Approach to Evaluating Health-Related Scientific Evidence and Expert Opinion to Support Disability Benefit Decision-Making

Jim Thompson MD CCFP(EM) FCFP
Medical Advisor
Research Directorate, Policy Division
Veterans Affairs Canada

26 October 2012

Acknowledgement: I wish to express my appreciation to Dr. Margaret Boswall and Dr. Linda VanTil at VAC for their expertise and assistance in developing this approach.

Contents

Citation:
Thompson J. Approach to evaluating health-related scientific evidence and expert opinion to support disability benefit decision-making. 26 October 2012.

1. Introduction

This note documents my approach to support disability benefit entitlement decision-making when a VAC decision-maker consults me to provide them with an opinion when they are dealing with uncertainty in health-related expert opinion and scientific evidence.

1a. Background

Eligible serving and former service members can apply to Veterans Affairs Canada (VAC) for entitlement to disability benefits, which are a key gateway to financial awards and a variety of supports. Entitlement to disability pensions and disability awards requires that the person have a medical diagnosis of a health condition connected to military service and a related permanent medical disability.

VAC's authority for disability benefit entitlement decision-making is established in federal legislation: the *Pension Act* (entered into force in 1919) and the *Canadian Forces Members and Veterans Re-establishment and Compensation Act* (New Veterans Charter; entered into force in 2006):

> Pension Act: "*21(1) In respect of service rendered during World War I, service rendered during World War II other than in the non-permanent active militia or the reserve army, service in the Korean War, service as a member of the special force, and special duty service, a) where a member of the forces suffers disability resulting from an injury or disease or an aggravation thereof that was attributable to or was incurred during such military service, a pension shall, on application, be awarded to or in respect of the member in accordance with the rates for basic and additional pension set out in Schedule I...*"
>
> Pension Act: "*21(2) In respect of military service rendered in the non-permanent active militia or in the reserve army during World War II and in respect of military service in peace time, (a) where a member of the forces suffers disability resulting from an injury or disease or an aggravation thereof that arose out of or was directly connected with such military service, a pension shall, on application, be awarded to or in respect of the member in accordance with the rates for basic and additional pension set out in Schedule I...*"
>
> Canadian Forces Members and Veterans Re-establishment and Compensation Act: "*2. "service-related injury or disease" means an injury or a disease that (a) was attributable to or was incurred during special duty service; or (b) arose out of or was directly connected with service in the Canadian Forces ... 45(1) The Minister may, on application, pay a disability award to a member or a veteran who establishes that they are suffering from a disability resulting from (a) a service-related injury or disease; or (b) a non-service-related injury or disease that was aggravated by service.*"
>
> The Veterans Review and Appeal Board interpreted "*attributable to*" as "*caused by*".

Both acts define *disability* as "*the loss or lessening of the power to will and to do any normal mental or physical act*". Entitlement combines both the presence of a medical diagnosis of illness or injury and related impairments on the one hand, and disability that occurs when a person encounters barriers preventing normal functioning on the other. *Medical diagnosis* is a

health condition for which an eligible person is claiming entitlement to disability benefits. Physical and mental health conditions often confer some form of physical and/or mental *impairment* that can affect function. *Disability* clinically means not being able to function as a result of encountering barriers, either internal adaptive coping or external social and physical barriers. People who have no health conditions can encounter barriers that *dis-able* them, such as being unable to get a well-paying job without appropriate education, or being unable to walk outside a space vehicle. These barriers can be overcome with education or technology that *enables* the person. Disability is more likely to occur when a person has physical or mental impairment due to a health condition, and disability is likely to be more severe when the impairment is more severe.

Entitlement may be provided if there is sufficient evidence of a medical diagnosis related to service, and then degree of disability is assessed to determine compensation. In keeping with both legislation and current clinical thinking, VAC approaches disability compensation in two steps:

1. Step one *entitlement*: Determine whether the health condition (medical diagnosis) was incurred during, or aggravated by, or attributable to, or directly connected with service, and whether the health condition resulted in a permanent disability.
2. Step two *assessment*: Determine the degree of disability.

In certain circumstances (*the insurance principle*), it is sufficient to demonstrate only that a condition was incurred during certain types of service, such as while serving in a Special Duty Area. In other cases, such as when a condition arises during service that is not special duty (*the compensation principle*), or when a condition arises years after special duty service, it is necessary to determine whether the condition or aggravation of the condition was caused by military service activity.

Adjudicators, policy writers and program developers working on disability benefit issues must weigh a variety of types of evidence, including *health-related expert opinion and scientific evidence*. New scientific evidence is being published worldwide at an increasingly high volume and rate, often is technically challenging to evaluate, and characterized by inconsistency and uncertainty, so often it is difficult to explain reasons for decisions. Good disability benefit entitlement decision-making is *timely, legally sound, medically sound, fair, consistent, efficient and transparent inside and outside VAC*. VAC tools to help decision-makers deal with this type of evidence include legislation, the Table of Disabilities, policies and the Eligibility Entitlement Guidelines. These tools require ongoing maintenance as new scientific information emerges, and cannot cover all the questions that arise. Expert legal and medical assistance is routinely required at VAC to help them deal with this type of evidence.

