THE NATIONAL ACADEMIES REPORTS Assessing Military-Related Exposures and Health Outcomes

Statement of Dr. Karl Kelsey
Professor of Epidemiology, Professor of Pathology and Laboratory Medicine
Brown University.

before the
Committee on Veterans' Affairs
U.S. House of Representatives

Full Committee Legislative Hearing

May 5, 2021

Chairman Takano, Ranking Member Bost, and members of the Committee, thank you for the opportunity to testify today. My name is Dr. Karl Kelsey and I am a physician as well as Professor of Epidemiology and Professor of Pathology and Laboratory Medicine at Brown University. I'm speaking to you today in my capacity as a member of several committees formed by the National Academies of Sciences, Engineering, and Medicine (National Academies) that assessed the evidence between exposures encountered during military service and adverse health effects.

The National Academy of Sciences was created more than 150 years ago through a congressional charter signed by Abraham Lincoln in order to serve as an independent, authoritative body outside the government that could advise the nation on matters pertaining to science and technology. Every year, approximately 6,000 Academies members and volunteers serve pro bono on consensus study committees or convening activities. The National Academies do not advocate for specific policy positions. Rather, they enlist the best available expertise across disciplines to examine the evidence, reach consensus, and identify a path forward. National Academies reports, proceedings and other publications are available via the web in PDF form without charge.

The National Academies have a long history of advising the federal government on the health effects of military service in general and on the effects of in-theater exposures resulting from military activities in particular. In addition to the 12-report *Veterans and Agent Orange* (VAO) series and the 11-report *Gulf War and Health* series, there have also been several focused reports that have examined the effects of particular exposures in veterans of different eras.

I was asked to focus my testimony on the standards used for weighing and assessing epidemiologic evidence of exposure to a toxin or group of toxins and health effects. I will describe the differences between association and causation and give examples of how those standards were used in National Academies' reports on assessments of veterans health. I will also discuss related issues including quality of data and how this affects estimates of risk and estimates of how many people may be affected given a particular exposure.

Association vs Causation

As used technically, the criteria for causation are somewhat more stringent than those for association and are more difficult to satisfy. Such factors as temporality (exposure must be present before the health outcome), strength of association, dose–response relationships, consistency of evidence through replication of findings and other knowledge, specificity of the association, biologic plausibility, coherence of evidence, and consideration of alternate explanations may be considered when deciding whether an observed statistical association is causal. These factors are not a checklist that require each criterion to be satisfied; the only necessary criterion is temporality: that the exposure comes before the outcome. A positive statistical association between an exposure and an outcome does not necessarily mean that the exposure is the cause of the outcome. Causality is more than a "link"; it is a demonstration that an exposure(s) is responsible for specific health outcome(s). For outcomes that have multiple causes, some causes may contribute to a larger proportion of the total cases for that outcome. For every exposure–outcome relationship, there will always will be gradations of evidence and certainty. As many are familiar with the adage, correlation does not equal causation, observed links or associations can be due to many factors.

To determine whether there is an association between exposure and a health outcome, epidemiologists estimate the magnitude of an appropriate measure (such as the relative risk or the odds ratio) that describes the relationship between exposure and a health outcome in a defined population or group. In evaluating the strength of the evidence linking toxin exposure with a particular outcome, considerations must be given to whether such estimates of risk might be flawed or might accurately represent true associations. Chance, bias related to errors in selection and measurement, and confounding are critical issues committee members consider as they review data to assess causation or association as these factors can create the illusion of causation or association, or hide it.

It has been the practice of all National Academies committees that produced the Veterans and Agent Orange and Gulf War and Health reports to evaluate all studies according to the same criteria and then to weight findings of similar strength and validity equivalently when drawing conclusions. Study committees that assessed military exposures and health outcomes not part of these series, generally followed the same practice. An absolute conclusion about the absence of association might never be attained because, as is generally the case in science, studies of health outcomes after an exposure cannot demonstrate that a purported effect is impossible, only that it

is statistically improbable.

