



National Audit Office

Report

by the National Audit Office

Cross-government

Evaluation in government

DECEMBER 2013

Our vision is to help the nation spend wisely.

Our public audit perspective helps Parliament hold government to account and improve public services.

The National Audit Office scrutinises public spending for Parliament and is independent of government. The Comptroller and Auditor General (C&AG), Amyas Morse, is an Officer of the House of Commons and leads the NAO, which employs some 860 staff. The C&AG certifies the accounts of all government departments and many other public sector bodies. He has statutory authority to examine and report to Parliament on whether departments and the bodies they fund have used their resources efficiently, effectively, and with economy. Our studies evaluate the value for money of public spending, nationally and locally. Our recommendations and reports on good practice help government improve public services, and our work led to audited savings of almost £1.2 billion in 2012.

Contents

Key facts 4

Summary 5

Part One

Coverage of evaluation evidence 12

Part Two

Quality of evaluation evidence 19

Part Three

Use of evaluation evidence 27

Part Four

Production, resources and barriers 36

Appendix One

Our audit approach 45

Appendix Two

Our evidence base 46

Appendix Three

Arrangements for evaluation
in government 48

The National Audit Office study team consisted of:
Anna Athanasopoulou, Phil Bradburn, Helen Hodgson, Anne Jennings and Thomas Williams under the direction of Michael Kell.

This report can be found on the National Audit Office website at www.nao.org.uk

For further information about the National Audit Office please contact:

National Audit Office
157–197 Buckingham Palace Road
Victoria
London
SW1W 9SP

Tel: 020 7798 7000

Enquiries: www.nao.org.uk/contact-us

Website: www.nao.org.uk

Twitter: @NAOorguk

Links to external websites were valid at the time of publication of this report. The National Audit Office is not responsible for the future validity of the links.

Key facts

£2.1bn

spent on government
R&D (2010-11)

£44m

spent on government
evaluation (2010-11)

102

FTE staff working
on evaluation in the
government

- 70 of 305** government evaluations between 2006 and 2012 have cost-effectiveness data
- 14 of 34** evaluations reviewed provide sufficient evidence of policy impact
- 4 of 15** chief analysts say cost-effectiveness evaluation is quite poor
- 15 per cent** of impact assessments in 2009-10 referred to evaluation evidence
- £3 million** cuts in spending on evaluation have been made since May 2010

Summary

1 An informed government collects high-quality information on context, expenditure, activities and results, and analyses this to expose issues or opportunities. It presents informed options to internal decision-makers, as well as candid assessments of plans and performance externally. Without this information, the government is not well placed to respond to funding cuts and longer-term challenges of providing sustainable, high-quality services and supporting economic growth.

2 Ex-post evaluation is the activity of examining the implementation and impacts of policy interventions, to identify and assess their intended and unintended effects and costs. Evaluation should be a key source of information on the cost-effectiveness of government activities, for accountability purposes and as a means to improve existing policies and to better design future policies. It is distinct from appraisal or ex-ante evaluation, which should be conducted before policy implementation.

3 Good-quality evaluation can provide evidence on attribution and causality – that is, whether the policy delivered the intended outcomes or impacts, and to what extent those were due to the policy. This involves developing a counterfactual and comparing the results with what would have happened without the intervention. Evaluation should complement other sources of information on cost-effectiveness, such as modelling and economic or financial analysis conducted for option appraisal, or data collected during policy implementation.

4 Evaluations are produced by analysts in government departments, by academics, consultancies, and other organisations commissioned by government. In some cases, the government has set up arm's-length bodies which commission or synthesise evaluation evidence, with varying levels of autonomy and independence. Recently, the government has set up a network of 'What Works' centres, which are responsible for synthesising evaluation evidence on the effectiveness of policy in a range of fields.

5 *Managing Public Money* sets out the main principles for dealing with resources used by public sector organisations in the UK.¹ It explains the importance of evaluating past initiatives, and emphasises that Parliament expects accounting officers to take personal responsibility for "ensuring that the organisation's procurement, projects and processes are systematically evaluated and assessed". *The Civil Service Reform Plan* explains that accounting officers "must be accountable for the quality of the policy advice in their department".²

¹ Available at: www.gov.uk/government/publications/managing-public-money

² HM Government, *The Civil Service Reform Plan*, June 2012. Available at: my.civilservice.gov.uk/reform/the-reform-plan/

6 The government's evaluation of its activities has often been criticised by the National Audit Office (NAO) and the Public Accounts Committee, and in other reports, including some by the government itself. These criticisms relate to:

- gaps in the coverage of evaluation evidence;
- poor-quality evaluation;
- insufficient use of evaluation evidence; and
- difficulties faced by independent researchers in accessing administrative data and other government data to conduct their own evaluations of government interventions.

Scope

7 Government guidance on evaluation distinguishes between process evaluations (how the policy was implemented); impact evaluations (what difference the policy made); and cost-effectiveness or economic evaluations (which measure and monetise the effects of a policy, relative to its costs).

8 This report focuses on impact and cost-effectiveness evaluation relating to government spending, taxation and regulatory interventions, across the 17 main departments and some of their bodies. We focus on these types of evaluation because they can help the government take decisions to improve the impact and the value for money.

9 Our approach and evidence base are set out in Appendices One and Two. The report aims to add to existing assessments of government evaluation by providing quantitative answers to four questions:

- What does existing impact and cost-effectiveness evaluation evidence cover?
- What is the quality of the existing evaluation evidence?
- How well does this evidence support strategic resource allocation, policy development and policy implementation?
- How much does the government spend in producing this evaluation evidence?

Key findings

Coverage of evaluation evidence

10 Government guidance sets out the expectation that all policies, programmes and projects should be subject to ‘proportionate’ evaluation. However, not all departments follow government evaluation requirements. Only two chief analysts said that they always followed the requirements set out in cross-government guidance and their own departmental guidance on evaluation (paragraphs 1.2 to 1.4).

11 It is difficult to establish the coverage of evaluation evidence, but it does not appear to be comprehensive. The government does not publish a comprehensive overview of evaluation evidence mapped against total government spending and other interventions. We were not able to map all existing evaluation evidence across government, but found evidence of significant gaps:

- Previous NAO reports have highlighted the lack of evaluation evidence in 12 of 17 main departments, and a lack of post-implementation reviews (a type of evaluation) of interventions covered by published impact assessments (paragraph 1.6).
- For this study, we reviewed almost 6,000 analytical outputs published on departmental websites between 2006 and 2012. We identified that 305 of these were impact evaluations, and 70 of those included assessments of cost-effectiveness. Of these 70 evaluations, 41 reports evaluated a total of £12.3 billion of government spending (paragraph 1.7).

12 Departmental chief analysts recognise that gaps exist, but few departments have plans in place to evaluate all of their major projects. Only four departments intended to evaluate all of their top five major projects. Plans to evaluate impact or value for money related to only £90 billion of £156 billion in major projects expenditure (paragraphs 1.8 to 1.9).

Quality of evaluation evidence

13 Departments’ own assessment varies regarding the overall quality of their evaluation evidence. Of the 15 chief analysts in our survey, five said that the quality of evaluation evidence in their department was “very good” or “quite good”, but four chief analysts considered the quality to be “quite poor” (paragraph 2.6).

14 Our assessment of the fitness for purpose of a selection of 34 evaluations from four departments finds significant variation. Only 14 evaluations were of a sufficient standard to give confidence in the effects attributed to policy because they had a robust counterfactual. The evaluations we reviewed covering spatial policy and business support were generally weaker than some of those covering labour market and education policies (paragraphs 2.11 to 2.13).

15 We found some evidence that evaluation reports that are weaker in identifying causality tend to be more positive in assessing what the intervention achieved. To the extent that less reliable evaluation studies provide bolder claims of policy impact, there is a clear risk that if the government allocates funding on that basis, it will spend on initiatives that give poor value for money (paragraph 2.15).

16 The quality of evaluations in some policy areas could be improved at relatively low cost. There could be wider application of some approaches used regularly in labour market and education evaluations, but which were not used in the business support and spatial policy evaluations that we assessed, such as the use of administrative data (paragraph 2.14).

Use of evaluation evidence

17 Our review of the documents provided to HM Treasury by three departments during the 2010 Spending Review found limited references to evaluation evidence, which underpinned only a small proportion of resources that they sought from the Treasury. Evaluation is not the only source of cost-effectiveness evidence, but we would expect to see more reference to evaluation evidence of previous phases, other similar programmes, or evaluations of unsuccessful approaches, to substantiate claims about expected cost-effectiveness (paragraphs 3.8 to 3.10).

18 Impact assessments of policies under consideration rarely include relevant learning from evaluation evidence. Of 261 impact assessments published in 2009-10, only 40 referred explicitly to evaluation within their evidence base (paragraphs 3.11 and 3.12).

19 Public Accounts Committee and NAO reports have criticised departments for absent or poor-quality evaluation. That evidence would help departments monitor impacts, modify policy and help Parliament hold departments to account (paragraphs 3.14 and 3.15).

20 There is little systematic information from the government on how it has used the evaluation evidence that it has commissioned or produced. However, departments were able to point to a few examples where evaluation evidence had clearly informed policy decisions; and some recent NAO reports have welcomed the government's use of evaluation evidence to inform policy design and implementation, such as Community Budgets (paragraph 3.13).

21 The government has acknowledged the problems service commissioners and providers face in accessing and using evaluation evidence. 'What Works' evidence centres for social policy are being set up. Their functions include synthesising evidence and promoting the absorption of evidence. This should help to support better and more informed use of evaluation evidence (paragraph 3.4).

Production, resources and barriers

22 Government departments use a wide range of models to commission and produce evaluations, but the rationale for this variation is not clear. Commissioning is usually done by departments (by policy teams and/or analysts), which can raise questions about the objectivity and credibility of the resulting evaluation. Some departments have set up bodies (which vary in their degree of independence and autonomy) to commission some of their evaluations. In 2011 the Department for International Development (DFID) set up the Independent Commission for Aid Impact (ICAI), which is an advisory non-departmental public body, funded by DFID. It reports to Parliament via the International Development Committee on its findings. The Department for Education established the Education Endowment Foundation (EEF), which is responsible for evaluating interventions aimed at disadvantaged schoolchildren in England. The EEF was intentionally set up with governance (and financial) arrangements removed from the direct influence of officials or ministers. It is not clear why these models have been developed for some areas of spending but not for others (paragraphs 4.2 to 4.5).

23 Information on staff time and budget spent on evaluation by departments is incomplete, so it is difficult for the government to take a view on whether the resources allocated are appropriate. Few departments were able to provide this information to us quickly. In most cases, the information was partial. Departments told us that in 2010-11 they spent £44 million in commissioning evaluation from external sources, and devoted 102 full-time equivalents (FTEs) to evaluation activity, which we estimate represents around £5 million in staff costs (paragraphs 4.10 to 4.13).

24 Independent evaluators outside of the government experience difficulties accessing a range of official and administrative data that can be used to evaluate the impact of government interventions. Given the current government drive to promote greater transparency and openness throughout the public sector, these concerns should be addressed (paragraphs 4.8 to 4.9).

25 Overall, there is a range of barriers to the production and use of evaluation evidence, on both the demand and supply sides. Chief analysts and their evaluation staff consider evaluation timescales and a lack of demand from policy colleagues are key issues. We believe a key factor is the lack of incentives for departments to generate and use evaluation evidence, with few adverse consequences for failing to do so (paragraphs 4.14 to 4.18).

Conclusion and areas for improvement

26 The government spends significant resources on evaluating the impact and cost-effectiveness of its spending programmes and other activities. Coverage of evaluation evidence is incomplete, and the rationale for what the government evaluates is unclear. Evaluations are often not robust enough to reliably identify the impact, and the government fails to use effectively the learning from these evaluations to improve impact and cost-effectiveness.

27 To ensure that there is better understanding of the coverage of evaluation evidence across government, and to encourage greater coverage in future:

- The government should publish a comprehensive overview of the impact and cost-effectiveness evidence that exists across its current interventions. This should be linked, as far as possible, to the corresponding business case and/or impact assessment.
- Departments should publish a list of significant evaluation gaps in their evidence base, and should set out and explain their priorities for addressing those gaps, in accordance with wider strategic priorities and the likely value of evaluation.
- To facilitate independent evaluation to help fill the gaps, departments should publish details of the datasets that they hold, and the support they will offer independent evaluators for research purposes. This should include clear processes for gaining access rights to data.

28 To ensure that evaluations are fit for purpose and provide a robust and reliable basis for decision-making:

- When new policies are announced, departments should explain how they intend to evaluate reliably those policy impacts, and to use the findings in decision-making. This should include an explanation of the policy design choices they have made to facilitate robust evaluation.
- Departments should publish all evaluations with a clear and concise summary of the findings, conclusions and costs of the intervention being evaluated. Reports should include full details of data collected, methods and an independent rating of robustness, using a consistent metric to help explain the reliability of findings.
- The government should review the arrangements that exist for commissioning and producing evaluation activity across the government, with a view to enhancing the robustness, credibility and impact of its evaluation activity.