2. Application

2a. Examples of Applications

This document explains the approach I use to formulate an opinion to support decision-making in the following VAC activities:

- Case-by-case decision-making for questions of disability entitlement and assessment.
- Decision-making about criteria used in the Entitlement Eligibility Guidelines and the Table of Disabilities.

- Decision-making about evidence-based statements in policies related to disability entitlement and assessment.
- Decision-making about evidence-based principles used in developing programs and services related to disability entitlement and assessment.

Many factors of military service have been claimed over the years (see section 3a), and a very large variety of physical and mental health conditions have been found related to military service by VAC *for the purposes of disability benefit entitlement*[1].

There is considerable variability in the kinds of questions that prompt referral by a VAC decision-maker. When an application hinges on causality, the decision-maker has to consider whether a health condition was caused or aggravated by a factor encountered in the applicant's service. This means the decision-maker considers evidence for:

1. Whether factor A causes condition B, or whether condition A causes condition B;
2. Whether the person was exposed in a sufficient manner to factor A to have caused condition B, or whether the person had condition A in manner that would cause condition B; and
3. Whether the latency period between exposure to factor A, or whether the presence of condition A and onset of the health condition was appropriate.

I also use this approach for dealing with questions related to the insurance principle, for example whether symptom A that was incurred in special duty service was part of condition B that developed later in life.

3. Terminology

3a. Evidence

Evidence: Any form of proof that is offered to substantiate a claim and/or to establish the existence or non-existence of any fact in dispute.

Health-Related Expert Opinion and Scientific Evidence: Health-related expert opinion and scientific evidence is a special type of evidence that is considered by decision-makers when they make a decision on a client's claim for disability benefit entitlement, or develop a guideline, policy or program. Expert advisors specialize in assisting decision-makers dealing with uncertainty in this type of evidence.

Scientific evidence can include results of scientific studies, critical reviews of multiple scientific studies, a client's health records and file reviews, depending on the nature of the question.

Expert opinion is informed judgement that fills gaps in scientific evidence. Examples include letters from client's health care practitioners, committee reports, professional guidelines, lists of risk factors and textbook entries based on author judgement.

Uncertainty: Uncertainty is the existence of doubt, controversy or lack of clarity in evidence.

[1] Pedlar DJ, Thompson JM. Research in the life courses of Canadian military Veterans and their families. In: A Aiken & SAH Bélanger (eds.): *Shaping the Future, Military and Veteran Health Research*. Kingston, Ontario: Canadian Defence Academy Press; 2011. p15-31.

At VAC, decision-makers consider all the available evidence when considering a claim for disability benefit entitlement, or when developing a guideline, policy and program related to disability benefit entitlement. An expert advisor's opinion about a body of health-related expert opinion and scientific evidence contributes to the decision, but is not the only determinant.

3b. Exposure and Latency

Decision-making for claims often revolves around whether exposure to a factor of military service caused the Veteran's health condition later in life.

Exposure has two meanings relevant to VAC policy: (1) a *hazard*, and (2) *contact with a hazard*.

With respect to the first meaning (*hazard*), a variety of factors have been connected to health and disability, including but not limited to:

- Weapons.
- Mechanical hazards.
- Physical, biological, chemical and radiological hazards.
- Environmental stress.
- Psychological stress (psychological trauma).
- Social stress.
- Illnesses and injuries that occur in military service (an illness or injury service can cause or aggravate certain conditions later in life).

The VAC exposure policy is limited to four hazard types: physical, biological, chemical and radiation[2].

With respect to the second meaning (*contact with a hazard*), exposure is characterized by *mode, extent, timing and biological effects*. *Mode* describes how the person was exposed, including *source* and the *route* or *pathway* taken by an exposure when it affects a person. *Extent* considers the *amount, frequency* and *duration* of exposure to the factor. *Timing* includes *latency*, which refers to the delay that occurs between exposure to a factor and manifestation of the health condition. "*Biological effects"* considers how body chemistry reacts to the exposure factor, including protective and adverse effects and variation between individuals.

3c. Association and Causality

There are two steps in evaluating evidence for causality between a factor and a health condition:

First, is there evidence of *association* between the factor and the condition?

Second, is there evidence of *causality* in the association between the factor and the condition?

Association means that a factor and a health outcome are said to be associated when the two appear to occur together. Associations can be explained by *chance*, *bias*, *confounding*, or *causality*:

[2] VAC Policy "*Hazardous Material and Radiation Exposure*", 2012.

