THEORY AND METHODS

Validity of predictions in health impact assessment

J Lennert Veerman, Johan P Mackenbach, Jan J Barendregt

....

See end of article for
authors' affiliations
....

Correspondence to:
L Veerman, Department of
Public Health, Erasmus
Medical Centre, PO Box
1738, 3000 DR Rotterdam,
The Netherlands; j.
veerman@erasmusmc.nl

Accepted 1 September 2006
....

Background: An essential characteristic of health impact assessment (HIA) is that it seeks to predict the future consequences of possible decisions for health. These predictions have to be valid, but as yet it is unclear how validity should be defined in HIA.
Aims: To examine the philosophical basis for predictions and the relevance of different forms of validity to HIA.
Conclusions: HIA is valid if formal validity, plausibility and predictive validity are in order. Both formal validity and plausibility can usually be established, but establishing predictive validity implies outcome evaluation of HIA. This is seldom feasible owing to long time lags, migration, measurement problems, a lack of data and sensitive indicators, and the fact that predictions may influence subsequent events. Predictive validity most often is not attainable in HIA and we have to make do with formal validity and plausibility. However, in political science, this is by no means exceptional.

There are various definitions of health impact assessment (HIA) and a wide variety of activities have been termed HIA. In agreement with Kemm[1] we would argue that in essence, HIA seeks to (1) predict the future consequences of possible decisions regarding projects, programmes or policies for health; and (2) inform policy decisions on the basis of these predictions. This restricts HIA to what some would call ''prospective'' HIA. Generally, the HIA procedure starts with screening for policies with potentially modifiable health consequences, followed by ''scoping'' to determine who should perform the assessment, and how. The results of the assessment are communicated to relevant parties, and finally evaluation and monitoring of health effects takes place. Discussions about evaluation of HIA to date tend to focus on their effects on decision making and participation, whereas less attention has been paid to the validity of predictions.[2 3] Notwithstanding the futility of the exercise if it exerts no influence on policy, in order to be valuable to policymakers and stakeholders, the predictions in HIA need to be valid. Validity is the expression of the degree to which a measurement measures what it purports to measure.[4] As yet, it is unclear how the validity of predictions should be defined in HIA, and how it can be assessed. In this paper, we discuss the assessment of the validity of predictions in HIA. The objective of this paper is twofold: it aims to (1) discuss the assessment of validity of predictions in HIA; and (2) propose a checklist to establish the validity of predictions in HIA. Where possible, we think that these predictions would preferably be quantified, but qualitative work can be judged by the same standards. The paper is structured in three sections. The first section presents the conceptual basis for predictions in HIA by referring to the work of Popper. In the second section, the concepts of validity are reviewed and discussed. In the third section, we apply Popper's logical structure to construct a checklist to establish the validity of predictions in HIA and critically discuss issues in assessing validity.

A CONCEPTUAL BASIS FOR PREDICTIONS

Before elaborating on the concept of validity for HIA, we will take a closer look at predictions. For an understanding of predictions, it is useful to look at the work of Popper.[5] He asserts that explanation, prediction and testing share three common elements: initial conditions, outcomes and a theoretical framework (fig 1). When we are testing a hypothesis, we accept our initial conditions and outcomes and make inferences about the theoretical framework. For explanation, the outcomes and the theoretical framework or the initial conditions are (provisionally) considered reliable, which allows us to make inferences about either the theoretical framework or the initial conditions. Prediction requires that the initial conditions and the theoretical framework are considered valid. (According to Popper, this is a case in which we apply our scientific results.) One cannot do two things at a time—that is, testing a theory and at the same time making predictions is logically impossible.

Thus, HIA assesses the initial conditions and policy plans, and uses the theoretical framework of science (eg, epidemiology, demography, physics and also the social sciences) to make predictions. In essence, HIA is a deductive activity: it uses (general) theory to make statements about specific situations. Though results may be extrapolated to similar policies in other situations, HIA is not, in the first place, concerned with generating knowledge that is generalisable; it aims to assess the effects of a specific policy on a specific population in a specified environment, using a theoretical framework that has resulted from previous scientific work. A HIA project can be compared with the daily weather forecast, which is not meant to check whether the model behind it is correct but to advise on what coat to wear tomorrow.

CONCEPTS OF VALIDITY IN HIA

When can we consider the predictions in HIA valid? We will examine different forms of validity and consider their applicability to HIA and illustrate this with reference to a study that we conducted on an aspect of the European Union (EU) policy on fruits and vegetables (FV). This policy guarantees producers a minimum price for their products by withdrawing FV from the market when prices drop below a specified level. On the basis of a simulation model with data on the EU agricultural policy, FV consumption and health, our study concluded that reform might, at maximum, result in modest health gains for the Dutch population: an estimated annual gain of 1930 disability-adjusted life years or an increase in life expectancy by 3.8 days for men and 2.6 days for women.[6]

....

