HM TREASURY

Review of quality assurance of Government analytical models:

final report

March 2013

HM TREASURY

Review of quality assurance of Government analytical models:

final report

March 2013

Official versions of this document are printed on 100% recycled paper. When you have finished with it please recycle it again.

If using an electronic version of the document, please consider the environment and only print the pages which you need and recycle them when you have finished.

Any queries regarding this publication should be sent to us at: public.enquiries@hm-treasury.gov.uk.

ISBN 978-1-909096-53-0
PU1404

Contents

Foreword

Modelling is essential to the work of government. From predicting the spread of pandemic flu to forecasting population growth, models underpin decisions which affect people's lives.

It is vital, therefore, that these models are fit-for-purpose. To that end, in October 2012 Sir Jeremy Heywood (Cabinet Secretary) and Sir Bob Kerslake (Head of the Civil Service) asked me to review the quality assurance (QA) of analytical models across government.

I see high quality analysis and use of evidence as fundamental to the civil service's effectiveness. The recent experience with the Intercity West Coast franchise competition underlines the importance of good quality assurance.

Over the past four months an inter-disciplinary team, drawn from across departments and based in the Treasury, has engaged with organisations across the public and private sectors to map the government's business critical models and current arrangements for their quality assurance, and identify best practice.

In December 2012 I published an interim report based on the review team's work to date. This summarised the approach and scope of the review, and set out initial findings from analysis of departmental returns. A copy of this can be found on the Treasury website.

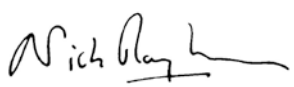
The work of the review team since that interim report has focussed on three areas. Firstly, we have refined our analysis of the current approach to quality assurance using the information received from departments. Secondly, we have identified a set of best practice principles for model quality assurance, through stakeholder engagement across the public and private sectors. Thirdly, we have identified recommendations for government.

The objective has been to ensure all models are of sufficiently high quality, and that their end users – Ministers and, ultimately, the public – can place their trust in them. Quality assurance is not the only factor which leads to robust models, but it is a key one. In working towards this goal, we have not passed judgment on individual models, something which remains the responsibility of departments. Rather, we have scrutinised the situation across government and produced recommendations to drive the spread of best practice.

The conclusions of the review are important and I commend them to Accounting Officers, Chairs and non-executive board members as well as those staff who will be tasked with implementing the recommendations.

As with all reviews, the efficacy of these recommendations will be dependent on their implementation. I therefore suggest an assessment of organisations' progress against the recommendations should take place in 12 months' time.

I would like to thank all the organisations that have helped with this review. In particular, I would like to thank the review team – Helene Radcliffe, Martha Goyder, Jennifer Bradley, Mark McDonnell, Colin Wilson, Declan Millin, Miles Elsden and Janos Suto – for their effort in pulling together this report within such a short timeframe.



Nick Macpherson

Executive summary

In October 2012, the Cabinet Secretary and the Head of the Civil Service commissioned a review of the quality assurance (QA) of analytical models that inform government policy. The review published an interim report in December 2012, setting out results of work to map business critical models and quality assurance in government. This is the final report of the review.

Hundreds of models are being used across government to influence policy, and it is vital that they are equal to this task. Quality assurance is a key means of ensuring this. To assure current arrangements are robust, the review team conducted three strands of work. Firstly, it interviewed public and private sector organisations and professional bodies, to identify best practice. Secondly, it collected and analysed data on departments' current QA practices. Thirdly, and drawing on the outputs from these workstreams, it developed recommendations.

The many components of best practice in QA fall under two headings: modelling environment, and process. The right modelling environment involves a culture where leaders value and recognise good QA. It requires adequate capacity, including specialist skills and sufficient time to conduct QA effectively. It also needs a set of controls, including a clear internal chain of responsibility and a route for challenge where analysts have concerns. The process side, on the other hand, is about a systematic approach to make QA accessible, easy and comprehensive. It requires clear guidance on QA, and clear documentation for every model.

The review found good signs in departments' current practice on QA. These include the broad spread across departments of important basic techniques like internal peer review, and the extent of internal guidance. Taken together, they indicate key elements of quality assurance are being widely applied.

Despite this, there is significant variation in the type and nature of QA used within, and between departments. Much of this is to be expected given the differences in organisations' remits, and the levels of risk in question. However, it is not certain that this is always the case. The review's work highlighted the benefits of a more systematic approach to creating a work environment that expects high quality QA – including allocating clear responsibility for key models and how they are used, and giving specialist staff adequate time to manage QA effectively. There is some good practice in guidance, but its nature and extent varies between departments.

