Choosing the Study Subjects: Specification, Sampling, and Recruitment

Stephen B. Hulley, Thomas B. Newman, and Steven R. Cummings

A good choice of study subjects serves the vital purpose of ensuring that the findings in the study accurately represent what is going on in the **population** of interest. The protocol must specify a **sample** of subjects that can be studied at an acceptable cost in time and money (i.e., modest in size and convenient to access), yet large enough to control random error and representative enough to allow generalizing study findings to populations of interest. An important precept here is that **generalizability** is rarely a simple yes-or-no matter; it is a complex qualitative judgment that depends on the investigator's choice of population and of sampling design.

We will come to the issue of choosing the appropriate *number* of study subjects in Chapter 6. In this chapter we address the process of **specifying** and **sampling** the *kinds* of subjects who will be representative and feasible (Figure 3.1). We also discuss strategies for **recruiting** these people to participate in the study.

■ BASIC TERMS AND CONCEPTS

Populations and Samples

A **population** is a complete set of people with specified characteristics, and a **sample** is a subset of the population. In lay usage, the characteristics that define a population tend to be
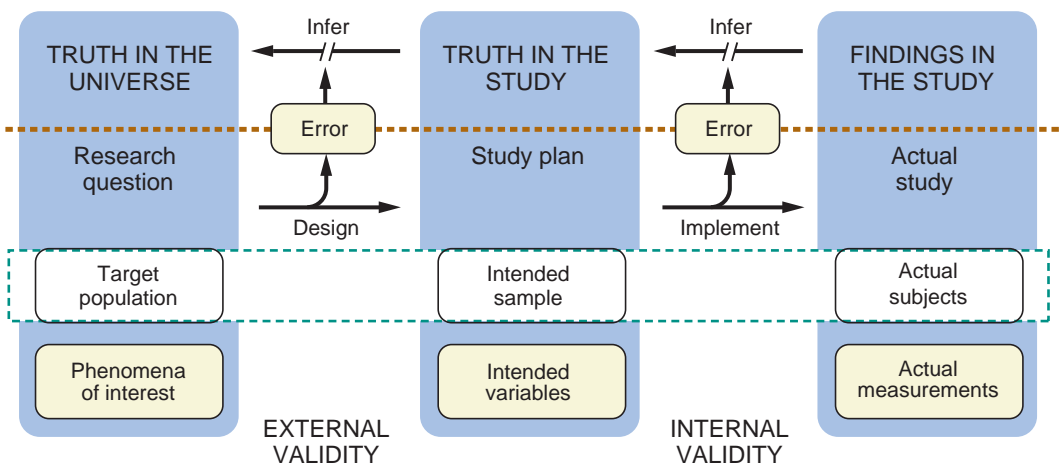

■ **FIGURE 3.1** This chapter focuses on choosing a sample of study subjects that represent the population of interest for the research question.

geographic—for example, the population of Canada. In research, the defining characteristics are also clinical, demographic, and temporal:

- Clinical and demographic characteristics define the **target population**, the large set of people throughout the world to which the results may be generalized—teenagers with asthma, for example.
- The **accessible population** is a geographically and temporally defined subset of the target population that is available for study—teenagers with asthma living in the investigator's town this year.
- The **intended study sample** is the subset of the accessible population that the investigator seeks to include in the study.
- The **actual study sample** is the group of subjects that does participate in the study.

Generalizing the Study Findings

The classic Framingham Study was an early approach to scientifically designing a study to allow **inferences** from findings observed in a sample to be applied to a population (Figure 3.2).

The sampling design called for identifying all the families in Framingham with at least one person aged 30–59, listing the families in order by address, and then asking age-eligible persons in the first two of every set of three families to participate. This "systematic" sampling design is not as tamperproof as choosing each subject by a random process (as discussed later in this chapter), but two more serious concerns were the facts that one-third of the Framingham residents selected for the study refused to participate, and that in their place the investigators accepted age-eligible residents who were not in the sample and volunteered (1).

