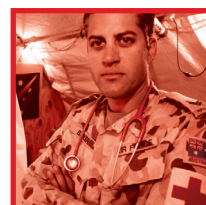




A Proposed Hierarchy of Evidence for Defence

**Lieutenant Colonel Phillip J. Hoglin,
Australian Army**

July 2013



A Proposed Hierarchy of Evidence for Defence

Lieutenant Colonel Phillip Hoglin, Australian Army

Not all evidence is created equal.¹

Introduction

Throughout any given year, Defence will develop, review and implement numerous policies that will have a direct impact on its personnel. Many are broad reaching and will affect every serving member, whereas others will only have an impact on a small section of the workforce. Regardless, it would be reasonable to assume that every new or revised policy has a strong and robust foundation in evidence. Unfortunately, this is not always the case and whether it is through the unavailability of evidence or the specific nature of the decisions that need to be made, the ADF is regularly at risk of formulating policy based on evidence that is less than ideal.

Distinguishing between good and bad—or relevant versus irrelevant—evidence is not always straightforward. For example, glossy and well-produced reports can mask poor research. In contrast, good research can be compromised by poor presentation, complex language or a lack of accessibility. Superficial barriers such as these are only the beginning of the problems that await policy makers as they attempt to sort through a spectrum of research documents related to their policy area. Although rarely acknowledged, sorting through the available evidence is a core task of policy makers, a skill that is not widely developed or even instilled outside our academic institutions.

To further compound the problem, it is not only policy makers who are confronted with the challenges of evidence-based decision making. Leaders in general, and military leaders in particular, are also vulnerable to overlooking, selecting the wrong, or placing unreasonable weight on one piece of evidence over another. This may be a consequence of the conditioning of military leaders to make relatively rapid tactical decisions in the absence of information. Although appropriate on the battlefield, the utility of the tactical decision-making process does not extend particularly well to the strategic policy-making realm and should be disregarded when individuals undertake these appointments. In its place, an evidence-based decision process that leverages from the available time and literature should be used, of which a critical component is the skill of identifying relevant evidence.

In this article, some guidelines are proposed to help personnel policy makers distinguish between evidence that is robust and able to support policy decisions and that which might be useful but should be treated with a healthy level of scepticism. After providing some context, discussion commences with a comparison and then a generic consolidation of existing hierarchies of research evidence being used in other disciplines in Australia, the US and the UK. The literature is then used as a basis for a proposed evidence hierarchy for use in Defence personnel policy development. Although the proposed hierarchy is intended for personnel policy makers, its scope is potentially broader and could be extended to other areas of strategic policy-making in Defence.

What is evidence?

The general sentiment of what is and is not evidence is broadly understood and is commonly associated with the notion of 'proof'. However, there is a subtle difference between proof and evidence, as the former usually refers to an undeniable and categorical truth, whereas the latter can provide for the construct of proof but is not necessarily proof in itself. The very definition of evidence is fertile ground for philosophers, particularly when the meaning varies among quantitative and qualitative research disciplines. Despite this, there is no need to complicate the definition of evidence here because it is unambiguously defined to mean 'the substantiation or confirmation that is needed in order to believe that something is true'.²

Why should we consider a hierarchy of evidence in policy making?

Not all evidence is created equal. Most policy makers would agree that it would simplify their research responsibilities if only one piece of evidence, based on an overwhelming consensus, was available, and that unreliable evidence was filtered before time and energy was consumed in reading it. Sadly, there is no pre-eminent research depository, which means that policy makers themselves are responsible for gathering all the evidence and considering each piece on its merits. But what are the merits of good evidence and how can policy makers identify them so that they can focus their efforts in the most effective way?

The current absence of guidelines about what evidence is appropriate or should be afforded the most weight leaves Defence's policy makers more exposed than necessary. There are obvious risks in an inability to isolate the best evidence, such as the development of policies that are inappropriate, ineffective, unnecessary or have negative unintended consequences. These risks can be reduced by the use of a hierarchy that allows a policy maker to assess where the evidence fits on the scale of reliability and quality, and the attention that should be afforded to it.

