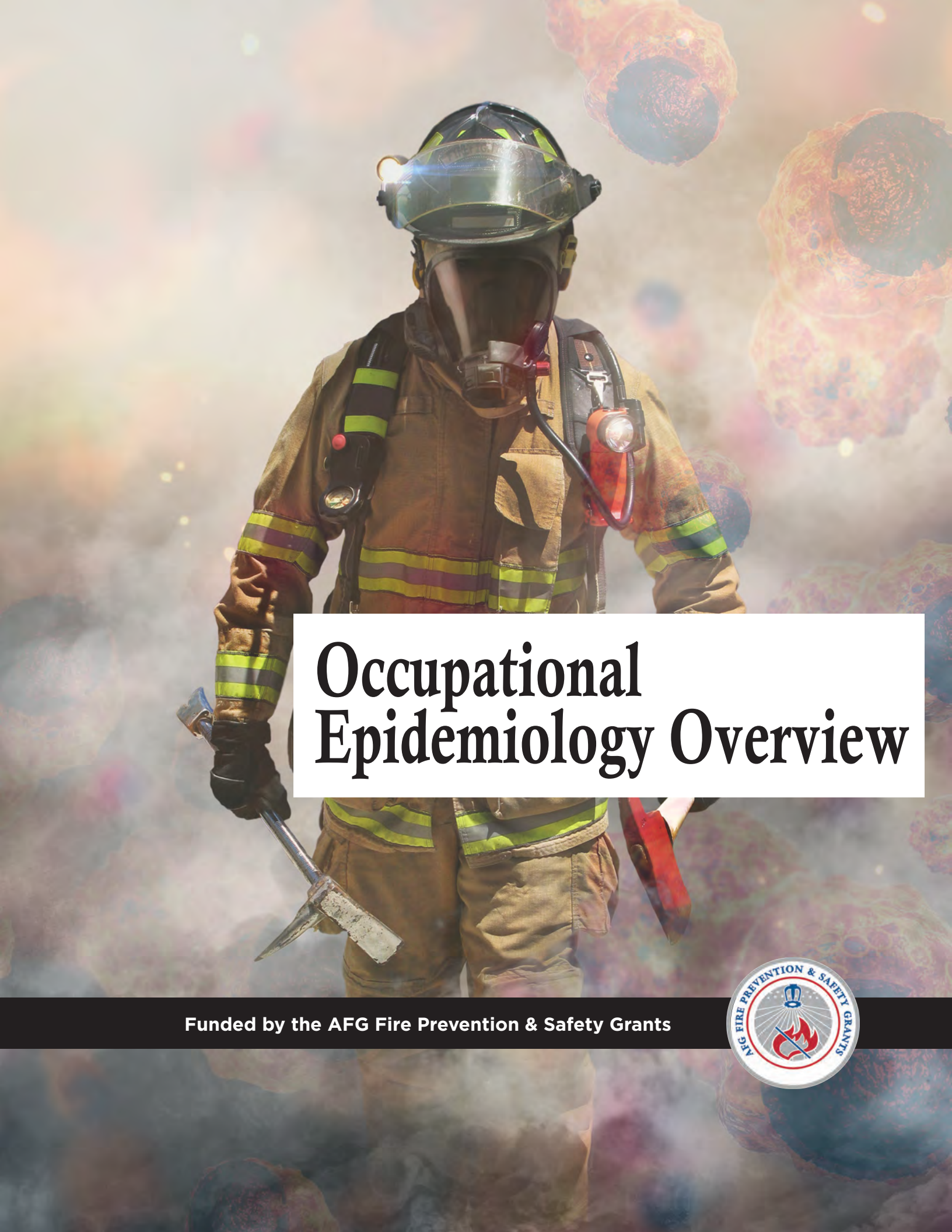

Occupational Epidemiology Overview

Funded by the AFG Fire Prevention & Safety Grants

2022

Authors:
Walker S. Carlos Poston, Ph.D., M.P.H., FACE
Sara A. Jahnke, Ph.D., FACE
Maria DH Koeppel, Ph.D.
Christopher K. Haddock, Ph.D., PStat®

