Considering evidence: The approach taken by the Hazardous Substances Advisory Committee in the UK

August 2015

1. Introduction

The Hazardous Substances Advisory Committee (HSAC) provides expert advice to UK Ministers, officials and other relevant bodies on the protection of the environment, and human health via the environment, from potentially hazardous substances and articles[1]. The Committee's membership is multi-disciplinary and independent, enabling it to approach the evidence from a range of different perspectives. Its Code of Practice requires members to observe the highest standards of impartiality, integrity and objectivity in relation to the advice they provide; the Code also includes clear provisions for handling conflicts of interest.

Hazardous substances are often the subject of controversy, on which individuals, and different groups in society, hold divergent views. In formulating its advice, the Committee needs to analyse, interpret and assess the available evidence, often in situations where the uncertainties may be considerable. This paper documents the different kinds of evidence that might be available to the Committee; the criteria that HSAC adopts in its assessments; and the wider perspectives and concerns that have a bearing on the issues at hand. It also proposes a process through which the Committee's judgements about the quality of the available evidence could be communicated in an accessible form.

2. Types of evidence

HSAC recognises that evidence varies in its source, robustness and defensibility, and that these factors will influence the degree of confidence that assessors can assign to any given 'piece' of evidence, or to a body of evidence as a whole. While most of the scientific evidence assessed by HSAC derives from experimental or epidemiological studies, or is based on modelling of some kind, observational and anecdotal evidence may also be considered (Table 1). Evidence in the last two categories sometimes provides a first indication that a phenomenon is worthy of further investigation, and can lead to more systematic studies. It is likely that the availability of less systematic evidence will increase with the evolution of social media. Statistical evidence is often grounded on hypotheses which have been tested to a certain degree. However, it takes time and resources to collect statistically robust data, so that such studies may not reflect rapidly changing circumstances and emerging problems.

[1] The HSAC Terms of Reference and its Code of Practice can be downloaded from https://www.gov.uk/government/groups/hazardous-substances-advisory-committee

Table 1. Categories of evidence

	Type of evidence				
	Experimental	**Model-based**	**Epidemiological**	**Observational**	**Anecdotal**
	Obtained through a methodological approach to experimental design and data collection.	*Computer modelling of effects or exposures to provide a measurement of impact.*	*Data based on studies of populations under real-world conditions.*	*Based on observations and experience.*	*Based on personal accounts of effects.*
	Possible to show causality/association	Infers causality	Infers association	Infers association	Hypothetical association - potentially identifies issues of concern, not yet addressed in scientific research
↑ Increasing quality	Repeated experiments with a high degree of replication and controls following internationally accepted standards (e.g. OECD Test Guidance Documents)	Approach informed by empirical evidence, all processes and parameters revealed to allow repetition by others	Follows published guidance (e.g. WHO) with clear methods and rationale for data inclusion or exclusion	Field observations made in a systematic way, but without a specific experimental design	Relatively high incidence of specific effects; consistency between unconnected accounts; different accounts carefully collated.
	Not meeting widely-accepted experimental protocols; un-tested method, poorly reported	Model without antecedents, parameters from assumptions not measurements, processes a black box, i.e. cannot be repeated by others	Un-tested method, badly reported, using non-standard measurements of impact	Circumstantial evidence random or 'one off' events or phenomena	Uncorroborated, unconfirmed anecdotes: 'a friend of a friend…'

There can be significant variations of quality within each type of evidence. Examples (not exhaustive) are given of what might be considered 'high' or 'low' quality evidence within each column; in practice, there will be a gradation. No simple (horizontal) quality continuum between different types of evidence is implied; see sections 3 and 4 below.

3. Judging quality: considerations to take into account

In reviewing the scientific evidence, HSAC considers the extent to which any given study meets the following, widely-accepted criteria. HSAC may attach particular weight to evidence that conforms to these criteria, though 'weaker' evidence (in these terms) should not be dismissed: it can be part of the bigger picture when different sources of evidence are combined.

- **Transparency of aims**. A study should have a clearly stated purpose, in terms of the problem to which it relates and the research questions to be addressed. Conventionally, this is achieved through the statement of a hypothesis. The hypothesis to be tested should preferably link to previous work, and the study should be clear about the ways in which it builds upon, or challenges, the evidence base. The nature of HSAC's work is such that the Committee is often focusing on substances that have not been subject to exhaustive scientific studies (nanomaterials would be one example). In this case the hypothesis may be that a suspected causal agent is responsible for harm and it is important to recognise that this is essentially an arbitrary formulation. In assessing the stated hypothesis, it has to be clearly structured so that it is properly testable and falsifiable. HSAC recognises that findings based on statistical evidence are conditional on the structure of the hypotheses, and also on a potentially arbitrary decision about significance levels (e.g. a 10% or 5% probability of Type 1 error – i.e. incorrectly rejecting a true null hypothesis) and confidence intervals selected by the researcher (e.g. a 90 or 95% probability of the true value lying within the interval).

