AN 'ODDS ON' APPROACH TO EVALUATION

Geoff White and Mark Matthews

In the December 2013 issue of *Evaluation Connections* we proposed a faster, smarter and possibly cheaper way of evaluating public policies. It is based on formulating succinct policy propositions against the available evidence and testing them as hypotheses in a structured, sequential manner as further evidence emerges. This would provide a stronger foundation for learning than the more audit-based approach – would be faster and could be cheaper.

In parallel, an assessment by the National Audit Office (NAO) of UK policy evaluations (2013) found that evaluation methods had not generally facilitated learning. The NAO highlighted a wide range of causes – narrow evaluation coverage, inappropriate methods, poor quality evidence, inadequate dissemination and limited use of the results.

We suggest the fault may run still deeper. The audit function has tended to dominate the learning role, with evaluations being bolted on after delivery rather than designed as an integral component of policy formulation and implementation. The result is that weak evaluation designs that do little to enable learning are often adopted. What is needed, especially when evaluation budgets are tight, is *innovation* in evaluation processes and methods – aiming for smarter and possibly cheaper ways of learning.

This article develops our earlier thoughts into what we call an 'odds-on' approach to evaluation since we take the view that policies rarely have a 100 % chance of success at the outset even when they are tried and tested (e.g. because the context has changed). Where the interventions are innovative and/or are set in complex and uncertain circumstances both risks and rewards are likely to be relatively high. Evaluation should provide the evidence to enable effective risk management and the intervention adapted where appropriate to strengthen the odds of success.

What we propose draws on the approach advocated by Patton (2010) to assist the development of policy in complex and uncertain contexts by setting up close to real-time feedback in a continuous learning loop which can incorporate both qualitative and experimental methods. It requires the forward-looking appraisal of a proposed policy intervention to be couched as a set of hypotheses about what it might achieve and then uses the monitoring and evaluation of evidence to test the relative strength of these hypotheses. At the start of this process available evidence about results secured in different contexts can be used to estimate the odds that each of these hypotheses will turn out to be correct. As experience unfolds and more evidence becomes available then these odds can be adjusted to reflect actual experience.

The use of competing hypotheses drawn from diverse theories of action and change within a single policy intervention or a set of similar interventions reduces the risk of misdiagnosis. These competing hypotheses can be tested and eliminated (or at least the odds of the hypotheses being correct can be reduced) via evidence gathering and analytical work. This may either confirm the initially assumed odds, or may lead us to revise these odds or, most usefully, develop updated hypotheses more consistent with the available data. The hypotheses testing would use diverse sources of evidence including case studies and other qualitative evidence building on the methods in Qualitative Comparative Analysis (Ragin, 2008) and process tracing (Collier, 2011) to draw robust inferences from such evidence sources.

