Chapter 7

DISCUSSION SUMMARY: DEFINING HEALTH OUTCOMES IN EPIDEMIOLOGICAL INVESTIGATIONS OF POPULATIONS DEPLOYED IN SUPPORT OF OPERATION IRAQI FREEDOM AND OPERATION ENDURING FREEDOM

JOSEPH H. ABRAHAM, SCD*

*Epidemiologist, Environmental Medicine Program, Occupational and Environmental Medicine Portfolio, Army Institute of Public Health, US Army Public Health Command, 5158 Blackhawk Road, Aberdeen Proving Ground, Maryland 21010-5403

INTRODUCTION

Identifying potential health implications associated with environmental conditions experienced during military deployment in support of Operation Enduring Freedom, Operation Iraqi Freedom, and Operation New Dawn is an ongoing effort of the US Department of Defense (DoD) and the US Department of Veterans Affairs (VA) researchers. Epidemiology is a primary tool used in this endeavor.[1] Epidemiology is "the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control health problems."[2] Epidemiological research often involves assessing the relationship between an event or trait (an exposure) and another event or trait (an outcome). Defining and ascertaining information on an outcome are primary tasks in the design and conduct of epidemiological studies. The outcome of an epidemiological study is a broad term representing a defined disease, state of health, or health-related event. Measurement and classification of health outcomes in epidemiological studies are an exercise in balance: minimizing errors and maximizing efficiency. Evaluating outcomes used in epidemiological assessments of airborne hazards was the primary focus of a work group at the VA/DoD Airborne Hazards Symposium held in August 2012. The work group identified limitations in current data systems (Exhibit 7-1), then focused on the question of whether the DoD and VA could, and should, standardize outcome definitions.

This chapter expands on the workshop discussion by reviewing concepts and consequences of outcome definitions used in epidemiological investigations of the health effects of airborne hazards among deployed and formerly deployed military personnel and veterans. The first section introduces concepts that could be helpful in formulating outcome definitions for use in epidemiological studies, as well as in evaluating a study's quality vis-à-vis outcome assessment. Consequences of operational outcome definitions on both individual study results and on inferences to be drawn from the body of epidemiological evidence regarding airborne hazards health effects are also discussed. The second section reviews outcome definitions used in selected published research studies relevant to the epidemiology of health effects potentially associated with airborne hazards in the deployment environment. Improving the state of science with respect to outcome definitions, including a discussion of the pros and cons of standardizing outcome definitions across studies being conducted by VA and DoD researchers, is the topic of the third section.

EXHIBIT 7-1

LIMITATIONS IN CURRENT DATA COLLECTION SYSTEMS IDENTIFIED DURING THE WORKSHOP

- Lack of integration between medical record systems (eg, AHLTA,* JMeWs†) and administrative medical encounter databases (eg, DMSS‡)
- Misclassification of health conditions and missing data
- Inaccuracy of health outcome data
- Subjective nature of health outcome assignment
- Imperfect or incomplete models of disease etiology
- Lack of clinical corroboration of symptom outcomes

*The Armed Forces Health Longitudinal Technology Application (AHLTA) is the electronic medical record system used by US Department of Defense medical providers.
†The Joint Medical Workstation (JMeWS) is the theater medical surveillance system that integrates medical information from separate health data collection systems for the US Army, Navy, Air Force, and Marine Corps.
‡The Defense Medical Surveillance System (DMSS) is the central repository of medical surveillance data for the US Armed Forces.

CONCEPTS OF HEALTH OUTCOME DEFINITION FOR EPIDEMIOLOGICAL STUDIES

Epidemiological study outcomes must be defined in advance of the conduct of the study, and they should be clear, specific, and measurable. These outcomes are often based on a combination of signs and symptoms, physical examinations, pathology, and diagnostic test results. Typically, outcome data consist of physical measurements, laboratory results, responses to self-administered questionnaires or interview questions, information garnered from

medical record reviews, or diagnostic codes abstracted from administrative databases.

In the textbook *Essentials of Epidemiology in Public Health* by Aschengrau and Seage,[3] they suggest that "it is best to use all available evidence to define with as much accuracy as possible the true cases of disease." An accurate measurement is one that is close to the true value. From a theoretical viewpoint, maximizing the accuracy of an outcome definition should be a primary objective because inaccurate outcome assessment can induce bias in both estimates of measures of association (eg, relative risks) and estimates of the precision of those measures (eg, confidence intervals). Practically speaking, however, the salient part of Aschengrau and Seage's phrase *with as much accuracy as possible* is *as possible*. Conducting gold standard tests in epidemiological studies, if such tests exist, may not be feasible based on logistical, technical, and ethical reasons.

For practical purposes, a number of elements (Exhibit 7-2) must be *jointly* considered in defining a health outcome for an epidemiological study. The researcher and those wishing to evaluate an epidemiological study must consider what is required of the measurement in judging its appropriateness. True health outcome status is often ambiguously defined and poorly measured in the context of an epidemiological study.

EXHIBIT 7-2

FEATURES TO CONSIDER IN DEFINING A HEALTH OUTCOME FOR USE IN AN EPIDEMIOLOGICAL STUDY

- Study hypothesis
- Objectives of the study
- Conceptually relevant health condition given a mechanistic, biological, or social model
- Accuracy of the outcome measurement:
  - Probability and magnitude of misclassification or measurement error
  - Consequences of misclassification or measurement error on estimated measures of association
  - Previous validation of the measurement
- Feasibility:
  - Logistical constraints
  - Available resources
  - Ethical considerations
  - Time constraints
- Magnitude of the hypothesized association
- Incidence versus prevalence as the appropriate metric of outcome frequency

Validity

Like all epidemiological investigations, epidemiological assessments of deployed and formerly deployed military personnel are susceptible to bias. In epidemiology, bias is defined as the difference between an expected estimate (eg, the average value of association estimates over many hypothetical study repetitions) and the true value, or the processes leading to such deviation.[4] Bias arises from systematic errors in the selection of study participants, confounding of exposure–outcome relationships, and errors in the ascertainment (measurement) of exposures and outcomes. Both measures of association and measures of variability can be biased. In the absence of bias, an epidemiological study is valid. On average, valid studies will produce an estimate of the true underlying association being assessed.