One of the main considerations needed to determine whether there is an association between exposure and a health outcome is how the exposure and health outcomes were measured. There have been numerous health studies of veterans of different eras, but most have been hampered by relatively poor measures of exposure and by other methodologic problems. For example, except in rare instances, no objective measurements of exposure to herbicides are available for most Vietnam veterans. Instead, and in accord with Congress' mandated presumption of herbicide exposure of all Vietnam veterans, VAO committees have used Vietnam-veteran status as a proxy for herbicide exposure when no more specific exposure information is available. Those committees have considered studies of populations of other groups potentially exposed to the constituents present in the herbicide mixtures used in Vietnam that had better or more objective measures of exposure. Similarly, research regarding exposure to open burn pits has been impeded by relatively poor measures of exposure including single-time measurements, unknown composition of burned materials, and monitors that have not covered the full range of chemicals known to be present in the emissions as well as other methodologic problems including inability to determine or account for the contribution of other sources of airborne hazards in the area.

As is often the case when reliable and accurate exposure information is not available for military populations, deployment to a particular area—which may be as nonspecific as a particular country or group of countries, (e.g., Southwest Asia theater)—is used as a proxy for exposure. Causal models and inference are dependent on high-quality data; poor exposure assessment and use of such proxies as deployed/nondeployed limits the ability to inform causality.

How health outcomes were collected or measured is also an important consideration when considering causality. Greater confidence is given for relationships that rely on outcomes that were objectively measured or tested or collected by a medical professional using a standardized exam or technique vs outcomes based on self-report. Validating a subset of self-reported responses with information contained in medical records may increase the confidence of self-reported health information. Many publications of health outcomes in veterans reviewed by NASEM committees have been based on self-reported responses from surveys administered by the Department of Veterans Affairs that may or may not have a validation component.

The evidence assessed to determine causality is continually evolving, In recent years, causality determination has become more complex as the scientific community learns more about how the totality of exposures influence health, including genetics, epigenitics, stress, psychosocial factors, and social determinants of health. Additionally new methods and technologic advancement create new data streams and push current approaches to incorporate new kinds of information including in vitro technologies, toxicogenomics and epigenetics, molecular epidemiology, and exposure assessment (NASEM, 2018a).

Categories and Standards Used by NASEM Committees to Assess Strength of Evidence Between Military Exposures and Health Outcomes

The categories developed and used by the VAO committees were adapted from those used by the International Agency for Research on Cancer in evaluating the evidence for carcinogenicity of various agents (IARC, 1977). Consistent with the charge to the Secretary of Veterans Affairs in the Agent Orange Act (PL 102-4), the distinctions between the categories are based on "statistical association," not on causality. The four categories are: sufficient, limited or suggestive, inadequate or insufficient, and no association. The classification of health outcomes are based on the committee's evaluation of the epidemiologic literature and evidence of biologic plausibility or mechanistic data, and reflect their judgment of the relative certainty of the association between the outcome and exposure to the herbicides used in Vietnam or to any of their components or contaminants. The distinctions in the category descriptions describe the completeness and quality of a body of evidence, and the degree of certainty about an association, or lack of evidence of association for the fourth category. For example, a health outcome placed in the sufficient category reflects the committee's judgment that a statistical association would be found in a large, well-designed epidemiologic study of the outcome in question in which exposure to herbicides or dioxin was sufficiently high, well-characterized, and appropriately measured on an individual basis. The default category for any health outcome is "inadequate or insufficient" until enough evidence has accumulated to reclassify it into a different category. For many conditions, however, particularly ones that are very uncommon, any association with the exposures of interest has remained unaddressed in the medical research literature; for these, the committee remains neutral based on the understanding that "absence of evidence is not evidence

of absence." In addition to the VAO Update committees, the use of these four categories of association have been used by NASEM committees responsible for reports on long-term health effects of antimalarial drugs when used for prophylaxis (NASEM, 2020a) and respiratory health effects of airborne hazards exposures in the Southwest Asia theater of military operations (NASEM, 2020b).