Existing non-military hierarchies

A hierarchy of evidence is not a new concept in the science and medical disciplines and it is no surprise that this is where it has been embraced the most. The consequence of poor use of evidence in medicine is obvious and potentially fatal; therefore, the attention afforded to a hierarchy is understandable. Why a hierarchy has not been adopted more broadly outside medicine and science could be symptomatic of a lack of classical science education, existence of political imperatives overriding the consideration of evidence, a social impression that evidence is not necessary for personnel policy development, and a general lack of policy decision culpability. It is a harsh reality that these reasons should not, but usually do, diminish the use of evidence by policy makers.

A useful start-point in developing a hierarchy for Defence personnel policy-making is to examine those already in use by other disciplines. Although the selected hierarchies shown in Table 1 use slightly different definitions, they are consistent in their themes, and for good reason explained later.³ For the purpose of consolidation, the hierarchy levels presented in Table 1 can be grouped into three simpler, non-scientific levels of evidence as follows.

- Level 1 - Evidence obtained from trials
- Level 2 - Evidence obtained from non-trial observations, and
- Level 3 - Evidence obtained from subject-matter experts.

Level 1 - Evidence obtained from trials

Some of the strongest evidence is obtained from a deliberate and planned trial of a policy, where some participants are exposed to the policy (the treatment group) and others are not (the control group).⁴ The method of randomly allocating people to either the treatment or control group helps ensure that each group is as similar as possible; therefore, any differences between the groups after the trial can be attributed to the policy. This method is referred to as a randomised controlled trial (RCT) and is universally considered to provide some of the best evidence.⁵

Evidence from RCTs can be further enhanced simply by the use of either multiple trials on different populations or, better still, the consolidation of the results from multiple studies in the same subject area. This latter method forms the basis of a class of research known as meta-analyses. Meta-analyses of RCTs are considered to be the strongest form of evidence, partly because they draw from a wider range of findings that have used different methodologies and/or different populations.

The key factor that differentiates RCTs from the other forms of evidence, which will be outlined soon, is that they are deliberately planned and designed to answer specific research questions from the outset. This also means that trials will typically have a dedicated research organisation tasked specifically to conduct research design, analysis and reporting, which usually results in directly relevant, robust and scientifically unbiased findings.

Level 2 - Evidence obtained from non-trial observations

The second level of evidence is usually obtained when RCTs are inappropriate or cannot feasibly be conducted. Evidence from non-trial observations of people and their behaviour can take many forms and comprises differing sub-levels of evidence in its own right. At one extreme, it is possible that a natural experiment can occur where control and treatment groups can be identified, thereby allowing analyses similar to those facilitated by RCTs. At the other extreme, observational data may consist simply of case studies without any control group.

Although natural experiments can provide evidence that rates almost as high as that obtained from RCTs and meta-analyses, in reality their occurrence is rare. This is because, in personnel research, it is difficult to identify a control group and treatment group which differ only in terms of their exposure to a policy. Fortunately, evidence obtained from other forms of non-trial observations is possible and these include research methods such as (in hierarchical order):⁶

- **Before-and-after studies.** Comparison of how people behave both before and after the implementation of a policy.
- **Time series.** Comparison of how the same people behave at different points in time where the introduction of a policy occurs at some point during the time series.
- **Cohort studies.** Comparison of how two cohorts behave where one cohort is subject to a policy and the other is not.
- **Case-control studies.** Comparison of people who exhibit certain behaviour with those whose behaviour is different to determine whether this might be predicted by exposure to a policy.
- **Case studies.** Simple description of people who exhibit certain behaviour without any comparison group.

Evidence obtained from these research methods, including that from natural experiments, have one common factor: they use data which is not necessarily designed to provide the answer to the research or policy maker's question. This is reflected in the hierarchies shown in Table 1, where there is a decreasing reliability of evidence obtained from each method, partly because of the diminishing likelihood that findings can be attributed to a policy rather than some other unobserved or random factor. In essence, the more confident policy makers are that findings are a result of the policy and not some other unobserved factor, the more confidence they can have in the evidence.

