

Epidemiologic Research: A Primer for the Clinical Exercise Physiologist

Gregory W. Heath, DHSc, MPH¹, Nicholas Boer, PhD, MPH¹

ABSTRACT

An understanding of epidemiologic methods is important for the clinical exercise physiologist to assess the complex relationships between physical activity and health and disease. It is essential to the practice of preventive and rehabilitative care to understand the web of causation and complex interactions among agent (exercise), host (individual), and environment (affects transmission of agent source to host) in relationship to disease/injury and clinical outcomes. Application of the epidemiologic measures of disease/injury occurrence, variations in occurrence, and statistical measures of attributable risk and population attributable risk contribute to the clinician's skill level in assessing potential cause-effect relationships reported in the literature about exercise medicine, physical activity, and public health. By becoming familiar with the study methods used in epidemiology, the clinical exercise physiologist will be better positioned to assess criteria for a cause-effect relationship as well as to critically evaluate the assessment efforts used across a variety of study designs and applications of epidemiology in clinical research and practice. *Journal of Clinical Exercise Physiology*. 2020;9(1):29–39.

Keywords: study methods, disease causation, physical activity, surveillance, measurement

INTRODUCTION

In the practice of clinical exercise physiology, epidemiologic studies are important for understanding the complex relationships that exist between physical activity and health and disease. Epidemiology is the study of the occurrence of health-related events and their determinants in human populations (1). The goal of epidemiology is to discover facts essential or contributory to the occurrence of disease or a state of health in a population. The objective of this discipline is to develop effective methods of disease prevention that act directly on causal agents, factors, or determinants. Epidemiology includes the study of where a disease or injury occurs, who has the disease or injury, and who is at risk for developing the disease or sustaining the injury. The importance of epidemiologic research is highlighted by the great impact that disease and injury have on society, an impact that is measured in loss of life, disability, emotional anguish, social dysfunction, and economic loss. Because health care expenditures in the United States now exceed 17.9% of the gross national

product (2), it is imperative that epidemiologic research be used to advance knowledge of the causes of disease and injury to reduce their consequences on the individual and society. The domain of epidemiologic research is groups of people—not individuals—with emphasis on identifying specific subgroups, explained by age, sex, race/ethnicity, socioeconomic status, educational level, and so forth, and focusing on factors of time, place, and person. One of the first and classic examples of the application of epidemiologic methods to the question about association of physical activity/exercise and health examined the association of physical activity with risk for coronary heart disease (CHD) among London bus conductors and bus drivers. This study, carried out by Dr. Jeremy Morris in 1954 (3), examined the CHD experience among bus conductors (those who gathered tickets from patrons on double-decker London buses) compared with the drivers who sat throughout their shifts driving the buses. When age standardized rates were calculated, it was found that among male bus conductors and drivers aged 35 to 64 years, the incidence rate of CHD was 2.0/1,000 per year

¹Public Health Program, Department of Health and Human Performance, University of Tennessee Chattanooga, Chattanooga, TN 37403 USA

Address for correspondence: Gregory W. Heath, DHSc, MPH, Guerry Professor, Public Health Program, Department of Health and Human Performance, Dept. #6606, University of Tennessee at Chattanooga, 615 McCallie Avenue, Chattanooga, TN 37403; e-mail: Gregory-Heath@utc.edu.

Conflicts of interest and sources of funding: None.

Copyright © 2020 Clinical Exercise Physiology Association

among the conductors compared to 2.7/1,000 per year among the drivers, a statistically significant difference (3). This marked the first modern-era link between physical activity and a specific health outcome. With this example as a backdrop, the purpose of this article is to provide an overview of epidemiologic research applications, specifically for clinical exercise physiologists and exercise scientists engaged in health-related research and practice.

Epidemiologic Methods

Epidemiologists practically apply the scientific method to problems of the health sciences. The discipline borrows from many specialties and combines them into a viable science. The application of epidemiology requires knowledge of causal factors relative to the occurrence of specific health events. This knowledge is often revealed through measures of association that declare suspected causal factors and specific health events are either statistically significantly associated or not statistically significantly associated; that is, either independent or not independent of one another.

For those associations that are statistically significant, several interpretations are possible. A noncausal or secondary association is one possibility if the two events are very common and thus, associated, but not in a causal fashion. A statistically associated significant relationship may be direct or indirect. In the case of a direct causal association, one event is the cause of a second event. On the other hand, an indirect causal association is characterized by an event that may be associated with a third event that is really the causal event.

The concept of the web of causation is an important principle in epidemiology (1). This web can be defined such that the effect may be the result of a complex interaction of causes, with the understanding that not every effect is the result of a single cause (4). An important tenet of epidemiology is that for effective public health interventions to be carried out, they do not require a complete knowledge of the web of causation. It is also understood that the web may be sufficiently deformed by an attack at one link to oftentimes render prevention efforts less effective. Finally, it is possible that additional unexpected side effects may occur, obscuring the relationships within the web of causation.

Fundamental to understanding the complex nature of a web of causation is the assessment of the critical classes of causal factors, which include the classic paradigm of agent, host, and environment. The agent in the context of a causation web or multifactorial outcome can be one of any number of factors including, but not limited to, factors that are genetic, physical (e.g., sunlight, fire, radiation, seat belts), nutritive (excess, deficiency), exogenous chemical (e.g., inhalation, ingestion, skin contact), physiological or psychological, as well as those that are invasive organisms (e.g., host manifests any number of personal attributes that may be linked to increased/decreased susceptibility to any agent or occurrence of disease/injury; these personal attributes include age, sex, immune status, behavioral attributes [e.g., smoking status, physical activity], race/ethnicity, social class, and

genetic predisposition). The environment is critical to both the host and agent. Characteristics of the external environment that may influence host/agent interaction can include physical (e.g., climate, altitude, urban, rural), biological (e.g., food supply, other living things), and social (e.g., population distribution, culture, access to recreation, access to health care). Clearly the internal environment of both the host and agent also are important and relevant to the interactions of the host and agent with their external environment.