Similarly, study committees of the *Gulf War and Health* series applied a similar categorical framework as the VAO committees based on association but also included a fifth category of "sufficient evidence of a causal relationship" as they were not constrained by law. Among all of the health outcomes considered in relation to military service in the 1990-1991 Gulf War, only posttraumatic stress disorder has satisfied criteria to be placed in the category of "sufficient evidence of a causal relationship." The approach of four association and a fifth causation categories was used by the Gulf War and Health committees as well as by other National Academies' committees, including health effects associated with burn pit emissions (IOM, 2011). EPA has adapted the use of these categories for its integrated scientific assessments for criteria air pollutants since 2008. It was developed to be flexible enough to be applicable to a range of various exposures and incorporate epidemiologic, toxicologic, and mechanistic data. Those criteria have been applied to a variety of agents, including sarin, depleted uranium, vaccines, pesticides, and environmental chemicals.

The committees responsible for each National Academies report are independent and may decide to change a word or phrase in the category description, but the intent of the classification is the same in providing the degree of certainty about an association. These changes reflect the study committees' needs to address specific issues, such as a single exposure as compared with a mix of exposures encountered throughout deployment, or to clarify language.

Although not specific to military exposures, some NASEM committees have used categories to assess strength of causality, most notably a series of reports on adverse effects of vaccines. For those reports, given that details on exposure were available, including dose, frequency, timing, etc., a robust evidence base of epidemiologic studies, randomized controlled trials, and mechanistic studies were available, causality standards could appropriately be applied. The four causality categories used were: evidence convincingly supports a causal relationship; evidence favors acceptance of a causal relationship; evidence is inadequate to accept or reject a

causal relationship; and evidence favors rejection of a causal relationship.

The 2008 IOM report, Improving the Presumptive Disability Decision-Making Process for Veterans, proposed a different set of categories describing the likelihood of a causal relationship:

1. Sufficient: The evidence is sufficient to conclude that a causal relationship exists.

2. Equipoise and Above: The evidence is sufficient to conclude that a causal relationship is at least as likely as not, but not sufficient to conclude that a causal relationship exists.

3. Below Equipoise: The evidence is not sufficient to conclude that a causal relationship is at least as likely as not, or is not sufficient to make a scientifically informed judgment.

4. Against: The evidence suggests the lack of a causal relationship.

However, that report had a much different purpose than the VAO, Gulf War and Health, and other committees tasked with assessing the strength of evidence for specific exposure and health outcome relationships. The 2008 study committee was charged with describing the current process for how presumptive decisions are made for veterans who have health conditions arising from military service and with proposing a scientific framework for making such presumptive decisions in the future. Presumptions are made in order to reach decisions in the face of unavailable or incomplete information. They address the gaps in evidence that introduce uncertainty in decision-making. In trying to assess whether a particular health problem in veterans can be linked to their exposures in the military, a presumption might be needed because of missing information on exposures of the veterans to the agent of concern or because of uncertainty as to whether the exposure increases risk for the health condition. A presumption might also be made with regard to the link between an exposure and risk for a disease, while the evidence is still uncertain or accumulating as to whether the exposure causes the disease. As noted by the 2008 authoring committee, policy decisions are based on more than the scientific evidence alone, and "there are social, economic, political, and legal factors beyond the scope of scientific evidence that may influence the presumptive disability decision-making process for veterans and the presumptive decisions that are established by Congress and VA" (IOM, 2008, p 22). Some of the gaps identified by the 2008 report committee, including lack of information on exposures received by military personnel and inadequate surveillance of veterans for service-

related illnesses; a failure to quantify the effect of the exposure during military service, particularly for diseases with other risk factors and causes; and a general lack of transparency of the presumptive disability decision-making process, continue to persist today, nearly 13 years after this report was published. However, the root issue is the lack of available and accurate exposure information, and without that information, causality cannot be determined, and therefore trying to impose standardized strength of evidence categories based on causation would be premature.