Level 3 - Evidence from subject-matter experts

Although Defence relies heavily on expert opinion, evidence from this source is the weakest, and in some hierarchies it is not even included as a source of reliable evidence. The National Health and Medical Research Council guidelines, for example, emphasise that 'no formal level of evidence is attached to expert opinion, ... expert working parties or anecdotal information'.⁷ It is accepted here, that in some cases, subject-matter expert opinion may be the only evidence available and should be included in the hierarchy. However, it should also be acknowledged that it is also the most likely to be weak, biased and unsubstantiated, and should not take precedence over higher levels of evidence where it is available.

Table 1. A selection of commonly used evidence hierarchies.*

| | Centre for Reviews and Dissemination (UK)⁸ | Scottish Intercollegiate Guidelines Network (UK)⁹ | National Health and Medical Research Council (AS)¹⁰ | US Preventative Services Task Force (US)¹¹ | Joanna Briggs Institute (AS)¹² |
|--|---|---|--|--|---|
| Level 1 - Evidence from trials | | 1++ High quality meta-analyses, systematic reviews of RCTs. | I. Systematic review of all relevant randomised controlled trials. | | Meta-analysis of experimental studies or one or more large experimental studies with narrow confidence intervals. |
| | Randomised controlled trials (RCTs). | 1+ Well conducted meta-analyses, systematic reviews of RCTs. | II At least one properly designed randomised controlled trial. | I. At least one properly randomized, controlled trial. | One or more smaller RCTs with wider confidence intervals or quasi-experimental studies. |
| | | 1- Meta-analyses, systematic reviews or RCTs. | | II-1 Controlled trials without randomisation. | |
| Level 2 - Evidence from non-trial observations | Quasi-experimental studies. <ul style="list-style-type: none"> • Non-randomised controlled studies. • Before-and-after study. • Interrupted time series. | 2++ High quality systematic reviews of case controlled or cohort studies. | III-1 Pseudo-randomised controlled trails. | | Cohort studies, case-controlled, observational studies. |
| | Observational Studies. <ul style="list-style-type: none"> • Cohort study. • Case-control study. • Case series. | 2+ Well conducted case control or cohort studies. | III-2 Comparative studies with concurrent controls and non-randomised allocation, cohort studies, case-control studies, or interrupted time series with a control group. | II-2 Cohort or case-controlled analytic studies, preferably from more than one centre or research group. | |
| | | 2- Case controlled or cohort studies. | III-3 Comparative studies with concurrent controls: historical control study, two or more single arm studies, interrupted time series without a parallel control | II-3 Multiple time series with or without the intervention. | |

| | | | | | |
|---------------------------------|--|---|--|---|-------------------------------|
| | | | group. | | |
| | | 3 Non-analytic studies, eg case reports, case series. | IV Evidence obtained from case series, either post-test or pre-test/post-test. | | |
| Level 3 - Evidence from experts | | 4 Expert opinion. | | III Opinions of respected authorities, based on clinical experience; descriptive studies and case reports; or reports of expert committees. | Expert opinion or conjecture. |

* Some descriptions have been abbreviated. Refer to the original source, as endnoted, for a complete explanation of each evidence level.

A hierarchy of evidence for Defence personnel policy-making

Application of a hierarchy to policy-making is not necessarily a new concept, although it is far less common than in the science disciplines. For example, in his paper for the Australian Treasury's 'Economic Roundup', Leigh proposed a possible evidence hierarchy for Australian policy makers,¹³ similar to that used by the UK Cabinet Office,¹⁴ as shown below.