Center for Fire, Rescue & EMS
Health Research
NDRI-USA, Inc.
New York, NY

The project team would like to acknowledge and thank Captain Robert Webb (Ret., Fort Worth Fire Department) who has been committed to supporting firefighters in understanding their cancers and sharing empirical evidence with them in their time of need. His assistance and review were appreciated by the team and his dedication to supporting his brothers and sisters is an inspiration.

Epidemiological Research Designs and Inferring Causality

While the "gold standard" in research for establishing causal relationships between exposures and disease outcomes is the randomized experiment[1], this approach would be impossible both practically and ethically when evaluating whether different occupations (and the things one is exposed to in those occupations) increase risk for disease and death. Scientists cannot randomly assign people to certain occupations to evaluate occupational disease risks. Similarly, it would be unethical to purposely expose some people to known carcinogens that are common in certain occupations while preventing exposure in others to study occupational diseases.

When confronted with these types of challenges, epidemiologists have most commonly utilized a research design termed the *longitudinal cohort study* to evaluate disease risks associated with different types of exposures. One of the best-known areas of research in which this approach was used was to understand the impact of tobacco use (particularly smoking) on the development of lung cancer and cardiovascular disease. Tobacco research demonstrates some basic principles in epidemiology for understanding how exposure to a potential toxin, in this case cigarette smoke, was determined to be a risk factor for the lung cancer, cardiovascular disease, and death. Cohorts (large groups) of individuals were assembled, their tobacco use status was determined (and other factors also assessed), and then they were followed over extended periods of time to ascertain disease development or death. It should be noted that this approach is imperfect because other factors can "confound" the theorized relationship between an exposure (e.g., smoking) and outcome (e.g., lung cancer). Confounding occurs when an association between two variables, *A* and *B*, is affected by some third variable that is itself associated with both *A* and *B*. For instance, smokers, aside from their tobacco use, also may engage in other unhealthy behaviors that may increase their risk for developing cancer or heart disease. Thus, longitudinal cohort studies are typically designed to address these differences and to employ the following principles to evaluate associations between exposures and disease outcomes or death when using this quasi-experimental approach[2,3]:

1. *Strength of the association* – Is risk of the disease much higher among exposed than unexposed persons?

2. *Consistency of the association among studies, especially when conducted using different techniques and in different samples/populations* – Is the association observed in various populations by various investigators using various approaches?

3. *Specificity of outcome* – Is a single putative cause producing a specific effect?

4. *Exposure precedes disease* outcome (temporal sequence) – Does the exposure precede the disease?

5. *Dose-response relationship/Biological Gradient* – If the exposure admits of gradation, is the amount of exposure associated with amount of risk?

6. *Plausibility of a biological mechanism* – Is the exposure linked to pathogenesis of the disease in some reasonable manner?

7. *Coherence of chain of evidence* – Is the association compatible with existing theory and knowledge?

8. *Experimental association* – Are there toxicological studies (e.g., animal studies, in vitro, etc.) using experimental approaches that support the association? Is there evidence that the condition can be altered (prevented or ameliorated) by an appropriate experimental regimen (e.g., does smoking cessation result in lowered cancer risk among ex-smokers when compared to current smokers?).

9. *Analogy to similar effect produced by a similar agent* – If one causal agent is known, does the second causal agent that is similar result in a similar association?

In addition to all the above, it also is necessary to determine the extent to which the existing body of research has taken other possible explanations into account and has effectively ruled out such alternate explanations. These epidemiological principles (aka the Bradford Hill Criteria[2,3]) for ascertaining whether an association between an exposure and outcome is causal are not absolute and should be applied rigorously and as a group. Thus, the more that appear to be satisfied, the more likely it is presumed that the association observed is truly causal. However, it is important to understand that the criteria can only be applied to disease risk in a population and not any individual case.