Epidemiologic Measures

Disease/Injury Occurrence

Several measures are used to quantify disease occurrence in the population. The incidence, or incidence rate, is the number of new cases of disease or injury during a defined period, divided by the product of the number of persons monitored during the time period (5). Incidence is usually expressed as the number of new cases occurring in a year among a specified population (6). Cumulative incidence is the risk of developing a disease over a defined time period, such as 1 year. Prevalence is another proportional measure; it is expressed as the number of existing cases of disease/injury divided by the total population, with the occurrence of the disease/injury measured at a specific point in time rather than over a certain time period (5). The prevalence of a disease/injury is influenced both by its incidence and persistence (i.e., how long people have a disease/injury before cure or death) (6). See Table 1 for a list of selected measures and their definitions.

Variation in Occurrence of Disease/Injury

A comparison between two groups to reveal the relative frequency of a health-related event is expressed as the ratio of the two rates and is referred to as a rate ratio. A specific type of rate ratio is the relative risk (RR), which is a comparison of the rate in a population subgroup exposed to an agent that is believed to cause a disease, injury, or death with the rate in a population subgroup not exposed (5,6). A similar measure specific to case-control studies (see epidemiologic study methods), often considered synonymous with the RR, is the odds ratio (OR) or cross-product ratio, which provides an approximation of the RR (5,6). In the instance of both the RR and OR, ratios exceeding 1 are interpreted as conveying greater risk or a greater likelihood of an outcome. The confidence interval (CI) is a measure used in association with rate ratios to declare their likelihood of not being due to chance alone. This represents a range of values for a rate-ratio or other variable constructed so that this range has a specified probability of including the true value of the rate-ratio. The end points of the confidence interval are called "confidence limits" and, when applied to the RR and OR, they determine statistical significance when 1.0 is not inclusive of the lower 95% or upper 95% range (1).

Measures of Prevention

Another statistic used to judge the strength of association between a risk factor and disease, injury, or death is the

TABLE 1. Selected epidemiologic measures.

Measure	Definition
Prevalence rate	$\frac{\text{Number of persons who have attribute/disease at particular time}}{\text{Population at risk of having attribute/disease at this point in time}}$
Incidence rate	$\frac{\text{Number of new events in a specified time period}}{\text{Number of persons exposed to risk during this period}}$
Relative risk	$\frac{\text{Risk of disease/injury or death in the exposed population}}{\text{Risk of disease/injury or death in the unexposed population}}$
Odds ratio	$\frac{\text{Number of exposed cases/Number of unexposed cases}}{\text{Number of exposed controls/Number of unexposed controls}}$

Population Attributable Fraction (PAF) is the proportion of cases that would be prevented if the risk factor were eliminated:
PAF formula = $\text{Pe}(\text{RR}-1)/1 + (\text{Pe}[\text{RR}-1])$, where Pe is the prevalence (percent) of those at risk (e.g., physically inactive) in the study population and RR is the relative risk.

attributable risk. The most common form of attributable risk is the population attributable risk, which indicates the proportion of cases in a population that occurred in a subgroup having the risk factor of interest (5–7). Epidemiologists assume that the proportion of disease found in that subgroup is the result of, or attributable to, that risk factor. Population attributable risk or fraction is a function of the both the RR of a factor and the frequency or prevalence of that factor in the population. Population Attributable Fraction is computed by the equation found in Table 1.

An estimate of the potential percentage reduction in the risk of disease, injury, or death for people who change from the exposed group to the unexposed group is the clinical attributable risk. This risk is the proportion of all cases in the exposed group attributable to the factor that defines the exposed group and is computed as $1-(1/\text{RR})$, where the denominator includes the RR of the exposure factor.

The relationship between RR and the population attributable risk is illustrated in Tables 2 and 3. Table 2 lists the measured RRs from the literature for each of the selected risk factors of smoking, obesity, and physical inactivity for selected noncommunicable diseases (NCD) including coronary heart disease, type 2 diabetes, breast cancer in women, colon cancer, and total mortality from work conducted by Lee et al. (7). Note the similarity of RR for physical inactivity for each NCD outcome. When expressed as an attributable risk (Table 3), taking into consideration the prevalence of each risk factor across the globe, the public health implications for each of the risk factors compared with one another are more fully appreciated.

Often, to fully understand the etiology of a condition, we need to understand the relationship between two or more exposures associated with the disease/injury. This type of relationship is referred to as effect modification and occurs when the effect of one exposure on disease risk is modified by the presence of another exposure (8). An example from the smoking literature is the interaction of cigarette smoking and asbestos exposure in relationship to lung cancer. Exposure to smoking alone carries a RR of 10.8 compared with nonsmokers, while exposure to smoking plus exposure to asbestos carries a RR of 53.2, increasing the risk of lung

cancer almost 50-fold compared with nonsmokers and 5-fold compared to smokers (9).

In addition to the impact that interaction has in influencing epidemiologic studies, bias also plays a role. Bias occurs if the observed estimate of a measure tends to deviate from its true value; this deviation then obscures the true relationship between an exposure variable and an outcome. Various types of bias that can be introduced in epidemiologic studies are listed in Table 4. Examples of bias include selection bias, where only certain subjects from a community are enrolled in a study when broader representation is desired. An example of information bias is where errors are made in classification of risk status or disease status. Misclassifying physical activity exposure by either use of an incomplete assessment measure or an unreliable instrument would produce significant errors in any study examining the role of physical activity and health outcomes.