These findings suggest the need to extend best practice across the whole of government – to ensure a sufficiently high standard everywhere. To this end, the review sets out the following headline recommendations for departments and their Arm's Length Bodies (ALBs) (see Chapter 4 for full recommendations):

- Recommendation 1: All business critical models in government should have appropriate quality assurance of their inputs, methodology and outputs in the context of the risks their use represents. If unavoidable time constraints prevent this happening then this should be explicitly acknowledged and reported;
- Recommendation 2: All business critical models in government should be managed within a framework that ensures appropriately specialist staff are responsible for developing and using the models as well as quality assurance;
- Recommendation 3: There should be a single Senior Responsible Owner for each model ("Model SRO") through its lifecycle, and clarity from the outset on how QA is

to be managed. Key submissions using results from the model should summarise the QA that has been undertaken, including the extent of expert scrutiny and challenge. They should also confirm that the Model SRO is content that the QA process is compliant and appropriate, that model risks, limitations and major assumptions are understood by users of the model, and the use of the model outputs is appropriate;

- Recommendation 4: The Accounting Officer's governance statement within the annual report should include confirmation that an appropriate QA framework is in place and is used for all business critical models. As part of this process, and to provide effective risk management, the Accounting Officer may wish to confirm that there is an up-to-date list of business critical models and that this is publicly available. This recommendation applies to Accounting Officers for Arm's Length Bodies, as well as to departments;
- Recommendation 5: All departments and their Arm's Length Bodies should have in place, by the end of June 2013, a plan for how they will create the right environment for QA, including how they will address the issues of culture, capacity and capability, and control. These plans will be expected to include consideration of the aspects identified in Box 4.A in Chapter 4 of this report;
- Recommendation 6: All departments and their Arm's Length Bodies should have in place, by the end of June 2013, a plan for how they will ensure they have effective processes – including guidance and model documentation – to underpin appropriate QA across their organisation. These plans will be expected to include consideration of the aspects identified in Box 4.B of Chapter 4 of this report. To support this recommendation, succinct guidance setting out the key, generic issues that drive effective quality assurance will be added to "Managing Public Money" – which offers guidance on how to handle public funds properly;
- Recommendation 7: To support the implementation of these recommendations, the review recommends establishing an expert cross-departmental working group to continue to share best practice experience and to help embed this across government; and
- Recommendation 8: Organisations' progress against these recommendations should be assessed 12 months after this review is published. HMT will organise the assessment, possibly with support from another department.

1 Introduction

Definitions and scope of the review

1.1 This review was commissioned to examine the quality assurance of government analytical models which are used to inform policy.

1.2 As set out in the interim report, a model is a mechanism for analysing or investigating some aspect of the real world. It is usually a quantitative method, system or approach which applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates. There are typically three parts to a model:

- inputs – in the form of data and assumptions;
- a processing component – often through calculations; and
- outputs – the key figures as well as the risks and limitations of the models.

1.3 Throughout this report, any use of the term model should be read as encompassing inputs, processing and outputs, and each of these component terms should be taken to include all items defined above.

1.4 Models are used for a huge variety of purposes in government, and a significant part of the review's work has been to take stock of the business critical models government uses. To help structure the returns from departments, and to provide an analytical framework, the review defined seven areas where government routinely uses models, as set out in Table 1.A below.

Table 1.A: Table defining models by their purpose

Model type	Purpose	Examples
Policy simulation	Appraisal of policy options, analysis of impact on people, finances, etc	Intra Government Tax Benefit Model
Forecasting	Assessing the future, perhaps to provide base information for policy development or financial planning	State Pension expenditure forecast
Financial evaluation	Assessment of liability or future cost	Pension liabilities, higher education loan repayment model
Procurement and commercial	Evaluation of VfM or affordability and award of contracts	Awarding of rail franchises
Planning	Planning current actions based on future forecasts	Teachers, NHS
Science-based	Understanding and forecasting natural systems	Climate change
Allocation	Distribution of funding across organisations responsible for service delivery	Police allocation formula

1.5 Given this breadth of uses and purposes, and the complexity of some models, it is essential to ensure models are robust – a subject explored in relation to micro-economic government models by a government report on improving analysis and modelling 'Adding it Up', in 2000[1].

1.6 Quality assurance refers to processes which can help ensure the model's inputs and outputs meet its quality requirements, manage risk of errors and ensure the model is fit-for-purpose. It is a key means of ensuring models are robust. Private organisations as well as the public sector apply a number of QA techniques. These range from review by a peer in the same organisation, to full external model audit. Box 1.A below sets out some key types.