- *Chance*: The association was due to random variation.
- *Bias*: The association was due to flaws in study design, sample recruitment, data collection, analysis or interpretation which led to favouring conclusions that deviate from the truth.
- *Confounding*: The association was due to the presence of unrecognized variables related to the factor and/or the health outcome.
- *Causality*: The relating of causes to the effects they produce; a relationship between a factor and a health condition, where exposure to the factor earlier in life results in the health condition later in life, as in a "causal relationship". The association was due to a causal relationship between the factor and the health condition such that the factor caused or aggravated the health condition. The term "causal association" is inappropriate. While there is debate about the definition of causality, several authorities have pointed to the importance of having this type of practical definition when public policy solutions are required[3].

Criteria for Causality

Several criteria drawing on multiple lines of evidence need to exist to support the conclusion that causality exists. In the 1960s, a set of principles for determining causality call the "Hill criteria" emerged and were widely accepted. Since then, thinking about criteria for causality have been refined and evolved (Table 1)[4,5,6,7,8,9,10,11].

[3] Parascandola M, Weed DL. Causation in epidemiology. J Epidemiol Community Health. 2001 Dec;55(12):905-12.
[4] Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. Evidence-based toxicology: a comprehensive framework for causation. Hum Exp Toxicol. 2005 Apr;24(4):161-201.
[5] Susser MW. What is a cause and how do we know one? A grammar for pragmatic epidemiology. Am J Epidemiolg 1991; 133:635-648.
[6] Samet JM, Bodurow CC. Improving the presumptive disability decision-making process for Veterans. Committee on evaluation of the presumptive disability decision-making process for Veterans. Institute of Medicine. 2007 Aug;789p.
[7] Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. Evidence-based toxicology: a comprehensive framework for causation. Human & Experimental Toxicology 2005;24:161-201.
[8] Hill AB. The environment and disease: association or causation? Proceedings of the Royal Society of Medicine. 1965;58:295-300.
[9] Lynch RM, Henifin MS. Causation in occupational disease: Balancing epidemiology, law and manufacturer conduct. Risk: Health & Environment. Summer 1998;259-270.
[10] Ward JD, Donal KJ. Statements of Principles: evidence-based compensation for Australian Veterans and serving defence personnel. ADF Health. 2004;5:89-93.
[11] Kaldor J. Critical appraisal and causal inference. In: Proceedings of the 2008 Repatriation Medical Authority Forum, Canberra, Australia. 2008;45-56.

Table 1. Criteria for Causality.

Epidemiological evidence:

- *Temporality*: Exposure to the exposure factor precedes onset of the health condition.
- *Numerical strength*: Statistical measures of association such as relative risk and odds ratios are sufficiently strong. When assessing causation, statistics based on *incidence* provide a better estimate of risk than *prevalence* where disease duration is combined with risk.
- *Lack of confounding:* Whether any other exposure factor explains the association.
- *Presence of dose-response:* Whether more people have the health condition when exposed to more of the exposure factor.
- *Specificity*: Whether the exposure factor causes only the health effect. Lack of specificity does not rule out causality.
- *Experimental control*: Whether randomized controlled trials and other types of direct evidence show that exposure to the factor causes the health condition. This type of evidence also rules out *reverse causality* where the health condition causes the exposure factor. Experimental control evidence is rarely available for Veterans' entitlement questions, for obvious reasons.

***Existing knowledge*:**

- *Coherence:* Whether causality fits with existing theory.
- *Biological and mechanistic plausibility*: Whether it makes sense biologically that the exposure factor could cause the health condition. This type of evidence comes from clinical, laboratory and animal research.

Strength of evidence:

- *Quality, Quantity and Consistency*: The degree to which studies are methodologically sound and adequately control for chance, bias and confounding, and expert opinion is well informed, qualified, reliable and credible. A sufficient number of good quality studies support rather than refute causality.

3d. Risk Factors

Risk factor can have two meanings: a factor *associated* with increased probability of an outcome but not necessarily causal[12]; or a factor that *causes* the increased probability of an outcome, also called a *determinant*[13,14]. The term "risk factor" is loosely used and often it is not clear whether there is sufficient evidence for a causal relationship. The criteria for causation described above can be used to differentiate between these meanings.

4. Epidemiological Studies

Different health study designs generally lie on a hierarchy of evidence for causality based on causality criteria (section 3b), from strongest to most limited:

[12] Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. Evidence-based toxicology: a comprehensive framework for causation. Hum Exp Toxicol. 2005 Apr;24(4):161-201.
[13] Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. Evidence-based toxicology: a comprehensive framework for causation. Hum Exp Toxicol. 2005 Apr;24(4):161-201.
[14] Porta M. A Dictionary of Epidemiology, 5th Edition. Edited for the International Epidemiological Association. Oxford University Press, 2008.

1. Randomized controlled trial (RCT).
2. Cohort study.
3. Case-control study.
4. Multiple time series of cross-sectional studies.
5. Individual cross-sectional study.