Standard Clinical Measures, the Electronic Health Record (EHR), and Data Sources

The clinical exercise physiologist or an exercise facility should regularly collect physical activity and physical fitness data on participants. The data variables and methods of data collection will differ between sites. The needs of the

TABLE 2. Relative risks of coronary heart disease (CHD), type 2 diabetes, breast cancer, colon cancer, and all-cause mortality associated with physical inactivity.^a

Outcome	Unadjusted Relative Risk (95% confidence interval)	Adjusted Relative Risk (95% confidence interval)
CHD	1.33 (1.18–1.49)	1.16 (1.04–1.30)
Type 2 diabetes	1.63 (1.27–2.11)	1.20 (1.10–1.33)
Breast cancer	1.34 (1.25–1.43)	1.33 (1.26–1.42)
Colon cancer	1.38 (1.31–1.45)	1.32 (1.23–1.39)
All-cause mortality	1.47 (1.38–1.57)	1.28 (1.21–1.36)

^aAdapted from Lee et al. (7)

TABLE 3. Summary of population attributable fraction (PAF) of physical inactivity for CHD, T2DM, CCa, and BCa for the world.^a

	Total Mortality	CHD	T2DM	CCa	BCa
PAF RR _{unadj} , % (CI)	14.2 (7.2–13.4)	10.4 (10.8–22.8)	18.1 (6.8–15.5)	11.6 (6.8–15.1)	11.8 (8.3–18.0)
PAF RR _{adj} , % (CI)	9.4 (5.1–12.5)	5.8 (3.2–7.8)	7.2 (3.9–9.6)	10.1 (5.6–14.1)	10.4 (5.7–13.8)

^aCHD = coronary heart disease; T2DM = type 2 diabetes; CCa = colon cancer; BCa = breast cancer; RR_{unadj} = relative risk unadjusted for confounding factors; RR_{adj} = relative risk adjusted for confounding factors (e.g., smoking, obesity, hypertension); CI = confidence intervals. Adapted from Lee et al. (7)

facility, the target population, available time, and the cost of testing will be factors that determine how data is collected and which variables are assessed. When collecting data, basic pieces of information should always be collected. This includes date of data collection, age, height, weight, race/ethnicity, and an assessment of physical activity per week. (Ideally, subjects should also have a data identifier; this is necessary if serial data is collected.) Collecting this information will provide a basis of comparison between different data sets. Different sites and systems should coordinate with each other to ensure that data variables and/or collection measures are as consistent as possible. For example, in multisite randomized-controlled trials this is often accomplished using a system called a “core lab.” The core lab educates all sites on proper data collection and provides oversight of this process.

Physical Activity

The assessment of physical activity should be completed using a verified tool. Survey questions should provide a basic assessment of exercise volume per week. Ideally, surveys or activity recall tools should assess multiple aspect of fitness (aerobic, strength, and flexibility). Table 5 provides a brief list of validated physical activity survey tools as reported by Dishman et al. (3) and can be found on the physical activity surveillance tools Centers for Disease Control and Prevention (CDC) website: (<https://www.cdc.gov/physicalactivity/data/surveillance.htm>) (10). Using surveillance tools can provide the exercise facility with a means to compare collected data to existing databases, which are available for public use.

The facility should determine which is more appropriate: a shorter assessment using a few questions or a more precise assessment using a larger number of questions. Asking a few questions may be preferred when the complete content of the survey is longer and when the focus of data collection is not specifically associated with physical activity. Longer surveys with more precise measures of physical activity will help tease out specific associations with physical activity type, volume, and health outcomes. The method by which the survey is delivered (e.g., led by an interviewer, self-reported, paper and pencil, electronic device) can also affect the precision of the data collection.

Fitness Measures

Exercise and wellness facilities will complete a variety of fitness evaluations on participants. Fitness variables can be

collected for aerobic fitness, muscular strength, muscular endurance, body composition, flexibility, and movement. A list of measures are contained in Table 6. This list is not necessarily exhaustive; therefore, fitness facilities should adjust the evaluations based on their needs and resources.

The methods by which fitness is assessed will vary, although a description of standard methods can be found in professional exercise testing textbooks (11,12). (Note: Care should be taken when using online sources for exercise testing. Websites and video posting sites will post methods that are not validated. Therefore, the accuracy of the information posted is often not known.) Fitness measures should be determined as a complete (holistic) entity. The collection of fitness variables (i.e., strength, muscular endurance, movement, and flexibility) should be as complete and well-rounded as possible. This can be a challenge. For example, strength measures are often completed for a bench press (upper body exercise) and leg press (lower body exercise). However, these exercises do not address many upper body muscle groups (such as the latissimus dorsi, and biceps for a bench press) or lower body muscle groups (such as the hamstrings and adductors muscles for a leg press). Therefore, developing a composite list of exercises that address the entire body is preferred. Similarly, muscular endurance tests often address upper body movements (such as curl-ups and push-ups) without addressing lower body muscular endurance. Flexibility assessments often consist of a sit and reach test, which does not address many potential areas of inflexibility in the body. It is accepted that developing a holistic physical fitness evaluation system may not always be possible. However, the goal should be to develop a well-rounded system of physical fitness evaluations that address the entire body and multiple aspects of physical fitness.

Completing physical fitness evaluations are mired in a lack of consistent methodology. This can be a significant factor when trying to make data comparisons. There are three main factors that can affect the ability to compare data from one facility to another: (1) *Different testing*

TABLE 4. Types of bias in epidemiologic studies.

Selection	Information	Measurement
Ascertainment	Interviewer	Validity
Detection	Recall	Reliability
Response	Reporting	

TABLE 5. Current US government surveillance systems that collect physical activity behavior among US adults.