Sensitivity, Specificity, and Misclassification

In epidemiology, outcomes are often classified dichotomously (ie, individuals either have the outcome or they do not have the outcome). Such binary outcome measures are susceptible to two types of systematic error: (1) outcome-free individuals incorrectly classified as having the outcome (false positives); and (2) individuals having the outcome, but incorrectly classified as being free of it (false negatives). Two statistical measures of a binary classification function—sensitivity and specificity—correspond to these two types of errors in assessments of an outcome in epidemiological studies.[5] For example, consider a health condition of interest (D), for which individuals can have a disease (D+) or not (D–). Consider an epidemiological classification function of the disease status (O) that divides individuals into two groups: (1) positive for the outcome (O+) or (2) negative for the outcome (O–). Consider a study assessing the risk of developing asthma (D) and defined as self-report of a physician's diagnosis of asthma (O).

Sensitivity refers to the proportion of individuals with the disease (D+) who are correctly identified as having the outcome (O+); it is the probability of individuals being classified as having the outcome among those who truly have it (Sensitivity = Prob[O+|D+]). An outcome metric with high sensitivity has minimal false negatives. In the asthma example, it is the percentage of individuals with asthma who are correctly identified as having asthma and are considered as having asthma for the purpose of the study. Sensitivity less than 100% implies that a proportion of individuals with asthma (1 – Sensitivity) will be incorrectly classified as not having asthma in the study (false-negative proportion).

Specificity refers to the proportion of those who are truly free of the disease (D–) and are correctly identified as negative for the outcome (O–). It is the probability of individuals being classified as not having the outcome among those who

truly do not have it (Specificity = Prob[O–|D–]). An outcome metric with high specificity has minimal false positives. In the asthma example, it is the percentage of individuals who do not have asthma who are correctly identified as nonasthmatic. Specificity less than 100% implies that a proportion of individuals who are free of asthma (1 – Specificity) will be incorrectly classified as having asthma in the study (false-positive proportion).

For any given classification function of disease status, there is usually a trade-off between sensitivity and specificity. For example, increasing the proportion of asthmatic individuals correctly classified as asthmatic results in a decrease in the proportion of nonasthmatic individuals correctly classified as nonasthmatic.

Error in disease classification of a study outcome is referred to as *outcome misclassification*. The consequences of outcome misclassification on study results are two-fold: (1) bias in the observed measures of association and (2) bias in the estimates of the variance of the observed measures of association (eg, confidence intervals). The magnitude of bias from outcome misclassification depends on the sensitivity, specificity, outcome prevalence, and the magnitude of the true association. Some epidemiologists[6,7] have argued that specificity is more important in evaluating bias in relative risks resulting from misclassified health outcome data. However, sensitivity is also important, and bias in relative risk is actually dependent on a third property of a classification function of disease status: the positive predictive value (or PPV). The PPV = Prob[D+|O+], which, in contrast to sensitivity and specificity, is dependent on the prevalence of the disease in the population being assessed.[8]

Note that estimating sensitivity, specificity, and the PPV all require knowledge of true disease status (although for the latter, true disease status must be known only among the subgroup of participants who are deemed positive by the classification function O+). Although it is preferable, in many cases it is not possible to obtain estimates of sensitivity, specificity, and PPV in a population of interest, primarily because information on true outcome status is unobtainable. Thus, critical readers of epidemiological studies are most often not privy to this validation data. Nonetheless, one can use the concepts of sensitivity, specificity, and PPV as heuristics to better understand the potential strengths and weaknesses of an outcome metric with respect to validity in the context of a given epidemiological study.

If the probability of misclassifying a study subject's outcome is independent of exposure status, the classification error is referred to as *nondifferential* or *random* outcome misclassification. In contrast, differential (nonrandom) outcome misclassification occurs when the probability of incorrectly assigning outcome status is *not* independent of exposure status. A paradigmatic example of differential misclassification occurs in case-control and retrospective studies when there is a greater level of accuracy in reporting outcome status among exposed study participants relative to unexposed subjects. A recall bias may occur because individuals who have the health condition under study may have more thoroughly considered their exposure history in an attempt to answer the question, "Why me?" The bias of observed estimates of association from differential misclassification can generate either overestimates or underestimates of the true association. The direction of bias of observed estimates of association from nondifferential misclassification is discussed herein. As a general rule, outcome assessment should be conducted in such a manner that is comparable with all exposure levels (ie, independent of exposure). Often, this maxim is implemented by "blinding" those who are responsible for outcome assignment to the exposure status of the study participants, especially when the outcome measure is subjective in nature.

Under certain circumstances, the bias from outcome misclassification is theoretically predictable.[9,10] If the misclassification of a dichotomous outcome is exactly nondifferential, and, additionally

- the misclassification error is independent of errors in other variables in the analysis,
- there is no interaction between the misclassification and other sources of bias (eg, selection bias, confounding), and
- there is no random sampling variability,

then nondifferential outcome misclassification will result in a bias of measures of association toward the null value for the association (eg, toward 1 for the relative risk). Technically, the direction of the bias from nondifferential misclassification refers to the average error across hypothetical study repetitions, (ie, ignoring random variability in the misclassification). Because these conditions are rarely met in practice, nondifferential outcome misclassification does not always lead to an underestimation of measures of association. In practice, an observed measure of association estimated in the presence of nondifferential misclassification is a joint function of the misclassification error, the true association, other sources of bias, and random sampling variability.

Bias in estimated measures of association from outcome misclassification is arguably less of an issue in assessing the strong effects of environmental exposures. However, as epidemiologists attempt to evaluate associations of increasingly small magnitude, the importance of preventing or decreasing classification errors increases.[11]

Repeatability

Repeatability is important in the assessments of outcomes that are not expected to change over time (eg, outcomes for chronic diseases). Assessing repeatability may be useful in characterizing an outcome metric in the absence of a gold

standard (diagnostic) test wherein one can compare an epidemiological outcome indicator. Therefore, an outcome that is repeatable is not necessarily valid. The repeatability of an outcome assessment, also known as test-retest reliability or stability, is a joint function of intrasubject variability, interobserver variability, and intraobserver variability. An outcome measure with no variability arising from these sources is repeatable. If, however, an outcome measure is random in nature (ie, independent of exposure status), its departure from perfect repeatability will result in an attenuation of relative measures of association.

Incidence, Prevalence, and Mortality

In planning a study, epidemiologists are often faced with a choice regarding whether to assess incidence or prevalence of the outcome, and, at times, they additionally consider whether or not to assess mortality as the outcome of interest. Generally speaking, incidence measures of outcome are more useful in investigating outcome etiology, as both prevalence and mortality can be influenced by determinants of outcome duration and survival in addition to causes of the outcome.[12] This limitation must be considered against the logistical efficiency gained by assessing prevalence or mortality as outcomes and the fact that—depending on the circumstance—prevalence and mortality may be able to be measured with less error, relative to the corresponding measure of incidence.