29 To improve the use of evaluation evidence in developing programmes and other interventions:

- Departments should publish their management response to published evaluation reports. This should explain the degree to which they accept the findings, what they have changed in response, and further action they intend to take.
- Accounting officers should publish the arrangements they have in place for ensuring value for money, and the role of evaluation evidence within that. This would help them to deliver against their responsibilities for the quality of policy advice, as set out in the *Civil Service Reform Plan*.
- HM Treasury should ask departments to provide evaluation evidence in the context of strategic resourcing decisions such as spending reviews, and also incentivise its use in business-as-usual decision-making in government.
- The government should consider how evaluation evidence can be used to support greater scrutiny by and accountability to Parliament, with a view to enhancing the robustness, credibility and impact of its evaluation activity.

30 To improve the transparency and prioritisation of evaluation resources:

- The government should carry out a strategic review and prioritisation of evaluation resources across departments. It should map resources against current evaluation gaps and requirements, to assess if changes should be made to improve the impact and value for money of evaluation resources.

Part One

Coverage of evaluation evidence

1.1 Departments ought to understand what evaluation evidence exists. Where there are gaps, departments should be open about them, prioritise which gaps will be filled and when, and address those priorities. This part:

- examines the government's stated requirements for evaluation;
- sets out our findings on the gaps in the evaluation evidence base; and
- assesses what the departments say about their intention to address gaps.

Requirements to evaluate

1.2 The requirements on departments to evaluate are set out in several government guidance documents.

- *Managing Public Money* states that “one of the essentials of effective internal decision-making is after-the-event evaluation of policy, project and programme outputs and outcomes.”
- The *Treasury Green Book* states that “When any policy, programme or project is completed or has advanced to a pre-determined degree, it should undergo a comprehensive evaluation. Major or ongoing programmes, involving a series of smaller capital projects, must also be subject to ex-post evaluation.”
- The *Magenta Book* complements the Green Book with detailed best practice on evaluation methods. It states that “all policies, programmes and projects should be subject to comprehensive but proportionate evaluation, where practicable to do so.” It goes on to state that there are “a number of formal requirements to evaluate”, listing three examples “when an evaluation might be a requirement”.³
- The *Impact Assessment Guidance* explains that “measures that include a statutory review provision should be formally reviewed within five years of enactment, and then regularly on a five-year cycle.⁴ The main vehicle for this review should be a post-implementation review (PIR). PIRs should be proportionate, ranging from light-touch to economic evaluation”.

³ These are: where a formal impact assessment was required and which are subject to post-implementation review; regulations containing a sunset clause or a duty to review clause; and projects subject to a post-implementation review as part of the Gateway review process.

⁴ Department for Business, Innovation & Skills, *Better regulation framework manual: practical guidance for UK government officials*, July 2013. Available at: www.gov.uk/government/publications/better-regulation-framework-manual

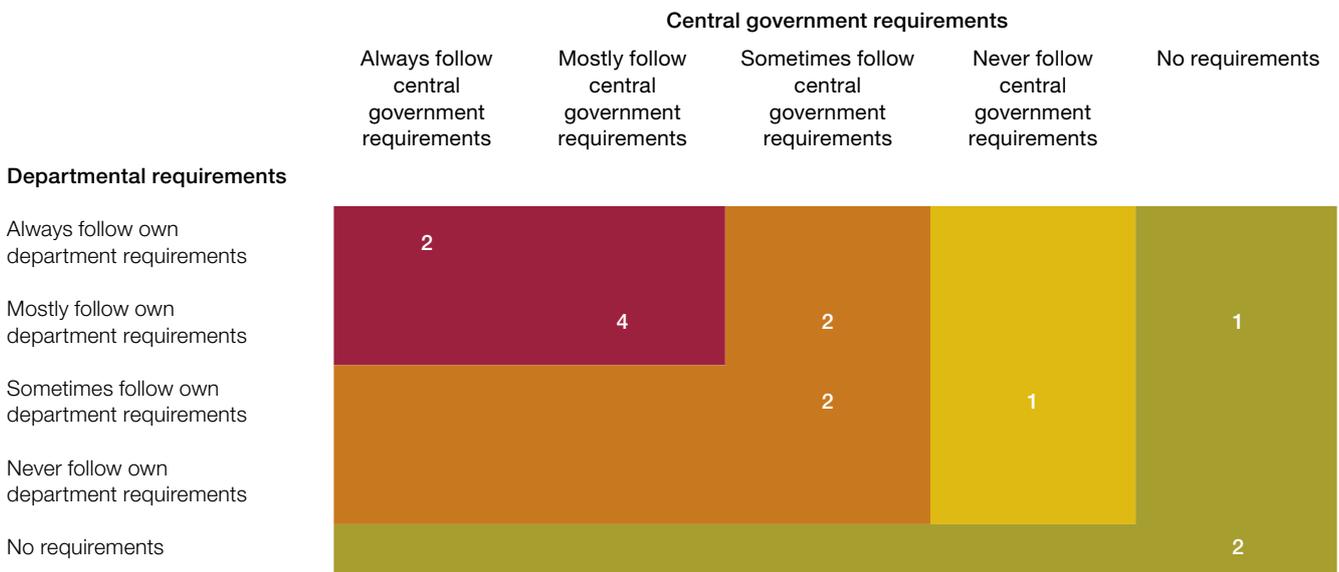
1.3 Evaluation incurs costs, which must be set against the likely benefits. Government evaluation guidance provides little practical guidance on how to implement the principle of proportionality in what and how to evaluate.⁵ It does, however, set out situations where greater resources for evaluation can be justified. These include:

- where policies are high-risk, high-profile or large-scale;
- where policies have a high degree of uncertainty or variation of impact;
- where policies are pilots and may be repeated or rolled out more widely; or
- where evaluation can be particularly influential in developing future policy, or can improve knowledge where existing evidence is weak.

1.4 Some departments have their own guidance on what to evaluate, and how it should be carried out. Examples include the Departments for Energy & Climate Change, Education and International Development. Responses from 15 departmental chief analysts who completed our survey show that departments vary in the extent to which they follow central and department-specific requirements on cost-effectiveness evaluation (**Figure 1**). Although requirements do exist, one department chief analyst said they never follow them, and two believed there were no such requirements.

Figure 1

Extent to which cross-government and departmental requirements are followed



Note

1 Figures show the number of departments in each category.

Source: National Audit Office survey of chief analysts

Gaps in evaluation

1.5 Government does not publish a comprehensive list of evaluation evidence and therefore it is difficult to establish what evaluation evidence is available across the full £700 billion of public expenditure. But we did find evidence of significant gaps.

1.6 First, we looked at the coverage of PIRs. National Audit Office (NAO) work⁶ identified that in 2005, only 50 per cent of impact assessments (IAs) committed to a PIR, and only half of those were subsequently completed. There are more recent positive signs. We reviewed impact assessments published in 2009-10, and found that 81 per cent of IAs committed to a PIR, and 9 per cent committed to ongoing monitoring as part of established processes. Ten per cent made no commitment to PIR or monitoring, and most of those either stated that a PIR was unnecessary or left a blank entry. Of 15 departments, seven had at least one IA in that category.

1.7 Second, we reviewed nearly 6,000 analytical and research documents published between 2006 and 2012 on 17 main government department websites. Only 305 of these were impact evaluations; and of these, only 70 made an assessment of cost-effectiveness. We were able to identify £12.3 billion of overall programme expenditure (in cash terms) evaluated by 41 of those evaluations.

1.8 Third, chief analysts told us that they recognise that gaps exist in their cost-effectiveness evaluation evidence, and 30 per cent of evaluation analysts consider there to be major gaps. Most departments told us that they regularly review the gaps in their evaluation evidence base. Three departments said that they do this at least every six months, and seven departments do so at least annually (**Figure 2**).

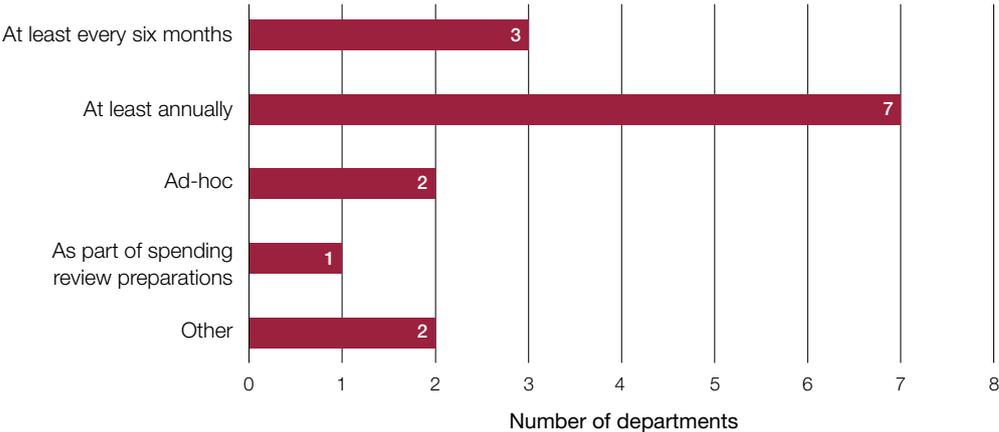
1.9 Fourth, we established that there is a lack of evaluation in progress or planned for the major projects identified by each department in their business plans (**Figure 3**).⁷ Departmental chief analysts told us that they intend to evaluate only 27 of the 71 major projects. Only £90 billion of spending will be evaluated, from a total of £156 billion. The Department for Work & Pensions (DWP), Department for Energy & Climate Change (DECC), and Department for Transport (DfT) account for the majority of this.

⁶ National Audit Office, *Post-implementation review of statutory instruments: analysis of the extent of review by government departments*, December 2009. Available at: www.nao.org.uk/report/briefing-for-merits-of-statutory-instruments-committee-analysis-of-the-extent-of-the-review-by-government-departments

⁷ Major projects defined in quarterly data summaries.

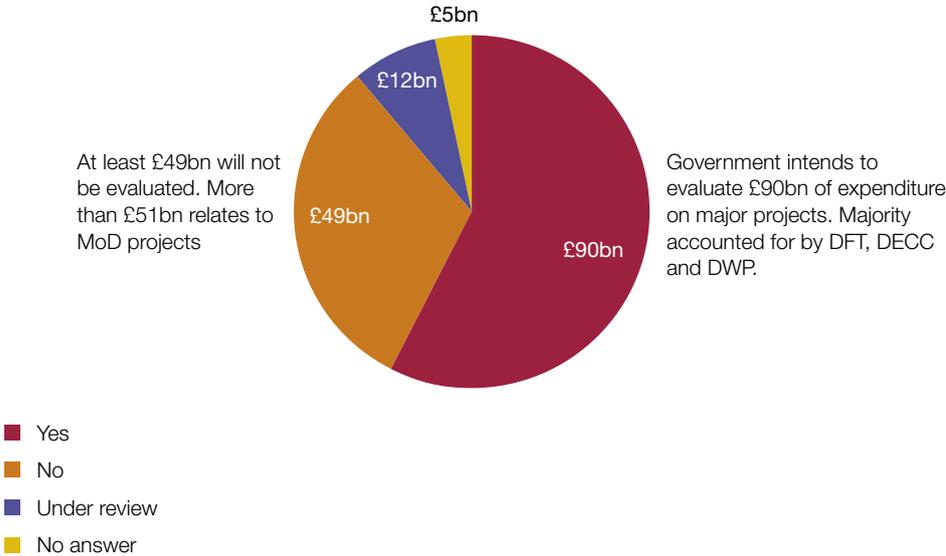
Figure 2
Review of evaluation evidence gaps in government

How often do you review the key gaps in your cost-effectiveness evaluation evidence?
(15 departments)



Source: National Audit Office survey of chief analysts

Figure 3
Government intentions to evaluate major project spend (£bn)



Note
1 Based on analysis using major project spend for each department.

Source: Analysis of chief analyst survey and quarterly data summaries

Evaluation strategies and plans

1.10 Departments need to have a clear strategy for their evaluation activities and plans for what they will evaluate. Government Social Research guidance⁸ suggests that departments publish an annual research strategy detailing proposed projects, details of contracts once awarded, or when in-house work is due to start. **Figure 4** provides examples.

1.11 We found that eight departments had a strategy that covers evaluation, but six departments do not. Most of these do not have plans to publish such a strategy in the next year (**Figure 5**).

1.12 Departments should have specific plans in place to produce evaluation evidence ahead of key decision-making. Of the departments who responded to our survey, 12 had a forward plan for delivering evaluations. Some publish plans, others include them within published business plans, while others have only internal plans. Two did not have a plan and did not intend to produce one. We reviewed structural reform plans from 17 departments and found that eight included evaluation activities within their plans. These set out the key actions the department will take to implement its coalition priorities, and are published as part of departmental business plans. Published evaluation plans are not always implemented; the Department for Business, Innovation & Skills (BIS), for example, has published only three of the five evaluations that it committed to in its 2010 strategy.

Figure 4

Summary of departmental evaluation strategies in BIS and DfT

Department for Business, Innovation & Skills (BIS) – The evaluation strategy (published 2010) sets out the principles by which BIS will evaluate its policies, based around ensuring a robust governance framework, embedding an evaluation culture and evaluation of key policies. BIS says it will ensure that evaluations are designed early in the development of new policy, and are undertaken when appropriate and in a proportionate manner. To take this forward, BIS plans to develop an evaluation programme incorporating the principles of the evaluation strategy. The evaluation programme will outline specific policies to be reviewed and the timescales for such evaluations. It is envisaged that the strategy and programme will be reviewed every three to five years. However, as of October 2013, no evaluation programme had been published.