Because respondents are often healthier than nonrespondents, especially if they are volunteers, the characteristics of the actual sample undoubtedly differed from those of the intended sample. Every sample has some errors, however, and the issue is how much damage has been
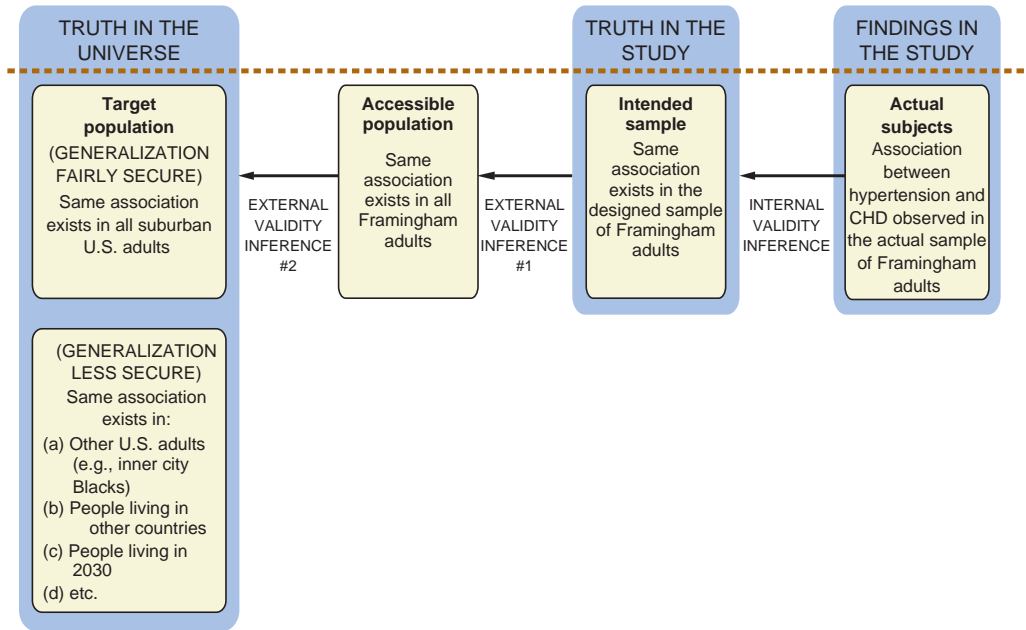
■ FIGURE 3.2 Inferences in generalizing from the study subjects to the target populations proceed from right to left.
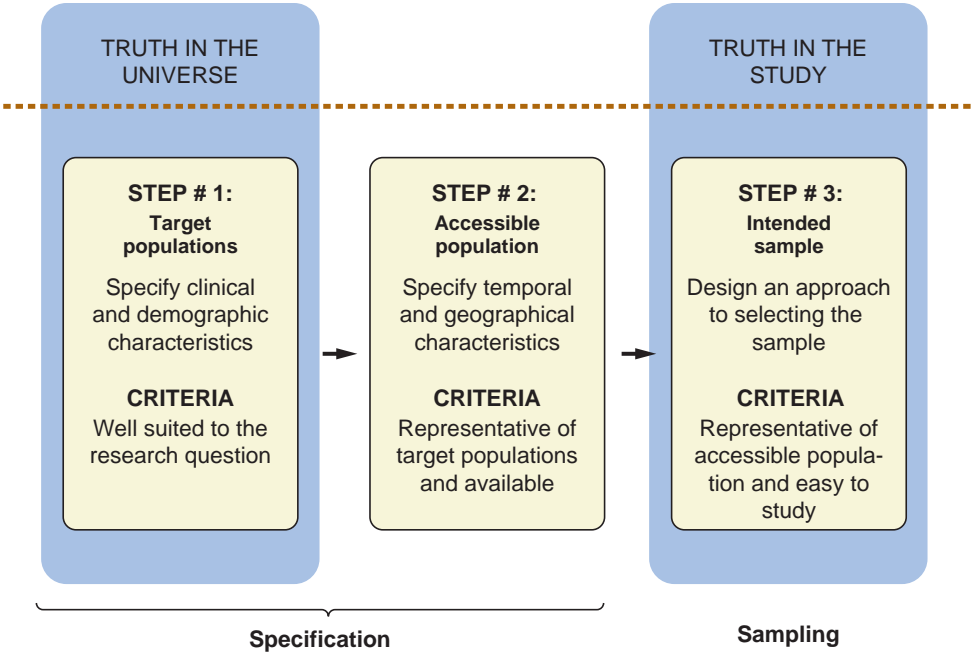
■ **FIGURE 3.3** Steps in designing the protocol for choosing the study subjects.

done. The Framingham Study sampling errors do not seem large enough to invalidate the conclusion that risk relationships observed in the study—for example, that hypertension is a risk factor for coronary heart disease (CHD)—can be generalized to all the residents of Framingham.