Objective Versus Subjective Outcome Measurements

In evaluating outcome assessment in epidemiological studies, epidemiologists often distinguish between assessments that are objective in nature and those that are subjective in nature. Objective assessments are often preferred for the following reasons, both of which are plausible, but neither of which is necessarily true: (1) objective measures are less susceptible to misclassification; and (2) misclassification of objective measures, compared with that of subjective measures, is less likely to be influenced by the exposure status of the study subjects. For example, survey respondents may self-report their health differently, depending on their understanding of the survey question, other questions asked in the survey, their health-related aspirations and expectations, their access to and use of health care, as well as their social and socioeconomic context. All of these may be associated with the exposure under study. In addition, respondents' answers may be directly influenced by their exposure status.

In many instances, however, subjective outcome assessment measures are preferred, often for their logistical efficiency. In many contexts, subjective measures have been validated to be strong predictors of logistically more complex objective outcome measures. Moreover, it is often *perceived* health that drives the use of healthcare resources. Finally, it is worth reiterating that objective assessments are not immune to measurement error, both differential and nondifferential.

Balancing Advantages With Costs

As alluded to earlier, it is often not feasible to implement an ideal outcome assessment (eg, a diagnostic evaluation with perfect sensitivity, specificity, and reliability) in the context of a large-scale epidemiological study. A more practical strategy to outcome assessment may include the use of gold standard outcome assessments in a subset of the larger study population. This strategy has the dual advantages of providing evidence of the degree to which the less resource-intensive outcome assessment methodology misclassifies true outcome status and providing information to facilitate analytic techniques to correct for the bias induced by the misclassification. Another strategy for balancing advantages with costs is to use a combination of outcome indicators (eg, signs and symptoms), examination results, medication use, and test results to increase the accuracy of outcome measurement.[3]

Evaluating Health Outcome Measurement

Consider the following three questions when critiquing a study's outcome assessment:

1. What is the conceptual outcome (ie, the outcome of true interest to the researchers)?
2. What is the operational outcome definition (ie, how the researchers define and ascertain information on the outcome)?
3. What impact does any discrepancy between the conceptual outcome and the operational outcome definition have on the estimates of association and the inference that can be drawn from the study?

The first question relates to the conceptually relevant outcome. The second question relates to the operational definition of the outcome as it is implemented in the study. The third question invites the individual evaluating the study to consider the impact of the outcome definition on internal validity (from bias) and external validity (limitations regarding generalizability).

Consider the following research question: Does exposure to particulate matter (PM) air pollution increase the risk of asthma (outcome) among deployed military personnel? The outcome (asthma) is readily understood by the reader. Asthma is a chronic inflammatory disorder of

EXHIBIT 7-3

EXAMPLES OF ASTHMA* DEFINITIONS USED IN EPIDEMIOLOGICAL STUDIES

- A postbronchodilator response of 12% improvement in either FEV_1 or FVC relative to baseline spirometry
- Appearance of ICD-9 diagnosis code 493 in a medical record
- Reported presence of persistent respiratory symptoms (cough, wheezing)
- Spirometric evidence of airway obstruction
- A positive answer to the question, "Do you have asthma?"
- A positive answer to the question, "Has a doctor ever told you that you have asthma?"
- Self-report of the presence of asthma symptoms (eg, recurrent cough and wheeze, apart from upper respiratory illness)
- Identified use of asthma-control medications
- Reported improvement of asthma symptoms with use of asthma-control medications
- Combinations of the above definitions

*According to the National Heart, Lung, and Blood Institute and the National Asthma Education and Prevention Program, asthma is defined as a chronic inflammatory disorder of the airways with generally reversible airflow obstruction and airway hyperresponsiveness causing episodic respiratory symptoms.
FEV_1: forced expiratory volume in 1 second; FVC: forced vital capacity; ICD-9: *International Classification of Diseases*, Ninth Revision
Data source: National Asthma Education and Prevention Program, NHLBI, National Institutes of Health. *Asthma Expert Panel Report 3. Guidelines for the Diagnosis and Management of Asthma.* Bethesda, MD: National Institutes of Health; 2007. NIH Publication 08-5846.

the airways with generally reversible airflow obstruction and airway hyperresponsiveness causing episodic respiratory symptoms.[13] However, asthma is variously defined for use in epidemiological studies. Although partially a reflection of the fact that asthma is a heterogeneous outcome, the number of definitions is also an indication of the trade-offs inherent in using different approaches to identify asthma as a health outcome (Exhibit 7-3).[14] A Danish study comparing three of these operational definitions (participant report of doctor diagnoses of asthma, hospitalization diagnosis codes for asthma, and use of antiasthmatic medication) observed kappa statistics in the range of 0.21–0.38, indicating only fair agreement.[15,16] One can think of these, and other, operational definitions as outcome metrics. They are *indicators* of true, underlying asthma. There is no one correct outcome metric, because each has its advantages and disadvantages in the setting of a given epidemiological study. Some may be more accurate than others, some more specific, some more sensitive, and some more efficient to implement.

HEALTH OUTCOMES USED IN DEPLOYMENT EPIDEMIOLOGY

This section provides a summary and discussion of the health outcomes used in a select group of six recently published articles on the topic of epidemiological assessments of health effects of airborne hazards in the deployment environment. The appropriateness of each study's health outcome was evaluated by considering the following:

- the conceptually relevant health condition;
- logistical constraints;
- previous validation of the measurement;
- probability and magnitude of misclassification;
- consequences of misclassification on estimated measures of association; and
- scientific context of the study, including standardization with outcome definitions and methods used in previous work.

Selected Studies

This section reviews the health outcomes of the six major studies.

Study 1

Smith B, Wong CA, Smith TC, et al. Newly reported respiratory symptoms and conditions among military personnel deployed to Iraq and Afghanistan: a prospective population-based study. Am J Epidemiol. *2009;170:1433–1442.*

The Millennium Cohort Study (MCS), the largest DoD prospective cohort study ever conducted, has enrolled more than 150,000 military personnel.[17] Launched in 2001, prior to the September 11 terrorist attacks, the MCS was designed to investigate the effects of military service, specifically

deployment-related exposures, on long-term health.[18] The study relies on mailed and web-based questionnaires to ascertain self-reported medical conditions.