4a. Randomized Controlled Trial

Randomized controlled trials meet the most criteria for causality. Subjects are followed over time to establish temporality. The RCT design allows measures of risk using *incidence* (new cases in a population over a defined period). The experimental nature of RCTs allows for several levels of exposure to the factors of interest, and evaluation of a treatment or other intervention. The quality of a RCT is assessed by looking at eligibility criteria (whether the findings generalize to the Veteran population of interest); the use of appropriate statistics to eliminate chance as responsible for the association; and the limitation of bias through the use of blind controls and placebos, minimal loss to follow-up, and random selection from the eligible subjects to ensure similarity between the treated and control groups.

4b. Cohort Study

Cohort studies account for fewer criteria for causality than RCTs. Subjects are followed over longitudinally over time to establish temporality, and this design allows measures of risk using *incidence* (new cases in a population over a defined period). The design often includes several levels of exposure to the factors of interest. The quality of a cohort study is assessed by looking at the eligibility criteria (does this generalize to the Veteran population of interest?); the use of appropriate statistics to eliminate chance as being responsible for the association; design that limits bias through the use of minimal loss to follow-up; and measures of confounders (age, sex, socio-economic status, smoking) to ensure similarity between the groups of differing exposures.

4c. Case-Control Study

Case-control studies meet even fewer criteria for causality. This design is not the best for establishing temporality, since it provides a retrospective history of exposure to the factors of interest. The selection of cases determines whether a case-control study measures *prevalence* (existing cases in a population at a point in time), or *incidence* (new cases in a population over a defined period). The quality of a case-control study is assessed by looking at the eligibility criteria for both cases and controls (Are they the same? Do both cases and controls generalize to the Veteran population of interest?); the use of appropriate statistics to eliminate chance as responsible for the association; design that limits bias through the use of blind controls; addressing recall bias since the exposure is retrospective; and measures of confounders (e.g., age, sex, socio-economic status, or smoking) to ensure similarity between the groups of differing outcomes. In spite of this design's short-comings, it is appropriate for rare conditions, chronic diseases and other long term effects of exposure to the factors of interest. Case-control studies are often used to study cancer; if cases are selected from a suitable cancer registry, this will allow the study to measure risk as incidence, which is more useful than prevalence in understanding causation.

4d. Cross-Sectional Study

Cross-sectional studies, by far the most common type of epidemiological study, meet the least criteria for causality. This design does not establish temporality, but provides a snapshot in time.

A series of cross-section studies conducted over time can provide some indication of temporality but not proof. This design measures *prevalence* (existing cases in a population at a point in time), which is a measure of burden, not risk. The quality of a cross-sectional study is assessed by looking at the eligibility criteria and measures of confounders (whether the findings generalize to the Veteran population of interest), and the use of appropriate statistical methods to eliminate chance as responsible for the association.

5. Dealing with Uncertainty in Health-Related Expert Opinion and Scientific Evidence

Consider the fictitious finding of a statistical association where more people with cigarette-stained fingers had lung cancer than those with unstained fingers. Does this mean that staining of fingers by cigarette smoke causes lung cancer? Or does it mean that the association is not causal, occurring instead merely as a result of bias, chance, or confounding?

When decision-makers are uncertain about a question related to disability benefit entitlement or assessment, they can turn to advisors to help clarify a body of expert opinion and scientific evidence. The expert advisor contributes a review of health-related expert opinion and scientific evidence to the process, but does not make the final decision.

5a. The "Q-4As" Model

A solution common to Veterans' and workers' compensation agencies in Canada and around the world has been developed to deal with uncertainty. Standard approaches encourage standard practices for conducting reviews of bodies of health-related expert opinion and scientific evidence, and using language for expert advisors to communicate subjective judgements about the strength of evidence and certainty of opinion. There are five steps in the process of dealing with uncertainty in expert opinion and scientific evidence[15]:

1. *Question:* The VAC decision-maker frames a question (adjudication, policy or program development) about disability benefit entitlement or assessment and communicates it to the expert. The decision-maker and expert might work together to refine the question.
2. *Acquire*: The expert gathers a body of health-related expert opinion and scientific evidence relevant to the question. The expert decides how much and what type of evidence is necessary for them to draw a conclusion and form their opinion.
3. *Assess*: The expert weighs the strength of evidence using standard principles of epidemiology and evidence review.
4. *Adapt*: The expert draws conclusions to form an opinion sufficient to answer the question, makes a subjective judgement about the strength of evidence and degree of certainty, and then communicates this opinion to the VAC decision-maker.
5. *Apply*: The VAC decision-maker considers all the evidence and makes the decision. The decision-maker can consult the advisor during this stage if clarification about the nature of health-related expert opinion and scientific evidence is required.

[15] Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. Evidence-based toxicology: a comprehensive framework for causation. Hum Exp Toxicol. 2005 Apr;24(4):161-201.