- **Methodology and results.** For experimental evidence, the Bradford Hill features or characteristics of causal associations [1] provide an excellent starting point for investigating causality. These include: *temporality, strength of the association, consistency of the observations, biological plausibility of the effect and evidence for recovery following diminution of the agent suspected of causing stress.*

Within each experimental study HSAC would also have regard to the following:

Methodology. For given data sources the methods used should have a sound scientific basis and should be fully described, capable of repetition and appropriate to the aims of the study. The reproducibility of the method should be tested by statistical examination of the replicates where the variability should ideally be low. The risks of bias in data collection should have been considered in the study design, and the efforts made to minimise any recognisable bias should be declared. There should be evidence of sound laboratory procedure, such as the use of controls and analytical blanks.

Results and interpretation. Results should be presented in a transparent way and should have appropriate statistical validity and power (for example, the data set should be of a suitable size, and appropriate confidence intervals and significance levels should be used). The caveat to any study is '*under the experimental conditions described'*, so it is important that the conditions are relevant to the problem under investigation. The interpretation of the data should consider potential sources of error in the study, and the extent to which these affect the degree of uncertainty assigned to the findings and conclusions. The *null and alternative hypotheses* should be carefully constructed so that the study gives robust findings, allowing researchers to be confident about their result.

- **Completeness**. A study should be sufficiently complete to enable third parties to review it and arrive at an independent interpretation of its findings, which may or may not coincide with that of the original authors. As already noted, the authors should themselves attempt to identify uncertainties and weaknesses in a given study, though it may not be easy (or even possible) to be comprehensive in this respect (see Sections 4 and 5).

- **Independent review.** The source of a study, and the likelihood of bias, are important considerations when assessing the quality of evidence (see also Section 5). Peer review (that is, review of a study by those regarded as having expertise in the field) is critical in this respect, even if it is an imperfect process. Greater confidence is also gained as other independent scientists replicate the original findings.

- **Accessibility**. Studies should be published or available in archival form in the public domain, so that the evidence can be readily examined. Even if they are not freely available, the costs of access to cited studies should not be prohibitive. Ideally, the raw data on which the study is based should also be available in a comprehensible form, so that its use can be assessed by others.

A number of formalised approaches, including, for example, the 'Klimisch criteria' and Harris *et al.* [2, 3], suggest further attributes of what might be considered 'high quality evidence', and these can be incorporated into a 'weight of evidence' approach [4] and systematic review. HSAC's remit means that it may be called upon to assess and report on a wide range of areas and potential hazards, and the specific approach used needs to be selected on a case-by-case basis. HSAC seeks to ensure that its recommendations are fully and transparently described.

It is important to be aware of the limitations of criteria such as those described above. Even studies that 'tick all the boxes' may, for example, be the subject of unconscious bias or 'groupthink'; research questions and design may be influenced by experimenters' prior beliefs, or driven by particular fashions or pressures to publish. Researchers may form their conclusions relatively quickly, using heuristics or 'rules of thumb' (quick decision-making devices that can be useful but can also lead to biased assessments of evidence) and when these approaches are applied they may or may not have a reasonable empirical basis. Further, it is in the nature of certain forms of bias that they seem normal and unbiased to those who hold them. HSAC considers it important, therefore, always to reflect on research questions and assumptions when considering scientific evidence and to ask, for example, 'how has the study been framed?', and 'what might have been missed?' One means of addressing such concerns is to bring people of different views together to define any study before it is commenced, in a form of adversarial collaboration.

4. The wider context

Criteria for assessing the quality of evidence can be surprisingly difficult to distinguish from those that individuals and groups deploy, consciously or not, in judging how much credence to attach to evidence in a particular case. The difficulty lies in identifying a benchmark from which to determine whether evidence is 'sound', in some wholly impartial, objective sense. In assessing the available evidence, and offering an opinion, HSAC is conscious that the science alone, while of fundamental importance, is unlikely to settle issues of deep controversy. Rather, those presented with new evidence (scientists and advisors, as well as pressure groups, publics and decision makers) tend to be influenced by:

- **Trust in the source of the evidence.** Important considerations are whether the evidence comes from an individual or institution seen to have authority from the recipient's perspective and whether the source has a known, explicit or inferred bias. Obvious interests in the issue at hand can reduce the trustworthiness of evidence ('they would say that wouldn't they?'). Conversely, evidence that takes an unpredictable or unexpected line can sometimes be persuasive.