Unfortunately, the classic longitudinal cohort study approach is not practical in occupational epidemiology because many of the exposures of interest are rare in general population samples and/or very specific to particular occupations. Thus, when evaluating disease and/or mortality risk associated with a specific occupation, it is most common to conduct a *historical* cohort of occupational exposures study, which is one in which some or most of the outcomes have occurred before the epidemiological investigation has been initiated[1]. Historical cohort studies clearly have an advantage with respect to time in that the investigators don't have to wait for years while new cases of the outcome occur; however, they are more susceptible to selection and information bias than concurrent/prospective longitudinal cohort studies. Another issue with this type of study is the fact that there is no “unexposed” group (i.e., using smoking as an analogy, there would be no non-smokers to compare with smokers with respect to disease risk) because, typically in occupational epidemiology studies, the focus is on one occupational group so everyone is exposed. This is the typical approach for firefighter health studies, where all participants are exposed because they are firefighters. Recently, in more sophisticated occupational epidemiology studies among firefighters, levels of exposure based on time in the career, run/call volumes, etc. have been used to estimate levels or intensity of exposures to toxins. However, there is still the issue of whom should firefighters be compared with to evaluate if their risk of disease or death is elevated? How can we determine whether employment as firefighter is associated with increased risk of cancer incidence or death? Increased in comparison to what other group? The most common approach in occupational epidemiology has been to ask whether the observed frequency of cancer cases or deaths in a cohort of firefighters is greater than the frequency that would be expected if the firefighters had experienced the same rate as the general population of persons of the same age, sex, and place of residence[1]. This is where standardized mortality ratios (SMR's) and standardized incidence ratios (SIRs) are useful and used to estimate morbidity and mortality risk.

Measures of Association: SMRs and SIRs

The SMR is the observed number of cases of mortality (death) divided by the number expected if the study population had experienced the same age-specific death rates as the

standard population. It often is used to assess the risk associated with working in a particular occupation or living in some particular location. For this purpose, the standard population should be as similar as possible to the study population except for the exposure, e.g., rates for persons of the same age, sex, and race in the same state at the same calendar time as persons in the study population. The following data are needed to calculate an SMR:

1. The death rates for the ***standard*** population, specific for age, sex, and race;
2. The distribution of the ***study*** population by the same categories of age, sex, and race. (This is best done with person-years, but can be done with just persons.); and
3. The total number of deaths in the ***study*** population.

The first two items are used to calculate the expected number of deaths in the study population, i.e., the number of deaths that would have occurred if the study population had experienced the same specific death rates as the standard population. SMR's are generally used when data needed to calculate accurate age-specific rates are not available.

Similar to the SMR, the SIR is a useful tool for screening incidence data (i.e., the occurrence, frequency, or rate of some disease of interest) and generating leads for occupational health research with respect to understanding the impact of occupational exposures. The SIR is the observed number of incident disease cases divided by the number expected if the study population had experienced the same age-specific rates as the standard population. Similar to SMRs, this data usually comes from state or national sources. In the case of firefighter cancers, it is not uncommon to use some general population rate from state or national cancer registries. The following data are needed to calculate an SIR:

1. The cancer incidence rates for the ***standard*** population, specific for age, sex, and race;
2. The distribution of the ***study*** population by the same categories of age, sex, and race. (This is best done with person-years, but can be done with just persons.); and
3. The total number of incident cancer cases in the ***study*** population.

The first two items are used to calculate the expected incidence in the study population, i.e., the number of incident cancer cases that would have occurred if the study population had experienced the same specific death rates as the standard population. Like SMR's, SIRs are used when data needed to calculate accurate age-specific rates are not available.