Example 1 - Australian Treasury's 'Economic Roundup'

- Systematic reviews (meta-analyses) or multiple randomised trials
- High quality randomised trials
- Systematic reviews (meta-analyses) of natural experiments and before-after studies
- Natural experiments using techniques such as differences-in-differences, regression discontinuity, matching or multiple regression
- Before-after studies
- Expert opinion and theoretical conjecture

Example 2 - UK Cabinet Office

- Randomised controlled trial
- Quasi-experimental study
- Pre-post study

These two examples, combined with those from Table 1, can be extended for use by Defence. An enhancement to these hierarchies is also proposed by including two dimensions, one referring explicitly to the level of evidence in a list similar to those already mentioned, and a second referring to the quality of evidence.¹⁵ This second dimension is not strictly necessary outside the Department of Defence because the quality of research is usually implicit in the level of research. But it is added here to explain how the hierarchy can be further subdivided. The proposed hierarchy for the level of evidence dimension is in the left panel of Figure 1 and the quality of evidence dimension is in the right.

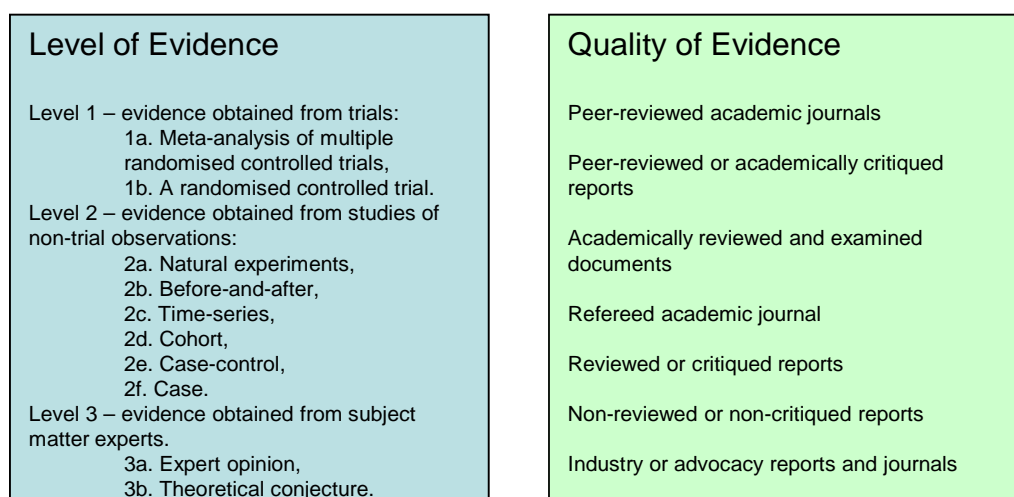


Figure 1. Proposed level and quality of evidence hierarchies for Defence

Dimension 1 – the level of evidence

The ADF has traditionally been hesitant to conduct randomised trials of personnel policy. One possible reason is that trials are not always appropriate and can sometimes be considered unethical or unfair,

particularly where one group might be exposed to a certain policy and the other is not, such as a trial of a particular condition-of-service.¹⁶ Other barriers may include a lack of awareness of the strength of evidence obtained through RCTs, under-resourcing of any specific organisation capable of conducting them and a lack of familiarity and expertise in their conduct. These barriers are not insurmountable and the Australian Defence Human Research Ethics Committee, for example, has been established to assist with ethical concerns, experimental design and the approval of trials involving people.¹⁷

In all likelihood, there are situations in Defence where randomised policy trials are appropriate and could be conducted if there was momentum to do so. For example, randomised trials could be used to determine the best policy for alcohol, smoking and drug awareness in Defence, participation in sport, effective physical rehabilitation methods, benefits of career education or even harassment awareness. However, there have been no RCTs in these or other policy areas. Consequently, the ADF remains broadly unaware of the effectiveness of any related personnel policy because of a lack of robust evidence.¹⁸

This lack of use of RCTs means that the second level of evidence, obtained through the use of non-trial observations, is often all that is available to Defence's policy makers. By chance, the ADF's existing data sources can sometimes provide the type of data necessary for research by supporting methods such as before-and-after studies. However, this is a sub-optimal and often accidental means of gathering data.

If options for RCTs and non-trial observations have been exhausted—and existing sources of data do not contribute to the research objectives—then the third level of evidence, or expert opinion, may be the only remaining source of evidence. Expert opinion is widely considered to be the lowest form of evidence¹⁹ and there is an obvious reason for this: opinions do not always have a basis in fact. However, there is some grey area when an opinion is offered by a subject-matter expert with a strong track record of robust research in a relevant field. In this case, the use of such an expert to represent a consensus derived from the available literature can be appropriate,²⁰ although this obviously does not absolve the policy maker of responsibility for the correct use and representation of the available evidence.