Department for Transport (DfT) – In March 2013, DfT published a monitoring and evaluation strategy. The strategy sets a framework for good-quality monitoring and evaluation evidence, and sets out three objectives including establishing a monitoring and evaluation programme, a robust governance framework and embedding a culture of monitoring and evaluation.

The strategy explains the importance of monitoring and evaluation evidence to accountability, decision-making and wise investment of public funds. It sets out the criteria for how DfT will establish priorities. DfT published a Monitoring and Evaluation Programme in October 2013 which sets out its plans for the interventions it will evaluate. This will be updated annually.

Sources: Department for Business, Innovation & Skills, *Evaluation strategy: the role of evaluation in evidence-based decision-making*, August 2010. Available at: www.gov.uk/government/publications/using-evaluation-to-inform-decisions-about-policy, accessed 14 November 2013. Department for Transport, *Monitoring and evaluation strategy*, March 2013. Available at: www.gov.uk/government/publications/monitoring-and-evaluation-strategy, accessed 14 November 2013. Department for Transport, *Monitoring and evaluation programme*, October 2013. Available at: www.gov.uk/government/publications/dft-monitoring-and-evaluation-programme-2013, accessed 14 November 2013

8 Government Social Research Unit, *Publishing research in government: GSR publication guidance*, January 2010. Available at: www.civilservice.gov.uk/networks/gsr/publications

Figure 5

Departments' evaluation strategies and forward plans

Department	Evaluation strategy?	Updates being prepared?	Forward plan of evaluation?
DFID	Yes	Recently published	Yes
DfT	Yes	Recently published	Yes
DECC	Yes	In progress	Yes
DfE	Yes	In progress	Yes
MoD	Yes	Planned	Yes
BIS	Yes	Planned	Yes
MoJ	Yes (internal)	In progress	Yes
DCMS	Yes	In progress	No
DCLG	No	In progress	No
HO	No – but in business plan	No plans	Yes
FCO	No – but in business plan	No plans	Yes
HMRC	No	No plans	Yes
DEFRA	No	No plans	Yes
DH	No	No plans	Yes
DWP	No	No plans	Yes

Note

- 1 MoJ = Ministry of Justice; DECC = Department of Energy & Climate Change; DFID = Department for International Development; DfT = Department for Transport; BIS = Department for Business, Innovation & Skills; DfE = Department for Education; HMRC = HM Revenue & Customs; DCLG = Department for Communities and Local Government; MoD = Ministry of Defence; HO = Home Office; FCO = Foreign & Commonwealth Office; DEFRA = Department for Environment, Food & Rural Affairs; DH = Department of Health; DCMS = Department for Culture, Media & Sport; and DWP = Department for Work & Pensions.

Source: National Audit Office survey of departmental chief analysts

1.13 The Government Office for Science conducted Science and Engineering Assurance Reviews of eight central government departments between 2010 and 2012 (**Figure 6**). For four of those eight departments, it made recommendations for improvements relevant to evaluation activity. These recommendations covered leadership and resources, better prioritisation and design of evaluation, and the use of a range of sources of evaluation evidence in a more integrated way, to learn lessons.

Figure 6 Recommendations of Science and Engineering Assurance Reviews

Resources

- Strengthen the research and evaluation leadership (DWP)
- Additional evaluation staff needed (DfT)
- Earlier engagement of analysts in policy, evaluation and delivery (DWP).

Prioritisation

- Better alignment of research/evaluation planning with business plans (DfT)
- Include stakeholders in developing analysis and innovation strategy (DfE)
- Prioritise and design evaluation carefully to ensure good value for money (DWP)
- Resources should target pilot and other policy evaluations that will yield useful information about impact and cost-effectiveness. (DfE).

Better use of evidence

- Periodically review the evidence – including evaluation (DfE)
- Greater use of evaluation evidence in approvals processes (DfT)
- Greater use of evaluation evidence from other countries (DfE)
- Use of evaluation to validate methodologies to inform policy-making (DfT)
- Learn lessons from evaluation and use it across the organisation (DFID).

Note

- 1 DfE = Department for Education; DfT = Department for Transport; DWP = Department for Work & Pensions; DFID = Department for International Development.

Source: National Audit Office analysis of SEA reviews published between 2010 and 2012. Available at: www.bis.gov.uk/go-science/science-in-government/reviewing-science-and-engineering/completed-reviews

Part Two

Quality of evaluation evidence

2.1 *The Civil Service Reform Plan* says that “Permanent Secretaries must be accountable for the quality of the policy advice in their department and be prepared to challenge policies which do not have a sound base in evidence or practice”.⁹

2.2 In this part, we focus on the quality of the ex-post evaluation evidence that the government commissions and publishes. Evaluations which aim to identify policy impacts should provide convincing evidence that the impacts can be attributed to the intervention, so that resource allocation, policy and implementation decisions can be properly informed.

2.3 This part:

- explains what we mean by quality and how it can be assessed;
- reports the evidence we have found on quality, and how that is related to the claims made for the efficacy of the policy; and
- discusses departmental arrangements for quality assurance.

Quality of evaluation in government

2.4 The quality of evaluation design and its implementation has consequences for the reliability with which policy impacts can be determined. There are inevitable trade-offs in terms of time to produce the evaluation, its cost, and quality. We recognise that government evaluations may not always achieve the highest levels of robustness because of these trade-offs and constraints.

9 HM Government, *The Civil Service Reform Plan*, June 2012. Available at: my.civilservice.gov.uk/reform/the-reform-plan/

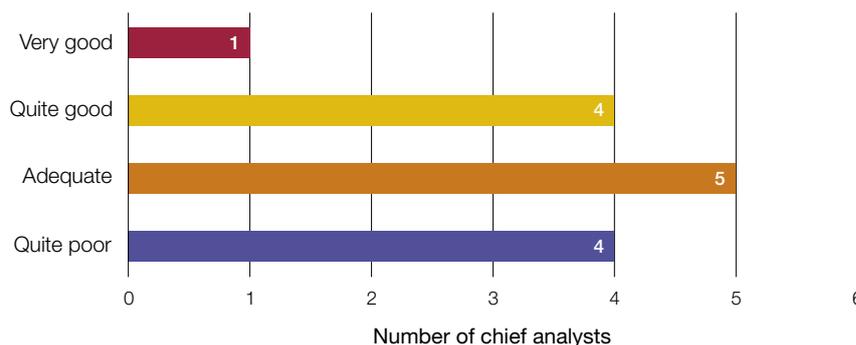
2.5 Twenty-one National Audit Office (NAO) reports published between May 2008 and October 2012 covering most departments commented on the weakness of evaluations. Key issues include weakness in comparison groups, the use of self-evaluation of performance without external scrutiny, and the use of monitoring rather than evaluation (i.e. there were no comparisons against a counterfactual, which examines what would have happened without the intervention). For example:

- *The management of adult diabetes services in the NHS* noted that service evaluation compared the performance of neighbouring primary care trusts rather than comparable peers.¹⁰
- *Partnering for school improvement* highlighted the use of self-evaluation in 75 per cent of cases, and noted that only 1 per cent of the partnerships surveyed were monitored or evaluated by the school's governors.¹¹
- *Implementing the government ICT strategy* noted that there were no clear criteria for measuring business outcomes.¹²

2.6 In general, departmental chief analysts do not consider their evaluation evidence to be strong. While ten consider theirs to be adequate or better, four consider the quality of their cost-effectiveness evaluation evidence to be “quite poor” (**Figure 7**). A significant proportion of departmental analysts who responded to our survey consider evaluation evidence in their departments to be not fit for purpose (**Figure 8**).

Figure 7

Chief analysts – “How would you best describe the quality of cost-effectiveness evaluation evidence in your department?”



Note

1 Responses from 14 departments.

Source: National Audit Office survey of chief analysts

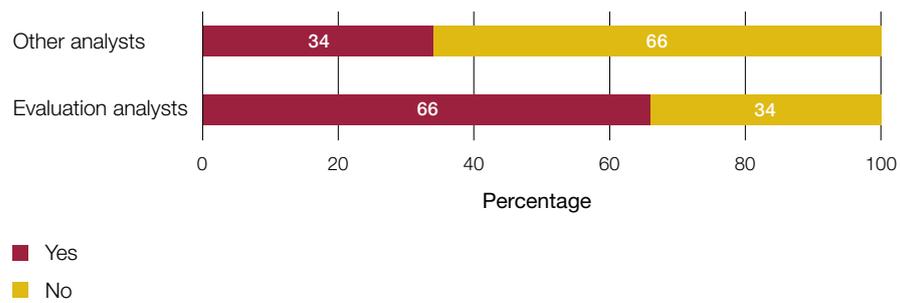
10 Comptroller and Auditor General, *The management of adult diabetes services in the NHS*, Session 2012-13, HC 21, National Audit Office, May 2012.

11 Comptroller and Auditor General, *Partnering for school improvement*, Session 2008-09, HC 822, National Audit Office, July 2009.

12 Comptroller and Auditor General, *Implementing the government ICT strategy: six-month review of progress*, Session 2010-12, HC 1594, National Audit Office, December 2011.

Figure 8

Analysts – “In your opinion, is evaluation evidence in your department fit for purpose?”

**Note**

1 Responses from 110 analysts.

Source: National Audit Office survey of analysts in government

Assessment of fitness for purpose of evaluations

2.7 The quality of evaluations can be assessed against a scale that focuses on the quality of the counterfactual against which policy is compared. A robust counterfactual is important because otherwise the impact attributable to the intervention may be overstated.

2.8 One framework is the Maryland Scale. This is a five-point scale designed by the University of Maryland to classify the strength of evidence. It is also used as the basis of a toolkit¹³ published by the Government Social Research Service (GSR), which allows users to assess research evidence. The creators of the Maryland Scale state that only studies with a robust comparison group design (level 3 and above) can provide evidence that a programme has caused the reported impact.

2.9 A similar framework is provided in *Quality in policy impact evaluation*.¹⁴ This was published by the government as supplementary guidance to its evaluation guidance. The document sets out the strengths and weaknesses of evaluation designs. It explains that higher-quality research design helps to more reliably attribute observed outcomes to policy. **Figure 9** overleaf provides a summary.

¹³ Available at: www.civilservice.gov.uk/networks/gsr/resources-and-guidance

¹⁴ Available at: www.gov.uk/government/uploads/system/uploads/attachment_data/file/190984/Magenta_Book_quality_in_policy_impact_evaluation__QPIE_.pdf

Figure 9

Measuring the robustness of evaluation

Maryland Scale

Strong research designs in the measurement of attribution

Level 5 – Random assignment and analysis of comparable units to programme and comparison groups.

Level 4 – Use of statistical techniques to ensure that the programme and comparison group were similar and so fair comparison can be made.

Level 3 – Comparison between two or more comparable groups/areas, one with and one which does not receive the intervention.

Weaker/riskier research designs in the measurement of attribution

Level 2 – Evaluation compares outcomes before and after an intervention, or makes a comparison of outcomes between groups or areas that are not matched.

Level 1 – Evaluation assesses outcomes after an intervention – but only for those affected. No comparison groups used.

Government Guidance (QIPE)

Random allocation/experimental design.

Individuals or groups are randomly assigned to either the policy intervention or non-intervention (control) group and the outcomes of interest are compared. There are many methods of randomisation from field experiments to randomised control trials.

Quasi-experimental designs

Intervention group vs well-matched counterfactual. Outcomes of interest are compared between the intervention group and a comparison group directly matched to the intervention group on factors known to be relevant to the outcome.

Strong difference-in-difference design. Before and after study which compares two groups where there is strong evidence that outcomes for the groups have historically moved in parallel over time.

Intervention vs unmatched comparison group.

Outcomes compared between the intervention group and a comparison group.

Predicted vs actual – Outcomes of interest for people or areas affected by policy are monitored and compared to expected or predicted outcomes.

No comparison group – A relationship is identified between intervention and outcome measures in the intervention group alone.