Survey (website)	Collection Method (mode)	Physical Activity Domains
National Health Interview Survey (NHIS): Cancer Supplement (http://www.cdc.gov/nchs/nhis.htm)	Survey questions (in-person); national representative sample of US adults 18 years and above	Active transportation
National Health Interview Survey (NHIS): Sample Adult Survey (http://www.cdc.gov/nchs/nhis.htm)	Survey questions (in-person); national representative sample of US adults 18 years and above	Leisure-time
National Health and Nutrition Examination Survey (NHANES) (http://www.cdc.gov/nchs/nhanes.htm)	Survey questions (in-person); national sample of US population, non-Hispanic whites, non-Hispanic blacks, and Mexican-Americans age 2 months old and above	Active transportation Leisure-time Work (paid/unpaid)
National Health and Nutrition Examination Survey (https://www.cdc.gov/Nchs/Nhanes/2003-2004/PAXRAW_C.htm)	Physical activity monitor (accelerometer); national sample of US population, non-Hispanic whites, non-Hispanic blacks, and Mexican-Americans ages 2 months old and above	Objective assessment (Actigraph)
Behavioral Risk Factor Surveillance System (BRFSS) (http://www.cdc.gov/brfss/)	Survey questions (telephone); national sample conducted in every state and territories metropolitan micropolitan statistical area-level data available	Non-occupational
American Time Use Survey (ATUS) (http://www.bls.gov/tus/)	One-day activity diary (telephone); national sample ages 15 years and above	Active transportation Leisure-time Household Occupational
National Household Travel Survey (NHTS) (http://nhts.ornl.gov/)	One-day travel diary survey question (telephone); national sample ages 5 years and above	Active transportation
American Community Survey (ACS) (http://www.census.gov/acs/)	Survey questions (online/mail/in-person); national representative ages 16 years and above	Active transportation

TABLE 6. Fitness variables and assessment methods.

Fitness Variable	Definition	Potential Measures
Aerobic fitness	Ability of the circulatory and respiratory system to supply oxygen during maximal (aerobic) activity	Estimated or measured VO_{2peak} through a maximal, submaximal, or field test using an aerobic mode of exercise (walking, running, cycling, etc.)
Muscular strength	Ability to exert force	A one repetitional maximal (RM) lift (or an estimate of 1 RM based on a 2–5 RM) divided by the individual's body weight
Muscular endurance	The ability of muscle to continue to perform without fatigue	The number of repetitions that a subject can complete moving a submaximal load (typically body weight) or the amount of time an individual can hold an isometric pose (e.g., holding a plank for time)
Body composition	The relative amounts of muscle, fat, bone, and other vital parts of the body (organs)	The percent body fat as determined by a number of accepted methods (underwater weighing, air plethysmography, DEXA, ^a skinfold, and bio-electrical impedance), body mass index, waist circumference, and waist (circumference) to hip (circumference) ratio
Flexibility	The range of motion available at a joint	Degrees that a joint structure can move in a given plane (as measured with a goniometer), the distance that an individual can move for a specific motion (e.g., sit and reach test), and the distance that limbs will be apart from each other (back scratch/shoulder mobility test, where the distance between hands is determined)
Movement	The ability to complete an exercise movement with proper ergonomic form	The functional movement screen (FMS) score, which is determined by scoring the subject's proficiency of completing exercise movements with the proper body mechanics and a squat test, which evaluates different body components when completing a squat and holding a dowel overhead

^aDEXA = dual-energy x-ray absorptiometry

modalities—The testing modality provides data that cannot be easily compared. This is the case when an individual completes a maximal aerobic exercise test on a treadmill versus other pieces of aerobic equipment that will produce lower maximal aerobic capacities (cycle ergometer, rowing ergometer). In such cases, care needs to be taken to apply the proper correction factor to make appropriate comparisons. This is also an issue when collecting body composition variables, where the comparison of percent body fat, body mass index, and waist circumference is difficult. These variables are correlated. However, if sites only collect one of these variables, comparisons are difficult (i.e., percent fat versus waist circumference versus body mass index). (2) *Different testing equipment*—This is particularly a problem when completing strength tests and different strength training machines are used. Differences between sites may be due to the use of equipment that has a different lever arm (this is typical with the leg press). Therefore, one population may seem stronger and fitter than another due to the use of different equipment, when there is not necessarily a true difference between the groups. (3) *Test administration*—The skill of the individual completing the measures may vary from facility to facility. Since the fitness industry is not regulated, individuals at a facility can be highly trained and skilled in performing fitness testing measures or the individuals can have very little professional training and knowledge of how to perform a test properly. Individuals with a high degree of training should complete fitness measures with a high degree of validity and reliability. However, the validity and reliability of individuals who do not receive professional training is likely to be lower. It is expected that testing requiring a significant degree of skill (such as measuring skinfolds or rating functional movement screen movements) or requiring participants to complete movements with a proper form (push-up, sit-up tests) would not be completed at a similar level and thus scores may be inconsistent based on the test administrators who implement the fitness testing measure. This would affect the comparison of fitness variables between facilities and the relationship between a fitness measure and disease outcome.

Importance of Improving Data Collection

Physical activity data have been regularly collected from national surveillance systems. However, data associated with fitness measures need to be collected and evaluated. Data from commercial fitness facilities are rarely published. Data from hospital and corporate wellness facilities are often collected and evaluated. However, the results may not necessarily be published. Currently, many facilities regularly perform some level of fitness testing (even if the testing is incomplete). Therefore, considerable data are being collected on the population. However, data collection methods need to be more consistent and comply to increased rigor.