In 2009, Smith et al published an assessment of respiratory symptoms and conditions among MCS participants who deployed to Iraq and Afghanistan.[19] The objective of the study was to assess whether respiratory symptoms and conditions were associated with deployment among MCS participants. The conceptually relevant health conditions were incident respiratory symptoms, incident asthma, and incident chronic bronchitis or emphysema (ie, chronic obstructive pulmonary disease or COPD). To ascertain disease status at baseline, participants were asked, "Has your doctor or other health professional ever told you that you have any of the following conditions?" Asthma, chronic bronchitis, and emphysema were listed as possible responses, and the question was repeated on annual follow-up assessments. To identify respiratory symptoms at baseline, participants were asked, "During the last 12 months, have you had persistent or recurring problems with any of the following?" Cough and shortness of breath were offered as possible responses. On follow-up assessments, participants were asked the same questions, but with the timeframe extended from "during the last 12 months" to "in the last 3 years." The study outcomes were defined as a positive endorsement regarding the corresponding symptom or condition on a follow-up assessment without such indication on the baseline questionnaire.[19]

As with many survey-based research initiatives, this study relies heavily on participants' ability to correctly recall their medical history. In this study, it is possible that respondents' errors in recall of their health conditions are likely, but independent of their deployment status (ie, nondifferential misclassification). If, however, previously deployed MCS participants more accurately recall incidence of respiratory symptoms relative to their nondeployed peers, the resulting differential misclassification would result in an upward bias of the association between deployment and respiratory symptoms.

In 2008, MCS researchers published an article identifying some of the challenges of using self-reported medical conditions in the conduct of epidemiological studies.[20] After considering using the kappa statistic and measurements of sensitivity and specificity to evaluate agreement between self-reported and diagnostic codes in the medical record, the authors decided to use measures of positive and negative agreement.[21] In comparisons of self-report of medical conditions to electronic medical record data, they observed "near-perfect negative agreement and moderate positive agreement" for 38 diagnoses of "less-prevalent conditions."[20,21] Positive agreement (percentage) among self-reported diagnoses for asthma, emphysema, and chronic bronchitis was 42.0%, 2.7%, and 12.9%, respectively; negative agreement among self-reported diagnoses for asthma, emphysema, and chronic bronchitis was 97.1%, 99.6%, and 96.7%, respectively. The high negative agreement percentages indicate that self-report of health conditions may be an effective means of ruling out a history of a particular condition—a finding similar to a classification scheme with high specificity and low false-positive proportion. However, the low positive-agreement percentages suggest that the self-reporting had low sensitivity and a high proportion of false negatives. In their discussion, the authors conclude that their results speak to the importance of using multiple data sources, when possible, to assess health outcomes.

Study 2

Szema AM, Peters MC, Weissinger KM, et al. New-onset asthma among soldiers serving in Iraq and Afghanistan. Allergy Asthma Proc. *2010;31:67–71.*

In 2010, researchers at the Northport Veterans Affairs Medical Center (VAMC) in Long Island, New York, published a study using medical record data.[22] The objective of the study was to assess whether former US military personnel with a history of deployment to Iraq or Afghanistan have higher risk of asthma compared with veterans who did not deploy to these regions. The conceptually relevant health condition was incidence of asthma among US military veterans. This contrasts with the outcome metric used to evaluate the hypothesis: the appearance of an *International Classification of Diseases*, Ninth Revision (ICD-9), diagnostic code for asthma (ICD-9 code 493) in a veteran's VAMC medical record. The authors enumerated the three clinical guidelines used by the VAMC clinic to diagnose asthma: (1) evidence of recurrent episodes of respiratory symptoms (cough, wheezing, dyspnea, and exercise-induced shortness of breath), (2) spirometric evidence of airway obstruction, and (3) improvement of symptoms after administration of a bronchodilator. These guidelines differ from those of the National Institutes of Health and the National Heart, Lung, and Blood Institute (NHLBI), which recommend that an asthma diagnosis be based on a combination of clinical symptoms consistent with airway obstruction, spirometric evidence of reversible airway obstruction (rather than control of symptoms), and exclusion of alternative diagnoses.[23] Consistent with the NHLBI criteria for diagnosis of asthma, the authors presented spirometric results, including data on reversibility of airway obstruction (and not improvement of symptoms) after bronchodilator administration.

The authors identified 290 veterans with asthma for inclusion in the study. Although the VAMC clinic guidelines for asthma diagnosis include spirometry, lung function data were presented for only 45 patients (approximately 16% of the veterans identified as having asthma based on the presence of an asthma diagnosis code in their medical record). The authors did not explain why lung function data were only available for such a small subset of subjects in the

study, nor did they deliberate on the representativeness of the spirometry results they did obtain. They did note that spirometric measurements were obtained while patients were taking medication to control their asthma.

In the discussion, the authors identify reliance on ICD-9 diagnosis codes as a limitation of the study, although they do not discuss the potential impact of the use of ICD-9 codes on the study results. They do suggest that more sensitive outcome measures—including methacholine challenge, cardiopulmonary exercise testing, impulse oscillometry, exhaled breath condensate nitric oxide levels, and skin prick testing for aeroallergens—may be helpful, but did not elaborate on the point. These outcomes may have better sensitivity in identification of pulmonary (and other) impairments, relative to outcomes defined on the presence of ICD-9 diagnostic codes in medical records.

The authors identify a lack of baseline (predeployment) spirometry measurements as a limitation of the study. The utility of longitudinal assessments of lung function among military personnel is currently a matter of debate,[24] and studies of longitudinal changes may inform this discussion. Longitudinal changes in lung function parameters that incorporate predeployment assessments may serve as potentially sensitive markers of exposure to deployment-associated airborne hazards. No longitudinal assessments of lung function that incorporate baseline testing among military personnel who deploy to southwest Asia have been published. Without baseline spirometry, evaluations of lung function often rely on comparisons of lung function among those with a history of deployment to a nondeployed reference population that may not be comparable with the exposed population. As performed in this study, such comparisons often use general population average lung function values. Similarly, lung function parameters can be expressed as a percentage of gender-, age-, and race-specific predicted values[25] and then used in conjunction with a cut-point (eg, 80% of predicted values) to indicate impaired lung function. The sensitivity and specificity of these methods depend on the definition of the cut-point. They are generally less sensitive, relative to measures of longitudinal change in pulmonary function.

The ICD-9 code-based outcome assessed in the study is subject to several sources of misclassification. Misdiagnosis of asthma because of reliance on evidence of symptom control after bronchodilator use as specified by the diagnostic criteria used by the VAMC clinic—rather than spirometric evidence of reversibility of airway obstruction—is possible (although likely not substantial), especially considering the authors reported evidence of reversibility (not control of symptoms) in their results. The discrepancy between the number of asthma outcomes identified and the number of patients with lung function data are potential evidence of a larger problem: the asthma diagnoses, as evidenced by the appearance of asthma ICD-9 codes in the medical records, may not have been based on functional parameters (ie, indicators of airway obstruction) assessed by spirometry, despite the diagnosis guidelines used by Northport VAMC.