The next concern is the validity of generalizing the finding that hypertension is a risk factor for CHD from the accessible population of Framingham residents to target populations elsewhere. This inference is more subjective. The town of Framingham was selected not with a scientific sampling design, but because it seemed fairly typical of middle-class white communities in the United States and was convenient to the investigators. The validity of generalizing the Framingham risk relationships to populations in other parts of the country involves the precept that, in general, analytic studies and clinical trials that address biologic relationships produce more widely generalizable results across diverse populations than descriptive studies that address distributions of characteristics. Thus, the strength of hypertension as a risk factor for CHD is similar in Caucasian Framingham residents to that observed in inner city African Americans, but the prevalence of hypertension is much higher in the latter population.

Steps in Designing the Protocol for Acquiring Study Subjects

The inferences in Figure 3.2 are presented from right to left, the sequence used for interpreting the findings of a completed study. An investigator who is planning a study reverses this sequence, beginning on the left (Figure 3.3). She begins by specifying the clinical and demographic characteristics of the target population that will serve the research question well. She then uses **geographic** and **temporal criteria** to specify a study sample that is representative and practical.

■ SELECTION CRITERIA

If an investigator wants to study the efficacy of low dose testosterone supplements versus placebo for enhancing libido in postmenopausal women, she can begin by creating selection criteria that define the population to be studied.

Establishing Selection Criteria

Inclusion criteria define the main characteristics of the target population that pertain to the research question (Table 3.1). Age is often a crucial factor, and in this study the investigator might decide to focus on women in their fifties, speculating that in this group the benefit-to-harm ratio of the drug might be optimal; another study might make a different decision and focus on older decades. The investigator also might incorporate African American, Hispanic, and Asian women in the study in an effort to expand generalizability. This is generally a good idea, but it's important to realize that the increase in generalizability is illusory if there is other evidence to suggest that the effects differ by race. In that case the investigator would need enough women of each race to statistically test for the presence of **effect modification** (an effect in one race that is different from that in other races, also known as "an interaction"; Chapter 9); the number needed is generally large, and most studies are not powered to detect effect modification.

Inclusion criteria that address the geographic and temporal characteristics of the accessible population often involve trade-offs between scientific and practical goals. The investigator may find that patients at her own hospital are an available and inexpensive source of subjects. But she must consider whether peculiarities of the local referral patterns might interfere with generalizing the results to other populations. On these and other decisions about inclusion criteria, there is no single course of action that is clearly right or wrong; the important thing is to make decisions that are sensible, that can be used consistently throughout the study, and that can be clearly described to others who will be deciding to whom the published conclusions apply.

TABLE 3.1 DESIGNING SELECTION CRITERIA FOR A CLINICAL TRIAL OF LOW DOSE TESTOSTERONE VERSUS PLACEBO TO ENHANCE LIBIDO IN MENOPAUSE

	DESIGN FEATURE	EXAMPLE
Inclusion criteria (be specific)	Specifying populations relevant to the research question and efficient for study:	
	Demographic characteristics	Women 50 to 59 years old
	Clinical characteristics	Good general health Has a sexual partner Is concerned about decreased libido
	Geographic (administrative) characteristics	Patients attending clinic at the investigator's hospital
	Temporal characteristics	Between January 1 and December 31 of specified year
Exclusion criteria (be parsimonious)	Specifying subsets of the population that will *not* be studied because of:	
	A high likelihood of being lost to follow-up	Alcoholic Plans to move out of state
	An inability to provide good data	Disoriented Has a language barrier*
	Being at high risk of possible adverse effects	History of myocardial infarction or stroke

*Alternatives to excluding those with a language barrier (when these subgroups are sizeable and important to the research question) would be collecting nonverbal data or using bilingual staff and questionnaires.

Specifying clinical characteristics for selecting subjects often involves difficult judgments, not only about which factors are important to the research question, but about how to define them. How, for example, would an investigator put into practice the criterion that the subjects be in "good health"? She might decide not to include patients with any self-reported illness, but this would likely exclude large numbers of subjects who are perfectly suitable for the research question at hand.

More reasonably, she might exclude only those with diseases that could interfere with follow-up, such as metastatic cancer. This would be an example of "**exclusion criteria**," which indicate individuals who meet the inclusion criteria and would be suitable for the study were it not for characteristics that might interfere with the success of follow-up efforts, the quality of the data, or the acceptability of randomized treatment (Table 3.1). Difficulty with the English language, psychological problems, alcoholism, and serious illness are examples of exclusion criteria. **Clinical trials** differ from observational studies in being more likely to have exclusions mandated by concern for the safety of an intervention in certain patients; for example, the use of drugs in pregnant women (Chapter 10). A good general rule that keeps things simple and preserves the number of potential study subjects is to have as *few exclusion criteria* as possible.