Both SMRs and SIRs represent a method of standardization using the indirect method because it answers the following question: How many deaths or incident cases would have been expected in the study population if it had experienced the same mortality or disease incidence as the reference population? Thus, this indirect method of standardization refers to the SMR and/or SIR being adjusted to control for the potentially confounding effects of one or more variables (e.g., differences in age, gender, etc.) by calculating a weighted average of category-specific rates by applying mortality or incidence rates from a standard population to weights of exposure time in the occupational group of interest (e.g., firefighters).

Interpreting SMRs and SIRs

SMRs and SIRs are generally interpreted in the following manner. An SMR of 1.0 (or 100% if the 100% multiplier is used, which is not uncommon presentation format in some studies) means there is no difference in mortality or disease risk for firefighters when compared to the general population. SMRs/SIRs less than 1.0 (100%) therefore imply lower risk and those greater

than 1.0 (100%) imply greater risk. SMRs and SIRs often are computed with 95% Confidence Intervals, which can be interpreted as that we are 95% confident that the calculated interval encompasses the true population parameter. When 95% Confidence intervals for SMRs/SIRs contain 1.0, then they are statistically not significant.

It should be noted that SMRs/SIRs below 2.0 (i.e., a doubling of risk) are not uncommon in occupational epidemiology studies due to a number of factors that are discussed below in the section on factors that may lead to underestimation of risk[4]. However, we believe that statistically significant associations should be considered as potentially important even when below risk doubling because, for one, statistical significance implies that it would be rare for an effect size of that magnitude to exist if there were actually no risk in the population. In addition, diseases like cancer are complex and many factors play a role in the etiology, course, prognosis, and mortality risk, so any significant elevation in risk, given all the factors discussed later that likely result in the underestimation of risk, should be considered as a potential health concern for firefighters. This is particularly important for firefighters, as noted by Guidotti[5]:

> "They may be trained to manage these risks and to protect themselves, but the working environment cannot be made safe because they deal with situations that are inherently dangerous and may lose control. In the interests of society and as safety professionals, however, they essentially waive the right to refuse dangerous work and routinely accept the risk, like a soldier sent into battle to defend the country. It is, by this logic, ultimately in society's interest to compensate for this risk because the work has to be done." (pg. 6).

Factors that Likely Underestimate Disease and Mortality Among Firefighters

Cancer risk estimates are usually in the form of standardized mortality ratios (SMRs) or standardized incidence ratios (SIRs). These risk estimates are derived by standardizing the age-distribution of the corresponding study population to some standard, most often the general population. This is because most occupational epidemiology studies of firefighters' cancer risks are designed as occupational cohort studies where there is no internal unexposed group. This type of study design is the most practical and avoids ethical issues that would be inherent in conducting an experimental study (i.e., we can purposely expose some firefighters to carcinogens) or traditional observational studies where there is an unexposed group in the cohort (e.g., comparing cancer risks for smokers vs. non-smokers.

Therefore, SMRs/SIRs from different study samples/populations are not directly comparable in the same manner as rates that were derived using direct standardization (i.e., when the death or disease rates of some group are applied to a standard population, the opposite of how SMRs and SIRs are computed) because study samples may have underlying differences that can affect mortality or disease rates, such as differences in age. For example, even in the case of one sample of firefighters where SMRs could be computed based on categories of years of experience as a proxy exposure variable, the SMRs may not be comparable if the age distributions of firefighters with more years of service substantially differs from those with few years of service.

SMRs/SIRs, when computed using the general population, likely underestimate risk because standard populations contain both exposed and unexposed individuals (e.g., disease or death rates from the US general population or state vital statistics will include firefighters, so firefighters in the denominator will suppress the risk ratio). Other factors that may lead to

underestimation of death or disease risk, particularly for firefighters, include exposure misclassification, the "Healthy Worker Effect", and when studies focus on rare outcomes.