Box 1.A: Types of Quality Assurance

Developer testing – use of a range of developer tools including parallel build and analytical review or sense check;

Internal peer review – obtaining a critical evaluation from a third party independent of the development of the model, but from within the same organisation;

External peer review – formal or informal engagement of a third party to conduct critical evaluation, from outside the organisation in which the model is being developed;

Use of version control – use of unique identifier for different versions of a model;

Internal model audit – formal audit of a model within the organisation, perhaps involving use of internal audit functions;

Quality assurance guidelines and checklists – model development refers to department's guidance or other documented QA processes (e.g. third party publications);

External model audit – formal engagement of external professionals to conduct a critical evaluation of the model, perhaps involving audit professionals;

Governance – at least one of planning, design and/or sign-off of model for use is referred to a more senior person. There is a clear line of accountability for the model;

Transparency – model is placed in the wider domain for scrutiny, and/or results are published; and

Periodic review – model is reviewed at intervals to ensure it remains fit for the intended purpose, if used on an ongoing basis.

1.7 The aspects of QA above are important not for their own sake, but because they help ensure sufficiently high quality models. This is their ultimate goal.

The work of the review team

1.8 This report reflects work undertaken by the review team between October 2012 and February 2013. The work involved three main elements:

- analysing information provided by departments – the review team asked departments to submit details of all models used by the department and its Arm's Length Bodies (ALBs) that they considered to be business critical. The purpose of this was to understand the scope of modelling in government. The review also

[1] http://webarchive.nationalarchives.gov.uk/+/http:/www.cabinetoffice.gov.uk/upload/assets/www.cabinetoffice.gov.uk/strategy/coiaddin.pdf

asked for information on the key aspects of quality assurance which applied to these models. This was to provide a snapshot of the extent and type of quality assurance undertaken by departments as of late 2012, and to observe any patterns which might inform the review. As there are many factors which determine what QA is carried out – not least the degree of risk and complexity – this data cannot be used to assess whether the QA of a particular model is the most appropriate. As discussed in Chapters 2 and 3, the review would expect there to be a wide range of approaches to QA, reflecting the circumstances surrounding a particular model;

- engaging with public and private sector organisations – to identify and define best practice. As well as engaging with government departments, the team interviewed private sector organisations from a range of industries including finance, consultancy, accounting, academia, investment banking, engineering, international financial institutions, research and professional bodies. The team also conducted desk research to identify and distil principles of best practice. This included analysis of existing departmental guidance on QA. A list of organisations who gave their time is at Annex B; and
- developing recommendations – the team consulted with departments and their ALBs from across government in developing recommendations.

Structure of this report

1.9 The rest of this report is structured as follows. Chapter 2 defines the key elements of best practice in quality assurance, as drawn from discussions with organisations across the public and private sectors and with professional bodies. Chapter 3 sets out the current extent and nature of modelling and QA across Government. Chapter 4 makes practical recommendations for how departments and their ALBs should move forward, to achieve best practice.

2 Best practice in model quality assurance

The key elements of quality assurance

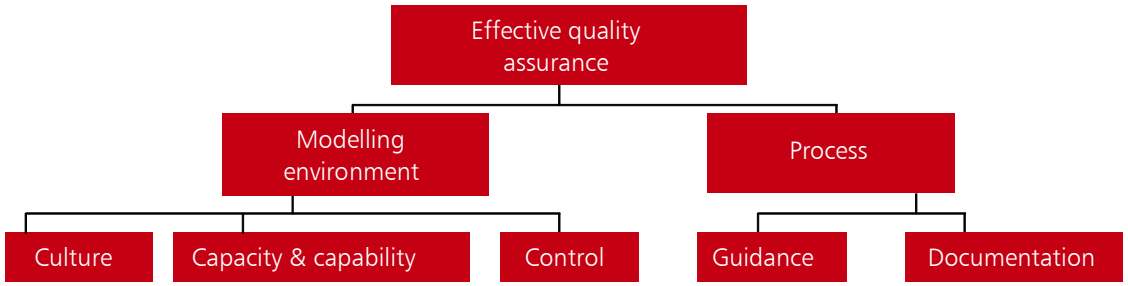
2.1 Quality assurance (QA) provides decision makers with key information about how a model works, and its risks and limitations. This is essential if a model's outputs are to be used with genuine understanding and confidence. As such, QA is a key aspect of the effective risk management of business critical models, and the decisions they help inform.

2.2 The work of the review team indicates that, sitting above the many principles and techniques which ensure good QA, there are two main requirements:

- modelling environment: creating the conditions in which QA processes can operate effectively, including through a culture that values QA and welcomes effective challenge, a well understood chain of responsibility and sufficient time for QA; and
- process: establishing a clear process for every stage of the model life-cycle. This includes working alongside the customer to ensure there is a shared understanding about the purpose and any limitations of the model.