In epidemiological studies, internal and external validity are important concepts.[7][8] Internal validity indicates the degree to which results of research support or refute a causal relationship between the dependent and independent variables. A HIA must be based on a theoretical framework that ultimately rests on research that is internally valid. HIA itself, however, is not primarily intended to investigate causal relationships; as we have seen, these simply have to be assumed valid in order to make prediction possible. Therefore the concept of internal validity does not directly apply to HIA.

External validity refers to the degree to which the theoretical knowledge resulting from research can be generalised to other populations. However, in a HIA, we are trying to do the reverse: established generalisable knowledge is applied to a specific population. HIA is, as it were, at the receiving end of external validity: the theory used to make the predictions must of course be relevant for the population concerned. External validity of a HIA itself is not of primary concern. Nonetheless, conclusions may be generalisable to similar situations. For example, if the validity of our study on the EU FV policy is accepted, it would be reasonable to expect a similar effect of abolishing withdrawal support on life expectancy in the UK. (However, since the British consume less FV they might benefit slightly more.)

In psychometric research, Cronbach's concepts of validity are in wide use. In contrast with internal and external validity, they are more appropriate for measurement instruments than for complete studies. Cronbach distinguishes face validity, content validity, criterion validity and construct validity. Face validity or plausibility is the degree to which an observer deems that the theoretical framework is understandable, applicable and plausible.[9][10] It is closely related to credibility, which is the confidence that (potential) users have in a theory or model. Plausibility should clearly be considered relevant to HIA. The causality of the relationships in a HIA must be credible, both in qualitative terms (is there a likely mechanism between cause and effect?) and in quantitative terms (is the strength of the association plausible?). In our example of the EU FV policy, the question would be whether abolishing withdrawal support would result in a higher consumption of FV. We postulated that, at maximum, the amount currently withdrawn would enter the market and that this would lower prices, which in turn would increase sales and ultimately consumption. But alternative scenarios are conceivable—for example, if producers decide to produce something else—and the assumption that consumption rises in equal proportion to the amounts now taken off of market can only be an overestimate.

Plausibility may look vague and arbitrary, but then so is causality. Causation cannot be proven, but ultimately rests on judgement, properly supported by evidence. Two generally accepted minimum conditions for a causal relationship are that the cause precedes the consequence and that there is a correlation between the two.[11] A third requirement is that there is a plausible mechanism. The difficulty with this requirement is: who decides what is plausible? De Groot[12] posits that, ultimately, the forum of the scientific community decides. The requirement of plausibility can therefore be translated as the obligation to convince one's peers, and this can be done by arguments based on logical inference and empirical data. A HIA therefore has to present evidence as to why the predictions are likely to be correct, especially as it is not intended to test hypotheses but to inform a policy process of (accepted) scientific knowledge. Plausibility in this definition clearly is not a superficial matter. Its synonym face validity seems to suggest otherwise and is therefore best avoided.

Criterion validity is the degree to which outcomes are confirmed by a "gold standard". For HIA studies as a whole, there are no such standards, but there may be for measurement instruments used in HIA. Criterion validity is sometimes subdivided into "predictive" and "concurrent", depending on when the gold standard is measured.[13] Predictive validity is the degree to which predictions are confirmed by facts. We would propose to turn things around and argue that the concept of criterion validity is redundant if one accepts the idea of predictive validity. Predictive validity should be established, and this can be performed using gold standard tests to the degree that these tests accurately measure the concept of interest.

Content validity is concerned with the question of whether all aspects of the phenomenon to be measured are represented in the appropriate proportions. Translated to HIA, the question is whether all the relevant determinants and health effects have been included in a plausible order of magnitude. This is a matter of judgement and can therefore be considered part of plausibility, removing the need for content validity as a separate form of validity in HIA.

Construct validity is the degree to which the outcomes correlate with those of other instruments that purport to measure the same construct. It applies to hypothetical concepts that cannot be measured directly. HIA should reflect the current scientific understanding, and so would, in principle, avoid using methods or concepts of which the construct validity has not been established in other research. We do not see an important role for construct validity in HIA as such.

Formal validity concerns how well an argument conforms to the rules of logic to arrive at a conclusion that must be true, assuming that the premises are true.[14] Though not always explicitly, formal validity plays a role in any research. Besides argumentation, it is also about the correctness of calculations and other methodological aspects of scientific endeavour. Applied to HIA, formal validity is concerned with the correct application of correct methods. Clearly, this must be in order for a prediction to be valid.