Efforts must be made to unify physical activity and fitness data collection methods, unify assessment variables, and create data repositories. Often, normative data are collected from a single source and presented as a single aspect

of fitness. Relatively large numbers of individuals are tested from a single location (such as fitness data developed by the Cooper Clinic, Dallas, Texas and known as the Aerobics Center Longitudinal Study) and data are presented for a single fitness variable (11). Normative data where populations from different geographic locations complete multiple fitness measures have not been presented. Thus, there are no data concerning multiple aspects of physical fitness and its relationship to health and longevity. It can be hypothesized that individuals who meet a certain threshold of physical fitness across multiple indices of fitness have improved health outcomes compared to individuals who perform poorly on multiple physical fitness indices. An American Heart Association scientific statement asserts that low levels of aerobic fitness are associated with an increased risk of cardiovascular disease and all-cause mortality (13). However, there are no current data to address multiple domains of physical fitness (e.g., aerobic, strength, flexibility, body composition) in an individual and their overall effect on health outcome.

With a consistent collection of physical activity and (complete) physical fitness measures, repositories need to be developed that couple these measures with electronic medical records to determine correlations. This will determine thresholds for physical activity and fitness that are associated with improved health outcomes and can guide clinical recommendations. Fitness thresholds have been determined for aerobic fitness (13). However, correlations and thresholds need to be determined for other measures of physical fitness and for multiple aspects of physical fitness in combination.

In summary, there is a need to consistently collect basic descriptive information on the study population. Physical activity should be assessed using validated survey tools. Physical fitness tests should assess the multiple aspects of fitness as previously listed. Fitness facilities need to work together to complete measures that are consistent and performed with a high degree of rigor. Repositories need to be created where physical activity, physical fitness, and health outcomes can be evaluated. Finally, hypotheses need to be evaluated that can guide clinical and public health policy.

Study Designs in Epidemiologic Research **Cross-Sectional Study**

When attempting to explain an association between a factor and a disease/injury, the least convincing design is cross-sectional. In this design, physical activity or fitness is measured simultaneously with a measure of the frequency of disease, injury, or death. Other risk factors may also be measured at the same time. Because this approach is analogous to the “snapshot” in photography (14), proper temporal sequence is not provided. An example of a physical activity study that uses a cross-sectional approach is The Iowa Farmers Study (15), which examined the association of physical activity with mortality; 62,000 all-cause deaths occurring from 1962 to 1978 in male residents of Iowa aged 20–64 years were examined. A randomly selected group of 95 farmers was compared with a group of 158 nonfarmers who

lived in a city. Farmers had a 10% lower rate of death due to CHD, and they were twice as likely to participate in strenuous physical activity compared with the nonfarmers. The farmers were also more fit as determined by lower exercise heart rate and longer endurance time on a treadmill test. Because the farmers had higher cholesterol and higher body mass index, the apparently protective effect attributed to their high activity and fitness could not be explained by lower cholesterol and body mass. In other words, when these known risk factors for CHD were controlled for, the farmers still appeared to benefit directly from their higher fitness levels (i.e., some evidence for independence of the effects of physical activity and fitness was present). However, the farmers had lower estimated body fat, as determined by skinfold thickness, and their rates of smoking and alcohol consumption were half that of the city dwellers. Therefore, the apparently protective effect of physical activity was not fully independent; it could just as likely be explained by the marked reduction among the farmers in the known risks of smoking and drinking alcohol. Thus, a conclusion that the active lifestyle of the farmers protected against CHD deaths must be accepted with an element of caution.

Another application of the cross-sectional study is determining the prevalence of selected conditions and health behaviors for the purpose of public health planning and surveillance. For example, through the Behavioral Risk Factor Surveillance System (BRFSS), which is maintained by the CDC in Atlanta, physical activity and other health behaviors are routinely assessed in each of the states (16). This information is useful in establishing the current physical activity patterns among demographic groups and geographic regions. The BRFSS has recently reported that, on average, about 30% of US adults are inactive during their leisure time, failing to meet the 2018 Physical Activity Guidelines for Americans (17). A wide variation in the prevalence of leisure-time physical inactivity among adults 18 years and older across the United States is shown in Figure 1, with states in the West reporting lower estimates of physical inactivity compared, for example, with states in the Southeast (18).

Relevant to the issue of validity among exposure or outcome measures used in epidemiologic studies is the use of the kappa statistic. This is a measure of the degree of non-random agreement between measurements of the same categorical variable (e.g., activity counters vs. physical activity questions). If the measures agree more often than expected by chance, kappa is positive; if concordance is complete, $\text{kappa} = 1$; if there is no more nor less than chance concordance, $\text{kappa} = 0$; if the measures disagree more than expected by chance, then kappa is negative (1). When developing measures to be used in studies and where a “gold standard” is lacking, the comparison of valid and reliable measures using the kappa statistic is advantageous.

Case-Control Study

When there are no clear suspected causes of a disease, an epidemiologist operates very much like a detective, attempting to piece together causes after the fact. In this situation,