Sources: Maryland Scale: Available at: www.civilservice.gov.uk/networks/gsr/resources-and-guidance; Government guidance (QIPE): Available at: www.gov.uk/government/uploads/system/uploads/attachment_data/file/190984/Magenta_Book_quality_in_policy_impact_evaluation__QPIE_.pdf

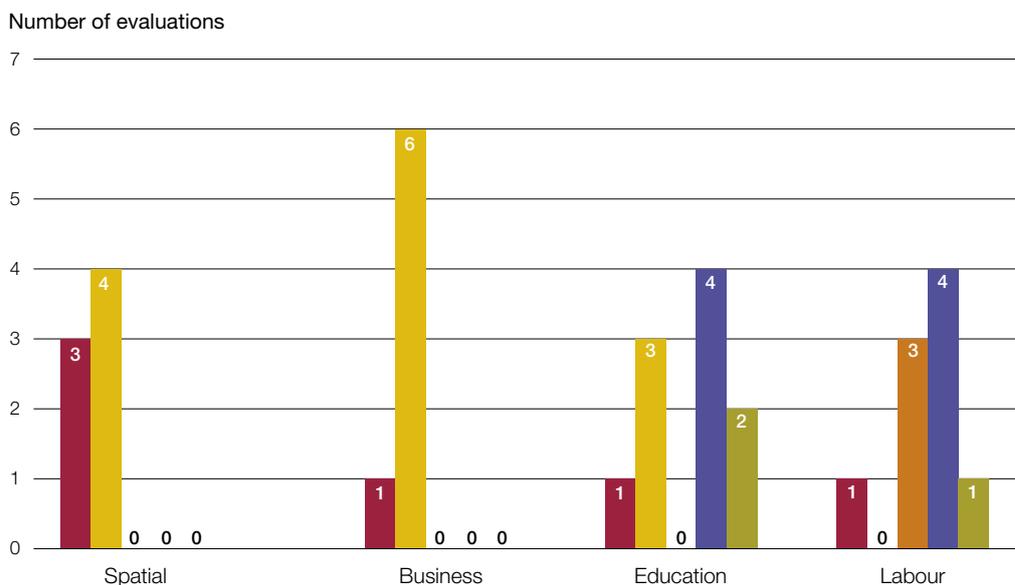
2.10 Government guidance explains that some of the requirements for a reliable impact evaluation may not always be met, and may be outside the control of the evaluator. It explains that there are measures that could be put in place before the policy starts, and in particular it emphasises that “the ability to obtain good evaluation evidence rests as much on the design and implementation of the policy as it does on the design of the evaluation”. This recognises that policy-makers have responsibility for securing good evidence, and that relatively minor adjustments in policy implementation can greatly improve the ability to obtain high-quality evaluation evidence.

2.11 Against this background, we have reviewed a selection of recent evaluations and assigned a score using the Maryland Scale, to understand the quality of published evaluations. We based this assessment on published outputs only (the *Magenta Book* explains that it is important for evaluation results, including methodological approaches, to be published for the purposes of public accountability and peer review, and to support learning over time). Our scoring does not reflect the challenges that government or their contractors may have faced in producing impact evaluations.

2.12 Our review covered 34 published evaluations across four policy areas: spatial policy, active labour markets, business support and education policy. The review was conducted by a team of evaluation experts from the London School of Economics (available at: www.nao.org.uk/report/evaluation-government/). The team focused on the extent to which the measured impacts can be reliably attributed to the policy being evaluated, and they used the Maryland Scale to summarise their assessments. The review identified good practice, weaknesses and recommended improvements.

2.13 The review found that the fitness for purpose of the evaluations it examined varied within and between the four policy areas, but was generally poor. There were some high-quality evaluations in the areas of active labour markets and education: six of nine education reports and eight of ten labour market evaluations were of a sufficient standard to have confidence in the impacts attributed to policy. Evaluations in the areas of business support and spatial policy were considerably weaker. None of 14 evaluations, when assessed against the Maryland Scale, were at the threshold (level 3) for evidence that the programme has caused the reported impact (**Figure 10** overleaf). In part, this finding may reflect the difficulty in implementing effective evaluation design.

Figure 10
Robustness and reliability of ex-post evaluation of policy impact



Maryland Scale, 5 is the highest level of robustness

- Level 1
- Level 2
- Level 3
- Level 4
- Level 5

Note

1 The numbers in the chart refer to the number of evaluation reports in each category of robustness.

Source: National Audit Office presentation of in-depth analysis. Selection of 34 evaluations

2.14 The differences between the four policy areas cannot be attributed to policy challenges alone. There is scope for government to improve the quality of evaluations at relatively low cost. For example:

- Departments could use administrative data, for example to supplement evidence from surveys to provide more objective evidence of policy impact.
- Studies in education and labour markets made good use of policy design, which rolled out implementation in different geographic areas over time. This helped to provide a robust counterfactual to reliably assess policy impact.

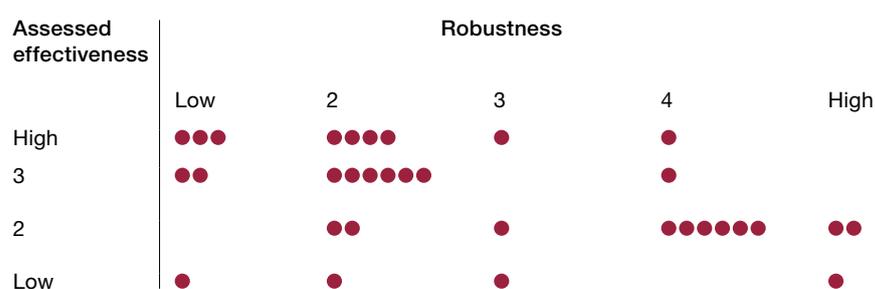
2.15 There are indications that some of the least robust evaluations were more positive in their assessment of effectiveness. This creates the risk that funds may be spent in the mistaken belief that those initiatives are effective. We reviewed the evaluations to identify the strength of the impacts claimed and the extent to which those reports noted caveats or uncertainties. **Figure 11** shows a cluster of reports that rate poorly in terms of robustness (1 and 2 on the Maryland Scale), while describing positive impacts with few caveats or uncertainties (scoring 3 and 4 on assessed effectiveness). There is a second cluster of reports that rate highly in terms of robustness (4 or 5 on the Maryland Scale), while being much more careful about the strength of impact and noting greater uncertainties or caveats (1 and 2 on the assessed effectiveness).

Ensuring quality in evaluation

2.16 Government guidance explains that independent scrutiny of outputs by peer review is a good way to ensure quality and demonstrate impartiality of findings.¹⁵ The routine publication of research and evaluation, along with methods, means that a wide range of external experts can scrutinise and challenge the findings.

2.17 Departments should have clear arrangements for quality assurance and governance in place, to deliver fit-for-purpose evaluation outputs. The majority of departments do have some arrangements in place (see Appendix Three).

Figure 11
Relationship between robustness and claimed impacts in evaluations



- Note**
- 1 Robustness assessed on Maryland Scale. Assessed effectiveness, rated low to high.
 - Low = Small or insignificant effects.
 - 2 = Mixed effects, positive for some, negative or insignificant for others.
 - 3 = Positive effects, with some caveats or uncertainties noted.
 - High = Significant positive impacts, no or only minor caveats or uncertainties noted.

Source: National Audit Office analysis of external assessment by London School of Economics

15 Government Social Research Unit, *Publishing research in government: GSR publication guidance*, January 2010. Available at: www.civilservice.gov.uk/networks/gsr/publications

2.18 There are some promising developments in departments that could be shared and built upon across government:

- The Department for Work & Pensions (DWP) requires potential contractors to submit a sample of previous work before they may bid for evaluation contracts.
- The Department for Communities and Local Government (DCLG) has internal research gateway processes, to scrutinise evaluation designs and plans before they are commissioned.
- The Department for International Development (DFID) ensures that analysts have the necessary skills to oversee and manage, or design and deliver evaluations, with appropriate accreditation schemes.
- The Department for Work & Pensions (DWP) and Department of Energy & Climate Change (DECC) use external peer-review and scrutinise evaluation outputs, during their delivery and before they are published.
- The government has attempted to introduce greater independence in evaluation in some areas.¹⁶
- 'What Works' centres will inform decisions on £200 billion of public spending (see paragraph 3.4).
- Business, Innovation & Skills (BIS) is using randomised control trials to test approaches to help small and medium-sized enterprises (SMEs) overcome barriers to achieving growth as part of the Growth Vouchers programme.¹⁷

¹⁶ See Figures 14 and 22.

¹⁷ Available at: online.contractsfinder.businesslink.gov.uk/Common/View%20Notice.aspx?site=1000&lang=en¬iceid=1048743&fs=true

Part Three

Use of evaluation evidence

3.1 In this part, we focus on how the government disseminates and uses evaluation. Government officials should take account of relevant evaluation evidence to provide advice and support to ministers and senior civil servants so they can take informed decisions. We recognise that those decisions are made in the context of political and operational considerations and that evaluation evidence on cost-effectiveness is only one input.

3.2 Recent government guidance on evaluation sets out the role of evaluation in supporting evidence-based policy-making.¹⁸

3.3 We have drawn on this guidance to identify three main uses for evaluation, which we assess in this part:

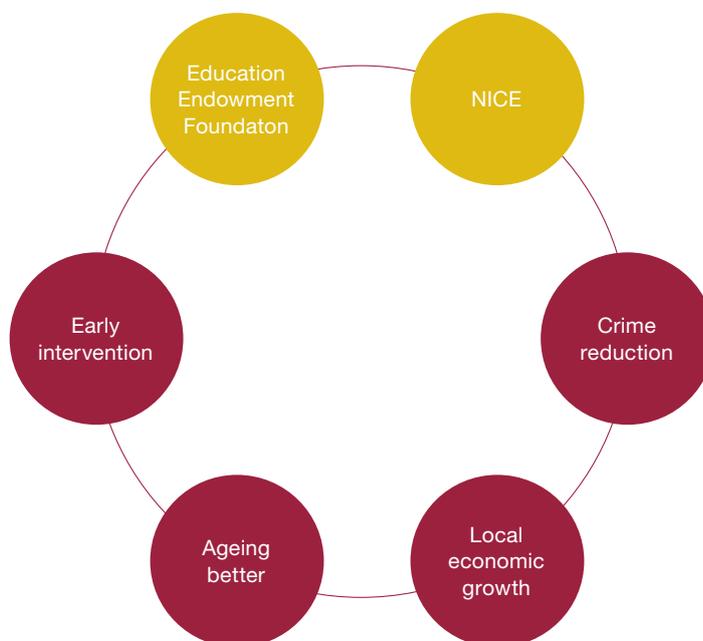
- to inform strategic resource allocations, such as in spending reviews;
- to inform decisions about policies and programmes, in terms of the design of new programmes, and improving or stopping existing programmes; and
- to support accountability, by demonstrating the costs and benefits of spending.

3.4 There are potential uses of evaluation evidence outside central government, including informing decisions by local service commissioners and providers. We have not assessed this as part of this study, but we note the announcement by the government to establish a number of 'What Works' centres.¹⁹ These are intended to ensure that local practitioners and commissioners can access and understand the relevant evidence base. The network includes the existing National Institute for Health and Care Excellence (NICE) and the Education Endowment Fund (EEF), and new centres covering crime reduction, local economic growth, ageing and early intervention (**Figure 12** overleaf).

¹⁸ HM Treasury, *Magenta Book: guidance for evaluation*, April 2011. Available at: www.hm-treasury.gov.uk/data_magentabook_index.htm

¹⁹ HM Government, *What works: evidence centres for social policy*, March 2013. Available at: www.gov.uk/government/publications/what-works-evidence-centres-for-social-policy

Figure 12
What Works' network



Each 'What Works' centre will:

- present and disseminate findings in a form that can be understood, interpreted and acted on;
- undertake systematic assessment of relevant evidence and production of synthesis;
- develop a common currency for comparing the effectiveness of interventions;
- advise interventions and projects to ensure they can be evaluated effectively;
- kite-mark and recommend interventions; and
- identify research and capability gaps and work with partners to fill them.

Source: Cabinet Office

Disseminating evaluation results

3.5 Evaluation results may be used by decision-makers inside and outside government, as well as a wider range of stakeholders for accountability purposes. To support the use of evaluation, these decision-makers need to have easy access to the evidence, and understand the reliability that they can place on the findings.

3.6 Government guidance states that the products of social research should be made publicly available and published within 12 weeks of departments receiving a final draft report.²⁰ We found evidence that research and evaluation reports are not always published in line with guidance:

- In our survey of government analysts, 17 per cent involved in evaluation said that they have never published a report on their website, and only 45 per cent said that this happens in all cases.
- In 2011 and 2012, the Department for Communities and Local Government (DCLG) published previously unreleased research, which had been commissioned under the previous administration. For transparency and accountability purposes, this is a valuable exercise and is welcome for those reasons. Those evaluation reports cost the government up to £1.1 million (see **Figure 13** overleaf).

3.7 Our survey of government analysts suggests that policy-makers do have access to findings, while ministers do not always see evaluation reports. Forty-five per cent of analysts said that they always shared evaluation results with ministers, 27 per cent do in most cases, 17 per cent in some cases, and 11 per cent said that they never shared reports with ministers. One-third of evaluation analysts said that interested parties outside the government cannot always access ex-post evaluations.

Informing strategic resourcing decisions

3.8 When carrying out spending reviews, the government takes decisions on strategic multi-year allocation of resources between departments. The most recent of these was completed in 2013. A recent National Audit Office (NAO) report looked at evidence on the cost-effectiveness of capital and resource spending requested by HM Treasury and provided by departments as part of the Spending Review 2010.²¹ The NAO report concluded that information on the value of resource spending, which represents nearly 90 per cent of controllable spending, was patchy and often hard to compare across programmes and departments.