Accurate exposure classification for fire service studies is generally relatively crude and often based solely on job title. However, fire departments, and even stations within departments, vary considerably with respect to their overall call volumes, the number of fires that they fight, the types of fires they fight, and how often they respond to hazardous materials situations. Thus, there can be tremendous variability between and within departments with respect to exposures to toxins in the environment. Complicating this picture even more is the variability in how departments implement and enforce use and maintenance of personal protective equipment (PPE). In addition, the changes in building materials over the last several decades has led to the development of products of combustion that have yet to be accurately characterized and measured. Thus, in the case of cancer incidence and/or mortality, it is difficult to truly capture "exposure" to carcinogens other than through very crude categorizations, such as job classification, years in the fire service, and departmental call volumes.

The "Healthy Worker Effect" refers to one of the most common types of selection bias in occupational epidemiology and refers to the fact that on average, people who work have relatively lower rates of disease and death than those who do not. This is likely even more so the case for firefighters who typically have to undergo extensive medical clearance evaluations and have to complete an occupational readiness test (e.g., the Candidate Physical Abilities Test) in order to qualify to even join the fire service. Once in the fire service, it is not uncommon for firefighters to undergo mandatory annual or semi-annual medical evaluations and some departments require ongoing fitness and occupational readiness testing, further creating a workforce that, on average, is likely to be healthier than the general public (whose rates often are used to make comparisons for the computation of SMRs and SIRs). Thus, disease and mortality risks may be underestimated when comparing firefighters to the general public.

Finally, some diseases that may be of concern for firefighters are relatively rare and may have extended latency periods. Given these issues, it is likely that only large studies with extensive follow-up will be able to adequately estimate risks associated with these conditions and that smaller studies, like those common in the early literature on firefighter cancer risks, with modest sample sizes and follow-up periods, would likely end up underestimating risks for rarer forms of cancer and mortality.

Statistical Issues to Consider

The following are issues to consider when evaluating the scientific literature on firefighting and cancer.

Sample versus Population – A population is the entire set of entities under study. For instance, one population could be all male firefighters who have worked urban departments in the United States for at least one year and who are exposed to a specific set of compounds for a defined amount of time. Because of practical barriers and expense, studies rarely include an entire population. A sample is a subset of a population. Sometimes samples are drawn at random or carefully selected groups from the population are studied. For instance, firefighters who regularly fight fires and are exposed to diesel exhaust from Philadelphia, PA and Oakland, CA may form the sample in a study. This sample will be used to draw inferences about the population, all urban firefighters in the US who are exposed to similar risks.

Study Characteristics – To determine if a study is relevant to a case, it is important to determine if the study sample is characterized by similar people, settings, exposures, and outcomes[6]. For example, say that the sample used in a study were male firefighters from urban areas in the United States who were exposed to fire, smoke, and diesel exhaust and whose health status was assessed using accepted, standardized methods. That study would provide high *external validity* if the case under question is a firefighter who works in an urban location in the US and who experienced similar exposures and was diagnosed using similar, standardized methods.

Precision of an Estimator – Statistical precision of an estimator is defined as "the closeness with which it can be expected to approximate the relevant population value. It is necessarily an estimated value in practice, since the population value is generally unknown."[7]. Precision can be measured using the standard error (SE), which provides a numerical estimate of how much fluctuation from the population parameter that we can expect in a sample estimate. For example, the standard error (SE) of an average or mean is:

$$SE_{Mean} = \sqrt{\frac{S^2}{n}}$$

Where S^2 is the variance (a measure of dispersion which equates to the average of the squared differences from the mean) of the values used to compute the mean and n is the number of values. As you can see, both the variability of the values used to compute the SE_{Mean} and the sample size determine its magnitude of the SE. Small sample sizes and larger variances lead to larger SEs and thus less precision. Relevant to this case, the SE for the standardized mortality ratio is:

$$SE_{SMR} = \frac{\sqrt{O}}{E}$$

Where O is the observed number of deaths in a specific group and E is the expected number of deaths from the comparison group.