The probability of outcome misclassification may not be independent of exposure if deployment itself was an indication for referral in the mind of the diagnosing physician. This differential misclassification would result in an upward bias of estimated odds ratios.

Study 3

Szema AM, Salihi W, Savary K, Chen JJ. Respiratory symptoms necessitating spirometry among soldiers with Iraq/Afghanistan war lung injury. J Occup Environ Med. *2011;53:961–965.*

The same researchers at the Northport VAMC conducted a follow-on study to evaluate the hypothesis that US military veterans with a history of deployment to Iraq and Afghanistan have higher rates of respiratory symptoms compared with veterans who had not deployed to those regions.[26] The conceptually relevant health condition is the incidence of persistent respiratory symptoms that indicate possible lung injury or pulmonary disease. To operationalize the conceptual outcome, the authors reviewed the Northport VAMC patient records to identify individuals referred for spirometric evaluation. The authors also presented lung function parameters assessed by spirometry, although these later data were not used to classify study subjects with respect to health outcome.

The authors focused on asthma as a condition potentially underlying respiratory symptoms. In this study, the indication for referral for spirometry was the presence of clinical symptoms consistent with a diagnosis of asthma. The set of qualifying symptoms was not specified.

In their discussion, the authors conflate evidence of clinical symptoms indicating referral for spirometry with the presence of pulmonary disease or injury; this is an unsupported inference rather than an actual outcome misclassification. However, if the conceptual outcome of interest was lung injury or disease, rather than respiratory symptoms, the outcome misclassification error is greatly exacerbated, primarily from false positives. Ambiguity in the relevant conceptual outcome definition may be fostered by the following three factors:

1. The VAMC physicians referring patients for spirometry must *specify the relevant symptoms* prior to the pulmonary function testing.
2. The VAMC physicians referring patients for spirometry must *specify a diagnosis* prior to the pulmonary function testing.
3. The clinical guidelines used by VAMC clinicians to diagnose asthma rely on clinical symptoms, rather than functional (lung function) parameters, to assess reversibility of airway obstruction.

Although they discuss the clinical guidelines for a diagnosis of asthma, the authors do not elucidate the set of diagnoses given to patients by the referring physician, nor do they provide the reader with an analysis of the diagnoses made following the spirometric assessments.

The study presents an interesting opportunity to contrast the findings of analyses drawn on medical records data with the assessments of functional parameters. The conclusions of the study, based on a higher proportion of previously deployed veterans presenting with respiratory symptoms and being referred for spirometric evaluations relative to veterans without a history of deployment, contrasts with the spirometric findings presented. The average forced expiratory volume in 1 second and the forced vital capacity parameters were higher among veterans with a history of deployment relative to veterans who had not deployed to Iraq or Afghanistan. The mean forced expiratory volume in 1 second and the forced vital capacity ratios between the two groups were not statistically different.

The authors did not report any validation of the diagnostic code-based outcome, either by evaluating faithfulness to the clinics' diagnostic guidelines or other guidelines that combine clinical assessment of symptoms with measurements of lung function and exclusion of other diagnoses.

Findings of the study are likely biased because the outcome classification (referral for spirometry) was not independent of the exposure of interest (deployment to Iraq or Afghanistan). Veterans with a history of deployment were sourced from VA clinics catering specifically to veterans of Operation Iraqi Freedom and Operation Enduring Freedom, and the referring physicians were (appropriately, from a clinical standpoint) aware of the exposure status of their patients. The authors' coining of the term "Iraq/Afghanistan War Lung Injury" to refer to the pulmonary complaints evaluated in the study may be an inadvertent allusion to this source of bias; the exposure is implicit in the definition of the outcome. The authors did not report their having evaluated the degree to which the history of deployment was an indication for referral for lung function testing, nor did they discuss the impact of this potential source of bias.

The outcome for the study was defined using an easily accessible, preexisting database. Correspondingly, the primary advantage of the outcome metric is logistical efficiency. Additional strengths include the fact that the administrative diagnostic codes culled from the patient record database correspond to both prespecified diagnostic criteria and clinically assessed functional parameters (ie, lung function); however, these strengths are more theoretical in nature than realized.

In 2012, researchers at the US Army Public Health Command (including the author) published three original research papers in the *Journal of Occupational and Environmental Medicine*.[27–29] All three evaluations leveraged diagnostic code data obtained from military medical records systems.

Study 4

Abraham JH, Baird CP. A case-crossover study of ambient particulate matter and cardiovascular and respiratory medical encounters among US military personnel deployed to southwest Asia. J Occup Environ Med. *2012;54:733–739.*

The aim of this study[27] was to evaluate the hypothesis that acute PM exposure precipitates the incidence of acute cardiovascular or respiratory events such as myocardial infarctions or severe asthma attacks. Both the occurrence and timing of the health event with respect to exposure were of interest in evaluating the hypothesis. The conceptually relevant health condition is the incidence of a serious cardiovascular or respiratory health event. This contrasts with the outcome metric used: the appearance of any one of a set of ICD-9 diagnostic codes in either of two medical record databases (the Joint Medical Workstation, and the Transportation Command Regulating and Command & Control Evacuation System). Case status was defined as having any one of the qualifying cardiovascular (ICD-9 diagnosis codes 390–459, Diseases of the Circulatory System) or respiratory (ICD-9 diagnosis codes 460–519, Diseases of the Respiratory System) outcomes. The date of the medical encounter was defined as the incidence of the health event. The case definition was not validated as part of this study, nor has it been validated in other studies of military personnel.

The outcome assessed in this study is subject to several sources of misclassification. Health events that otherwise may have qualified for inclusion in the study may not have appeared in the medical records obtained, either because personnel did not seek medical attention or because medical encounters were not entered into the electronic medical record. These situations result in false negatives. Case status was defined to *cast a wide net* to increase case ascertainment (ie, to increase sensitivity and decrease the number of false negatives). However, some of the diagnostic codes may indicate encounters for conditions that are not biologically relevant; any cases with these codes could be considered false positives. Misclassification of the health event's *timing* is also likely, because the date of the medical encounter may not have been the date of incidence of the conceptually relevant health condition.

The probability of outcome misclassification resulting from both design choices and imperfect capture of otherwise qualifying medical events was likely independent of exposure. If an association between acute PM exposure and acute cardiovascular events truly exists, bias from this nondifferential outcome misclassification would have attenuated odds ratios in the direction of no association.[6,30] This study was unable to reject the null hypothesis of no association between ambient PM levels and cardiorespiratory medical encounters. Nondifferential outcome misclassification in this study is thus a potential noncausal explanation for the null findings.