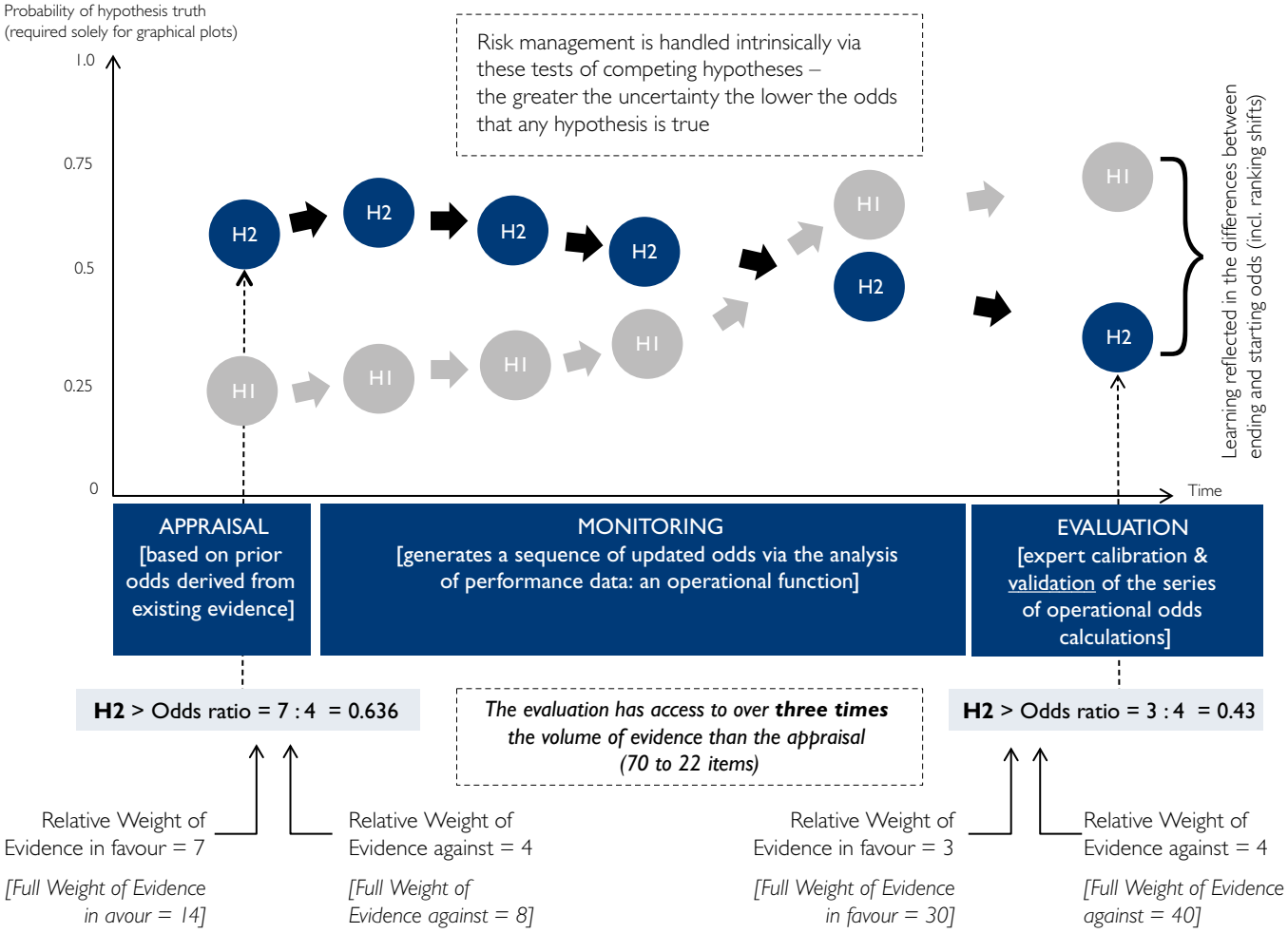
The broad principle, that the odds of different hypotheses being true is updated when new information is received, underpins what is known as Bayesian inference. Whilst this principle is easily grasped, its expression in formal statements of method and analysis has often been impenetrable to non-specialists due to the overly complex mathematical approaches used. There are, however, simplified approaches to Bayesian updating based on the use of 'natural frequency' data that directly express the relative incidence of observations in a manner that is more readily intelligible and lends itself to odds-based expressions of risk, (Gigerenzer, 2002). As in clinical medicine, this approach allows the overall prevalence of diagnostic errors in hypothesis tests (i.e. false positive and false negative results) to be factored into specific evaluations, opening up an avenue to apply the diagnostic methods used in psychology and in signal processing[1] to policy development and evaluation.

In this approach, the odds of different hypotheses being correct given currently available information are calculated by determining the extent to which information supports, or does not support, each of a set of specific alternative hypotheses. These odds are calculated by summarising all available evidence in a binary (true or false) manner in regard to each alternative hypothesis.

For example, the public funding for science and technology through institutes or programmes could be characterised as having a range of different 'yield' hypotheses depending on whether the anticipated spin-off process works through company formation, attraction of inward investment, increased licence income, new skills acquisition and development etc. In each case, and as experience unfolds, the goal is to be able to calculate the odds that each (or none) of these hypotheses is correct using currently available information – and expressed in betting terms as odds *for* relative to odds *against* (i.e. 3:1 for or 2:1 against etc).

If it turns out that the licence income hypothesis is exhibiting the most favourable odds of being the best, then this insight can be used to reshape the existing and future interventions. If the odds do not look good for any of the intervention yield hypotheses, then a case be-

1 Signal processing is concerned with identifying errors in communication and processing.



The diagram above depicts how the odds-updating process could operate.

gins to emerge for either complementary support or indeed cancelation of the programme – with consequent lessons for the future.

The advantages of the proposed approach are three-fold:

1. Seamlessly connecting appraisal, monitoring and evaluation – which are too often approached as separate activities.
2. Integrating risk and uncertainty assessments into appraisal and evaluation rather than treating risk management as a bolted-on compliance activity.
3. Potentially providing faster and more cost-effective evaluations than the traditional audit approach. Evaluations can be updated frequently and at low cost because results can be expressed in a tabular form that summarises the evidence for and against each competing hypotheses as odds. This succinct and fast-to-deliver format makes it easier to brief senior officials and ministers on progress achieved.

Political imperatives are driving devolution of policy capacity to the national and local levels in the UK. This could provide a useful opportunity to develop these faster, cheaper and integrated approaches to appraisal, evaluation and risk management. We are keen to discuss possible pilot implementations of this new approach with fellow evaluators and government departments, agencies and devolved authorities.

References

Collier, D. (2011). Understanding process tracing. *Political Science and Politics* 44, No 4.

Gigerenzer, G. (2002), *Reckoning with risk*. Penguin.

National Audit Office (2013), *Evaluation in government*, December.

Patton, M. Q. (2010), *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford Press.

Ragin, C. C. (2008). *What is Qualitative Comparative Analysis?* Retrieved November 2015 from http://eprints.ncrm.ac.uk/250/1/What_is_QCA.pdf.

Swets, J.A., Dawes, R.M., Monahan, J. (2000) Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*. 1, 1–26.