5b. Question: Clarify the Question

The advisor is asked for their opinion about conclusions that can be drawn from a body of health-related expert opinion and scientific evidence relevant to a question, and the degree of certainty. Typical disability benefit entitlement questions posed to advisors deal with *causation, exposure* and *latency*. In some cases the question posed might deal with only a narrow aspect, such as whether factor A causes condition B, or whether a person was exposed sufficiently to factor A to have caused condition B. In other cases the question might be more comprehensive, such as whether this person's condition B was caused by exposure to factor A.

Getting the question clear is the first step. In routine adjudication cases, questions might follow standard formats and usually require no clarification. For complex questions that are not routine, and for supporting decision-making tools like the Entitlement Eligibility Guidelines and the Tables of Disabilities, and for policy and program development, it might be necessary to revisit the question both before and during the work.

5c. Acquire: Search for Expert Opinion and Scientific Evidence

The kind and amount of evidence that will be acquired depends on the nature and context of the question, and the expert's own familiarity with the related field.

Expert opinion. If expert opinion is required for the task, options range from verbally checking with an expert in the field, to searching for published expert opinions from agencies like the U.S. Institute of Medicine Committees, specialty associations, or scientific panels at workers' compensation boards.

Scientific evidence usually is confined to peer-reviewed publications in credible scientific journals and credible textbooks based on such literature. In rapidly advancing fields, textbooks can be out of date by several years at publication. Scientific papers include reports of single studies, or meta-analyses and critical reviews of multiple studies.

Search methods. In most routine case-by-case decision-making methods for searching for expert opinion and scientific evidence might be very limited. Questions that are less routine questions or have broach program and policy implications might require more formal and exhaustive approaches. Systematic searches for expert opinion and scientific evidence have the following characteristics[16]:

- *Goal*. The goal should fit the question.
- *Inclusion criteria*: Depending on the task, searches might be limited to peer-reviewed papers published in credible journals, and to formally developed, reviewed and published expert consensus opinions from credible organizations.
- *Exclusion criteria*: Searches might for example exclude websites that lack credibility and reliability, outdated textbooks, and unpublished manuscripts.
- *Search strategy*: Searches can include checking local reference collections, conducting computer-assisted searches opportunistically or systematically, or engaging a professional librarian. Search strategies can be opportunistic or exhaustive. Search strategies can include checking current textbooks, one-time literature searches using an

[16] Liberati A et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Int Med. 2009.

online citation database, or a step-wise methodology, depending on the nature of the question posed and availability of evidence.

5d. Assess: Evaluate the Evidence

The next step is to assess each piece of evidence individually for both findings and strength of evidence. A body of evidence might include several individual studies, one or more published literature reviews, and various forms of expert opinion. There are important quality considerations to consider for each.

5d1. Evaluating Individual Studies

Evaluating a body of health-related expert opinion and scientific evidence begins by reviewing each piece of evidence separately. This is not always necessary when sound alternative approaches are available, such when a good literature review is available, or when the expert is very familiar with the subject.

The goal is to evaluate the strength of evidence represented by the study. Reviewers evaluate the key elements of *relevance*, *study design*, and *quality*[17] and, for causality questions, *strength of association*[18].

Relevance:

Are the study's research questions relevant to the question posed to the expert? The review begins with determining the relevance of the study to the question posed. Sometimes a title or abstract might suggest that the study is relevant, but closer reading shows this is not the case.

How were outcomes and exposures measured? Were they relevant to the research questions? In many epidemiological studies, exposure to a factor of interest is assessed by self-report or proxy, rather than direct measures. Strong studies use direct, quantified measures of both outcomes and exposures that are relevant to the study's research questions.

Do the eligibility (inclusion and exclusion) criteria allow for generalization to the Veteran population of interest? The study would be less relevant to the question asked of the expert if it was done on a different population, so the findings would be less likely to apply to the question.

Study Design:

What study design was used? Understanding the quality of the study is different for each study design (section 4d). For the question of causation being asked, the design has important implications. The RCT design rarely can be used to establish causation, since it is unethical to deliberately expose subjects to suspected hazards. The case-control design is most useful for questions of causation if incidence is measured, and the exposure to the hypothesized factor is more frequent among cases than controls when other factors are held constant[19]. The cross-

[17] GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004 Jun 19;328:8 p.
[18] Kaldor J. Critical appraisal and causal inference. In: Proceedings of the 2008 Repatriation Medical Authority Forum, Canberra, Australia. 2008;45-56.
[19] Evans AS. Causation and disease: The Henle-Koch postulates revisited. Yale J Biol Med 1976; 49:175-195.

sectional design can be used to generate hypotheses for causality based on the criteria in Table 1 if the prevalence of disease is higher among those exposed to the hypothesized cause than in those not exposed.

Study Quality:

Was there good control for bias? Bias is systematic error introduced into study. For example, do the eligibility criteria introduce selection bias? There are many other types of bias to consider.