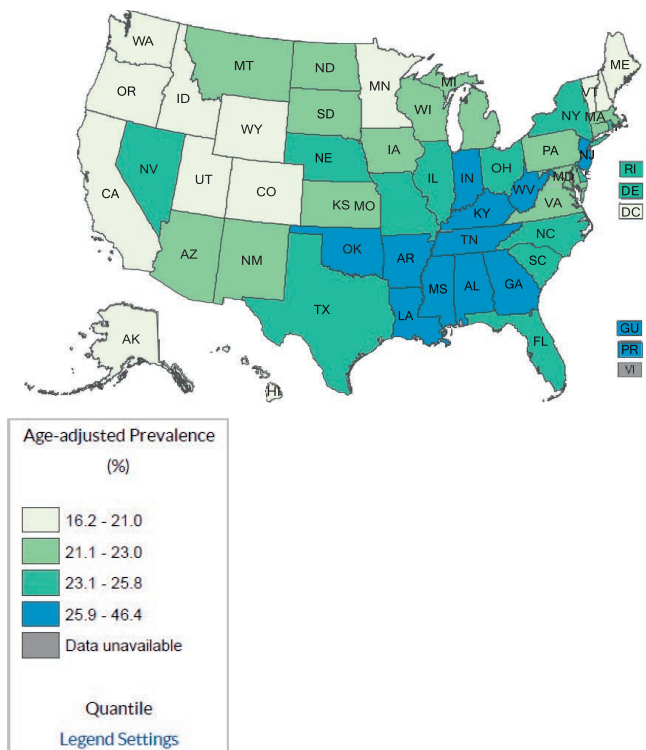


FIGURE 1. Prevalence of physical inactivity among women, United States, 2017. Data source: Behavioral Risk Factor Surveillance System (BRFSS), Centers for Disease Control and Prevention (16).

the most common design is the retrospective case-control study. This approach can be compared to a flashback in cinematography (14). An example of a retrospective case-control study is the Seattle Heart Watch Study (19), which examined 1,250 cases of sudden cardiac death among men and women aged 25 to 75 years living in the Seattle area to determine the association of physical activity habits with risk of sudden cardiac death. Of these cases, 163 were selected in which subjects had appeared risk-free prior to the time of their fatal heart attack. Spouses were interviewed about the decedent's physical activity at work and during leisure time during the year preceding death. Each case was paired with a randomly selected control who had similar age, smoking habits, and blood pressure. Low activity on the job and low or moderate activity during leisure were unrelated to death rate. However, people in the top 50% of participants in vigorous physical leisure activities as determined by using the metabolic equivalent of task unit categories that were equivalent to 60% (i.e., vigorous intensity) of aerobic capacity or higher (jogging, climbing stairs, chopping wood, swimming, singles tennis) had just 40% of the risk for sudden cardiac death when compared with cases who spent no leisure time performing vigorous physical activities (18).

Prospective Cohort Study

A prospective study permits observation of the characteristics and behaviors of a group or cohort of people across time. It permits a natural history of physical activity, fitness, and health-related events to be chronicled as they occur, much

like a motion picture. Because it is longitudinal, a prospective cohort design enables an investigator to measure physical activity and health-related events at multiple points in time and consequently test whether an association between physical activity and a low rate of disease is persistent.

The Aerobics Center Longitudinal Study measured physical fitness defined as endurance time on a treadmill test in over 10,000 men and 3,000 women at the time they participated in a preventive medical examination (20). The men and women were re-examined about 8 years later. During the period of observation, 240 deaths among men and 43 deaths among women occurred after about 110,000 person-years of exposure. Age-adjusted death rates (per 10,000 person-years of exposure) from all causes were lower with each successive level of fitness in men from the least fit (64 deaths) to the most fit (19 deaths), and similarly in women from the least fit (40 deaths) to the most fit (9 deaths) (20). The effects of higher fitness were independent of age, smoking, cholesterol concentration, systolic blood pressure, blood sugar, and parental history of coronary heart disease. Much of the decreased death was explainable by reduced rates of cardiovascular disease and all-site cancers (20).

The Randomized Clinical Trial (RCT)

This study design is considered the “gold standard” to determine whether associations uncovered in epidemiologic observations represent cause-and-effect relations. The validity of the RCT depends on having a representative population sample and matching treatment and control groups with respect to characteristics thought to affect outcome. The random assignment of subjects to the treatment or control group is essential to equally distribute known and unknown confounding variables between groups.

Examples of a randomized study design in exercise science are secondary prevention trials among heart attack survivors and persons with chronic heart failure to determine whether exercise training reduces recurrence rates of morbidity and premature death.

For example, in a study performed in Finland (21), 375 men and women who had survived a myocardial infarction at the time of hospitalization were randomized into either a multiple risk factor intervention group or a control group. The intervention group (which included exercise) had a significant reduction in total cardiovascular mortality and cardiac sudden death, but not in reinfarction. However, because there was no evidence of improved physical fitness on bicycle ergometer testing in the intervention group, the independent effect of exercise was not clearly demonstrated (21).

In the United States, part of the multicenter HF-ACTION (Exercise Training Program to Improve Clinical Outcomes in Individuals with Congestive Heart Failure) trial sought to evaluate the influence of baseline physical activity levels on responses to aerobic exercise training and clinical events in outpatients with chronic systolic heart failure.

Changes among 742 patients in exercise capacity, all-cause mortality, cardiovascular mortality, and hospitalization

were evaluated as a function of the baseline tertiles (three evenly divided groups) of physical activity (22). At baseline, the highest physical activity tertile showed greater peak oxygen uptake, cardiopulmonary exercise test duration, and 6-minute walk test distance than the other two physical activity tertiles. Compared to the lowest physical activity tertile, the middle tertile had an 18% lower risk of cardiovascular disease and hospitalizations, and the upper tertile showed a 23% lower risk of cardiovascular disease death and heart failure hospitalizations. The investigators concluded that patients with chronic heart failure exposed to aerobic exercise training significantly improved exercise test duration to a similar extent across all baseline physical activity tertiles. Despite these differences, there were no significant differences in event rates within each physical activity tertile comparing the subgroups randomized to exercise training versus usual care (22).

Assessment of Causality

Community-based or clinic-based interventions are established on the presumption that the associations found in epidemiologic studies are causal rather than occurring by chance or because of bias. However, in most instances in which epidemiologic methods are used to observe health events in the population, the circumstances do not permit the investigator to absolutely prove that an association is causal. That said, several cardinal principles or criteria have been used in epidemiologic research for judging the strength of inference drawn from studies about the cause-and-effect relationship between a factor such as physical inactivity and a disease or injury. These criteria were initially developed by Sir Bradford Hill (known as the Hill Criteria) and are often cited as a checklist for causality in epidemiologic studies (23; see Text Box).