For Defence, when policy makers leverage from the professionalism and knowledge of its personnel, it is reasonable to expect that resulting opinions would have a sound basis in experience. However, there is no guarantee of the validity of an opinion or any practical means with which to test it. Furthermore, Defence subject-matter experts can be challenged by an alternative view from another subject-matter expert which, without the benefit of published research, can leave the policy maker with little to differentiate between two or more conflicting expert opinions.

Ideally, policymakers would have considered the need for evidence, likely sources of data, and appropriate assessment methods well before a policy was considered for implementation.²¹ This would allow for a formal analysis and reporting regime that would also provide evidence for future policy initiatives. Unfortunately, despite attempts such as the introduction of the Workforce Initiative Assessment Framework in 2011, the ADF has not yet overcome this evidence-gathering problem/barrier and, as a result, current policies are often not well reported and future policies are not developed with a sound evidence base.

Dimension 2 – the quality of evidence

As stated near the beginning of this article, not all evidence is created equal, and there is a simple method policy makers can use to identify good quality research by using the means through which the evidence was presented or published.²² The means of publication can be used as a proxy for its quality because somebody else, for example a reviewer, examiner or editor, has already assessed the quality on the policy maker's behalf.²³ Of course, this is not absolute, and good research exists outside peer reviewed journals—and poor research can sometimes slip through different layers of vetting. Nevertheless, the proposed quality of evidence dimension to a hierarchy can be a useful first-cut approach to assessing evidence quality.

Policy makers should remain mindful that the quality dimension is still no substitute for formal assessment of the quality of evidence; it serves only to provide an overlay to the level of research

dimension in its absence. The use of this dimension is relatively straightforward. At the extremes, evidence derived from peer-reviewed academic journals should be considered of high quality, whereas it should be assumed that evidence from industry journals or magazines would have an inherent bias and would be far more likely to misrepresent and/or misreport important information, and is therefore of lower quality. Other publications and sources fit in between these two extremes on the quality scale. Defence policy makers should always take a moment to assess the source and quality of the evidence and grade it accordingly or, if they are unable, seek appropriate guidance.²⁴ The potential consequence of not doing this is obvious: an acceptance of evidence as being stronger than it would otherwise deserve.

It is an unfortunate reality that for many of Defence's policy decisions, there is little in the way of evidence available through existing journal articles or peer-reviewed reports, simply due to the specialised nature of many of the policy decisions that must be made. Consequently, Defence has often been required to find much of the necessary evidence itself through the conduct of its own research. Increasingly, the ADF has also needed to commission external consultants to conduct its research, primarily due to limited internal capability and capacity but also due to the need for perceived impartiality.

However, commissioned research should not necessarily be granted a high status of report integrity, as an assumption that it will produce unbiased, credible and robust evidence able to withstand academic critique is, at best, misguided.²⁵ Although commissioned research could potentially fit high on the quality scale, in order to do so it should first be subjected to peer and/or academic review, otherwise its findings should be treated with a level of scepticism.

The quality of evidence dimension may be suitable in the short term, however, in the longer term, Defence should look toward establishing a more robust and professional approach to its assessment of research quality. This follows from the rationale that if Defence is unable to maintain the capacity and capability to conduct personnel research itself then it must, as a minimum, possess the expertise and systems necessary to academically review the reports submitted to, or obtained by, it.

Although it is outside the scope of this article, there is a strong justification for an internal centre of excellence to undertake this role. This could be established relatively easily through the consolidation of Defence's existing research expertise and/or partnerships with academic institutions. It is likely that academics and informal partnerships already exist and coordination of these resources might not be a cumbersome activity.