²⁰ Government Social Research Unit, *Publishing research in government: GSR publication guidance*, January 2010. Available at: www.civilservice.gov.uk/networks/gsr/publications

²¹ Available at: www.nao.org.uk/report/managing-budgeting-in-government/

Figure 13

Evaluations not published when complete; released after a delay

Evaluation	Cost (£)	Commissioned	Published
International Migration and Rural Economies	24,275	2009	March 2011 ¹
Condensation risk – impact of improvements to Part L and robust details on Part C	158,560	2003	May 2011 ²
Long-term evaluation of local area agreements and local strategic partnerships: Final report	47,898	2007	June 2011 ³
Evaluation of inspiring communities: scoping report including the theory of change and outcomes framework	40,898	2009	June 2011 ³
Evaluation of the Enhanced Housing Options Trailblazers programme	406,000	2009	October 2011 ⁴
Assessment of the Decent Homes programme – Final report	67,900	2009	October 2011 ⁴
Quirk asset transfer demonstration programme	49,570	2005	March 2012 ⁵
Process evaluation for Communitybuilders	40,345	2009	March 2012 ⁵
Sharing data to improve local employment outcomes: Evaluation of the local data-share pilots	56,000	2009	March 2012 ⁵
Evaluation of the REACH National Role Model Programme	97,080	2008	March 2012 ⁵
Unlocking Capacity – lessons learned from four Connecting Communities areas	94,873	2005	March 2012 ⁵
Total	1,083,399		

Notes

- 1 Available at: www.gov.uk/government/speeches/unpublished-research-reports-immigration-the-economy-and-regeneration
- 2 Available at: www.gov.uk/government/speeches/unpublished-research-reports-building-and-the-environment
- 3 Available at: www.gov.uk/government/speeches/unpublished-research-reports-housing-and-local-government
- 4 Available at: www.gov.uk/government/speeches/unpublished-research-reports-housing
- 5 Available at: www.gov.uk/government/speeches/unpublished-research-reports-communities

Source: Department for Communities and Local Government

3.9 Departments may draw on a range of evidence including ex-ante appraisal evidence, evaluation evidence and other analysis to inform their bids. Chief analysts explained that, in some cases, the contribution of evaluation evidence in spending reviews and policy-making may not always be fully documented.

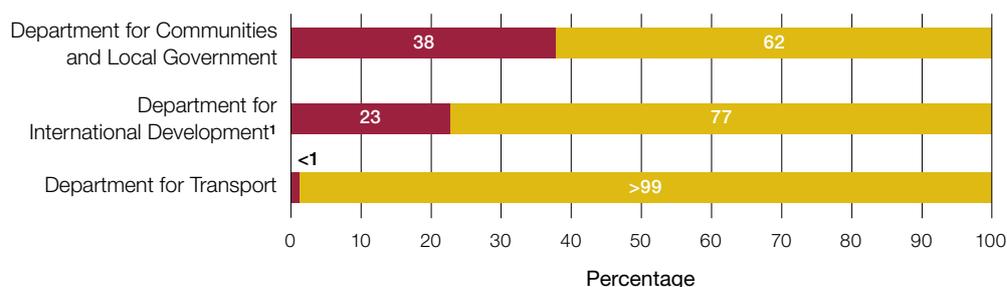
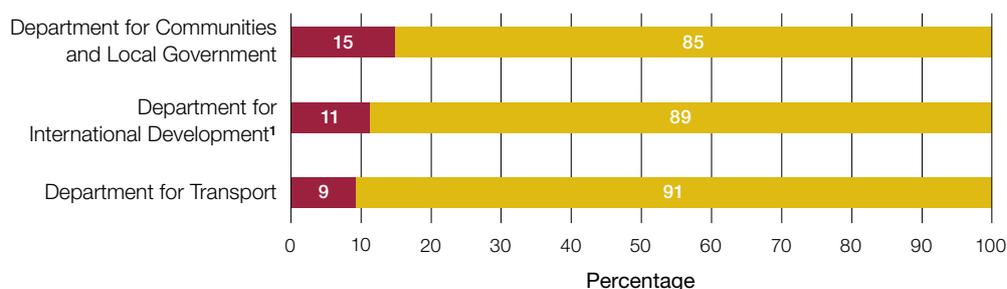
3.10 Using written evidence available for three departments covered by our 2012 report, we examined the extent to which the documents submitted to HM Treasury for Spending Review 2010 referred to evaluation evidence. **Figure 14** overleaf shows that a small fraction of bids by DCLG, the Department for International Development (DFID) for its bilateral programme, and the Department for Transport (DfT) explicitly referred to evaluation evidence. We found wide variation among the three departmental submissions we examined.

- In the case of DFID, we found that references to evaluation evidence in allocating its bilateral aid expenditure was highly variable between country plans and thematic areas. Only five of 13 thematic areas of spending referred to any ex-post evaluation evidence. We found that of 25 country-specific operational plans, 17 did not refer to any evaluation evidence. However, three countries had plans where over 35 per cent of spending was underpinned by ex-post evaluation evidence. In 2011, the NAO's *DFID Financial Management Report*²² commented positively on changes that DFID had made in its approach to strategic allocation of resources.²³
- DfT is generally considered to be strong at ex-ante option appraisal, which we note in our report on reducing costs in DfT.²⁴ However, we identified references to ex-post evaluation evidence only in the areas of sustainable transport, cycling, bus subsidy and road safety. Those evaluations underpinned around 5 per cent of DfT's proposed budgets submitted to HM Treasury. However, DfT's chief analyst told us that officials also reviewed with their Secretary of State how well evaluation evidence of roads schemes confirmed the reasonableness of the evidence from appraisals.
- In the case of DCLG, we found that 38 per cent of capital spending and 15 per cent of resource spending was underpinned by documents that explicitly referenced cost-effectiveness evaluation evidence.

²² Available at: www.nao.org.uk/report/department-for-international-development-financial-management-report/

²³ We did not analyse the Department's bids for its core funding of multilateral organisations as part of our review of evaluation in government. Those bids were informed by the Department's multilateral aid review and we concluded in 2012 that it provided a much improved basis for deciding how to allocate funding.

²⁴ Available at: www.nao.org.uk/report/reducing-costs-in-the-department-for-transport/

Figure 14Spending review bids¹ underpinned by evaluation evidence**Capital Expenditure****Resource Expenditure**

■ Yes

■ No

Note

1 Covers DFID bilateral aid spending and does not cover multilateral aid spending.

Source: National Audit Office analysis of spending review documents from Department for Communities and Local Government, Department for International Development and Department for Transport

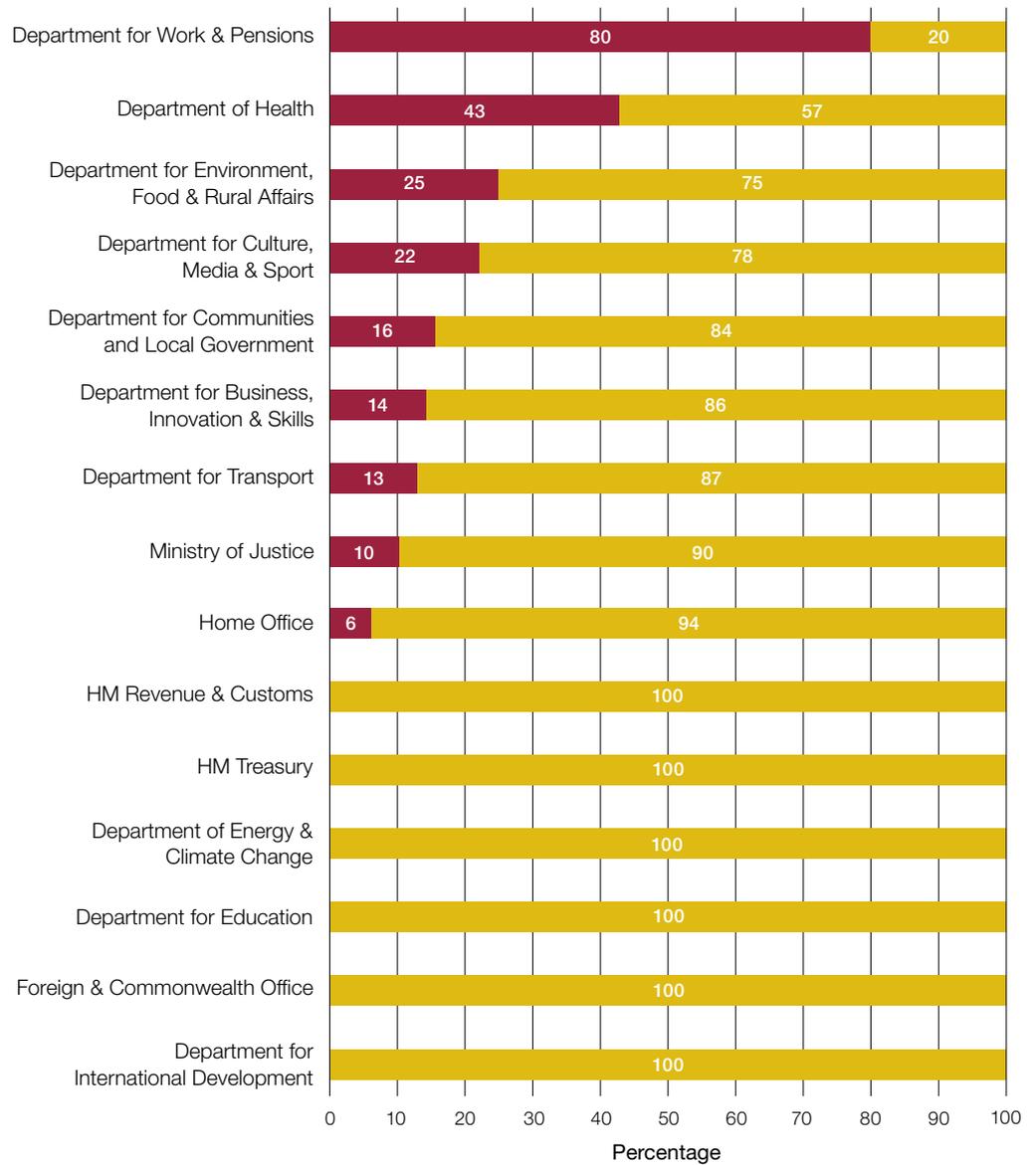
Informing policy decisions

3.11 Government evaluation guidance explains that informing policy decisions is the main purpose of evaluation. An impact assessment (IA) is generally required for all UK government regulatory interventions, and is intended to assess and present the likely costs, benefits and risks of proposals. Departments may draw on a range of evidence, including available evaluation evidence. We reviewed IAs and surveyed departmental chief analysts to understand how evaluation evidence influences policy decisions.

3.12 Departments vary in the extent to which their IAs refer to evaluation evidence. For this study, we reviewed all of the 261 final IAs published in 2009-10. We found that only 40 referred explicitly to evaluation evidence. Six departments that published final IAs in 2009-10 did not include any references to evaluation evidence (**Figure 15**). This fits with the evidence from government analysts: 21 per cent said that they 'frequently' use the evidence for this purpose; 51 per cent said they 'sometimes' use it, and the remaining 28 per cent said they 'rarely or never' use the evidence in this way.

Figure 15
Use of evaluation in impact assessments (2009-10)

Use of evaluation evidence in 261 impact assessments (2009-10)



- Impact assessments which reference evaluation evidence
- Impact assessments which do not reference evaluation evidence

Note

1 DECC was established in 2008, which may have some bearing on the use of evaluation evidence in their impact assessments.

Source: National Audit Office analysis (BRE impact assessment library)

3.13 A number of departmental chief analysts gave examples of how evaluation evidence has been used to inform development, modification or termination of specific policies. **Figure 16** shows a selection of examples where they felt confident that they could draw a direct link between the evidence and subsequent policy decision.

Informing accountability

3.14 Evaluation evidence can help Parliament and taxpayers hold the government to account. Many NAO reports have highlighted the lack of evaluation evidence. Specifically, 42 value-for-money (VFM) reports (of 252 reports published between May 2008 and October 2012) made criticisms. Of these:

- thirty-one criticised the lack of evaluation evidence, relating to 12 departments;
- ten further reports have criticised departments for failing to collect data for future evaluation; and
- one report criticised both the lack of evaluation evidence and the failure to collect data, which will make future evaluation difficult.

Figure 16

Results from evaluation have been used to inform policy decisions

Discontinuing policies

- HMRC: Evaluation of stamp duty holiday for first-time buyers (published November 2011) was 'key' to the Chancellor's decision to discontinue the policy (announced November 2011).
- BIS: Evaluation of Regional Development Agencies (RDAs) identified poorly performing projects (published March 2009) and contributed to scaling back RDA budgets before abolition (decision taken in June 2010).

Expanding policies

- BIS: Evaluation on the economic benefits of different further education courses/qualifications (published March 2011) informed the decision to expand apprenticeships and cut back on other spending (announced February 2011).
- DWP: Interim evaluation of mandatory activity for long-term unemployed (published 2006) informed the decision to extend it to those aged over 50 (from June 2007).

Note

- 1 HMRC = HM Revenue & Customs; BIS = Department for Business, Innovation & Skills; DWP = Department for Work & Pensions.