2.3 Chart 2.A below summarises these key prerequisites for effective QA:

Chart 2.A: Elements of effective quality assurance



2.4 These elements were common to the review's conversations with a wide range of organisations – across the private and public sector as well as professional bodies. Together they can help empower and incentivise model developers to prevent errors.

2.5 Within Chart 2.A above, the right modelling environment and process are essential to create a sound QA framework. They need to remain in place whatever the type and complexity of the model. The detailed mechanisms for checking the model's reliability and accuracy, however, will vary depending on the model and the risks inherent in the model and its use. The circumstances in which different levels of QA are appropriate are discussed at the end of this chapter, which recognises not all types of QA will be appropriate all of the time and for all models. In all cases, QA needs to be proportionate, and the resources employed should represent value for money.

2.6 It is worth noting that the elements in Chart 2.A are all inputs to effective quality assurance. These inputs are not valuable for their own sake, but because they contribute to effective models.

2.7 The next sections of this chapter discuss the key principles which can deliver effective environment and process, in more detail.

An environment for effective QA

2.8 Many of those the review spoke to emphasised that the modelling environment is fundamentally important to the quality of the models produced. The review has grouped the modelling environment into three categories: culture, capacity, and control.

Culture

2.9 Almost all studies of organisational culture confirm the importance of clear leadership from the top of the organisation. Organisations the review spoke to also referred to the importance of model risk being recognised as a Board level risk.

2.10 It is vital that all levels in an organisation understand the value attached to models and quality assurance. Some QA experts expressed a belief that the resulting expectations of quality are more important in shaping behaviours than detailed processes designed to achieve such quality. Leadership is also about expecting and facilitating effective challenge. A key judgement for complex models is how to secure this challenge, and whether some form of external scrutiny or review is the best way to engage effectively with relevant experts.

2.11 Ultimately the purpose of models is to help decision makers make better decisions. Good models provide insights and understanding, but only if they accurately reflect the policy environment and are used correctly.

2.12 Successful modelling is therefore not just a matter of modellers accurately building models. Decision makers also need to understand the strengths and limitations of the chosen modelling approach. Departments' cultures should reflect this by minimising barriers between policy and analytical professions, and encouraging mutual understanding and respect, as well as emphasising the importance of communication skills.

2.13 Incentives for staff should align with this approach, so they understand the value of quality-assured outputs as well as timely delivery. Some stakeholders described the power of substantial reputational or financial consequences for responsible individuals if QA is found to be lacking. For example, one public sector organisation referred to the impact of QA on annual appraisals and promotion boards, while a private sector organisation referred to the impact on staff bonuses.

2.14 Several stakeholders emphasised the importance of analysts, whether model developers or model users, being empowered to say "no" where necessary, for example if more weight is attached to model outputs than can be justified by the robustness of the modelling process or if there is insufficient time or data to produce outputs of sufficient quality. There was universal agreement across stakeholders – from industry to academics – that if there are caveats these need to be clearly communicated, and if modelling is not possible within the given constraints, analysts should have the support and means to say so.

2.15 Communicating and understanding uncertainty in model outputs is therefore vital. For example, a research organisation told us that it was crucial that users of their models were aware of the confidence intervals around their model forecasts, although they also recognised that sometimes users just wanted to know a single figure.

2.16 A "no-blame" culture which encourages **transparency** regarding models, modelling approaches and limitations is more likely to enhance the quality of models and their outputs than one in which issues are hidden. One department described to the review the benefits they gained from regular meetings amongst modellers where each has to bring an example of an error that has occurred and explain what went wrong. This encourages collaboration between teams and promotes a culture of learning from mistakes.

2.17 Transparency is important because it facilitates effective scrutiny. Publishing all or some details of a model can therefore be a powerful quality assurance tool. Box 2.A below gives an example of a particularly transparent government model.

Box 2.A: The 2050 Calculator – Department of Energy and Climate Change

The 2050 Calculator is a scenario testing tool that allows users to explore different ways of reducing UK emissions by 2050. It was developed in-house by the Department of Energy and Climate Change in 2010. To date over 150,000 unique users have accessed the tool.

The 2050 Calculator sets a new standard for **transparency**. Both the model and its assumptions are published on the internet, and during development DECC published several "Calls for Evidence" and worked with hundreds of stakeholders. Users are impressed by the open, honest assessment of uncertainty, improving trust in the model and its insights. Journalists are enthusiastic; the Guardian calling it '…probably one of the most open and transparent pieces of policy-making ever undertaken by the British government'.