Clinical Versus Community Populations

If the research question involves patients with a disease, hospitalized or clinic-based patients are easier to find, but selection factors that determine who comes to the hospital or clinic may have an important effect. For example, a specialty clinic at a tertiary care medical center attracts patients from afar with serious forms of the disease, giving a distorted impression of the features and prognosis that are seen in ordinary practice. Sampling from primary care practices can be a better choice.

Another common option in choosing the sample is to select subjects in the community who represent a healthy population. These samples are often recruited using mail, e-mail, or advertising via Internet, broadcast, or print media; they are not fully representative of a general population because some kinds of people are more likely than others to volunteer or be active users of Internet or e-mail. True "population-based" samples are difficult and expensive to recruit, but useful for guiding public health and clinical practice in the community. One of the largest and best examples is the National Health and Nutrition Examination Survey (NHANES), a **representative** sample of U.S. residents.

The size and diversity of a sample can be increased by collaborating with colleagues in other cities, or by using preexisting data sets such as NHANES and Medicare data. **Electronically accessible data sets** from public health agencies, healthcare providing organizations, and medical insurance companies have come into widespread use in clinical research and may be more representative of national populations and less time-consuming than other possibilities (Chapter 13).

■ SAMPLING

Often the number of people who meet the selection criteria is too large, and there is a need to select a **sample** (subset) of the population for study.

Nonprobability Samples

In clinical research the study sample is often made up of people who meet the entry criteria and are easily accessible to the investigator. This is termed a **convenience sample**. It has obvious advantages in cost and logistics, and is a good choice for some research questions.

A **consecutive sample** can minimize volunteerism and other selection biases by consecutively selecting subjects who meet the entry criteria. This approach is especially desirable, for example, when it amounts to taking the entire accessible population over a long enough period to include seasonal variations or other temporal changes that are important to the research question.

The validity of drawing inferences from any sample is the premise that, for the purpose of answering the research question at hand, it sufficiently represents the accessible population. With convenience samples this requires a subjective judgment.

Probability Samples

Sometimes, particularly with descriptive research questions, there is a need for a scientific basis for generalizing the findings in the study sample to the population. Probability sampling, the gold standard for ensuring generalizability, uses a random process to guarantee that each unit of the population has a specified chance of being included in the sample. It is a scientific approach that provides a rigorous basis for estimating the fidelity with which phenomena observed in the sample represent those in the population, and for computing statistical significance and confidence intervals. There are several versions of this approach.

- A **simple random sample** is drawn by enumerating (listing) all the people in the population from which the sample will be drawn, and selecting a subset at random. The most common use of this approach in clinical research is when the investigator wishes to select a representative subset from a population that is larger than she needs. To take a random sample of the cataract surgery patients at her hospital, for example, the investigator could list all such patients on the operating room schedules for the period of study, then use a table of random numbers to select individuals for study (Appendix 3).
- A **systematic sample** resembles a simple random sample in the first step, enumerating the population, but differs in that the sample is selected by a preordained periodic process (e.g., the Framingham approach of taking the first two out of every three families from a list of town families ordered by address). Systematic sampling is susceptible to errors caused by natural periodicities in the population, and it allows the investigator to predict and perhaps manipulate those who will be in the sample. It offers no logistic advantages over simple random sampling, and in clinical research it is rarely a better choice.
- A **stratified random sample** begins by dividing the population into subgroups according to characteristics such as sex or race, and taking a random sample from each of these "strata." The Stratified subsamples can be weighted to draw disproportionately from subgroups that are less common in the population but of special interest to the investigator. In studying the incidence of toxemia in pregnancy, for example, the investigator could stratify the population by race and then sample equal numbers from each stratum. Less common races would then be over-represented, yielding incidence estimates of comparable precision from each racial group.
- A **cluster sample** is a random sample of natural groupings (clusters) of individuals in the population. Cluster sampling is useful when the population is widely dispersed and it is impractical to list and sample from all its elements. Consider, for example, the problem of interviewing patients with lung cancer selected randomly from a statewide database of discharge diagnoses; patients could be studied at lower cost by choosing a random sample of the hospitals and taking the cases from these. Community surveys often use a two-stage cluster sample: A random sample of city blocks is drawn from city blocks enumerated on a map and a field team visits the blocks in the sample, lists all the addresses in each, and selects a subsample of addresses for study by a second random process. A disadvantage of cluster sampling is the fact that naturally occurring groups are often more homogeneous for the variables of interest than the population; each city block, for example, tends to have people of similar socioeconomic status. This means that the effective sample size (after adjusting for within-cluster uniformity) will be somewhat smaller than the number of subjects, and that statistical analysis must take the clustering into account.