Was there good control for chance? Appropriate statistical methods control for chance. Statistics calculate the probability (*P* value) that the results observed by the study could have occurred by chance under the null hypothesis of "no difference". Statistical significance is often designated by $P < 0.05$ in many studies. Alternately, the precision of clinically relevant rates are designated by the *95% confidence interval* that estimates the 95% probability that the true value of the rate is contained within the interval's range.

Was there good control for confounding? Good studies take into account other potential factors that could explain the association between outcomes and exposure to a factor. Were measures for outcomes and exposures used that account for confounding? Studies that use modeling analysis to account for the influence of many variables at the same time are better able to control for confounding than studies that describe a list of variables one at a time.

Were the study's conclusions supported by the findings? Problems can arise if the authors use a study design that was limited for the conclusions they drew.

Strength of Association:

If a statistically significant association was detected, what was the strength of association between exposure and outcome? This is usually calculated as a statistic generated by regression modelling that controls for confounding variables. The two most common are *relative risk* and *odds ratio*[20].

Relative risk (RR) is the ratio of the risk of disease among the exposed to the risk among the unexposed. RR is calculated for RCT and cohort studies using incidence of disease, and calculated for cross-sectional studies using prevalence of disease. It is not calculated for case-control studies.

Odds ratio (OR) is the ratio of two odds, usually generated by a logistical regression model. Use of OR is most appropriate when calculated for case-control studies of a rare outcome, where the odds of exposure in cases compared to controls is an approximate estimate of the RR. For cohort and cross-sectional studies the OR is difficult to interpret, in part since it may be calculated using odds of exposure, disease or prevalence.

Risk Factor does not provide information on the strength of association, or the criteria considered for causality. Generally, risk factors based on incidence are more likely to be causal factors than those based on prevalence.

[20] Porta M. A Dictionary of Epidemiology, 5th Edition. Edited for the International Epidemiological Association. Oxford University Press, 2008.

5d2. Evaluating Multiple Studies

When several research studies are applicable to the question, the expert advisor pools the information from all of them. Most methods for evaluating the strength of a body of expert opinion and scientific evidence for questions of causality consider the *relevance* of the evidence to the question and the *quality, quantity and consistency* of the evidence[21,22,23]

Relevance refers to the degree to which the studies pertain to the question.

Quality of individual studies is evaluated as described in section 5d1.

Quantity refers to the number of research studies available. For example, multiple high quality studies add to the strength of evidence, while a few low quality studies may indicate insufficient evidence. There is no magic metric for evaluating quantity in making a judgement about causality and strength of evidence.

Consistency refers to the degree to which the findings or opinions in a body of evidence are similar. In the case of scientific evidence, consistency refers to the degree of conformity between the findings of studies conducted by different investigators under different circumstances. Say there are six papers applicable to a causality question, where the first step is to determine whether there is an association. Two report relative risks below 1 (exposure to the factor appears protective), one reports a relative risk of about 1 (exposure to the factor appears not associated with the outcome), and three reports relative risks above 1 (exposure to the factor appears hazardous). This suggests a degree of inconsistency in the evidence about association. However, the advisor might assign more weight to the finding of hazardous risk if the three studies finding a relative risk above 1 were of much higher quality than the studies finding no association or a protective association.

Quality, quantity and consistency all need to be considered together, not in isolation. Finding RR or OR exceeding 1 is not in itself sufficient evidence of causality. It is important to consider effect size (the degree to which 1 was exceeded), the quality of the studies that produced the measures (section 5d1), and other criteria of causality (Table 1).

The principles of evaluating the quality of literature reviews (section 5d3) apply to conducting reviews of multiple studies. Advisors rarely need to conduct comprehensive reviews that strictly follow those principles.

5d3. Evaluating Critical Reviews

There are principles for evaluating published critical reviews of a body of scientific evidence. High quality reviews use systematic methods to gather, weigh, analyse and synthesize scientific evidence, and make statements about strength of evidence. The U.S. Institute of Medicine (IOM) committees produce such reports on a variety of issues related to Veterans' health issues.

---

[21] GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004 Jun 19;328:8 p

[22] Agency for Healthcare Research and Quality. Systems to rate the strength of scientific evidence. US Department of Health and Human Services. Evidence Report/Technology Assessment Number 47.

[23] Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. Evidence-based toxicology: a comprehensive framework for causation. Hum Exp Toxicol. 2005 Apr;24(4):161-201.

There are two broad types of critical reviews: meta-analyses and literature reviews. Meta-analysis combines statistical data from several studies to perform a new analysis. Literature reviews evaluate a body of publications.

Literature Reviews

Like epidemiological studies, literature reviews range in the strength of evidence they represent. Reviews can be limited by the way they were conducted, even if there are strong research publications for the authors to review. Literature reviews that opportunistically gather studies and review them using a narrative methodology are more limited evidence than reviews that are conducted and analyzed using a more rigorous methodology.