We therefore propose that three types of validity are relevant for HIA: plausibility, formal validity and predictive validity, whereby plausibility broadly refers to the subject matter, formal validity to the method and predictive validity to direct empirical evidence. Other types of validity can either be considered redundant or are unimportant in HIA.

ESTABLISHING VALIDITY OF PREDICTIONS IN HIA

The predictions in a HIA can be considered valid if plausibility, formal validity and predictive validity are in order. In the appendix, we use Popper's logical structure (fig 1) to construct a list of aspects of a HIA that need to be examined in order to determine their validity. This checklist helps to systematically examine a HIA study by subsequently focusing on a number of questions regarding the plausibility and formal validity of the assessment of the initial conditions and of the theoretical framework that was applied, and on the predictive validity of the study. A web-only supplement illustrates the use of the checklist by highlighting some of the points that an independent assessment of the validity of the EU FV study could focus on (available at http://jech.bmj.com/supplemental).

Assessing validity is not a problem for plausibility and formal validity. Formal validity can be checked, though it often requires time, effort and expertise. For example, in the case of the EU FV study, the data and calculations in the spreadsheet

Figure 1 The logical structure of Popper's unity of method in all theoretical or generalising sciences. Initial conditions are linked to outcomes by a theoretical framework. When two of the three are provisionally assumed valid, the third can be inferred.

What is already known

- Health impact assessment (HIA) seeks to influence policy by predicting the consequences of decisions for health outside the healthcare field.
- To have a positive effect on health and for the long-term credibility of HIA, these predictions have to be valid.
- However, it is unclear as to how validity should be defined and assessed.

What this paper adds

- This paper provides a logical framework for predictions, based on Karl Popper's work.
- We argue that three types of validity are relevant for health impact assessment (HIA): plausibility, formal validity and predictive validity, and present guidelines for establishing the validity of predictions in HIA.

Policy implications

- This paper supports health impact assessment practitioners in making valid predictions by providing a theoretical framework and a corresponding method of validating predictions.
- It may also be of interest to other researchers since making prediction possible is one of the fundamental purposes of all scientific endeavour.

used could be checked, though this would require some understanding of life table analysis. Plausibility will superficially be assessed by policymakers and stakeholders, but should really be checked by independent experts in the relevant scientific discipline (which will increase face validity for the first group as well). For the study on the EU FV policy, this would include specialists in agricultural economics and epidemiology.

In operational research, which is concerned with building simulation models of (military, industrial and economic) processes to optimise outcomes, an established approach to validation of models is "independent verification and validation"(IV&V).[10] In IV&V, an independent party examines the model and judges its validity. Verification should be understood as checking the formal validity (are all calculations correct?), whereas validation refers to the examination of plausibility and predictive validity. Although a complete IV&V can be costly, it could well be considered a means to establish formal validity and plausibility in HIA.

In HIA, the predictive validity of entire studies usually cannot be established. This would require outcome evaluation of completed HIA studies, which is difficult for a number of reasons.

In the first place, there is often a long time lag between a change in policy and the corresponding health effects. For example, a decision to stop using asbestos in The Netherlands was taken decades ago, but the incidence of mesothelioma is still rising.[15] Besides having to wait a long time before measurement is at all possible, this time lag makes it expensive and increases the loss to follow-up. The latter is made worse by migration, which may be differential: those experiencing most inconvenience by a development may be the first to move away.[16]

Second, many factors influence the same outcomes that the HIA considers, often to a much greater extent. This may obscure any effect of the policy decision under consideration. For example, a trend in smoking may obscure any effect of changes in the EU FV policy on cardiovascular disease and cancer.

Third, many health problems are hard to measure. Routine data may not be available at the appropriate geographical scale or may not be measured frequently enough to pick up changes due to the policy under scrutiny.

Fourth, HIA intends to influence policy, but if successful it invalidates its own predictions. This is what Popper[5] called the

"Oedipus effect" after the mythical figure who killed his father whom he had never seen because of a prophecy that had caused his father to abandon him as a child. (With regard to the EU agricultural policy, there is little risk of this effect occurring, if only because no direct communication with stakeholders took place—which distinguishes that study from a HIA exercise.)

Finally, in most cases a control group is lacking. One intervention group and no control does not make for a strong research design.[16]

When complete assessment of predictive validity is not possible, sometimes partial predictive validation is feasible by focusing on intermediary outcomes.[17] For example, the effect of a price change on the consumption of FV is measurable shortly after a tax reduction, whereas the outcome of interest (cardiovascular disease and cancer) is unlikely to ever be measurable.