Hypothesis Testing and Error – When conducting a statistical test there are typically two hypotheses under consideration. The null hypothesis typically (although not necessarily) states that there is no relationship between two factors or no association among groups. For instance, in a study of firefighters the null hypothesis may be that there is no relationship between firefighting and the incidence of cancer. The alternative hypothesis is contrary to the null hypothesis and states that the findings are the result of a real effect. An alternative hypothesis could be that the mortality from cancer is greater in firefighters than in the general public.

Tests are based on probabilities and thus no conclusion is 100% certain. When you are testing a hypotheses two types of errors are possible. A **Type 1 error** occurs when you reject a null hypothesis that is true. The probability of a Type 1 error is α (alpha), which is the level of significance you set for your test. For instance, an α of 0.05 indicates that you are willing to accept a 5% chance that you are wrong when you reject the null hypothesis. A **Type 2 error** will occur when you fail to reject a false null hypothesis. The probability of a Type 2 error is denoted by β, which depends on the power of a test. Factors impacting the power of a test include the size of the effect under study (e.g., the greater the association of cancer and firefighting), the

variability of the measure used (i.e., S^2), and the sample size used. The probability of correctly rejecting a false null hypothesis is 1-β.

Statistical Significance – A p-value is the probability of attaining a result at least as large as the one found given that the null hypothesis is true. In other words, a p-value of 0.05 would tell us that if the null hypothesis were true that only 5% of values would be this large or larger. In statistical hypothesis testing, statistical significance is attained when the p-value of the test is less than the significance level set for the study. The significance level is the level of α we select based on our tolerance to making a Type 1 error. If our attained p-value is lower than the α level we set for the test we say that the test is statistically significant.

Confidence Interval – One frequent goal of statistics is to estimate a population parameter, such as the risk of cancer in firefighters, with a sample. Because we don't know the actual value of the population parameter (otherwise we wouldn't need to estimate it with a sample), we can construct a confidence interval which reflects the degree of uncertainty associated with a sample estimator. A confidence interval means that if we construct a confidence interval repeatedly over a large number of samples, we expect (100-α)% of them to include the true population parameter and α% of them not to. For a 95% confidence interval, if we collected a large number of samples and constructed confidence intervals, we would expect 95 of those intervals to include the true population parameter. Statisticians prefer to use confidence intervals over just point estimates because they indicate (1) the precision of the estimate and (2) the uncertainty of the estimate. The most practical way to reduce the width of a confidence interval is to increase the size of the sample in your study.

Why Sample Size is Important – From the discussion above, it follows that, all other things being equal, larger samples are more informative than smaller samples. Larger samples lead to greater precision of estimates, makes it more likely that you will detect small but meaningful effects, and are more representative of the full diversity of the population. Compared to smaller samples, larger samples have more information and thus reduce the uncertainty of our conclusions.

References

1. Checkoway H, Pearce N, Kriebel D. *Research Methods in Occupational Epidemiology*. Oxford University Press; 2004.

2. Lucas RM, McMichael AJ. Association or causation: evaluating links between “environment and disease”. *Bull World Health Organ*. 2005;83(10):792-795.

3. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerg Themes Epidemiol*. 2015;12:14. doi:10.1186/s12982-015-0037-4

4. Guidotti TL. Evaluating causality for occupational cancers: the example of firefighters. *Occup Med Oxf Engl*. 2007;57(7):466-471. doi:10.1093/occmed/kqm031

5. Guidotti TL. Evaluating Causation for Occupational Cancer Among Firefighters: Report to the Worker’s Compensation Board of Manitoba. Published online 2005. Accessed November 18, 2016. http://webcache.googleusercontent.com/search?q=cache:AMBXfqyphGQJ:www.aph.gov.au/DocumentStore.ashx%3Fid%3D05ea1e2c-543c-45f1-b110-211ea30d4970+&cd=1&hl=en&ct=clnk&gl=us

6. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. 2nd ed. Houghton Mifflin Company; 2001.

7. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2 edition. Routledge; 1988.

DetecTogether