The transparency of the UK calculator led not only to free quality assurance from global experts in the field, but also tangible diplomatic benefits. For example, the Chinese Government published their own version of the 2050 calculator – a major breakthrough in transparency and Sino-UK climate change co-operation.

The team was also formally recognised, winning the Science, Engineering and Technology Civil Service Award in 2010.

2.18 A further benefit of an open approach is increased re-use of models or model components, i.e. **sharing** or collaboration across teams or departments. As well as increasing efficiency, re-use of tried and tested models can enhance quality assurance.

2.19 Making models as intuitive as possible can help drive transparency. Consultancy and accounting firms emphasised this point. They pointed to a number of techniques they employ, which include providing a guide upfront of what the model does, in prose not numbers; clearly structuring presentation of the model with key findings and graphs; and a logic map of the model. This makes the model easily accessible to reviewers, and so facilitates scrutiny.

Capacity and capability

2.20 As well as a culture that encourages high quality QA, organisations need a basic set of tools to carry out the task well.

2.21 A strong, and common, message from the private sector, academics and research organisations was that there is no substitute for **expertise and experience**. This is essential in building the judgement needed to gauge risk and spot errors. For an organisation as a whole, a key element of risk management is ensuring that models are developed, managed and maintained by appropriately skilled and experienced staff. This should include ensuring the model user is fully capable of using the model and understanding its outputs.

2.22 Several organisations talked about the value of experts whose experience enables them to recognise when results are inconsistent, and one quoted a figure of 10,000 hours to become such an expert. It is interesting to note that many accountancy firms have highly expert partners whose key role is quality assurance. In these organisations, expertise in model development and quality assurance is highly valued as a key professional discipline. Equally, many noted the role of professional standards such as CIMA, CIPFA, ACA and others.

2.23 In some cases it may be appropriate for those with the relevant skills to be 'in-house'; for other organisations this is not realistic, and they will need to buy-in expertise. In all cases, it is the ability to access and deploy the experience and expertise that is important, wherever this may originate. A **diversity** of backgrounds and experience in the team may help get the best out of individuals, helping teams to avoid group think and use individuals' judgment effectively. This can help counter situations in which a set of common assumptions prevent individuals from spotting simple errors.

2.24 The review saw many examples where a separate specialist teams conducted the QA, but others where the relevant specialists were embedded in other teams. There is no right answer here, although a key factor to consider – as raised specifically by one research organisation – is the ability to retain suitably experienced staff.

2.25 As well as the capability to achieve effective quality assurance, it is also necessary to have the appropriate **capacity**; that is, sufficient staff available and adequate time for the quality assurance process. Many stakeholders expressed the view that the biggest single impediment to achieving effective QA in practice was the allowance of insufficient time, and that this must be addressed as part of the planning process. A project and programme management approach is important here.

2.26 Readily available information or **guidance** on how to carry out effective QA, and the most common likely problems, can also contribute to an efficient and effective process.

Control

2.27 The third and final key factor in ensuring an appropriate environment for good QA is sufficient control, to ensure and verify that QA has been completed effectively.

2.28 The message from professional services firms was that this control element is essential, however strong the culture, because models are inherently prone to error. This is because of the degree of accuracy required in a mathematical model, where a misplaced 'plus' or 'minus' can transform the results. As one firm put it to us, a typing error in a prose document is unlikely to change its entire meaning, but with a model it could have profound consequences.

2.29 A key element of best practice involves establishing a single individual with overall responsibility for each model in development or each use of a model. This contributes to effective QA by creating a sense of ownership and **accountability**. In consulting and accountancy firms, it is the norm for a senior partner to sign-off on models prior to external release. Partners would undertake their own checks and seek comfort from the team that undertook the modelling. Some departments also seek to identify a clear chain of responsibility at the outset, reflecting the importance of the model. It is vital that organisational structures enable suitable individuals to be appointed to these roles.

Box 2.B: The role of the model SRO

In the public sector setting, ownership and accountability for specific models can be implemented by appointing a model SRO. The key prerequisites are that this should be a named individual with sufficient seniority to take responsibility for the model throughout its life cycle and sign-off that it is fit-for-purpose, prior to its use[1]. The model SRO may have either a policy or technical background.

The SRO must ask the right questions and satisfy themselves that appropriate QA is being undertaken – but they do not need to be a specialist to ask these questions. Instead, when assigning roles and responsibilities, departments must give careful thought as to the nature of the project, and ensure that the SRO is sufficiently senior to take responsibility for the business critical model in question.