Sources: Survey of chief analysts; Department for Work & Pensions impact assessment. Available at: www.gov.uk/government/publications/mandatory-work-activity--2

3.15 Between 2010 and 2012 the Public Accounts Committee has highlighted poor practice and a lack of evaluation in government departments:

- DFID lacked understanding of the costs and benefits of its programmes, which means that it cannot compare value for money across its portfolio and reallocate resources to the most effective interventions.²⁵
- The Department for Education (DfE) cannot focus resources on the most effective measures for recruiting teachers because it does not have the evidence from evaluation.²⁶
- The Ministry of Justice (MoJ) had limited evidence of what interventions work in youth justice, meaning that it is difficult to achieve better value for money, and there is a risk that the most successful interventions may be cut.²⁷
- DfT does not give sufficient attention to evaluation of major projects. If it does not complete an evaluation of High Speed 1, then it risks not learning lessons from the project – specifically the impact on regeneration.²⁸
- The Department for Work & Pensions (DWP) did not properly evaluate pilots before launching Pathways to Work. The flawed evaluation gave too positive a view of expected performance.²⁹
- DCLG did not set up a rigorous monitoring and evaluation framework when it introduced the New Homes Bonus, which meant that the Department could not identify impacts (including unintended consequences) or adjust its implementation.³⁰

25 Available at: www.hm-treasury.gov.uk/d/hmt_minutes_52_55_57_61_reports_cpas_feb2012.pdf

26 Available at: www.hm-treasury.gov.uk/d/minutes_14_18_reports_cpas_march2011.pdf

27 Available at: www.hm-treasury.gov.uk/d/minutes_19_21_reports_cpas_may2011.pdf

28 Available at: www.hm-treasury.gov.uk/d/hmt_minutes_82_1-4_6-10_reports_cpas_nov2012.pdf. DfT have subsequently commissioned an evaluation and expect to publish it in 2014.

29 Available at: www.hm-treasury.gov.uk/d/minutes_1_2_reports_cpas_dec2010.pdf

30 Available at: www.publications.parliament.uk/pa/cm201314/cmselect/cmpublic/uc114-i/uc114.pdf

Part Four

Production, resources and barriers

4.1 In this final part, we:

- explain the arrangements for commissioning and producing evaluations;
- discuss how the government enables independent evaluation, particularly by making data available;
- present how much the government spends on evaluation; and
- consider the barriers to the production and use of evaluation evidence.

Commissioning and delivery

4.2 In most cases, departments decide what is to be evaluated, how, and by whom. Most departments have a form of central scrutiny panel, to challenge the interventions being evaluated and to provide quality assurance.³¹

4.3 Evaluations are commissioned and delivered by departments in a number of ways (see Appendix Three). Evaluations are produced by in-house analysts or tendered out to external researchers. Analysts in some departments combine financial data with impact data from external researchers to estimate cost-effectiveness (see **Figure 17**). There are likely to be strengths and weaknesses in these evaluation arrangements. In-house analysts who produce evaluation findings may be more able to influence policy development with emerging findings, while externally commissioned work may have greater credibility externally because it is produced at arm's length from the department. There appears to be no clear rationale for the differences.

4.4 The Institute for Government's report, *Making policy better*, has questioned the credibility of these arrangements.³² It argues that departments have incentives and opportunities to tone down critical evaluation findings, or to influence those they have commissioned to do the evaluation. It has concerns that evaluations have often focused on narrow, department-specific questions, with less focus on cross-departmental lessons.

³¹ Arm's-length bodies often have their own evaluation budgets and arrangements. In some cases, the parent department coordinates and oversees cross-cutting evaluations, e.g. the Department for Culture, Media & Sport's 2012 Olympics evaluation.

³² Available at: www.instituteforgovernment.org.uk/sites/default/files/publications/Making%20Policy%20Better.pdf

Figure 17

Arrangements for delivering evaluation in government

Evaluation delivered by:

External organisations	In-house staff and commissioned from external organisations	In-house staff
DCLG, DEFRA, DCMS, BIS	HMRC, MoD, DfT, MoJ, HO, DWP, DECC	FCO

DfE, DH – In addition to in-house staff and external organisations producing evaluation evidence for these two departments, each has arm’s-length organisations (EEF and NICE) that are responsible for aspects of prioritising, commissioning, delivering and synthesising evaluation evidence.

DFID – In addition to evaluations commissioned by teams that manage aid programmes, in 2011 DFID set up the Independent Commission for Aid Impact (ICAI). ICAI is an advisory non-departmental public body, funded by DFID. It is responsible for examining and reporting on all UK Official Development Assistance. It reports to Parliament via the International Development Committee on its findings.

Note

- 1 DCLG = Department for Communities and Local Government; DEFRA = Department for Environment, Food & Rural Affairs; DCMS = Department for Culture, Media & Sport; DECC = Department of Energy & Climate Change; BIS = Department for Business, Innovation & Skills; HMRC = HM Revenue & Customs; MoD = Ministry of Defence; DfT = Department for Transport; MoJ = Ministry of Justice; HO = Home Office; DWP = Department for Work & Pensions; FCO = Foreign & Commonwealth Office.

Source: Analysis of chief analyst survey evidence

4.5 In some policy areas, departments have responded to concerns about the credibility of ‘marking their own homework’ by delegating the responsibility for commissioning evaluation to other organisations, with varying degrees of autonomy and independence. Examples include the Independent Commission for Aid Impact (ICAI), the Education Endowment Fund (EEF) and the National Institute for Health and Care Excellence (NICE) (see **Figure 18** overleaf). The government has stated that the ‘What Works’ centres (see Figure 12) will advise local delivery bodies on how their interventions can be evaluated effectively.

Enabling others to produce evaluation evidence

4.6 Departments can enable independent researchers, such as academics or think-tanks, to evaluate government interventions. This can be done through providing ‘core’ funding not tied to specific evaluations; for example, the Department for Business, Innovation & Skills (BIS) and the Welsh Assembly Government currently co-fund (with the Economic and Social Research Council) the Spatial Economics Research Centre, based at the London School of Economics. *The Civil Service Reform Plan (2012)* introduced a fund of up to £1 million per year to be used by ministers to commission policy work from other organisations.³³

33 HM Government, *The Civil Service Reform Plan*, June 2012. Available at: my.civilservice.gov.uk/reform/the-reform-plan/

Figure 18**Government-established bodies involved in evaluation****Educational Endowment Foundation (EEF)**

Who they are and what they do: Independent grant-making charity dedicated to breaking the link between family income and educational achievement, and ensuring that children from all backgrounds can fulfil their potential and make the most of their talents. The EEF's role is to identify, develop, support and evaluate projects to raise the achievement of disadvantaged children in the country's most challenging schools. The EEF aims to make grants to projects which can be robustly evaluated, and to organisations that it can effectively support.

Funding and governance: Funded by a £125 million grant from the Department for Education (DfE). With investment and fundraising income, the EEF intends to award up to £200 million over its 15-year life. The EEF was intentionally established with governance (and financial) arrangements removed from the direct influence of officials or ministers.

National Institute for Health and Care Excellence (NICE)

Who they are and what they do: The National Institute for Health and Care Excellence was set up in 1999 to reduce variation in the availability and quality of NHS treatments and care. Its role is to develop evidence-based guidelines on the most effective ways to diagnose, treat and prevent disease and ill-health. NICE designs its technology appraisals programme to ensure that people across England and Wales have equal access to new and existing medicines that are deemed clinically effective and cost-effective, reducing the risk of a 'postcode lottery' of care.

Funding and governance: £59 million grant-in-aid and parliamentary funding (2011-12). Topics are referred by the Department of Health. Guidance is created by independent advisory committees.

Independent Commission for Aid Impact (ICAI)

Who they are and what they do: ICAI has been set up by the Department for International Development (DFID) and reports to Parliament through the House of Commons International Development Select Committee (IDC). It does not report to ministers. ICAI's strategic aim is to provide scrutiny of UK aid spending, to promote the delivery of value for money for taxpayers, and to maximise the impact of aid. Its website states that it will: publish between 10 and 15 reports each year; report to the IDC; advise DFID and other departments on the effectiveness of its expenditure; and champion the use of independent evidence to help the UK to spend aid on what works best.

Funding and governance: £2.6 million (2012-13 budget) provided by DFID to cover the commissioners and secretariat, and associated costs. Advisory non-departmental public body. ICAI provides the IDC with an annual report on ICAI's activities from the preceding year. This includes key findings from evaluations, reviews and investigations.

Sources: Education Endowment Foundation: educationendowmentfoundation.org.uk/about/; National Institute for Health and Care Excellence: www.nice.org.uk/aboutnice/; and Independent Commission for Aid Impact: <http://icai.independent.gov.uk/about/>

4.7 The government committed to provide easier access to a wider range of data held by departments. This should help independent evaluators. The government's white paper on open data aims to facilitate data-sharing with research bodies.³⁴ The civil service white paper states that the civil service should make more data available freely for experts to test and challenge policy approaches. All main government departments published an 'Open Data Strategy' in 2012 but only three mention evaluation.³⁵

34 HM Government, *Open data white paper: unleashing the potential*, Cm 8353, June 2012. Available at: data.gov.uk/sites/default/files/Open_data_White_Paper.pdf

35 The Department of Health, which included evaluation to improve services, and the Department for International Development and Department for Work & Pensions (DWP), which committed to publishing research.

4.8 We asked external researchers and bodies which fund evaluation about their experiences of accessing government data (including administrative data) that can be used by independent evaluators. They said that they find it difficult to consistently gain access to departmental data, and that it has become harder in recent years. They noted that this was possibly as a consequence of sensitive child benefit data, which were lost in 2007.³⁶ More specifically, external researchers and funders said:

- It is not clear what administrative data are in principle available for use, and data used or produced during evaluations are not usually made available.
- Some departments do not routinely archive their datasets, and overseas researchers are not permitted to access data.
- There is often a piecemeal approach to gaining approval to use data and there are inconsistencies in decisions and the interpretation of data protection laws.

4.9 These findings are consistent with evidence collected by the Administrative Data Task Force, which is led by the Economic and Social Research Council (ESRC), the Medical Research Council and the Wellcome Trust. Its report *Improving access for research and policy* concludes: “Despite their considerable value as research resources, access to and linking between relevant administrative datasets has often been inhibited by issues relating to the legality of re-use and linkage for research and policy purposes. In some instances, research plans have been abandoned after funding has been agreed.”³⁷

Resources

4.10 The government deploys staff and spends resources (from research and programme budgets) to carry out evaluation. The cost of an evaluation is likely to depend on whether data collection is required, the scale of policy intervention, methods, and timescale involved. **Figure 19** overleaf provides examples.

4.11 The Office for National Statistics publishes data on Research and Development (R&D) expenditure, albeit with a significant lag. In 2010-11, the government spent £2,130 million on R&D; of this, £492 million was ‘research to support policy’. Part was used for evaluation, but not separately identified in the statistics.

³⁶ Hansard HC, 20 November 2007, vol. 467, col 1101. Available at: www.publications.parliament.uk/pa/cm200708/cmhansrd/cm071120/debtext/71120-0004.htm#07112058000527

³⁷ Available at: www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf

Figure 19

Examples of evaluations in government (by cost of evaluation)

Title	Objective	Type	Cost (£)
New Deal for Communities (DCLG)	Assesses role and impact of NDCs in improving their local neighbourhoods, and to develop knowledge about the effectiveness of community-based partnerships in delivering neighbourhood renewal. Covers £2 billion of spend.	Process, impact	£8.90m (2005–2008)
Employment Retention and Advancement (DWP)	Uses randomised control trial approach. Includes: a) process study intended to provide insight into possible reasons for the programme's impacts or lack of impacts; b) an impact study to compare outcomes for participants in a control group; c) a cost-benefit study: examines the net economic gains or losses (or net present value). Spend n/a.	Process, impact, cost-effectiveness	£0.78m (2011)
Regenerating the English Coalfields (DCLG)	Review of the literature; an analysis of secondary data sources since 1998, an assessment of regeneration programme documentation and monitoring data; six case studies reviewing the changing conditions. Covers £772 million of spend.	Process, impact, cost-effectiveness	£0.27m (2006)

Sources: NDC evaluation – www.rmd.communities.gov.uk/project.asp?intProjectID=12614, and ERA evaluation – www.gov.uk/government/publications/breaking-the-low-pay-no-pay-cycle-rr765; Coalfields evaluation – www.rmd.communities.gov.uk/project.asp?intProjectID=12170

4.12 We found it difficult to obtain reliable, accurate information from departments on overall spending on evaluation, because departments either said they did not have this information, or that it would only be available at disproportionate cost. We identified £44 million of expenditure on externally commissioned evaluations in 2010-11, with staff input of around 100 full-time equivalents (FTEs) at an estimated cost of £5 million.³⁸ A number of departments were unable to provide information. **Figure 20** provides a summary.

4.13 A freedom of information (FOI) request revealed that since 2010-11 four departments have reduced evaluation resources. Four³⁹ have cancelled or curtailed 25 evaluations between May and December 2010.⁴⁰ Eleven ongoing evaluations were cancelled before completion, reducing spending by more than £3 million. A further 14 evaluations were cancelled.

38 Assuming 102 FTEs at median salary of Grade 6/Grade 7 (£53,430). Source: *Annual civil service employment survey*.

39 Department for Work & Pensions; Department for Education; Department for Communities and Local Government; and Department for Business, Innovation & Skills.