Notably, no other NASEM committee tasked with assessing the evidence of health effects related to exposures encountered during military service has adopted the use of these categories.

An IOM report for the US Army noted that the causal categories used to assess vaccines and support presumptive disability decision-making are more appropriate when assessing mostly epidemiologic evidence whereas more diverse bodies of evidence composed of toxicologic and mechanistic data may require the added flexibility built into the association-based categories used by VAO and Gulf War and Health committees (NASEM, 2018).

Limitations in Use of Causal Standards

Without information on the extent of exposure among most veterans regardless of era and quantitative information about the dose–time–response relationship for each health outcome in humans, estimation of the risks experienced by veterans exposed to the compounds of interest is not possible. Although record keeping has improved and exposure estimation has been incrementally better on the whole, there are still few instances when assessing effects of military exposures when it makes sense to use causality standards.

In the Veterans and Agent Orange series, for each association between a specific health outcome and exposure to the chemicals present in the herbicides used by the military in Vietnam, the study committees were asked to consider the increased risk of disease among those exposed to herbicides during service in the Republic of Vietnam during the Vietnam era. The requisite information to assign risk estimates continues to be absent despite concerted efforts to model the exposure of the troops in Vietnam, to measure the serum dioxin concentrations of individual

veterans, and to model the dynamics of retention and clearance of TCDD in the human body. Accordingly, each VAO Update committee has been unable to derive quantitative estimates of any increased risks of various adverse health effects that Vietnam veterans may have experienced in association with exposure to the herbicides sprayed in Vietnam.

In general, the committees that have examined health effects related to military exposures have concluded that it would be inappropriate to use quantitative techniques, such as meta-analysis, to combine individual study results into a single summary measure of statistical association because of the many differences among studies in definitions of exposure, health outcomes considered, criteria for defining study populations, correction for confounding factors, and degree of detail in reporting results. The appropriate use of meta-analysis requires more methodologic consistency across studies, especially in the definition of exposure, than is present in the literature reviewed by the committees. It is more informative to include a detailed discussion of the results from individual studies with a thorough examination of each study's strengths and weaknesses. In general, the committees did not consider case reports, case series, or other published studies that lacked control or comparison groups.

Thank you for the opportunity to testify. I would be happy to address any questions that you might have.

References

Institute of Medicine. 2008. Improving the Presumptive Disability Decision-Making Process for Veterans. Washington, DC: The National Academies Press.

Institute of Medicine. 2011. Long-Term Health Consequences of Exposure to Burn Pits in Iraq and Afghanistan. Washington, DC: The National Academies Press.

International Agency for Research on Cancer (IARC). 1977. Some Fumigants, the Herbicides 2,4-D and 2,4,5-T, Chlorinated Dibenzodioxins and Miscellaneous Industrial Chemicals. IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man, Vol. 15. Lyon: IARC. 111-138, 273-299.

National Academies of Sciences, Engineering, and Medicine 2018a. Advances in Causal Understanding for Human Health Risk-Based Decision-Making: Proceedings of a Workshop in Brief. Washington, DC: The National Academies Press.

National Academies of Sciences, Engineering, and Medicine. 2018b. Review of Report and Approach to Evaluating Long-Term Health Effects in Army Test Subjects. Washington, DC: The National Academies Press.

National Academies of Sciences, Engineering, and Medicine 2020a. Assessment of Long-term Health Effects of Antimalarial Drugs When Used for Prophylaxis. Washington, DC: The National Academies Press.
National Academies of Sciences, Engineering, and Medicine 2020b. Respiratory Health Effects of Airborne Hazards Exposures in the Southwest Asia Theater of Military Operations. Washington, DC: The National Academies Press.