Study 5

Abraham JH, DeBakey SF, Reid L, et al. Does deployment to Iraq and Afghanistan affect respiratory health of US military personnel? J Occup Environ Med. *2012;54:740–745.*

Another article published in the same issue of the *Journal of Occupational and Environmental Medicine* presented the results of a study that assessed the impact of deployment on the respiratory health of US military personnel.[28] The researchers' aim was to evaluate the association between postdeployment respiratory conditions and deployment to Iraq or Afghanistan. The conceptually relevant outcomes include the incidence of respiratory symptoms (eg, cough, wheezing, and shortness of breath) and a set of respiratory health conditions (eg, asthma, chronic bronchitis, and emphysema) after deployment. These outcome contrast with the outcome metrics assessed in the study: the appearance (postdeployment) of any one of a set of ICD-9 diagnostic codes in the hospitalization and outpatient medical encounter records of TRICARE beneficiaries. Specifically, the respiratory symptom outcome was defined as a single instance of "symptoms involving respiratory system and other chest symptoms" (ICD-9-CM [Clinical Modification] diagnosis code 786) in the medical record. Respiratory outcomes were similarly defined as a single instance of any diagnosis code in the broad category of "diseases of the respiratory system" (ICD-9-CM diagnosis codes 460–519). They were further categorized into six narrower ranges of diagnostic codes corresponding to healthcare encounters for

1. acute respiratory infections: ICD-9-CM diagnosis codes 460–466,
2. other diseases of the upper respiratory tract: ICD-9-CM diagnosis codes 470–478,
3. pneumonia and influenza: ICD-9-CM diagnosis codes 480–487,
4. COPD and allied conditions: ICD-9-CM diagnosis codes 490–496,
5. pneumoconiosis and other lung diseases due to external agents: ICD-9-CM diagnosis codes 500–508, and
6. other diseases of the respiratory system: ICD-9-CM diagnosis codes 510–519.

Sources of outcome misclassification are similar to those discussed for the in-theater assessment of cardiorespiratory encounters discussed previously, with some notable differences. The existence of false negatives is plausible because of, for example,

- individuals not seeking medical care,
- errors in the medical record,
- failure of medical providers to record health care encounters, and
- medical care received and paid for outside the TRICARE system.

The recording of medical conditions occurring in garrison is regarded as having much higher fidelity, relative to the systems used to record healthcare encounters during deployment. Miscoding of diagnoses and failure to record encounters likely occurred at lower frequencies in the medical systems that provided the outcome data in this study.

Usual patterns of healthcare utilization are disrupted during military deployment. Following troops' redeployment to garrison, it is likely that healthcare utilization, in general, spikes for a period of time. This is because personnel seek healthcare for conditions not addressed during their deployment (before returning to its normal level), except for the potential increase in healthcare required as a consequence of deployment experiences. This study did not discriminate between healthcare encounters that could be considered as part of the redeployment spike and those that are indicative of persistent increases in the respiratory health conditions of interest.

Regarding misdiagnoses and potential false-positive case assignment, this study found that 50% of diagnoses within the "COPD and allied conditions" category fall into the diagnostic code category of "bronchitis, not specified as acute or chronic" (ICD-9-CM diagnosis code 490). For reasons that can only be speculated on, military healthcare providers are frequently making use of this relatively uninformative, nonspecific diagnostic category. Furthermore, as discussed in Chapter 6 (Epidemiology of Airborne Hazards in the Deployed Environment) of this volume, the rate of medical encounters coded as "unspecified bronchitis" has increased in the military population (among both deployed and nondeployed personnel) from 2000 to 2011. It is as yet unclear whether personnel with encounters for unspecified bronchitis have true COPD conditions, such as asthma or chronic bronchitis.

Study 6

Baird CP, DeBakey S, Reid L, et al. Respiratory health status of US Army personnel potentially exposed to smoke from 2003 Al-Mishraq Sulfur Plant fire. J Occup Environ Med. *2012;54:717–723.*

Researchers in Study 3 assessed the postdeployment respiratory health of US Army personnel potentially exposed to smoke from a large sulfur fire in Iraq.[29] Again, outcomes were defined based on ICD-9 diagnosis codes abstracted from military medical records databases. Outcomes consisted of the following large diagnostic code categories:

- diseases of the circulatory system: ICD-9 diagnosis codes 390–459;
- diseases of the respiratory system: ICD-9 diagnosis codes 460–519; and
- symptoms, signs, and ill-defined conditions: ICD-9 diagnosis codes 780–799.

The authors also focused on the following smaller categories of respiratory disease diagnostic codes:

- COPD and allied conditions: ICD-9-CM diagnosis codes 490–496,
- asthma: ICD-9 diagnosis code 493,
- other chronic bronchitis: ICD-9 diagnosis code 491.8,
- pneumoconiosis and other lung disease from external agents: ICD-9 diagnosis codes 500–508,
- ischemic heart disease: ICD-9 diagnosis codes 410–414,
- other forms of heart disease: ICD-9 diagnosis codes 420–429,
- cerebrovascular disease: ICD-9 diagnosis codes 430–438,
- symptoms involving the cardiovascular system: ICD-9 diagnosis code 785, and
- symptoms involving the respiratory system: ICD-9 diagnosis code 786.

These outcome definitions are subject to the same errors (primarily nondifferential misclassification) as those discussed for Studies 4 and 5. However, outcome misclassification in this study may have been differential with respect to exposure if healthcare utilization among personnel exposed to the fire increased as a result of exposure concerns.

In addition to the case definitions based on diagnostic codes, the study also leveraged characterizations of health status as self-reported on DoD-mandated postdeployment health assessments (PDHAs). Errors in recalling health conditions on PDHAs likely occurred because of lapses in memory and complex determinants of respondents' motivation to report health concerns accurately. It has thus been wisely suggested that PDHA results be interpreted with caution.[31] Because differential misclassification bias would be induced if the quality (accuracy and completeness) of postdeployment survey data was not independent of exposure to the sulfur fire in this population. In light of evidence that rates of survey completion differed between exposure groups, the authors of this study chose not to compare the PDHA results between exposed and unexposed groups.