Strength of Association

The first criterion is that studies show a statistically meaningful association (i.e., not likely to be explainable by random or chance observation) between physical activity and lowered prevalence or incidence of disease. The stronger these associations, the less likely they are the result of confounding or bias.

Consistency of Association

Consistency is achieved when the association of increased physical activity or fitness with lower rates of disease is similar for different types of people, in different geographical regions, and when different measures or components of physical activity or fitness are used. Consistency of association makes bias an unlikely explanation for such a series of observations.

Specificity of Association

Even though a study may show a dose-response pattern between increasing levels of physical activity and decreased risk for disease, the pattern of reduced risk seen with increasing levels of physical activity must remain in the presence

and in the absence of other potential causes of the disease. An illustration of this criterion, taken from the physical activity and CHD literature, is the Harvard Alumni Study carried out by Paffenbarger (24).

Temporality

For a lower rate of disease or death associated with higher levels of physical activity or fitness to be interpreted as being possibly caused by activity or fitness, sedentary or unfit subjects must be similarly as healthy at the onset of the study as are subjects determined to be more physically active or fit. Also, the measurement of physical activity or fitness must precede the measurement of subsequent events of disease or death.

Biologic Gradient

If physical activity exerts a protective effect for reducing disease, injury, or death (or conversely, causes some kinds of injury), it should be possible to determine some systematic pattern of relationship between increasing levels of physical activity and altered rates of disease, injury, or death. If rate ratios vary randomly across levels of activity or across differing changes in physical activity, an attempt to explain that physical activity was causally responsible for the variation would be unconvincing. The most convincing pattern would be a linear gradient of decreased rate ratios that was proportional to each increment of increased physical activity or physical fitness. It is also possible that the dose-response relationship is curvilinear, such that each successive increment in physical activity or fitness corresponds with an accelerating change in the rate ratio of disease, injury, or death. A negatively accelerating dose-response would indicate an attenuation of benefit, meaning that the proportionately largest reduction in the rate ratio would be noted at relatively low levels, or across small increases, of physical activity or fitness, with reductions becoming progressively smaller at the higher levels, or at larger increases, in physical activity or fitness. A positively accelerating relationship would indicate the converse; in other words, larger benefits would occur at higher levels, or greater changes, of physical activity or fitness. Finally, it is possible that a threshold of response may exist rather than the typical graded pattern. That is, there may be some minimal level of physical activity or fitness that explains all, or nearly all, of the altered rate ratios. Once the minimal threshold is exceeded, no further change in disease, injury, or death would be observed.

Plausibility

Even when the preceding criteria have been met, the overall case established for cause-and-effect will remain weak if the association between increased physical activity and decreased disease or death cannot be explained. A convincing explanation requires evidence that physical activity or physical fitness induces biological changes that are coherent with the current etiology (i.e., understanding of the causes and course of development) and the pathophysiology (i.e.,

the process by which the function of cells and systems deteriorate) of a disease.

Coherence

Once the proper temporal sequence is established, it is still important to determine whether an association noted between physical activity or fitness and disease rates remains as time passes and that the evidence is not contradictory to the known biology and natural history of the disease.

Experimental Evidence

The most compelling evidence that increased physical activity reduces rates of disease or death would come from an experiment conducted in a large group of initially healthy people drawn randomly from a total population, with the participants randomly assigned to at least three levels of physical activity of differing intensity or amount, or to a control group that remained sedentary for several years. However, a study like this would be extremely costly, difficult to manage, and require tremendous resources, including funding.

In the absence of a population experiment, confirmation must come from studies of lower animals. Studies using rats, dogs, and nonhuman primates show favorable changes in the cardiovascular system after exercise training. In addition, many clinical studies with small groups of humans show that physical activity can reduce mild hypertension and alter blood lipids, blood sugar, clotting factors, and white blood cells in positive ways, as well as stimulate bone mineral density and reduce depression, to name a few benefits. Such clinical experiments are important for demonstrating the efficacy of physical activity for health-related outcomes and for building a stronger case for biologically plausible mechanisms. Nonetheless, they cannot demonstrate that the benefits observed for small select groups of people are generalizable to larger segments of the population.

Analogy

When many studies find associations between physical activity or fitness with reduced risks of disease or death, it is more likely that each study estimated the same true effect of physical activity. When studies do not agree, it must be determined whether the differences can be explained by study-related factors. An example is a study's potential use of different, or sometimes inaccurate, methods of measuring or defining physical activity, fitness, or health-related outcomes (i.e., measures were imprecise or not comparable among the studies). Another example is comparisons of physical activity or fitness levels without proper control or accounting of other factors that might have contributed to rates of disease or death more than did the differences in activity or fitness (i.e., independence was not uniformly assured among the studies). Sometimes studies are hampered by improper and misleading uses of statistical theory and tests (i.e., computations of rates or the conclusions reached from the rates were wrong). Finally, differences between physical activity/fitness and health-related measures may

exist because (a) different types of people were observed, (b) different characteristics/components of physical activity or fitness were measured/manipulated, or (c) different amounts/levels of physical activity or fitness were compared.

When all other criteria for judging the scientific strength of cause-and-effect evidence are satisfied, the number of studies finding results that agree determines the confidence with which it can be concluded that physical activity or fitness improves health or longevity. As is the

case for the other criteria, the number of studies that agree differs widely according to the disease or health outcome studied. An example of the application of the Hill Criteria is the review by Powell and co-workers (25) examining the relationship between physical activity and the incidence of CHD. These authors set forth a compelling case for the cause-and-effect relationship between increasing levels of physical activity and the prevention of CHD in their review (25).