These are the types of questions to ask in evaluating the strength of a literature review for questions of causality[24,25]:

1. What were the research questions? Were they relevant to the question posed to the expert and to the objectives of the literature reviewers?
2. What was the search strategy and how rigorous was it? Were inclusion and exclusion criteria specified? Was a computer search method used? Was the search exhaustive and systematic, or opportunistic? What supplementary search methods used?
3. Was a systematic approach used to evaluate studies and weigh and grade them? Were the principles of evaluating individual studies adhered to systematically? Were appropriate metrics used to quantify findings? Were the findings presented in a systematic and transparent manner?
4. Was a systematic approach used to synthesize the body of evidence? Were alternative conclusions considered?
5. Was the methodology replicable?
6. Were the conclusions supported by the findings?
7. Was a statement of strength of evidence and degree of certainty provided?
8. Was potential conflict of interest disclosed?

Meta-Analyses

In meta-analysis, researchers use statistical methods to pool findings from multiple similar studies. Individual studies might have small sample sizes, for example. While this might seem to be a compelling way to overcome limitations in small studies, for example when a particular health condition or exposure is rare, meta-analyses are subject to all the limitations of the individual studies, and to problems inherent in combining heterogeneous studies that have a variety of different problems controlling for chance, bias and confounding, and have varying relevance[26]. There are specialized techniques for evaluating the quality of meta-analyses[27].

---

[24] Agency for Healthcare Research and Quality. Systems to rate the strength of scientific evidence. US Department of Health and Human Services. Evidence Report/Technology Assessment Number 47.
[25] Mullen PD, Ramírez G. The promise and pitfalls of systematic reviews. Annu Rev Public Health. 2006;27:81-102.
[26] Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. Evidence-based toxicology: a comprehensive framework for causation. Hum Exp Toxicol. 2005 Apr;24(4):161-201.
[27] The Cochrane Collaboration. Systematic reviews: CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination, Published by CRD, University of York, 2009;294 p. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6. Updated September 2006;257 p.

5d4. Evaluating Expert Opinion

In questions of causality, exposure and latency, an expert sometimes has to weigh the opinions of other experts. For example, expert committees and textbook authors typically use qualitative judgement categories to communicate their evaluation of the strength of evidence for questions of association and causality. Risk factors listed in textbooks are qualitative judgements about whether a factor might cause a disorder. An individual expert like a client's health care practitioner might submit an opinion about causality, exposure or latency. While another expert can independently evaluate the scientific evidence considered by such experts, he or she cannot know exactly how they arrived at these judgements.

The weight that can be assigned to expert opinion varies considerably. The opinion of a client's physician might be based only on personal experience, at best citing only one or two "cherry-picked" references supporting the opinion. This opinion would have much lower weight than the consensus opinion of a formally convened panel of independent experts who used standard procedures to acquire, analyze and synthesize a body of evidence.

Judgements about expert opinion consider *relevance, credibility, reasonableness* and *reliability*[28].

- *Relevance* refers to whether the opinion answers the question posed and applies to the person's claim, or the population of interest.
- *Credibility* refers to the believability and plausibility of the expert's *opinion*, not the *person*. Credibility is judged for example by considering whether the opinion fits with other proven facts, and by assessing the scientific and other evidence considered by the expert.
- *Reasonableness* refers to the quality of being rational and having sound thinking and judgement.
- *Reliability* refers to the quality of being reliable, meaning trustworthiness and dependability. Reliability is judged by:
    a. Credibility and reasonableness;
    b. Whether evidence was given in a setting allowing questioning of the expert, like a hearing or a peer-reviewed publication process; and
    c. Assessment of the expert's objectivity, potential conflicts of interest, and degree of authority, (expertise, qualifications, special skill and knowledge).

5e. Adapt: Synthesize the Evidence and Communicate Opinion

In this stage, the expert adapts the evidence review to the decision maker's context, and communicates their opinion in language the decision-maker can use.

Synthesis is the process of distilling the analysis to draw a conclusion from available expert opinion and scientific evidence in response to the question, and to make a subjective judgement about the weight of evidence and degree of certainty[29,30].

[28]Toombs. Legislative Framework, Adjudication: Disability pension/award program. VAC Legal Services, Charlottetown. 08 May 2007;12 p.
[29] Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. Evidence-based toxicology: a comprehensive framework for causation. Hum Exp Toxicol. 2005 Apr;24(4):161-201.

The first step is to determine whether there is a statistical association between a factor and a health outcome, using the principles described in section 5d. The second step, depending on the question posed by the decision-maker, is to determine whether there is sufficient evidence for causality (Table 1, section 3b), or sufficient degree of exposure, or appropriate latency.

5e1. Subjective Judgements about Strength of Evidence and Degree of Certainty

Synthesis includes communicating degree of certainty about the conclusion in a manner that can be understood by VAC decision-makers, either for case-by-case claims adjudication, or for development of policies and programs. When dealing with questions of causality, the expert makes a *subjective judgement* about degree of certainty based on accumulation of criteria for causality.