Summarizing the Sampling Design Options

The use of descriptive statistics and tests of statistical significance to draw inferences about the population from observations in the study sample is based on the assumption that a probability

sample has been used. But in clinical research a random sample of the whole target population is almost never possible. Convenience sampling, preferably with a consecutive design, is a practical approach that is often suitable. The decision about whether the proposed sampling design is satisfactory requires that the investigator make a **judgment**: *for the research question at hand,* will the conclusions drawn from observations in the study sample be similar to the conclusions that would result from studying a true probability sample of the accessible population? And beyond that, will the conclusions be appropriate for the target population?

■ RECRUITMENT

The Goals of Recruitment

An important factor to consider in choosing the accessible population and sampling approach is the feasibility of recruiting study participants. There are two main goals: (1) to recruit a sample that adequately **represents** the target population, minimizing the prospect of getting the wrong answer to the research question due to systematic error (bias); and (2) to recruit a sufficient **sample size** to minimize the prospect of getting the wrong answer due to random error (chance).

Achieving a Representative Sample

The approach to recruiting a representative sample begins in the design phase with wise decisions about choosing target and accessible populations, and approaches to sampling. It ends with implementation, guarding against errors in applying the entry criteria to prospective study participants, and enhancing successful strategies as the study progresses.

A particular concern, especially for descriptive studies, is the problem of **nonresponse**.[1] The proportion of subjects selected for the study who consent to be enrolled (the response rate) influences the validity of inferring that the enrolled sample represents the population. People who are difficult to reach and those who refuse to participate once they are contacted tend to be different from people who do enroll. The level of nonresponse that will compromise the generalizability of the study depends on the nature of the research question and on the reasons for not responding. A nonresponse rate of 25%, a good achievement in many settings, can seriously distort the estimate of the prevalence of a disease when the disease itself is a cause of nonresponse.

The degree to which nonresponse bias may influence the conclusions of a descriptive study can sometimes be estimated during the study by acquiring additional information on a sample of nonrespondents. The best way to deal with nonresponse bias, however, is to minimize the number of nonrespondents. The problem of failure to make contact with individuals who have been chosen for the sample can be reduced by designing a series of repeated contact attempts using various methods (mail, e-mail, telephone, home visit). Among those contacted, refusal to participate can be minimized by improving the efficiency and attractiveness of the study, by choosing a design that avoids invasive and uncomfortable tests, by using brochures and individual discussion to allay anxiety and discomfort, by providing incentives such as reimbursing the costs of transportation and providing the results of tests, and by circumventing language barriers with bilingual staff and translated questionnaires.

Recruiting Sufficient Numbers of Subjects

Falling short in the rate of recruitment is one of the commonest problems in clinical research. In planning a study it is best to assume that the number of subjects who meet the entry criteria and agree to enter the study will be fewer, sometimes by severalfold, than the number projected

---

[1]Concern with nonresponse in the process of *recruiting* subjects for a study (the topic of this chapter) is chiefly a concern in descriptive studies that have a primary goal of estimating distributions of variables in particular populations. Nonresponse in the *follow-up* process is often a major issue in any study that follows a cohort over time, and particularly in a clinical trial of an intervention that may alter the response rate (Chapter 10).

at the outset. The approaches to this problem are to estimate the magnitude of the recruitment problem empirically with a pretest, to plan the study with an accessible population that is larger than believed necessary, and to make contingency plans should the need arise for additional subjects. While recruitment is ongoing it is important to closely monitor progress in meeting the recruitment goals and tabulate reasons for falling short of the goals. Understanding why potential subjects are lost to the study at various stages can lead to strategies for reducing these losses.

Sometimes recruitment involves selecting subjects who are already known to the members of the research team (e.g., in a study of a new treatment in patients attending the investigator's clinic). Here the chief concern is to present the opportunity for participation in the study fairly, making clear the advantages and disadvantages. In discussing participation, the investigator must recognize the ethical dilemmas that arise when her advice as the patient's physician might conflict with her interests as an investigator (Chapter 14).