The length of the follow-up period in this study (and others) is limited. The historical proximity of the deployments of interest to the time when the study was conducted places an obvious limit on the amount of postdeployment person-time that can be accrued by formerly deployed military personnel. However, there are other limitations that are surmountable. In this study, subjects' postdeployment person-time accrued only while they remained in military service and did not again deploy; follow-up was discontinued when subjects separated from military service or had a subsequent deployment. These study design decisions were implemented because a subject, once redeployed, is not served by TRICARE. Therefore, the administrative database used to identify medical encounters does not collect medical encounter data in such instances. Linking with other medical record databases (eg, in-theater medical encounter databases, VA databases, and databases of large private insurance/healthcare providers) can facilitate the extension of follow-up for health outcome incidence beyond the service connection. This improvement in outcome capture is particularly important in evaluations of the relationship between deployment-associated hazards and diseases of long latency (eg, many types of cancers).

IMPROVING THE STATE OF SCIENCE BY IMPROVING OUTCOME ASSESSMENT

Measurement and classification of health outcomes in epidemiological studies are exercises in balance: minimizing errors and maximizing efficiency. Although challenges will persist, epidemiologists assessing potential health effects of deployment-associated airborne hazards can better align their operational outcome definitions with conceptually relevant outcomes. Doing so requires not only adherence to epidemiological principles, but also ingenuity, cleverness, and hard work on the part of the investigators, in addition to their having adequate resources to conduct high-quality studies.

Exhibit 7-4 summarizes identified weaknesses in epidemiological studies of deployment-associated airborne hazards' health effects with respect to the health outcome assessment. There will always be health outcome measurement and classification errors in epidemiological studies, just as there will remain technical, logistical, ethical, and financial barriers to implementing better outcome assessment methods. For any given assessment instrument, a trade-off between sensitivity and specificity in defining outcomes will be present. Despite these realities, steps can be taken in the design of a study, the analysis of study data, and the discussion of study results that can either reduce bias because of outcome misclassification or at least diminish the impact of imperfect outcome classification on the inferences drawn from a study (Exhibit 7-5).

The best way to overcome bias in epidemiological measures of association and estimates of precision from errors in

EXHIBIT 7-4

SUMMARY OF IDENTIFIED WEAKNESSES IN EPIDEMIOLOGICAL STUDIES OF DEPLOYMENT-ASSOCIATED AIRBORNE HAZARDS' HEALTH EFFECTS WITH RESPECT TO THE HEALTH OUTCOME ASSESSMENT

Assessment Methods

- Failure to ascertain and compare multiple sources of health outcome information
- Absence of efforts to validate outcome classification within the context of a given study
- No efforts to correct for bias in measures of association from outcome misclassification
- Inadequate use of sensitivity analyses to evaluate impact of misclassification on inference
- Inadequate discussion of the sources of outcome misclassification in published papers
- Insufficient discussion of prior validation of outcome assessment instrument(s)
- Insufficient discussion of the impacts of outcome misclassification on estimated measures of association and associated measures of precision
- Overreliance on administrative sources of outcome data

outcome classification is to avoid the errors in the design of the study. This can be achieved by using the most valid and reliable outcome assessment methods available. In designing a study, researchers should ask themselves the following questions:

- Is the measured outcome to be used in the study a valid and reliable indicator of the conceptually relevant health outcome?
- What are the potential sources of error in outcome classification?
- Are the sources of error in outcome classification independent of subjects' exposure status?
- Can outcome be assessed in different ways, potentially using a gold standard assessment method to classify outcomes in at least a subset of the study population?

Health outcome metrics should be empirically tested for validity and reliability, ideally in the setting of a pilot study conducted in advance of a large-scale investigation. As discussed, avoiding errors entirely is rarely possible. In designing a study, investigators should take steps to prevent, in particular, outcome classification errors that are not independent of subjects' exposure status. In defining outcomes, investigators should further acknowledge the trade-off between sensitivity and specificity, and weigh the advantages and disadvantages of false-positive and false-negative outcome assignments.

Statistical methods have been developed to correct for bias of measures of association from outcome misclassification, in conjunction with validation parameters. If appropriate steps are taken to parameterize outcome misclassification in the study design, analytic techniques can be used to reduce bias of measures of association and correct spurious estimates of precision from misclassification.[32–35] These methods rely on validation data and/or a set of assumptions that may or may not apply in a given study setting. In the absence of formal correction for bias from outcome misclassification, sensitivity analyses should be performed to assess the impact of outcome misclassification errors on the estimated measures of association.[36,37]

EXHIBIT 7-5

RECOMMENDATIONS FOR IMPROVEMENT OF USE OF ADMINISTRATIVE DATABASES

- Reduce outcome misclassification and measurement error in the study design phase
- Conduct validation of outcome measures
- Correct for outcome misclassification
- Coordinate health outcome definitions across research groups
- Acknowledge and evaluate outcome misclassification and measurement error and resulting bias in estimated measures of association
- Increase funding for research efforts and technologies development for outcome assessment

Finally, the discussions of study limitations published in the literature do not sufficiently enumerate the potential sources of errors in health outcome assessment and the impact of these errors on the validity of study results. Robust discussion of the limitations and impacts of imperfect outcome assessment should be included in the reports of epidemiological investigations. Such disclosure facilitates the reader's task of drawing an appropriate inference from the study.

The epidemiological studies of deployment-associated airborne hazards' health effects published to date rely disproportionately on administrative health record databases as the sole source of outcome classification data. Although the advantages are clear in terms of costs and logistical efficiency, such reliance is not without consequences. The result of this overreliance is persistence of sources (and impact on inference) of outcome misclassification across the literature. Inference would be strengthened if different research groups arrived at consistent results after having assessed the same hypothesis using different methods of outcome ascertainment.

To the extent that different research groups rely on the same sources of outcome data, the strategies they use to define outcomes (using medical record data, for example) could be better coordinated within or between research groups. The symposium work group identified a need to establish a follow-on working group to develop outcome definition criteria. At the very least, the use of standard outcome definitions facilitates comparability across research efforts conducted at different institutions. At best, such standardization can ensure that outcomes are appropriately defined for a given research goal. As a starting point, the Armed Forces Health Surveillance Center (AFHSC) has developed a set of outcome definitions for medical surveillance to "facilitate comparisons of case counts performed in different populations by different public health agencies" and to "harmonize health surveillance and epidemiologic analyses throughout the Department of Defense."[38] These outcome definitions are being adopted by researchers at the AFHSC and the US Army Public Health Command. Several research groups have begun to incorporate evidence of persistence of chronic disease into case definitions, and this practice should be adopted by other groups if appropriate to their investigations. Requiring evidence of persistence, such as requiring two or more diagnostic codes for a related condition observed in a prespecified time period for a given individual, will increase the specificity of the case assignment scheme and correspondingly decrease the proportion of false positives. Note, however, that this comes at a cost, namely a decreased sensitivity of the outcome classification scheme and a relatively greater proportion of false negatives.