Predictive validity can also be supported by using knowledge from initial conditions and outcomes in the past: this allows testing of the theory (historical data validation).[10]

Historical data validation would be confronted with the same problems as predictive validity in general, except that it saves a long wait for the results. Although theoretically possible, we know of no example in the literature on HIA.

External reviews of HIA exercises have been carried out, but few focus on the validity of predictions. We know of no example in which the predictive validity was established. Formal validity and, to a lesser extent, plausibility have been assessed. In practice, these two forms of validity are closely connected. For example, the evaluation of the Alconbury Airfield HIA assessed formal validity among others by checking that reasons for including and excluding determinants are clearly stated. The recommendation to pay more attention to impacts that are hard to quantify can be considered as an element of plausibility.[18]

Finally, even in a properly validated HIA, unanticipated adverse effects may arise. Science cannot (ever) claim to provide all knowledge needed, so if not for establishing predictive validity, HIA studies should make recommendations for the monitoring of health outcomes to aid early detection.

CONCLUSION

Predictions are at the core of HIA, but predictive validity will most often prove unattainable. Instead, we have to make do with less than the gold standard and assess HIA studies and methods for plausibility and formal validity only. It may be of comfort to know that in political science this is by no means exceptional. Few decisions can be taken with the confident knowledge of relevant and thorough outcome evaluations.

Supplementary figure is available at http://jech.bmj.com/supplemental

....
Authors' affiliations
J Lennert Veerman, Johan P Mackenbach, Department of Public Health, Erasmus Medical Centre, University Medical Centre, Rotterdam, The Netherlands
Jan J Barendregt, School of Population Health, University of Queensland, Brisbane, Queensland, Australia

REFERENCES

1 **Kemm J**. Perspectives on health impact assessment. *Bull World Health Organ* 2003;**81**:387.
2 **Parry JM**, Kemm JR. Criteria for use in the evaluation of health impact assessments. *Public Health* 2005;**119**:1122–9.
3 **Quigley RJ**, Taylor LC. Evaluating health impact assessment. *Public Health* 2004;**118**:544–52.
4 Nieuwenhuijsen MJ, ed. *Exposure assessment in occupational and environmental epidemiology.* New York: Oxford University Press, 2003.
5 **Popper KR**. *The poverty of historicism.* 2nd edn. London: Routledge & Kegan Paul, 1957:130–43.
6 **Veerman JL**, Barendregt JJ, Mackenbach JP. The European Common Agricultural Policy on fruits and vegetables: exploring potential health gain from reform. *Eur J Public Health* 2006;**16**:31–5.
7 **Campbell DT**, Stanley JC. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally, 1963.
8 **Campbell DT**. Reforms as experiments. *Am Psychol* 1969;**24**:409–29.
9 **Cronbach LJ**. Test validation. In:Thorndike RL, editor.In:*Educational measurement.*2nd edn. Washington, DC: American Council on Education, 1971.
10 **Sargent RG**. Verification and validation of simulation models. *Proceedings of 1996 Winter Simulation Conference* Conference, Coronado, California, 1996:55–64.
11 **Rothman KJ**, Greenland S. Causation and causal inference. In: Rothman KJ, Greenland S, eds. *Modern epidemiology.* Philadelphia: Lippincott Williams & Wilkins, 1998:7–28.
12 **De Groot AD**. *Methodology: foundations of inference and research in the behavioral sciences.* The Hague: Mouton, 1969.
13 **Gliner JA**, Morgan GA. *Research methods in applied settings: an integrated approach to design and analysis.* Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
14 **Verlinden J**. Formal validity review, 1998. http://www.humboldt.edu/~act/HTML/tests/validity/review.html (accessed 29Jan 2007)..
15 **Swuste P**, Burdorf A, Ruers B. Asbestos, asbestos-related diseases, and compensation claims in The Netherlands. *Int J Occup Environ Health* 2004;**10**:159–65.
16 **Parry J**, Stevens A. Prospective health impact assessment: pitfalls, problems, and possible ways forward. *BMJ* 2001;**323**:1177–82.
17 **Lindholm L**, Rosen M. What is the ''golden standard'' for assessing population-based interventions?—problems of dilution bias. *J Epidemiol Community Health* 2000;**54**:617–22.
18 **Close N**. *Alconbury Airfield Development HIA evaluation report.* Cambridge: Cambridgeshire Health Authority, 2001.

APPENDIX

HOW TO ESTABLISH VALIDITY OF PREDICTIONS IN HIA: A CHECKLIST
Plausibility
Definition: degree to which an observer deems that the theoretical framework is understandable, applicable and plausible.

