trials* 1995;16:277–285.
7. Lee PY, Alexander KP, Hammill BG, et al. Representation of elderly persons and women in published randomized trials of acute coronary syndromes. *JAMA* 2001;286:708–713.
8. NIH Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research – Amended, October, 2001. http://grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm
9. Diabetic Retinopathy Study Research Group: Preliminary report on effects of photocoagulation therapy. *Am J Ophthalmol* 1976;81:383–396.
10. Diabetic Retinopathy Study Research Group. Photocoagulation treatment of proliferative diabetic retinopathy: the second report of diabetic retinopathy study findings. *Ophthalmol* 1978;85:82–106.
11. Wooster R, Neuhausen SL, Mangion J, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* 1994;265:2088–2090.
12. Patel MR, Mahaffey KW, Garg J, et al. for ROCKET AF investigators. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med* 2011;365:883–891.

13. Veterans Administration Cooperative Study Group on Antihypertensive Agents. Effects of treatment on morbidity in hypertension: results in patients with diastolic blood pressures averaging 115 through 129 mm Hg. *JAMA* 1967;202:1028–1034.
14. Veterans Administration Cooperative Study Group on Antihypertensive Agents. Effects of treatment on morbidity in hypertension: II. Results in patients with diastolic blood pressure averaging 90 through 114 mm Hg. *JAMA* 1970;213:1143–1152.
15. Hypertension Detection and Follow-up Program Cooperative Group. Five-year findings of the Hypertension Detection and Follow-up Program. 1. Reduction in mortality of persons with high blood pressure, including mild hypertension. *JAMA* 1979;242:2562–2571.
16. The CONSENSUS Trial Study Group. Effects of enalapril on mortality in severe heart failure. *N Engl J Med* 1987;316:1429–1435.
17. The SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med* 1991;325:293–302.
18. The SOLVD Investigators. Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. *N Engl J Med* 1992;327:685–691.
19. Vollmer T. The natural history of relapses in multiple sclerosis. *J Neurol Sci* 2007;256:S5–S13.
20. Sondik EJ, Brown BW, Jr., Silvers A. High risk subjects and the cost of large field trials. *J Chronic Dis* 1974; 27:177–187.
21. Ridker PM, Danielson E, Fonseca FAH, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med* 2008;359:2195–2207.
22. <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm332181.pdf>.
23. Darrow JJ, Avorn J, Kesselheim AS. New FDA breakthrough-drug category—implications for patients. *N Engl J Med* 2014;370:1252–1258.
24. McMurray JJV, Packer M, Desai AS, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N Engl J Med* 2014;371:993–1004.
25. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA*. 2003;290:1624–1632.
26. Thorpe KE, Swarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009;62:464–475.
27. Ridker PM and PREVENT Investigators. Long-term, low does warfarin among venous thrombosis patients with and without factor V Leiden mutation: rationale and design for the Prevention of Recurrent Venous Thromboembolism (PREVENT) trial. *Vasc Med* 1998;3:67–73.
28. Mooney MM, Welch J, Abrams JS. Clinical trial design and master protocols in NCI clinical treatment trials. [abstract]. *Clin Cancer Res* 2014;20(2Suppl):Abstract IA08.
29. Hakonarson H, Thorvaldsson S, Helgadóttir A, et al. Effects of a 5-lipoxygenase-activating protein inhibitor on biomarkers associated with risk of myocardial infarction: a randomized trial. *JAMA* 2005;293:2245–2256.
30. The U.S. Food and Drug Administration. Drugs. Table of pharmacogenomics biomarkers in labeling. www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm.
31. Mrazek DA. *Psychiatric pharmacogenomics*. New York: Oxford University Press, 2010.
32. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014;42 (Database issue): D980-5.
33. Mailman MD, Feolo M, Jin Y, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007;39:1181–1186.
34. Wilhelmsen L, Ljungberg S, Wedel H, Werko L. A comparison between participants and non-participants in a primary preventive trial. *J Chronic Dis*. 1976;29:331–339.
35. Smith P, Arnesen H. Mortality in non-consenters in a post-myocardial infarction trial. *J Intern Med* 1990; 228:253–256.

36. Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomized clinical trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 2002;324:71–86; correction *BMJ* 2002;324:141.
37. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med* 1989;321:129–135.
38. Peto R, Gray R, Collins R, et al. Randomized trial of prophylactic daily aspirin in British male doctors. *Br Med J* 1988;296:313–316.
39. Ridker PM, Cook NR, Lee I-M, et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *N Engl J Med* 2005;352:1293–1304.
40. Ikeda Y, Shimada K, Teramoto T, et al. Low-dose aspirin for primary prevention of cardiovascular events in Japanese patients 60 years or older with atherosclerotic risk factors. A randomized clinical trial. *JAMA*. Published online November 17, 2014. doi:10.1001/jama.2014.15690.
41. Berger JS, Roncaglioni MC, Avanzini F, et al. Aspirin for the primary prevention of cardiovascular events in women and men: a sex-specific meta-analysis of randomized controlled trials. *JAMA* 2006;295:306–313; correction *JAMA* 2006;295:2002.
42. Pedersen TR. The Norwegian Multicenter Study of timolol after myocardial infarction. *Circulation* 1983;67 (suppl 1):I-49-1-53.
43. CASS Principal Investigators and Their Associates. Coronary Artery Surgery Study (CASS): a randomized trial of coronary artery bypass surgery. Comparability of entry characteristics and survival in randomized patients and nonrandomized patients meeting randomization criteria. *J Am Coll Cardiol* 1984;3:114–128.
44. Kaariainen I, Sipponen P, Siurala M. What fraction of hospital ulcer patients is eligible for prospective drug trials? *Scand J Gastroenterol* 1991;186:73–76.
45. Benedict GW. LRC Coronary Prevention Trial: Baltimore. *Clin Pharmacol Ther* 1979;25:685–687.
46. Pitt B, Pfeffer MA, Assmann SF, et al. TOPCAT Investigators. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med* 2014;370:1383–1392.