Although resource-dependent, researchers need to advocate for initiatives to develop and validate novel indicators of outcomes and adopt the use of such indicators as they can. Adopting more refined outcome assessment technologies and strategies can reduce outcome measurement errors and the associated bias of measures of association.

SUMMARY

It is incumbent upon researchers conducting epidemiological studies to minimize sources and impacts of bias, where feasible, and to be aware of and disclose factors that impact the validity of their work. Sources of bias can be, at times, obvious and at other times obscure; engaging with an epidemiologist to evaluate the health outcome in the design phase may be a cost-efficient step in improving the quality of a study. Making inference using epidemiological study results requires an evaluation of the impacts of these errors. Prudent use of epidemiological study results then proceeds by contextualizing the results within the larger set of evidence, epidemiological and otherwise. As often as possible, the goal should be to measure what matters and measure it well.

REFERENCES

1. Weese CB, Abraham JH. Potential health implications associated with particulate matter exposure in deployed settings in southwest Asia. *Inhal Toxicol.* 2009;21:291–296.

2. Last JM, Spasoff RA, Harris, SS, eds. *A Dictionary of Epidemiology.* 4th ed. New York, NY: Oxford University Press; 1998.

3. Aschengrau A, Seage G, III. *Essentials of Epidemiology in Public Health*. Sudbury, MA: Jones & Bartlett Publishers; 2003.

4. Rothman KJ, Greenland S, Lash T. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.

5. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ.* 1994;308(6943):1552.

6. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol.* 1977;105:488–495.

7. White E. The effect of misclassification of disease status in follow-up studies: implications for selecting disease classification criteria. *Am J Epidemiol.* 1986;124:816–825.

8. Brenner H, Gefeller O. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. *Am J Epidemiol.* 1993;138:1007–1015.

9. Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol.* 2005;34:680–687.

10. Wacholder S, Hartge P, Lubin JH, Dosemeci M. Non-differential misclassification and bias towards the null: a clarification. *Occup Environ Med.* 1995;52:557–558.

11. Rothman KJ, Poole C. A strengthening programme for weak associations. *Int J Epidemiol.* 1988;17:955–959.

12. Hatch M, Thomas D. Measurement issues in environmental epidemiology. *Environ Health Perspect.* 1993;101(suppl 4):49–57.

13. Expert Panel Report 3 (EPR-3). Guidelines for the Diagnosis and Management of Asthma—Summary Report 2007. *J Allergy Clin Immunol.* 2007;120(suppl 5):S94–S138.

14. Pekkanen J, Pearce N. Defining asthma in epidemiological studies. *Eur Respir J.* 1999;14:951–957.

15. Hansen S, Strøm M, Maslova E, et al. A comparison of three methods to measure asthma in epidemiologic studies: results from the Danish National Birth Cohort. *PLOS ONE.* 2012;7:e36328.

16. Altman DG. *Practical Statistics for Medical Research*. London, England: Chapman and Hall; 1991.

17. Gray GC, Chesbrough KB, Ryan MAK, et al. The Millennium Cohort Study: a 21-year prospective cohort study of 140,000 military personnel. *Mil Med*. 2002;167:483–488.

18. Smith TC, Jacobson IG, Hooper TI, et al. Health impact of US military service in a large population-based military cohort: findings of the Millennium Cohort Study, 2001–2008. *BMC Public Health.* 2011;11:69.

19. Smith B, Wong CA, Smith TC, et al. Newly reported respiratory symptoms and conditions among military personnel deployed to Iraq and Afghanistan: a prospective population-based study. *Am J Epidemiol*. 2009;170:1433–1442.

20. Smith B, Chu LK, Smith TC, et al. Challenges of self-reported medical conditions and electronic medical records among members of a large military cohort. *BMC Med Res Methodol.* 2008;8:37.

21. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol.* 1990;43:551–558.

22. Szema AM, Peters MC, Weissinger KM, et al. New-onset asthma among soldiers serving in Iraq and Afghanistan. *Allergy Asthma Proc.* 2010;31:67–71.

23. National Asthma Education and Prevention Program, NHLBI, National Institutes of Health. *Asthma Expert Panel Report 3. Guidelines for the Diagnosis and Management of Asthma.* Bethesda, MD: National Institutes of Health; 2007. NIH Publication 08-5846.

24. Rose C, Abraham J, Harkins D, et al. Overview and recommendations for medical screening and diagnostic evaluation for postdeployment lung disease in returning US warfighters. *J Occup Environ Med.* 2012;54:746–751.

25. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med.* 1999;159:179–187.

26. Szema AM, Salihi W, Savary K, Chen JJ. Respiratory symptoms necessitating spirometry among soldiers with Iraq/Afghanistan war lung injury. *J Occup Environ Med.* 2011;53:961–965.

27. Abraham JH, Baird CP. A case-crossover study of ambient particulate matter and cardiovascular and respiratory medical encounters among US military personnel deployed to southwest Asia. *J Occup Environ Med.* 2012;54:733–739.

28. Abraham JH, DeBakey SF, Reid L, et al. Does deployment to Iraq and Afghanistan affect respiratory health of US military personnel? *J Occup Environ Med.* 2012;54:740–745.

29. Baird CP, DeBakey S, Reid L, et al. Respiratory health status of US Army personnel potentially exposed to smoke from 2003 Al-Mishraq Sulfur Plant fire. *J Occup Environ Med*. 2012;54:717–723.

30. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol.* 1980;112:564–569.

31. Mancuso JD, Ostafin M, Lovell M. Postdeployment evaluation of health risk communication after exposure to a toxic industrial chemical. *Mil Med.* 2008;173:369–374.

32. Duffy SW, Warwick J, Williams AR, et al. A simple model for potential use with a misclassified binary outcome in epidemiology. *J Epidemiol Community Health.* 2004;58:712–717.

33. Lyles RH, Tang L, Superak HM, et al. Validation data-based adjustments for outcome misclassification in logistic regression: an illustration. *Epidemiology.* 2011;22:589–597.

34. Magder LS, Hughes JP. Logistic regression when the outcome is measured with uncertainty. *Am J Epidemiol.* 1997;146:195–203.

35. Luo S, Chan W, Detry MA, Massman PJ, Doody RS. Binomial regression with a misclassified covariate and outcome. *Stat Methods Med Res.* 2012;March 15 [e-pub ahead of print].

36. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol.* 1996;25:1107–1116.

37. Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology.* 2003;14:451–458.

38. Armed Forces Health Surveillance Center. *AFHSC Surveillance Case Definitions*. Silver Spring, MD: AFHSC; 2013. http://www.afhsc.mil/caseSurveillanceDefs. Accessed August 22, 2013.