40 Available at: www.radstats.org.uk/details-revealed-uk-government-social-research-and-statistics-cuts/

Figure 20
Resources spent on evaluation

Expenditure in 2010-11 (£m)	R&D for policy	Of which evaluation (source: Departments)	Of which Impact and CEE (source: Departments)	FTEs evaluation (source: Departments)
BIS	10	1.2	1.2	4
CO	n/a	n/a	n/a	n/a
DCLG	23	0.4	0.2	1.8
DCMS	14	0.3	0.3	2
DECC	3	0.5	Minimal	3.2
DEFRA	100	15.1	n/a	17.3
DfE	27	17.7	n/a	9.2
DFID	209	2.5	2.5	24.4
DfT	16	n/a	n/a	2
DH (inc NHS)	32	n/a	n/a	n/a
DWP	28	6.3	4.9	30
FCO	3	In-house	In-house	n/a
HMRC	9**	0.3	0.3	8.5
HMT		n/a	n/a	n/a
HO	18	n/a	n/a	n/a
MoD		Minimal	Minimal	–
MoJ	n/a	n/a	n/a	n/a
	492	44.3	9.4	102.4

Notes

- 1 DCMS = Department for Culture, Media & Sport. R&D figure relates to DCMS broader group, and not the core Department. This figure is neither directed nor validated by DCMS.
- 2 DECC = Department of Energy & Climate Change. "DECC's evaluation resources have grown substantially since 2011-12, with considerable increases in both internal staff and external spend."
- 3 BIS = Department for Business, Innovation & Skills. Evaluation figure for BIS also only included those in the Central Evaluation Team.
- 4 MoD = Ministry of Defence. "MoD carries out Test & Evaluation activity to inform decisions on the suitability, safety and effectiveness of military capabilities. It is estimated roughly that overall expenditure is of the order of £1 billion per year."
- 5 MoJ = Ministry of Justice. Evaluation work is conducted alongside other analysis and analytical support for policy and operational colleagues. MoJ is not able to disaggregate the resources for evaluation from other analytical work.
- 6 DFID = Department for International Development spend includes that only by its Central Evaluation Department.
- 7 DCLG = Department for Communities and Local Government; DEFRA = Department for Environment, Food & Rural Affairs; HMRC = HM Revenue & Customs; DfT = Department for Transport; HO = Home Office; DWP = Department for Work & Pensions; FCO = Foreign & Commonwealth Office; HMT = HM Treasury; CO = Cabinet Office; DH = Department of Health.

Source: Data provided by departmental chief analysts. R&D data from Office for National Statistics

Barriers

4.14 A number of government and independent reports have discussed barriers to the production and use of evaluation.⁴¹

4.15 In terms of producing evaluation evidence, the key barriers cited are:

- difficulties in evaluating some government interventions, and repeated failures to design or pilot in such a way that enables rigorous evaluation;
- insufficient skills and capacity;
- difficulties in accessing and joining up administrative data;
- short electoral cycles, and high rates of ministerial and official turnover;
- the absence of consistent demand for evaluation from ministers and senior civil servants; and
- concerns about 'unhelpful' conclusions about policies' effectiveness.

4.16 The barriers in using evaluation results are attributed to a combination of:

- the time-lags in commissioning and delivering evaluation;
- a lack of the necessary analytical skills to act as an intelligent customer (i.e. sufficient technical knowledge of the research being provided);
- a lack of sanctions for failing to evaluate, or positive incentives such as HM Treasury clearly linking resource allocation to robust evidence on cost-effectiveness; and
- failure to synthesise and communicate evaluation findings in effective, digestible ways.

⁴¹ Institute for Government, *Policy making in the real world*, April 2011; Institute for Government, *Evidence and evaluation in policy making*, September 2012; *Evaluation book*, 2012; Performance and Innovation Unit, *Adding it up*, January 2000.

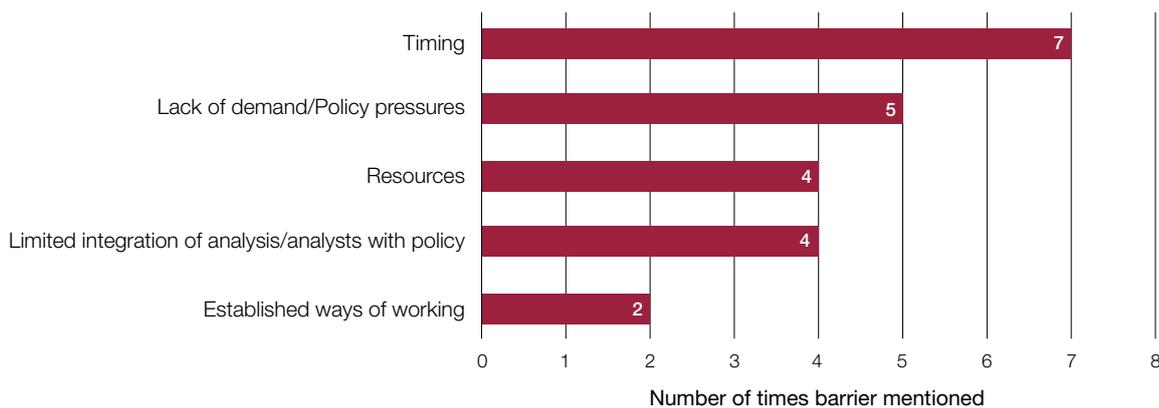
4.17 Departmental chief analysts explained in response to our survey what they considered to be barriers to better quality and use of evaluation (**Figure 21**). The two most important were mismatches in timing between production of evaluation evidence and policy decisions, and lack of demand. Two chief analysts highlighted established ways of working as a barrier – specifically, that technical people carry out evaluations but decisions are taken by ministers and policy-makers, and that the challenge is sometimes “convincing people to challenge long-established ways of working”.

4.18 Our survey indicated that 34 per cent of evaluation analysts believed evaluation findings are delivered too late (**Figure 22** overleaf). They highlighted difficulties accessing evaluation evidence, inconclusive or negative findings and lack of robust findings.

Figure 21

Key barriers to better quality and use of evaluation

For each of the following factors, please indicate whether this is a barrier in using ex-post cost-effectiveness evidence in your Department (responses from “frequent” and “sometimes” evaluators)



Note

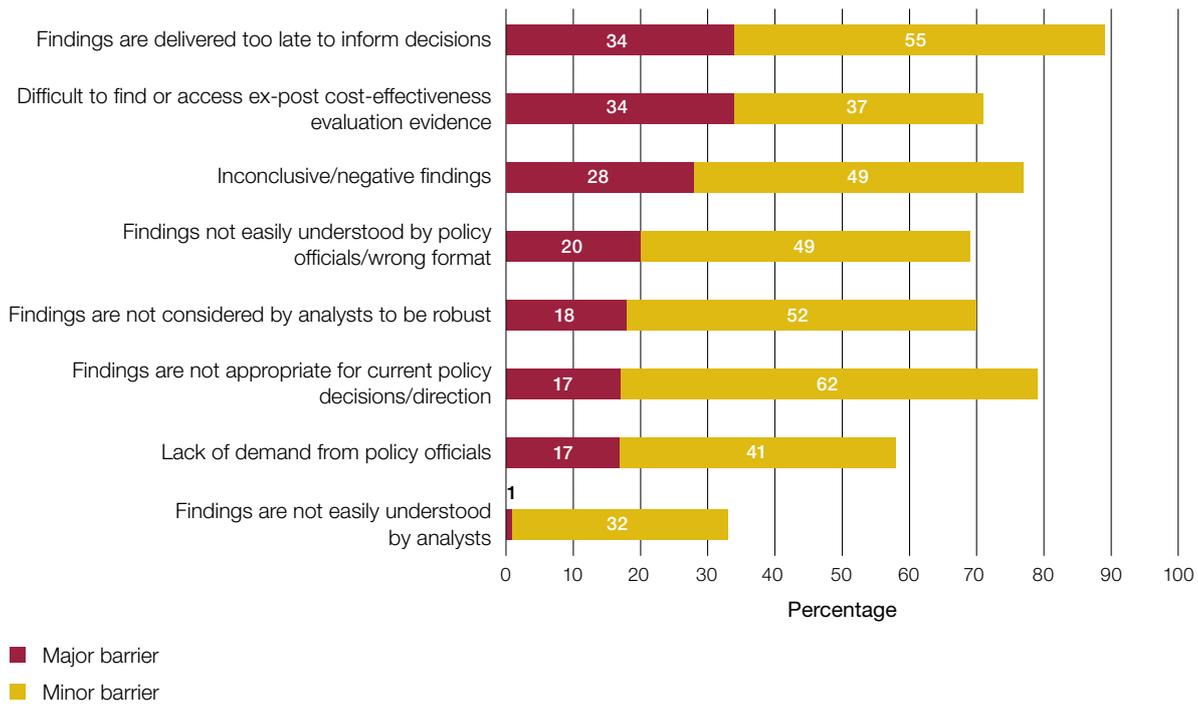
1 Multiple answers allowed.

Source: National Audit Office survey of departmental chief analysts

Figure 22

Analyst views of barriers to using evaluation evidence

For each of the following factors, please indicate whether this is a barrier in using ex-post cost-effectiveness evidence in your Department (responses from “frequent” and “sometimes” evaluators)



Note

1 Results presented for those most frequently involved in evaluation, 110 respondents.

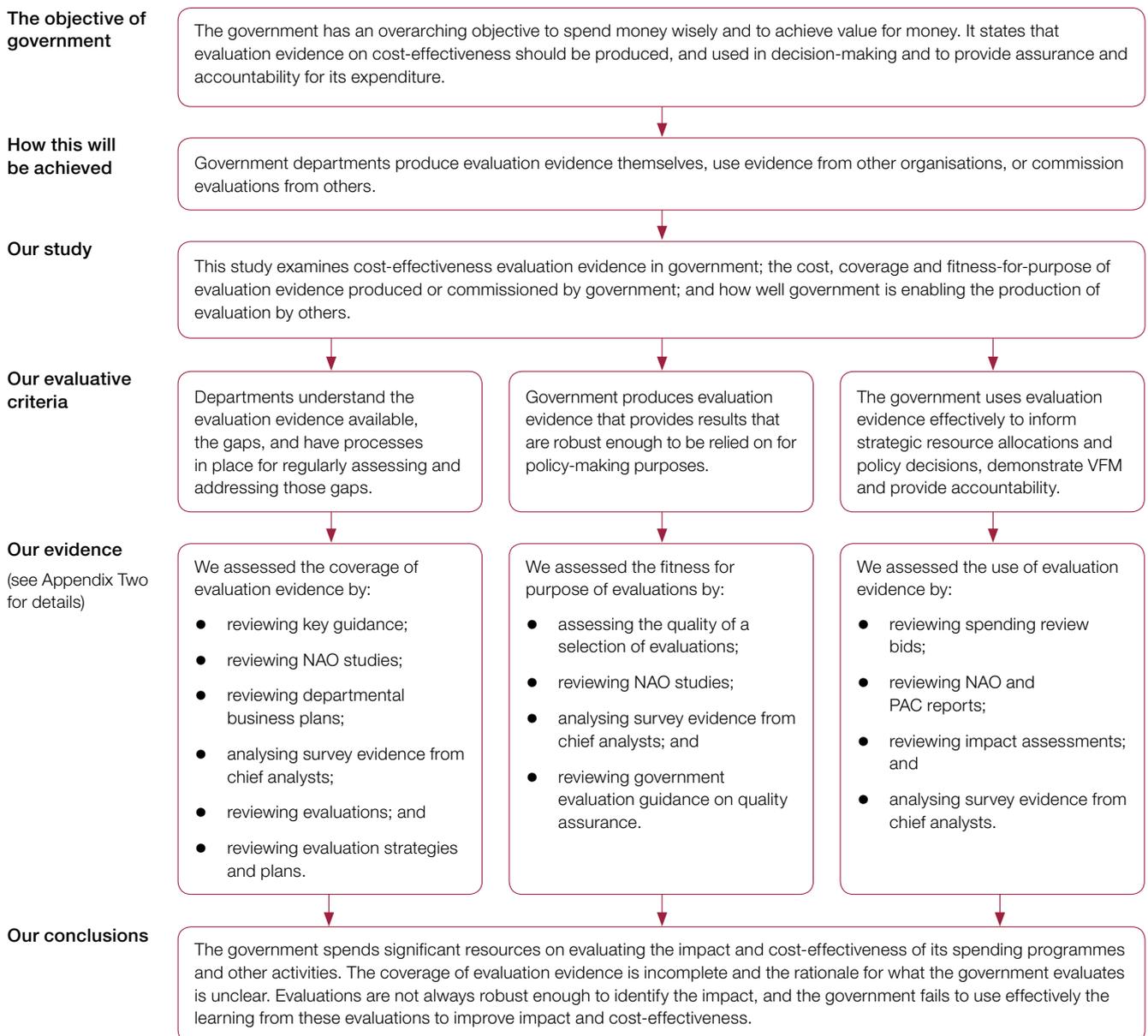
Source: National Audit Office survey of analysts

Appendix One

Our audit approach

Figure 23

Our audit approach



Appendix Two

Our evidence base

1 We formed our conclusions based on findings from our analysis of evidence reviewed between July 2012 and March 2013. Our audit approach is at Appendix One. Our study focused on the 17 central government departments.

2 We assessed the arrangements for evaluation, as well as the coverage, quality and use of evaluation in the government.

3 We reviewed the government's approach to evaluation by:

- reviewing recent publications that comment on the state of evaluation in the government, to identify criticisms;
- reviewing Committee of Public Accounts minutes to identify the issues it has raised in hearings, and the reasons why evaluation is important to public service delivery;
- reviewing internal government reviews of evaluation, including the Government Office for Science's *Science and Engineering Assurance Reviews*;
- gathering information from departments on their institutional arrangements for delivering evaluation evidence;
- interviewing a range of external evaluators and researchers, to understand their experience of accessing data to undertake independent evaluations; and
- reviewing evidence from the Office for National Statistics and departmental chief analysts regarding expenditure on research and development (R&D) and evaluations, and the number of staff working on evaluations. We received responses from 14 departments.