Using the two dimensions together

When the two dimensions are combined, the policy maker can make an overall assessment of the evidence. Recall that the rationale behind the hierarchy of evidence levels (Dimension 1) is that it reflects the confidence a policy maker can have in attributing a certain outcome to the policy being examined and not some other unobserved influence. Similarly, the hierarchy of evidence quality (Dimension 2) reflects the risk of bias or poor methodology in the report.²⁶

Using these dimensions, the interactions proposed in Figure 2 suggest that a meta-analysis of RCTs published in a peer-reviewed academic journal is the best evidence possible. In contrast, theoretical conjecture published in an industry report is akin to an opinion column in a newspaper, and evidence from such a source should be afforded little or no weight. Figure 2 shows other interactions between the level and quality of evidence.

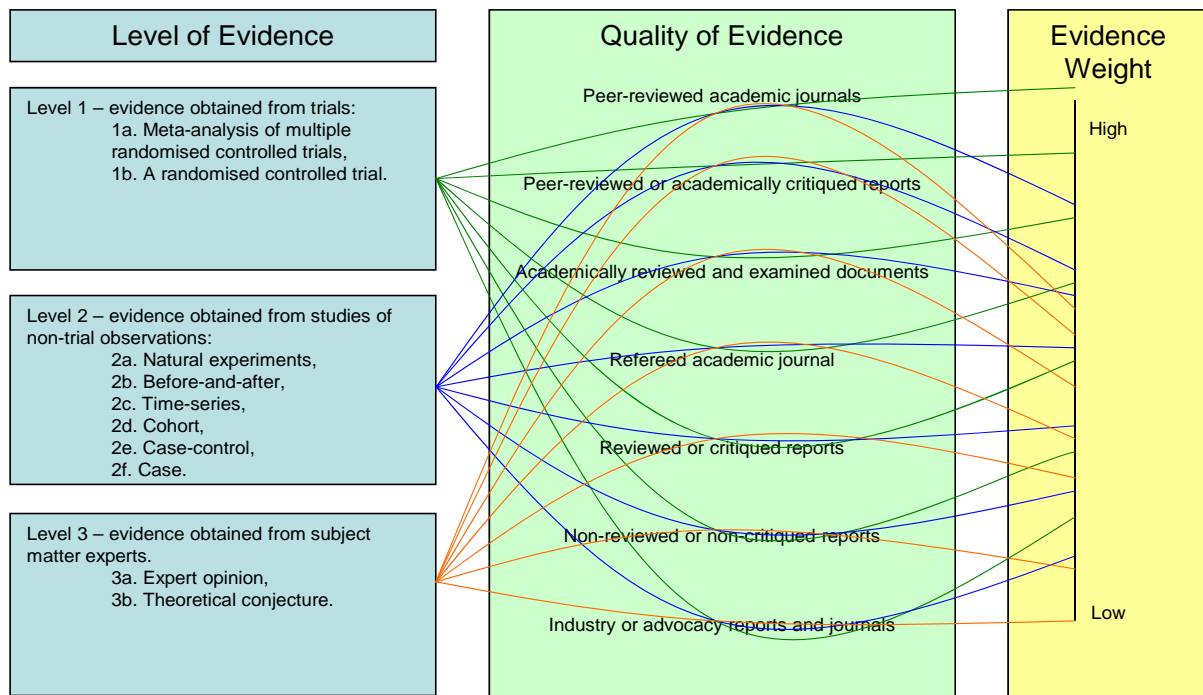


Figure 2. Interactions between the level and quality of evidence dimensions

Conclusion

The ability of the ADF to make sound policy decisions depends on its ability to identify and use the best possible evidence. For some policy proposals, there may be a large volume of evidence through which the Defence policy maker will need to sort in order to develop the best policy based on the best evidence. For other proposals, there may be very little existing evidence, in which case there is a need for the policy maker to deliberately source the evidence themselves using whatever resources are available. Regardless of whether evidence already exists or whether it needs to be developed, an understanding of what type of evidence is stronger, and therefore more able to support the policy decision, is important for Defence.