Projects that depend on highly complex and sophisticated models may choose an SRO with the ability to understand the technical or analytical aspects of the model and to "sense check" the outputs. Similarly, projects dependent on complex analytical or economic assumptions will require an SRO who can understand the sensitivities and uncertainties inherent in the policy area. The key requirement is that policy professionals and analysts work together closely to ensure the model SRO is able to ask the right questions, fully understands the uses and limitations of the model and is therefore able to sign-off to confirm it is fit-for-purpose.

In either case, the SRO's sign-off assures (based on the model SRO's individual accountability) that:

- the QA process used is compliant and appropriate;
- model risks, limitations and major assumptions are understood by the users of the model; and
- the use of the model output is appropriate.

The sign-off covers both model development and output use, and potentially straddles analytical and policy disciplines. Therefore the model SRO may need to seek appropriate assurances from the other disciplines, to ensure there is a single coherent confirmation.

Reconfirmation of some or all of these would be required if the model was subsequently used for a purpose other than that for which it was originally designed or if the circumstances surrounding its use have changed. Where a model is being used for a new purpose/project, the model SRO will need to confirm that the model is suitable for the new use. See paragraphs 2.55-2.59, for more detail about these circumstances.

If the model SRO cannot give their sign-off, this signals the model is not fit-for-purpose. In this case, the model should not be used until any specific issues are rectified. This may entail amending the model, undertaking further QA, or producing a completely new model that better supports the policy need.

[1] A definition of the SRO role in Government appears in an OGC report, 'Review of the Senior Responsible Owner Role in the Major Projects and Programmes of Government', September 2009. This is based on the recommended approach in Managing Successful Programmes (MSP): "The SRO is the individual responsible for ensuring that a project or programme of change meets its objectives and delivers the projected benefits. They should be the owner of the overall business change that is being supported by the project. The SRO should ensure that the change maintains its business focus, has clear authority and that the context, including risks, is actively managed. This individual must be senior and must take personal responsibility for successfully delivery of the project. They should be recognised as the owner throughout the organisation."

2.30 It is important that responsibility for the quality of models is not divorced from responsibility for efficient resource management. To represent value for money, QA should be proportionate and tailored to the level of risk inherent in each model and its use. This does not involve automatically applying the maximum level of QA in each case simply in order to minimise the risk of any weaknesses. For example, the review learned that some consultancy firms – and parts of government – explicitly undertake a risk assessment at the start of each engagement to ensure they understand and apply the appropriate level of QA from the start.

2.31 The governance process for models should also establish an effective control environment, which, for example, defines appropriate change control procedures and approval processes.

2.32 A checklist approach to control can be a useful tool. For example, HMRC's analysts use a checklist for QA that is well understood and used across all business critical models. This identifies a clear process and sets out an assessment reflecting each model's importance.

Process – the model development lifecycle

2.33 Alongside the model environment, the right process is essential. This process must be based on engagement with the customer to ensure there is a shared understanding about the purpose and limitations of the model. It is also about creating a system to ensure certain actions are always undertaken at the appropriate point, and the right questions asked. It is about embedding QA in model development, to reduce error.

2.34 Every organisation approaches process differently, based on its needs and the level of risk. But two elements are essential:

- guidance should set out exactly what a model developer should consider, at each stage of model development. It should be as simple as possible – as one analyst put it to us, you should be able to 'press a button and the QA machine starts'; and
- documentation should be created as the model is developed, to set out its purpose, limitations, risks, and QA undertaken. The aim is to ensure the model and its risks are transparent. This is important because it promotes effective control, and facilitates future use of the model.

2.35 Taken together, these two products can help prevent errors and, where they occur, ensure teams can pick them up quickly. The rest of this section sets out some key considerations that might be expected as part of guidance for each stage of a model's lifecycle. It includes reference to the stages at which documentation is necessary.

2.36 This section draws on the best practice from the variety of organisations the review team have spoken to. It aims to capture the key factors all parties should consider when commissioning, designing and building a model, and represents a sequential, step-by-step approach to model development.

2.37 As shown in Chapter 3 of this report, business critical models vary widely in complexity and risk. Any guidance should be proportionate to the organisation and the specific models in question. This process is also not meant to be prescriptive; for some models, certain steps may not be necessary or can be run in parallel. However, those involved in modelling work should consider the appropriateness of all the steps.

2.38 Where an existing model is being considered for a new purpose or in new circumstances, in either its current or modified form, not all of the stages below will be required. Some form of quality assurance, however, will still be vital. This situation is discussed further in "adapting this process", below. Similar considerations may apply when a new SRO is appointed for an existing model.

2.39 Chart 2.B sets out the four key stages of the model development process, which forms the structure for the rest of this chapter. In reality this process may not be strictly linear, and may need a degree of iteration.