- **Defensibility.** Recipients of evidence form views about whether it has been arrived at in a defensible way, and there is clearly overlap here with the standard, scientific criteria for 'good' evidence (Section 3). The wider point is that in matters of controversy, the quality of the evidence is itself likely to be a matter for dispute, because this becomes part of the process of questioning unwelcome findings. Assumptions, judgements and biases (which are always, and necessarily, present, sometimes hidden behind claims about 'objectivity') are likely to be exposed and questioned.

- **Conformity to the recipient's 'worldview'.** Worldviews may include beliefs about nature (for example, whether natural systems are fragile or robust) and positive/negative feelings about particular 'risky' activities; they may be shared within groups, communities and cultures. Even for the most 'objective' of recipients, evidence is likely to be filtered through a worldview.

- **Framing.** Evidence may be more persuasive if it relates to a meaningful framing of the problem from the recipient's perspective. For example, individuals may be unimpressed by evidence suggesting that activity X is 'safe' if the risk in question is not what really bothers them about that activity. Alternatively, they are likely to seize upon evidence exonerating X, if X is an activity that they want to promote. Even high quality evidence, according to criteria such as those in Section 3, will make little difference if it relates to an issue that is not, in fact, the primary issue of concern.

HSAC needs to be aware of these wider considerations, in addition to 'purely' scientific matters, when reviewing the available evidence and will aim to reflect them when presenting an opinion (see Annexes 1 and 2). HSAC's view is that awareness of context will enhance the utility of its opinions by decision-makers.

5. Towards a transparent assessment

HSAC addresses different kinds of questions, for which the evidence varies in terms of type, quality and amount. The Committee needs, therefore, to be flexible in its specific approach when weighing the evidence and arriving at an opinion (examples of published HSAC opinions are provided in Annex 2). Criteria of the kind outlined in Section 3 can be applied as appropriate to the scientific evidence but the body of evidence *as a whole* needs also to be considered, and HSAC will take account of important, wider questions such as those of problem framing. HSAC also reflects on its own perspectives (assisted by the diversity of its membership) and on the wider context within which the problem has been presented. The Committee's intention, in adopting this rounded approach (illustrated schematically in Annex 1), is to reach opinions that are robust, relevant and defensible. HSAC also considers how it might best reflect the overall strength of any given assessment and communicate the degree of confidence in its opinion. Quantitative measures, while attractive in some senses, can be open to misinterpretation, but a number of useful systems exist for indicating levels of confidence in a qualitative way [5-7]. HSAC will adopt (and adapt) one or more of these systems, as appropriate to the case in hand, when presenting its conclusions.

References

[1] A.B. Hill, The Environment and Disease: Association or Causation?, Proceedings of the Royal Society of Medicine 58(1965) 295–300.

[2] C.A. Harris, A.P. Scott, A.C. Johnson, G.H. Panter, D. Sheahan, M. Roberts, J.P. Sumpter, Principles of Sound Ecotoxicology, Environ. Sci. Technol. 48(2014) 3100-3111.

[3] H.J. Klimisch, M. Andreae, U. Tillmann, A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data, Regul. Toxicol. Pharmacol. 25(1997) 1-5.

[4] D.L. Weed, Weight of evidence: A review of concept and methods, Risk Anal. 25(2005) 1545-1557.

[5] Annual Report of the Government Chief Scientific Adviser. Innovation Managing Risk and Not Avoiding It (Evidence and Case Studies), (2014).

[6] T.R. Society, Resilience to Extreme Weather, Royal Society, London, 2014.

[7] C.B.F. Michael D. Mastrandrea, Thomas F. Stocker,, K.L.E. Ottmar Edenhofer, David J. Frame, Hermann Held, Elmar Kriegler,, P.R.M. Katharine J. Mach, Gian-Kasper Plattner, Gary W. Yohe,, a.F.W. Zwiers, Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties http://www.ipcc-wg2.gov/meetings/CGCs/index.html, 2010.

Further reading

Jasanoff, S. (1987) ‘Contested boundaries in policy-relevant science’, *Social Studies of Science* 17, 2: 195-230.

MacGillivray, B. (2014) ‘Heuristics structure and pervade formal risk assessment’, *Risk Analysis* 34, 4: 771-787.