To be established by researchers, external experts and stakeholders. (Stakeholders' judgement has to be interpreted with caution, bearing in mind the interests the stakeholder may have in the outcomes.)

Initial conditions
- Is the policy plan/project described accurately?
- Is the description of the baseline situation accurate?
- Has uncertainty in the initial conditions been assessed?
- How robust is the model to (foreseeable) changes in the initial conditions? For example if increases in air transport are likely, have these been included in the assessment of the health consequences of building housing near an airport?

Theoretical framework
- Is the causal web underlying the analysis valid according to the state of the pertaining scientific field?
- Is the order of magnitude of the causal relationships in concurrence with current scientific knowledge?
- Has the degree of certainty of the causal relationships been described?
- Are all exposures to determinants of health that are likely to result from the intended policy/project included in the analysis?
- Of the exposures included, have all plausible health outcomes been included?
- Have all populations likely to be affected by the policy been included in the analysis?
- If available, how do the results of similar exercises compare with the predicted effects in this HIA? Can any differences be satisfactorily explained by differences in the initial conditions (including intervening events during the period of analysis) or lack of formal validity of the previous analyses?

Formal validity (= verification)
Definition: the degree to which correct methods have been applied correctly.

To be established by researchers and external experts.

Initial conditions
Have the right methods been applied to:

(1) describe the policy proposal; and
(2) describe the baseline situation; and have both these sets of methods been applied correctly?

Theoretical framework
Have the right methods been applied to:

(1) construct the causal framework;
(2) estimate the order of magnitude of the causal relationships;
(3) estimate degree of certainty of the causal relationships;
(4) find all significant determinants of health of which the exposure changes as a result of the proposed policy;
(5) find all health outcomes that result from changes in exposure;
(6) identify populations likely to be affected by the policy been included in the analysis; and have these methods been applied correctly?

Predictive validity
Definition: the degree to which predictions are confirmed by facts.

To be established by researchers and external experts.

Historical predictive validity

- Are historical data on initial conditions and subsequent outcomes available on which the model underlying the HIA can be tested?
- If testing has been performed, how well does the model "postdict" these outcomes, and can any differences between model and empirical data be explained satisfactorily by

differences in the initial conditions or uncertainty in initial conditions (including intervening events during the period of analysis) and/or outcomes?

In retrospect

- To what extent did the predictions materialise?

SPEAKER'S CORNER..

Pharmaceutical advertising: from prospectus to marketing

Some 30 years ago, Henry Gadsen, director of Merck Pharmaceuticals, confided to *Fortune* magazine that he was worried about the potential market for his products being confined to sick people. Three decades on, the marketing strategy of the major pharmaceutical companies is increasingly being specifically aimed at healthy people.[1]

Advertising is one of the marketing strategies used by pharmaceutical companies to inform both the end user and doctors prescribing treatment. However, advertising messages are characterised as a mixture of information and persuasion (from prospectus to marketing), as they use rational sales discourse implemented through advertising language, possibly creating the need to use pharmacological treatment, which may not always be necessary or the most effective solution to the problem.

Although it is true that we cannot do away with advertising, it also needs to be read critically. This would allow us to take informed decisions regarding the quality of pharmaceuticals, based on the best scientific evidence available to us.

With the aim of regulating this situation, the World Health Organization[2] has made recommendations for the presentation of pharmaceutical information in advertising. This was because they saw that advertising strategies respond more to market interest (increasing the target public) than to informative functions. However, this function is needed by consumers in order to consume the advertised products in a correct and safe manner.

In recent years, advertising discourse has been subjected to several investigations. Many studies have concluded that advertisements should present information more accurately, as, at times, they offer information of little importance in the clinical trials or it is difficult, if not impossible, to recover the bibliographical references used to back up advertising campaigns.

Accordingly, doctors and users have to evaluate advertising and the information presented in advertising messages critically. On the other hand, manufacturers must present information accurately. To do this, they should take into account the recommendations of specialised international organisations. This could also be a means of strengthening the advertising strategy, by lending it credibility.

B Cambronero-Saiz

Correspondence to: Ms B Cambronero-Saiz, Carretera San Vicente del Raspeig s/n Alicante, Alicante 03690, Spain; belen.cambronero@ua.es

Reference

1 **Moynihan R**, Cassels A. *Selling sickness: how drug companies are turning us all into patients.* New York: Nation Books, 2005.
2 **Word Health Organization**. *Criteria for medicinal drug promotion.* Geneva: WHO, 1988, http://www.who.int/medicinedocs/collect/edmweb/pdf/whozip08e/whozip08e.pdf (accessed 29 Jan 2007).