At VAC, the subjective judgement categories shown in Table 2 are used to characterize the strength of health-related expert opinion and scientific evidence and certainty of conclusions about causality between a factor and a health condition[31,32]. These categories allow the author to express his conclusions and opinion in a way that makes sense to VAC decision-makers.

My opinion is not binding on the decision-maker.

Table 2. Categories of strength of evidence and degrees of certainty.

1. *More probable than not or greater that causality exists.*
   Health-related expert opinion and scientific evidence supports causality with a degree of certainty of more probable than not or greater.

2. *At least as likely as not that causality exists.*
   On balance, health-related expert opinion and scientific evidence is equally for and against causality and it cannot be determined which is stronger.

3. *Insufficient to support causality.*
   Health-related expert opinion and scientific evidence is not sufficient to conclude that causality exists without speculating; possible but not probable.

4. *More probable than not that causality does not exist.*
   Health-related expert opinion and scientific evidence supports the lack of causality with a degree of certainty of more probable than not or greater.

(Source: VAC Policy "Assessing and Categorizing Health-Related Expert Opinion and Scientific Evidence" 2012)

5e2. Background to the Four Categories

Principles for evaluating the strength and use of scientific evidence in decision-making have been evolving for more than 100 years. Since the 1950s, experts have used a variety of

[30] Samet JM, Bodurow CC. Improving the presumptive disability decision-making process for Veterans. Committee on evaluation of the presumptive disability decision-making process for Veterans. Institute of Medicine. 2007 Aug;789p.
[31] VAC Policy "*Assessing and Categorizing Health-Related Expert Opinion and Scientific Evidence"* 2012.
[32] VAC Policy "*Hazardous Material and Radiation Exposure*" 2012.

subjective categories for conveying strength of evidence and degree of certainty for questions of both causality and the efficacy and safety of interventions[33].

United States. In the U.S., the 1994 U.S. Institute of Medicine (IOM) "Veterans and Agent Orange" Committee used four categories of strength of evidence for association. Since then, more thinking has clarified that the question is about causation, not mere association. Subsequent IOM Committees have increasingly considered causality. The 2004 U.S. Surgeon General's report on smoking and 2006 IOM Committee on asbestos used four categories describing strength of evidence for causal relationships, not association. The 2007 IOM Committee on Veterans' presumptive disability benefit entitlement decision-making conducted a comprehensive review and concluded by consensus that four categories should be used to convey judgements about strength of evidence and degree of certainty for causation in questions of Veterans' disability questions. The Committee recommended that "equipoise" (at least as likely as not) be used as the threshold of evidence to infer causality and resolve reasonable doubt in favour of an applicant for this purpose.

Australia. In Australia, the Repatriation Medical Authority (RMA) establishes Statements of Principle that bind decision-makers. The Australian RMA uses two standards for two different types of Statements of Principles used by DVA entitlement adjudicators. Their "reasonable hypothesis" SOPs are used for operational and hazardous duty and have a lower threshold than the SOPs used for other types of service, which are based explicitly on balance of probability ("more probable than not")[34]. Their "reasonable hypothesis" legal standard is approximately equivalent to our "at least as likely as not". The legal meaning of "hypothesis" is not synonymous with the scientific meaning, and clarification of "reasonable hypothesis" occurred over several cases heard in Australia's federal courts. The test for reasonable hypothesis is that the evidence "indicates" that the hypothesis is true. "Indicate" is not the same as conclusive proof, allowing generous latitude in judgement. Under Australian law, to be "reasonable" in this instance means there must be something pointing to the hypothesis which appears to be true, using all the available evidence together.

Canada. In Canada, the categories shown in Table 2 correlate with legal standards of evidence that range from higher to lower degrees of certainty:

- Criminal law: beyond reasonable doubt.
- Civil law: more probable than not.
- Lower threshold: at least as likely as not.

Canadian Veterans' legislation allows finding in favour of an applicant when it is *at least as likely as not* that a service activity caused the health condition after considering all the evidence together. Canada does not have two evidentiary thresholds, instead recognizing hazardous duty by allowing for health conditions to have arisen in special duty service without requiring they be caused by service activities (the insurance principle versus the compensation principle).

[33] Samet JM, Bodurow CC. Improving the presumptive disability decision-making process for Veterans. Committee on evaluation of the presumptive disability decision-making process for Veterans. Institute of Medicine. 2007 Aug;789p

[34] Ward JD, Donal KJ. Statements of Principles: evidence-based compensation for Australian Veterans and serving defence personnel. ADF Health. 2004;5:89-93.

5f. Apply: Make a Decision

A VAC decision-maker adjudicates a claim for entitlement, not me. My opinion contributes to the evidence that may be considered by a decision-maker in a given client's case or when formulating policy. My opinion does not decide questions; rather my opinion is only one of the pieces of evidence considered by a VAC decision-maker.