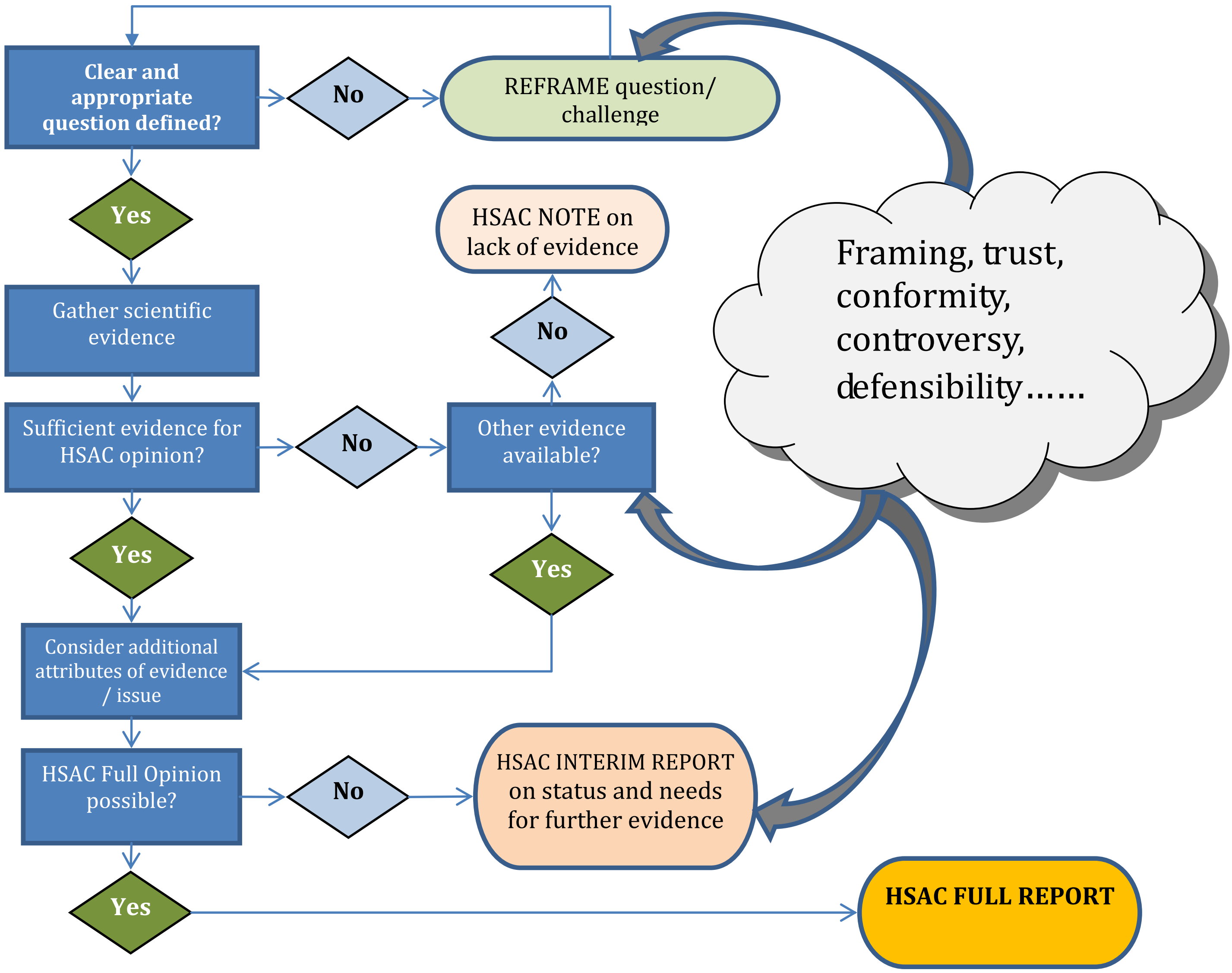
Pielke, Roger A. Jr. (2007) *The Honest Broker: Making Sense of Science in Policy and Politics*, Cambridge: Cambridge University Press.

Royal Commission on Environmental Pollution (2008) *Novel Materials in the Environment: The Case of Nanotechnology*, esp. Ch 4, ‘The challenges of designing an effective governance framework’, Twenty-seventh report, Cm 7468, London: TSO.

Stirling, A. (2007) ‘Risk, precaution and science: towards a more constructive policy debate’, *EMBO [European Molecular Biology Organization] Reports*, 8, 4: 309–15.

Sutherland WJ, Spiegelhalter D, Burgman MA. (2013) Twenty tips for interpreting scientific claims. *Nature*, 503: 335-337.