Chart 2.B: The four stages of the model development process

2.40 At all times, it is for the model developer and the model customer to agree – in discussion with the SRO – what constitutes a proportionate approach to both the model development and any supporting QA. A strong relationship between the customer and the developer is key to ensure both parties understand the requirements driving the model development and what the model can and cannot provide. Box 2.C, below, sets out the different roles which are likely to exist within a public sector organisation that develops and uses models.

Box 2.C: Roles within the model development process

Although details may vary according to the circumstances of individual projects, the review's work with departments indicates that there are generally three main parties concerned with the use of models in the public sector:

Model developers – these analysts build the models and normally undertake quality assurance (verification[2]) on the model itself.

Model users – these run the models to produce outputs and interpret the results and may undertake quality assurance on the model inputs and outputs (validation). They may be the same analysts as the model developers or may be separate.

Model customers – these use the results from the modelling as part of their decision-making process. They need to be aware of the model limitations and confident that the results are robust for the use that they are making of them, e.g. whether for procurement and commercial, forecasting or policy simulation uses. They will need to work closely with developers and modellers to agree the scope and specification.

As described in more detail in Box 2.B above, a model SRO should take overall responsibility for a model and its use. They will normally be drawn from the senior management of one of the groups above.

Scope, specify and design

2.41 There should be a clear understanding of the requirements and scope between the customer and the model developer at the commissioning stage. The modeller needs to have a

[2] The terms 'verification' and 'validation' (V&V) used in this document are consistent with international quality management system ISO9000. Verification is considered a quality control process used to assess whether a model meets the initial specifications. Validation is considered a quality assurance process used to establish, to the necessary degree of assurance, that a model meets its intended requirements. Verification is generally an internal process while validation often involves acceptance of fitness for purpose with end users and other stakeholders.

good understanding of the decision or policy question that is being posed and what the end use of the model output will be. Equally, the model customer needs to understand the constraints, limitations, risks and complexity involved in any proposed modelling. This clarity around the intended scope and use of the model was an important theme from professional bodies and professional service firms. One consultancy firm gave the example of a particular final report which devoted 25 pages to setting out the context of results and questions being addressed, with only five pages dedicated to the model outputs.

2.42 It is therefore important that customer and developer clearly agree a definition of the scope of the modelling task at the beginning of the process, and document it. This scope will be the basis for model development going forward, so it is important that both sides understand it. Agreeing these issues at this early stage allows both parties to capture and manage any risks. Once the modellers and policy or other customers have agreed the scope, they should produce a formal **specification document** before model development begins.

2.43 It is important that the **design stage** includes a clear understanding of the model structure and logic as well as the underlying assumptions, limitations, inputs required and outputs expected. The model SRO should at this stage check that the proposed design meets the organisation's requirements. They should check the assumptions, limitations, inputs and outputs to make sure they remain consistent with the intended use of the model, and discuss the most appropriate approach to QA.

2.44 Tables C.1 and C.2 in Annex C set out some of the issues to consider during the scoping, specification and design process, and in what circumstance.

Build/populate

2.45 The next stage is to build and populate the model based on the model design. Depending on the type, complexity and use of the model this may take the form of a spreadsheet, use of a suitable software modelling environment or writing dedicated computer code. The modelling team therefore needs to take an informed decision on the best build approach.

2.46 This is the stage where much of the verification testing takes place and will include QA for the model assumptions and input data, as these are critical to understanding the risks and limitations of the model outputs. It is important to consider these components at this stage, to ensure the model outputs are as robust as possible. This might include the methods outlined in Table C.3 in Annex C.

Test

2.47 At this stage the completed model should be available, together with a full set of quality controlled input data and details of the model's inputs' limitations or uncertainties.

2.48 It is important to develop a program of validation testing that is proportionate to the risk, complexity and novelty of the model under consideration. It is at this stage that the model SRO should ensure that the model is fit-for-purpose. A number of external stakeholders highlighted the importance of sense checking by an expert. The ability to understand if the model results are sensible is a key component of both testing and model use.

2.49 As with all stages in this process, the level of testing should be proportionate to the need. However, it is important that sufficient time and resource are available at the testing stage. Table C.4 in Annex C sets out examples of appropriate QA at the model test stage. Box 2.D below provides an example of one model, Pensim2, which outlines the developer testing involved in this complicated model.