THE HILL CRITERIA (23)

1. Strength: Stronger associations are less easily explained away by confounding than weak associations.
2. Consistency: Similar conclusions are found from among diverse methods of study and in different populations under a variety of circumstances.
3. Specificity: Exposure is linked to a specific effect or mechanism.
4. Temporality: Exposure *always* precedes the disease or outcome in time.
5. Biological Gradient: Increases in exposure dose translates into an increased dose-response in risk.
6. Plausibility: Appears worthy of belief—that is, the mechanism must be plausible in the face of known biological facts.
7. Coherence: The data/facts stick together to form a coherent whole.
8. Experimentation: Experimental evidence supports observational evidence—based on biological and/or clinical studies.
9. Analogy: Similarities are seen among observed things/data that are otherwise different (this is considered a weak form of evidence). Example: Before HIV was discovered, epidemiologists noticed that AIDS and hepatitis B had analogous risk groups, suggesting similar types of agents and transmission.

REFERENCES

1. Porta M, editor. A dictionary of epidemiology. 6th ed. New York: Oxford University Press; 2014.
2. Martin AB, Hartman M, Washington B, Catlin A. National health care spending in 2017: growth slows to post-Great Recession rates. *Health Affairs*. 2019;38(1):101377hth aff201805085.
3. Morris JN, Raffle PAB. Coronary heart disease in transport workers: a progress report. *Brit J Indust Med*. 1954; 11:260–4.
4. Dishman RK, Heath GW, Lee IM. Physical activity epidemiology. 2nd ed. Champaign, IL: Human Kinetics; 2013.
5. Redwood R, Remington PL, Brownson RC. Chronic disease epidemiology. In: Brownson RC, Remington PL, Wegner MV, editors. *Chronic disease epidemiology and control*. 4th ed. Washington, DC: American Public Health Association; 2016. p. 35–76.
6. Rothman KJ. Measuring disease occurrence and causal effects. In: *Epidemiology: an introduction*. 2nd ed. New York: Oxford University Press; 2012. p. 38–68.
7. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, for the Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet*. 2012;380:219–29.
8. Rothman KJ. Measuring interactions. In: *Epidemiology: an introduction*. 2nd ed. New York: Oxford University Press; 2012. p. 198–210.
9. Saracci R. The interactions of tobacco smoking and other agents in cancer etiology. *Epidemiol Rev*. 1987;9:175–93.
10. Centers for Disease Control. Physical activity surveillance systems. <https://www.cdc.gov/physicalactivity/data/surveillance.htm>. Accessed October 7, 2019.
11. Riebe D, Ehrman JK, Liguori G, Magal M. ACSM's guidelines for exercise testing and prescription. 10th ed. Philadelphia, PA: Wolters Kluwer; 2018.
12. Gibson AL, Wagner DR, Heyward VH. *Advances fitness assessment and exercise prescription*. 8th ed. Champaign, IL: Human Kinetics; 2019.
13. Ross R, Blair SN, Arena R, Church TS, Despres JP, Franklin BA, Haskell WL, Levine BD, Lavie CJ, Meyers J, Niebauer J, Sallis R, Sawada SS, Sui X, Wisloff U. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign. A scientific statement from the American Heart Association. *Circulation*. 2016;143: e653–99.
14. Paffenbarger RS Jr. Contributions of epidemiology to exercise science and cardiovascular health. *Med Sci Sports Exerc*. 1988;20:426–38.
15. Pomrehn PR, Wallace RB, Burneister LF. Ischemic heart disease mortality in Iowa farmers. The influence of lifestyle. *JAMA*. 1982;248:1073–6.
16. Centers for Disease Control and Prevention. Behavioral risk factors surveillance system. https://nccd.cdc.gov/BRFSSPrevalence/rdPage.aspx?rdReport=DPH_BRFSS.ExploreByTopic&irbLocationType=StatesAndMMSA&isIClass=CLASS15&isIToPic=TOPIC23&isIYear=2018&rdRnd=50828. Accessed October 16, 2019.
17. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, George SM, Olson RD. The physical activity guidelines for Americans. *JAMA*. 2018;320(19): 2020–8.
18. An R, Xiang X, Yang Y, Yan H. Mapping the prevalence of physical inactivity in U.S. states, 1984–2015. *PLoS ONE*. 2016;11(12):e0168175. DOI:10.1371/journal.pone.0168175.

19. Siscovick DS, Weiss NS, Hallstrom AP, Inui TS, Peterson DR. Physical activity and primary cardiac arrest. *JAMA*. 1982;248:3113–7.
20. Blair SN, Kohl HW, Paffenbarger RS Jr, Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality: a prospective study of healthy men and women. *JAMA*. 1989;262:2395–401.
21. Hamalainen H, Luurila OJ, Kallio V, Knuts LF, Arstila M, Hakkila J. Long-term reduction in sudden deaths after a multifactorial intervention programme in patients with myocardial infarction: 10-year results of a controlled investigation. *Eur Heart J*. 1989;10:55–62.
22. Mediano MFF, Leifer ES, Cooper LS, Keteyian SJ, Kraus WE, Mentz RJ, Fleg JL. Influence of baseline physical activity level on exercise training response and clinical outcomes in heart failure: the HF-ACTION trial. *JACC*. 2018;6(12):1011–9. DOI:10.1016/j.jchf.2018.09.012.
23. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295–300.
24. Paffenbarger RS, Hyde RT, Wing AL. Physical activity as an index of heart attack risk in college alumni. *Am J Epidemiol*. 1978;108:161–75.
25. Powell KE, Thompson PD, Caspersen CJ, Kendrick JS. Physical activity and the incidence of coronary heart disease. *Ann Rev Publ Health*. 1987;8:253–87.