ANNEX 1 HSAC workflow diagram



ANNEX 2 Considering evidence: Two HSAC case studies

Two examples of published HSAC opinions illustrate how HSAC considers evidence in practice: the 2010 opinion of the Advisory Committee on Hazardous Substances (ACHS) (HSAC's predecessor) on the flame retardant decabrominated diphenyl ether (deca-BDE) and the HSAC 2012 opinion on nanomaterials[2]. These illustrate how HSAC's commentary is constrained by limits on the quality and/or quantity of evidence, and how HSAC can inform the scientific community about the type of evidence that is useful in practice. The constraints on evidence are often significant, particularly with newly identified and innovative materials when the scientific community has not had the opportunity to conduct exhaustive studies. This annex briefly outlines these two examples to illustrate how HSAC balances different criteria when, as is often the case, the evidence base is incomplete.

Flame retardants

The ACHS opinion on the deca-BDE family of flame retardants reflected some of the problems that emerge when the evidence base is partial and incomplete. It also illustrates how Table 1 and HSAC's workflow diagram (see Annex 1) can be applied in practice. ACHS was asked to form an opinion about whether or not deca-BDE compounds "were of an equivalent level of concern to be a PBT [Persistent Bioaccumulative Toxic] substance for the purposes of the REACH Regulation?". The challenge was to assess the experimental evidence about impacts of deca-BDE collected in the face of a range of confounding factors, including poor aqueous solubility, and a wide range of potential breakdown products - some of which could arise from other parent materials. These confounding factors meant that degradation products were not identified in most studies. ACHS carefully considered the quality of the experimental evidence and one study was discounted because it was conducted under artificial, indoor conditions, making it difficult to extrapolate to natural light and outdoor conditions. Another study had a particular weight in ACHS's deliberations – specifically in forming an opinion about whether deca-BDE is a very persistent compound – ACHS's opinion was that it is, but "*ACHS recognises that this conclusion is, of necessity, based upon the results of a single study".* Also, given the limited experimental evidence base, circumstantial evidence also played a significant role in deliberations: *"Circumstantial evidence indicates that there is potential for deca-BDE to debrominate in the environment to substances that are also of concern (e.g. hexa- and hepta-BDE)".*

This report illustrates well how HSAC balances evidence. Two types of evidence as set out in Table 1 were used: experimental evidence, and circumstantial evidence. The experimental evidence was limited and mixed: for example ACHS judged one study to be particularly relevant, but its quality in terms of Table 1 was moderate: "[the finding]*, which has been reported informally but not yet in peer-reviewed papers, provides substantial evidence of transformation of deca-BDE to breakdown products in surface sediments*". The qualitative, circumstantial evidence was relatively good. This case also demonstrates the lack of any

[2] https://www.gov.uk/government/groups/hazardous-substances-advisory-committee#publications

easily-defined quality gradient between different types of evidence, and illustrates the HSAC workflow (Annex 1) in practice: the quantitative experimental evidence was insufficient to form a judgement, and so the committee considered whether other evidence, *viz.* qualitative circumstantial evidence, was available, and this played a significant role in ACHS's deliberation. Overall, the opinion was based on a careful balancing of different types of evidence: "*The existence of strong qualitative evidence, together with some quantification in experimental systems, has convinced the ACHS that deca-BDE has the potential to undergo environmental degradation*".

Nanomaterials

The HSAC 2012 opinion on nanomaterials sets out some ways in which HSAC can lead in improving the quality of scientific evidence. This opinion sets out a number of recommendations, and identifies "*some issues and difficulties …* [and a] *need for more consistent standards for nanoparticles*". The report illustrates how some of the criteria (accessibility, transparency of aims etc.) outlined in Section 3 of this report can be applied in practice. This report also illustrates how the weight of evidence can be balanced when conventional scientific experimental evidence is limited, and the uncertainties are particularly profound for innovative, manufactured materials that are relatively novel. In these cases it can also be difficult to define the substance of concern precisely. HSAC observed: "*There are significant knowledge gaps limiting our understanding of the human and environmental hazard and risks of nanotechnology*". The evidence base is mixed, and experimental evidence is patchy and of low quality "[the] *literature has been criticized due to a lack of quantification of nanomaterials and experimental systems*".

The HSAC report sets out a number of recommendations to show how the quality of experimental and model-based evidence can be improved – i.e. how to move the evidence from low to high quality (Table 1). Some examples include the need for external validity – that the scientific evidence should mirror real-world use: HSAC recommended that "*studies of the particular sample investigated must show similar (range of) key features to that used commercially*". Another recommendation focused on the need for transparency "*Validation and uncertainty in all metrics should be discussed*". The report focused on quantitative evidence and an interesting question to be explored in future HSAC judgments on nanomaterials and nanoparticles is how epidemiological and circumstantial evidence can be combined to improve HSAC opinions when the evidence base about innovative materials is limited.