Often recruitment involves contacting populations that are not known to the members of the research team. It is helpful if at least one member of the research team has previous experience with the approaches for contacting the prospective subjects. These include screening in work settings or public places such as shopping malls; sending out large numbers of mailings to listings such as driver's license holders; advertising on the Internet; inviting referrals from clinicians; carrying out retrospective record reviews; and examining lists of patients seen in clinic and hospital settings. Some of these approaches, particularly the latter two, involve concerns with privacy invasion that must be considered by the institutional review board.

It may be helpful to prepare for recruitment by getting the support of important organizations. For example, the investigator can meet with hospital administrators to discuss a clinic-based sample, and with community leaders, the medical society and county health department to plan a community screening operation or mailing to physicians. Written endorsements can be included as an appendix in applications for funding. For large studies it may be useful to create a favorable climate in the community by giving public lectures or by advertising through radio, TV, newspapers, fliers, websites, and mass mailings.

■ SUMMARY

1. Most clinical research is based, philosophically and practically, on the use of a **sample** to represent a **population**.
2. The advantage of sampling is **efficiency**: It allows the investigator to draw inferences about a large population by examining a subset at relatively small cost in time and effort. The disadvantage is the sources of **error** it introduces: If the sample is not sufficiently representative for the research question at hand the findings may not **generalize** well to the target population, and if it is not large enough the findings may not sufficiently minimize the role of **chance**.
3. In designing a sample, the investigator begins by conceptualizing the **target population** with a specific set of **inclusion criteria** that establish demographic and clinical characteristics of subjects well suited to the research question.
4. She then selects an appropriate **accessible population** that is geographically and temporally convenient, and defines a parsimonious set of **exclusion criteria** that eliminate subjects who are unethical or inappropriate to study.
5. The next step is to design an approach to **sampling** the population. A **convenience sample** may be adequate, especially for initial studies of some questions, and a **consecutive sample** is often a good choice. **Simple random sampling** can be used to reduce the size of the sample if necessary, and **other probability sampling** strategies (**stratified** and **cluster**) are useful in certain situations.
6. Finally, the investigator must design and implement strategies for **recruiting** a sample of subjects that is sufficiently **representative** of the target population to control systematic sources of error, and **large enough** to control random sources of error.

APPENDIX 3

This table provides a simple paper-based way to select a 10% random sample from a table of random numbers. Begin by enumerating (listing and numbering) every person in the population to be sampled. Then decide on a rule for obtaining an appropriate series of numbers; for example, if your list has 741 elements (which you have numbered 1 to 741), your rule might be to go vertically down each column in this table using the first three digits of each number (beginning at the upper left, the numbers are 104, 223, etc.) and to select the first 74 different numbers that fall in the range of 1 to 741. Finally, pick a starting point by an arbitrary process (closing your eyes and putting your pencil on some number in the table is one way to do it) and begin applying the rule. The modern approach, with a computerized series of random numbers, basically works the same way.

TABLE 3.2 SELECTING A RANDOM SAMPLE FROM A TABLE OF RANDOM NUMBERS

10480	15011	01536	81647	91646	02011
22368	46573	25595	85393	30995	89198
24130	48390	22527	97265	78393	64809
42167	93093	06243	61680	07856	16376
37570	33997	81837	16656	06121	91782
77921	06907	11008	42751	27756	53498
99562	72905	56420	69994	98872	31016
96301	91977	05463	07972	18876	20922
89572	14342	63661	10281	17453	18103
85475	36857	53342	53998	53060	59533
28918	79578	88231	33276	70997	79936
63553	40961	48235	03427	49626	69445
09429	93969	52636	92737	88974	33488
10365	61129	87529	85689	48237	52267
07119	97336	71048	08178	77233	13916
51085	12765	51821	51259	77452	16308
02368	21382	52404	60268	89368	19885
01011	54092	33362	94904	31273	04146
52162	53916	46369	58569	23216	14513
07056	97628	33787	09998	42698	06691
48663	91245	85828	14346	09172	30163
54164	58492	22421	74103	47070	25306
32639	32363	05597	24200	38005	13363
29334	27001	87637	87308	58731	00256
02488	33062	28834	07351	19731	92420
81525	72295	04839	96423	24878	82651
29676	20591	68086	26432	46901	20949
00742	57392	39064	66432	84673	40027
05366	04213	25669	26422	44407	44048
91921	26418	64117	94305	26766	25940

REFERENCE

1. www.framinghamheartstudy.org/about/background.html, accessed 7/23/12.