In this article, an evidence hierarchy based on two dimensions relating to the level and quality of evidence has been proposed. If adopted, the ADF will have an improved basis on which to make policy decisions using the available evidence. This will not only minimise the risk of poor policy development but will also assist in identifying gaps in knowledge and the subsequent need for additional research. Whether the hierarchy is adopted or not, several stark deficiencies need to be addressed: the capability and capacity to conduct Defence personnel research and the expertise with which to assess the level and quality of evidence.

Lieutenant Colonel Phillip Hoglin has been the senior Workforce Analyst for Army since January 2012. He graduated from RMC in 1994, having completed a Bachelor of Science (Honours) majoring in statistics. He has also completed a Master of Science in Management through the US Naval Postgraduate School and a Master of Philosophy in Statistics through the University of New South Wales. He has been involved in workforce analysis since 2004.

Acknowledgements

The author would like to acknowledge Dr Corinne Manning, Army Headquarters, and Angela Sturrock, Defence People Group, for their advice and guidance in preparing this article.

NOTES

- ¹ This is a play on words, linking to the phrase 'All men are created equal', attributed to Thomas Jefferson, as used in the US Declaration of Independence ([1776] and later stylised by Benjamin Franklin (and others): see <http://en.wikipedia.org/wiki/All_men_are_created_equal> accessed 10 May 2013 and <<http://www.archives.gov/exhibits/charters/declaration.html>> accessed 16 May 2013. The phrase was also stylised by G. Orwell, *Animal Farm*, Penguin: London, 1946 in the phrase 'All animals are created equal but some animals are more equal than others'.
- ² See Joanna Briggs Institute, *2011 Reviewers' Manual*, Joanna Briggs Institute: The University of Adelaide, Adelaide, 2011. Interested readers may also wish to follow the discussion offered in S. Miller and M. Fredericks, 'The Nature of "Evidence" in Qualitative Research Methods', *International Journal of Qualitative Methods*, Vol. 2, No. 1, 2003, p. 45.
- ³ There are numerous evidence hierarchies in use. In this article, a very small group which are representative of the academic literature on the topic has been selected. Refer to Table 1 for more detail.
- ⁴ The term 'policy' is used here; however, the literature may also refer to treatments, interventions etc.
- ⁵ For brevity, the description of Randomised Controlled Trials has been simplified. For a full explanation, see National Health and Medical Research Council, *How to Review the Evidence: systematic identification and review of the scientific literature*, National Health and Medical Research Council: Canberra, 1999.
- ⁶ Although before-and-after, time series and cohort studies all rate higher than case studies, there is no clear consensus on the hierarchical order of evidence obtained from these three types.
- ⁷ National Health and Medical Research Council, *A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines*, National Health and Medical Research Council: Canberra, 1998, p. 15.
- ⁸ Centre for Reviews and Dissemination, *Systematic Reviews: CRD's guidance for undertaking reviews in health care*, Centre for Reviews and Dissemination: University of York, Laverthorpe, York, 2009, p. 11.
- ⁹ Scottish Intercollegiate Guidelines Network, *A Guideline Developer's Handbook*, Scottish Intercollegiate Guidelines Network: Edinburgh, 2011. Also in A. Weightman, S. Ellis, A. Cullum, L. Sander and R. Turley, *Grading Evidence and Recommendations for Public Health Interventions: developing and piloting a framework*, Health Development Agency: London, 2005, p. 51.
- ¹⁰ National Health and Medical Research Council, *A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines*, p. 56. Also in T. Merlin, A. Weston and R. Tooher, 'Extending an Evidence Hierarchy to Include Topics Other Than Treatment: Revising the Australian "Levels of Evidence"', *BMC Medical Research Methodology*, Vol. 9, No. 34, 2009, p. 8.
- ¹¹ US Preventative Services Task Force, *US Preventative Services Task Force Procedure Manual*, Department of Health and Human Services, Agency for Healthcare Research and Quality: Rockville, Maryland, 2008, p. 36.
- ¹² Joanna Briggs Institute, *2011 Reviewers' Manual*, Joanna Briggs Institute, p. 150.
- ¹³ A. Leigh, 'What Evidence Should Social Policymakers Use?', in *Economic Roundup*, Australian Treasury: Canberra, 2009, p. 35.
- ¹⁴ UK Cabinet Office, *Think Research*, Social Exclusion Task Force: London, 2008, p. 14.
- ¹⁵ Four 'dimensions' are mentioned in National Health and Medical Research Council's *A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines*. These include level, quality, relevance and strength. Only the dimension of level and a variation of the quality dimension are addressed in this article.
- ¹⁶ It is also suggested there may be 'a perception of randomised policy trials as being unethical' in Leigh, 'What Evidence Should Social Policymakers Use?', p. 33.