4 We assessed the coverage of evaluation evidence by:

- reviewing key central government guidance on evaluation: the *Green Book*, the *Magenta Book*, *Managing Public Money* and guidance on impact assessments;
- reviewing previous NAO studies from between 2008 and 2012, to identify previous criticisms of departments with respect to the lack of evaluation;
- reviewing published departmental business plans to identify their plans to evaluate their major projects;

- gathering evidence from surveys of chief analysts and analysts working in government on aspects of evaluation, to determine their approach to identifying and addressing the gaps in evaluation evidence;
- reviewing published evaluation evidence on departmental websites to determine the quantity and extent of evaluations of different types; and
- reviewing evaluation strategies and plans published by departments, and reviewing the outputs to understand if commitments were fulfilled.

5 We assessed the quality of evaluation evidence in government:

- Commissioning Henry Overman from the London School of Economics to carry out a detailed assessment of the quality of 34 evaluations across four policy areas: education, business, spatial and labour. Further detail of their assessment can be found on our website at www.nao.org.uk/report/evaluation-government/.
- Reviewing our previous work across the government where we have commented on the quality of evaluation evidence.
- Analysing survey evidence from government chief analysts and evaluation analysts to understand the self-assessed quality of evaluation evidence in the government.
- Reviewing government evaluation guidance on the quality assurance of evaluation outputs, and the arrangements that departments have in place.

6 We examined the use of evaluation evidence:

- We carried out a review of three departments' submissions to HM Treasury as part of the 2010 Spending Review. The documents were gathered as part of a previous NAO study.⁴² This allowed us to quantify the proportion of bids (in monetary terms) that referred to evaluation evidence.
- We reviewed 261 final impact assessments completed in 2009-10 to identify where evidence from evaluation was used.
- We conducted a web survey of chief analysts and analysts, to gather evidence on how evaluation evidence is used in practice and how it has contributed to policy decisions. We received responses from 15 of 17 chief analysts. They did not all respond to every question and therefore some data does not sum to the full set of departments. We received responses from 110 analysts across government departments.

⁴² Comptroller and Auditor General, *Managing budgeting in government*, Session 2012-13, HC 597, National Audit Office, October 2012.

Appendix Three

Arrangements for evaluation in government

We asked departmental chief analysts about evaluation arrangements in their departments. The table below contains their description of the commissioning models, governance and quality assurance arrangements, and evaluation support in their departments.

Business, Innovation & Skills (BIS)⁴³

Commissioning model – Projects are required to set aside a budget for evaluation. More resources are committed to evaluations where policy is expensive, complex, large-scale, high-risk or a flagship programme.

Governance and quality assurance – The Evaluation Strategy Group (ESG) holds directors to account for whether and how key policies are evaluated, establishes whether lessons are learnt, and ensures that evidence feeds into policy. ESG reports to a senior Policy and Programme Board chaired by the accounting officer.

Support – A team of four in a central team. They provide support and advice on evaluation, and have a programme of work in place in support of the evaluation strategy.

Department for International Development (DFID)

Commissioning model – A mixed model. Some evaluations are commissioned by DFID, while the Independent Commission for Aid Impact (ICAI), established in 2011 and reporting directly to Parliament, scrutinises the impact of UK Official Development Assistance. DFID funds 3ie, which funds impact evaluations, and the World Bank for impact evaluations on human development programmes. DFID operational and policy teams also commission evaluations. Country-based offices are responsible for decisions on what to evaluate, allocating funds, procuring, overseeing and responding to evaluations. In 2012-13, 26 DFID-funded evaluations were published and 60 are expected to be published in 2013-14.

Governance and quality assurance – Country offices are accountable for ensuring that QA mechanisms are built in to ensure relevance and quality in the product. Evaluation policy sets out peer review requirements. The Investment Committee requires QA of all larger and strategically significant evaluations. Evaluation Department sets standards, advises and builds capacity.

43 These arrangements were in place until June 2013. New evaluation arrangements are being implemented.

Support – Evaluation Department has 15.7 FTE staff led at SCS level – 1 SCS, 3 G6 and 7.7 G7 plus support staff provide specialist help in designing, implementing and follow-up of evaluations. There are also evaluation specialists in Africa and Asia divisions and policy divisions, and up to 30 posts on evaluation in operational teams.

Department for Education (DfE)

Commissioning model – The Educational Endowment Foundation (EEF) administers a fund established by DfE for developmental and evaluative work of initiatives in schools. Evaluations are commissioned by DfE from external providers through competitive tender. Local partners (e.g. local authorities) evaluate some pilot initiatives.

Governance and quality assurance – The Research Scrutiny Group (RSG) and ministers approve external evaluators before evaluations are commissioned. Analysts conduct peer review processes and support individual teams taking forward evaluations. Academics are used on an ad hoc basis to consider proposals or evaluation design, or to review analysis.

Support – Analysts embedded in each policy directorate deliver this support role, drawing on a small central analytical resource for cross-cutting admin support and advice.

Home Office (HO)

Commissioning model – Evaluations are commissioned, managed and sometimes carried out by social scientists.

Governance and quality assurance – Project proposals or research designs are scrutinised by an internal panel to ensure the approach is fit for purpose, and are signed off by the chief scientist. Project plans include plans for QA during the project. Outputs are reviewed by internal panel or external experts.

Support – Home Office Science provides support across the department on evidence and evaluation.

Department of Energy & Climate Change (DECC)

Commissioning model – A director-level board chaired by the director of analysis has agreed 11 departmental evaluation priorities. Responsibility for evaluation of these and other evaluations lies with policy teams, with many having dedicated evaluation experts/teams to lead this work. Evaluation is carried out both internally and through externally commissioned work.

Governance and quality assurance – A rigorous ‘Evidence Framework’ is in place to ensure the quality of all DECC’s evidence, including evaluation. All evaluations have to have a QA plan, which specifies checkpoints for sign-off and approval to ensure the quality of both the design and execution of the work. All externally commissioned evaluations are scrutinised and approved by a multi-disciplinary R&D Approvals Board, chaired by the director of analysis. External peer review is conducted by DECC’s Social Science Expert Panel or by other relevant experts where appropriate.

Support – A central Policy Evaluation Team, headed by a G6 evaluation specialist and supported by a G7 specialist, supports and challenges evaluation across the department. An ‘Evaluation Practitioners Group’ offers peer support and development among those working on evaluation.

Foreign & Commonwealth Office (FCO)

Commissioning model – Evaluation is usually carried out in-house. Lead directorates decide how best to evaluate their areas of work. The Policy Unit and Communications teams have responsibility for monitoring policy impact and measuring progress towards improving capability, through the Diplomatic Excellence scoring.

Governance and quality assurance – Policy Unit and Finance Directorate assist the Board of Management to scrutinise and evaluate progress against foreign policy outcomes at mid- and end-year points. FCO Board of Management reviews policy impact monthly on the basis of recommendations from an internal panel, checked quarterly by external customers. The FCO’s Programme Evaluation Board (PEB) oversees the monitoring and evaluation of strategic programmes. Projects are continuously monitored and quarterly reports are provided to programme managers and SROs and a formal Annual Review of Programmes.

Support – Two groups complement directorate evaluation. The PEB evaluates all projects in excess of £500,000 and a sample of projects of £100,000 or more. High-value project evaluations are usually carried out by members of the FCO’s internal evaluation cadre. They are introducing a lesson-learning exercise, which will aim to evaluate policy formulation and implementation each quarter and identify the lessons.

HM Revenue & Customs (HMRC)

Commissioning model – Evaluation work is either undertaken in-house or commissioned by the central analytical unit.

Governance and quality assurance – All evaluation work is subject to internal peer review, and has access to external academics (e.g. on econometrics).

Support – Evaluation work is discussed regularly with key stakeholders and managed alongside other analytical work.

Department for Communities and Local Government (DCLG)

Commissioning model – Evaluations are commissioned and managed by analysts and generally carried out externally (although this balance is changing), with market relationship and procurement processes developed to ensure a range of expert providers can bid successfully for contracts. Specifications are developed by analysts who are expert in evaluation techniques and encourage providers to offer ‘best of kind’ current practice in devising the methods.

Governance and quality assurance – Research Gateway (comprising all the analytical heads of profession with representatives from Finance and Procurement) scrutinises proposals involving a spend of more than £20,000 to ensure they are methodologically sound and reviews research projects/evaluations at key milestones to ensure they are on track and delivering intended outputs.

Support – There is a strong ethos of peer support between analysts and DCLG analysts are closely involved in other external networks, including those with relevant academics and with analysts in other departments.

Department for Environment, Food & Rural Affairs (DEFRA)

Commissioning model – Evaluations are commissioned by analysts in the Department. A number of arm’s-length bodies also conduct or commission evaluations.

Governance and quality assurance – The director-level Evaluation Board has a work plan, which addresses the evidence gaps and enforces the quality standards.

Support – “A team in the chief economist’s office provides general support. In addition, the chief of social research, the social research group and other analysts in the Department provide advice.”

Department for Culture, Media & Sport (DCMS)

Commissioning model – The Evidence & Analysis Unit commissions evaluations. Arm’s-length bodies have their own research teams, although cross-cutting evaluations may be coordinated by DCMS.

Governance and quality assurance – Arrangements are ad hoc.

Support – DCMS has a central Evidence and Analysis Unit led at SCS level, which provides support across the Department on evidence and evaluation.

Ministry of Defence (MoD)

Commissioning model – Evaluations are commissioned by MoD, managed by the Defence Science and Technology Laboratory (Dstl) or MoD, and carried out by Dstl, consultants or academia.

Governance and quality assurance – Review of evaluations carried out across MoD is undertaken every six months. A framework process and investment to support military capability and evaluation.

Support – Defence Economics – formerly known as Division of Economic Statistics and Advice – produces guidance documents.

Department for Transport (DfT)

Commissioning model – Some evaluations are undertaken internally, while others are managed by DfT officials and delivered by evaluation practitioners and transport consultancies. Local authorities and delivery bodies are also responsible for evaluations (e.g. for local programmes), which will be overseen by the Department. Agencies have their own approaches and processes for evaluation.

Governance and quality assurance – DfT Strategy Committee oversees the evaluation strategy and its programme of evaluation (as of summer 2013). They are also enhancing quality assurance of evaluation plans and outputs and providing training and development for staff.

Support – A small central team provides technical support and advice.

Ministry of Justice (MoJ)

Commissioning model – Analysts commission and manage impact and cost-effectiveness evaluations. Some are carried out within MoJ, and others by external research and analysis organisations. Evaluations are often commissioned at a high level by policy colleagues and then commissioned or carried out by analysts.

Governance and quality assurance – Analytical quality assurance (AQA) applies to all Analytical Services Directorate projects throughout the project cycle. QA input is proportionate to the risks of the project: high-resource, business-critical, methodologically challenging projects will have the most QA time devoted to them. Analytical products are also internally and externally peer-reviewed before they are published.

Support – The central Analytical Services team produces and manages evaluation evidence within the Department and has a virtual 'evaluation group', which provides advice and support.

Department of Health (DH)

Commissioning model – The National Institute for Health and Care Excellence (NICE) has clearly defined rules over when and how to undertake assessments and evaluation of its decisions. This is not, largely, a discretionary decision for DH. Outside of NICE’s remit, the Department of Health Research and Development Directorate (RDD) commissions evaluations. RDD also run the National Institute for Health Research and the Policy Research Programme, through which the Department funds a number of policy research units including the Policy Innovation Research Unit (PIRU), which brings together leading health and social care expertise to improve evidence-based policy-making and its implementation across the National Health Service, social care and public health.

Governance and quality assurance – Both NICE and RDD have routine QA and peer review. The R&D Committee discusses proposals for research and evaluation. The committee has representatives from each directorate in the Department of Health as well as their arm’s-length bodies/executive agencies and makes recommendations on research priorities.

Support – Advice is provided by the R&D Directorate. Analysts provide advice to their policy colleagues where required.

Department for Work & Pensions (DWP)

Commissioning model – Lead analysts identify the need for evaluation and the appropriate evaluation strategy. They decide which parts should be done internally and which contracted out, select an appropriate contractor for any external work, and manage that work.

Governance and quality assurance – The Central Analysis Division (CAD) challenges the need for any particular project, the scale of it and the chosen methodology.

- **External:** contractors must provide evidence of the quality of their work before being accepted onto the DWP research framework. DWP analysts will review their work, calling as necessary on additional expertise within DWP, from other government departments or from external experts. This may be at an early stage as part of an advisory group, or to peer-review the products of the evaluation.
- **Internal:** work is peer-reviewed internally. Where there are significant technical issues, external experts are engaged to advise on methods. Open publication of all research is another part of the QA process.

Support – The CAD provides advice on evaluation methods and develops departmental standards (particularly on cost-effectiveness evaluation). Additionally, there is a team which develops and maintains their guidance on cost benefit analysis.

NB: HM Treasury did not complete a survey. Cabinet Office – we have received input from Cabinet Office including Efficiency & Reform Group, and the “What Works?” team.



National Audit Office