-
- 17 For further details on the Australian Defence Human Research Ethics Committee see: <<http://intranet.defence.gov.au/vcdf/sites/Research/ComWeb.asp?page=38869>> accessed on 26 February 2013.
- 18 This encourages a 'try it and see' approach rather than an evidence based approach to policy, or a willingness to accept lesser forms of evidence when better evidence could be obtained with foresight. See also The Economist, Science and Technology: 'Try it and see; Social science', *The Economist*, Issue 362, 2002, p. 8262.
- 19 Leigh, 'What Evidence Should Social Policymakers Use?', p.34, states that 'Lowest in the medical hierarchy (and not even rating a mention in the UK Cabinet Office's hierarchy) are expert opinions and descriptive case studies'. The National Health and Medical Research Council's *A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines*, p. 15, as cited in the text, states '[n]o formal level of evidence is attached to expert opinion, the findings of expert working parties, or anecdotal information'. More pragmatically, the Joanna Briggs Institute's *2011 Reviewers' Manual*, p. 107, states that '[w]hile rightly claimed not to be a product of "good" science, expert opinion is empirically derived and mediated through the cognitive processes of practitioners who have been typically trained in scientific method'. Although in Defence this is not necessarily a valid assessment of its subject-matter experts.
- 20 It can be argued that if an opinion is formed from consolidation of literature, then this is not strictly an opinion but a representation of available literature and would therefore be placed higher on the hierarchy of evidence by virtue of the literature from the 'opinion' was obtained.
- 21 The Workforce Initiative Assessment Framework (WIAF), developed by Workforce Planning Branch in Defence People Group, was an attempt to overcome this deficiency but, to date, it has not been widely adopted by the Services. Further details on the WIAF can be obtained through the Directorate of Workforce Intelligence.
- 22 Note that this paper applies a different definition to quality than the National Health and Medical Research Council which defines the 'quality of evidence' as referring to methods used by investigators to minimise bias in study design and in the conduct of a study. See National Health and Medical Research Council, *A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines*, p. 55. The Centre for Reviews and Dissemination explains that '[q]uality is a complex concept and the term is used in different ways' but that it is likely to consider the 'appropriateness of study design to the research objective, risk of bias, other issues related to study quality'. See Centre for Reviews and Dissemination, *Systematic Reviews*, p. 33.
- 23 An assessment of evidence differs from expert opinion. Although assessments are made by experts, the assessment itself is not 'opinion' *per se* but based on a defined set of assessment criteria. Reviewers' manuals, such as the Joanna Briggs Institute, *2011 Reviewers' Manual*, and CRD's, *Guidance for Undertaking Reviews in Health Care*, detail the requirements for assessment of evidence.
- 24 This assessment should consider the report in the context of its suitability for publication or likelihood of being accepted by a journal should it be submitted. Pragmatically, if research was professionally robust and could withstand the assessment process employed in academic and peer review, then it would be more likely to be published in an academic journal, in which case it would rise up the hierarchy. The ADF does not currently have formal guidelines for the assessment of evidence.
- 25 Although there is no shortage of recent examples of poor research commissioned by Defence, it would be unfair to list them here without a right of reply. Further information and examples can be obtained through the author. Commissioned research is not currently subject to any form of formal review or research.
- 26 Organisational/personal bias is differentiated from statistical bias in this article. The former is related to a human-factors component of research, whereas the latter can be created as a result of unavoidable deficiencies in the data or methodology.



<http://www.defence.gov.au/adc/publications/publications.html>