Box 2.D: Pensim 2 – Department for Work and Pensions (DWP)

Pensim2 is the DWP's in-house dynamic micro-simulation model for policy simulation of reforms affecting pensioner incomes, and is developed by a dedicated team of analysts in the Model Development Unit. The model produces distributional impacts of reforms and estimates the cost and impact of changes to pensions' policy to 2100. The model mainly uses administrative and survey data and is a complex model built in modules on a Genesis platform. Pensim2 is business critical as there is significant risk to government finances if estimates of the cost of pension reforms and pensioner income-related benefits are incorrect. There is also a large reputational risk to government if reforms are changed at a late stage due to modelling error. Pensim2 is therefore subject to a lot of quality assurance and undergoes a programme of continual checks and improvements. Particular care is taken when using the model for macro purposes, when external results-based checks and calibration may be carried out.

A number of quality assurance techniques are used, of which a key one is developer testing. Initial QA of any new modelling is undertaken by the developer and the impact of the change is examined by analysts before they sign-off the change. A detailed 'Change Control Matrix' is maintained by the development team that lists all the modules affected by a particular change. This reduces the probability of errors occurring when multiple developers are working on the same release. There are standard diagnostic and summary tools to help identify errors in coding and trace dependencies within the model. There is code to quickly produce 'standard outputs' that cover the whole range of outputs from the model so that developers and users can easily spot unexpected consequences of changes. A regular clean-up of code maintains transparency and usability. The underlying Genesis architecture is designed to ensure that the model is not a 'Black Box' and facilitates developer testing.

2.50 Transparency can be a powerful tool at this stage, as it allows the modelling team to harness the expertise of many third parties. Stakeholders often quoted external peer-review (whether through scientific publication or external model audit reports) as the gold standard of transparency. For example, in 2010 Met Office scientists published 263 papers, 80 per cent of which were co-authored with external partners, supporting the development of their Meteorological models.

Deliver and use

2.51 Once the model is fully tested and has a suitable set of documentation, the modellers should hand it over to the customer as agreed in the specification. The customer and the modeller should formally agree that the model meets the specification and the appropriate QA processes have been applied and that the model is fit-for-purpose. The model SRO will need to formally sign-off at this stage.

2.52 The formal deliverable will vary depending on the model; however there should be clear documentation as outlined at Box 2.E. This could be a quite brief, bullet-style list if the modelling is relatively straightforward or low risk. High risk, complex or novel models may need a more detailed set of documentation covering specification, design, build and testing.

2.53 Box 2.E, below, sets out the documentation that is likely to be needed at each stage of the model development process.

Box 2.E: Documenting QA – a best practice framework

At the design stage

Model design documentation to support the build phase describes the model, and should include the quality assurance strategy for the build and testing phases.

Some QA may be performed at this stage to provide assurance that the model structure, logic and assumptions are robust before the model is built. Review by either internal or external reviewers should be considered for complex models and an assessment of the suitability and availability of the inputs and outputs should be made.

At the build stage

The documentation at this stage accurately describes the model as developed (noting any differences from the design), any verification testing done and the test results.

Once the model is complete and has been subject to appropriate verification testing, a further validation testing phase should be conducted, and documented, to ensure the model is fit for the purpose.

At the test or 'deliver' stage

The documentation includes: a description of the tests run; the test results; any issues identified; and corrections made. If user documentation is needed it should also be developed and reviewed at this stage together with any required training material.

At all stages

The documentation should be comprehensive yet proportionate to the risk and complexity of the model. For example less complex, lower risk models may only require a short description of the model at the design stage. However, more complex or higher risk models would be likely to require a more formal approach to documentation.

2.54 Once a model is in use, the need for QA is not over. On an ongoing basis, the model SRO and model customer need to ensure the model use is appropriate. Particular care must be taken if the model is subsequently used for a purpose other than that originally intended or in changed circumstances, as discussed below. These considerations are also outlined in Annex C, at Box C.5.

Adapting this process and ongoing use

2.55 The process above relates primarily to situations in which a team is developing a new model, to support a specific policy goal. However, in some cases models need to support many policy goals, or existing models need to address new policy questions or be used in changed circumstances. Even in these situations, model developers should apply the underlying principles of good QA.

2.56 When a model is supporting a range of policy areas, the review's conversations with departments suggest the model customer should be responsible for ensuring that the model is fit-for-purpose for their specific policy needs. The policy team may not own the model, but the existing model SRO and model user should reach an understanding of the customer's needs and the capabilities, limitations and risks of the model in this context. The model SRO will need to confirm suitability for the model's new use. Equally, the model user should consider creating a specification document comparable to the original model description, as outlined in the delivery

phase of the process above. This can help identify any differences between the existing model capabilities and the needs of the specific policy question at hand.

2.57 The model SRO and policy customer then need to reach specific agreement as to whether to accept the risk associated with any differences, or commission modifications to make the model more suitable. If they decide to modify the model, they should consider a cut-down version of the QA process above, proportionate to the risk associated with any changes.

2.58 In the case where the model customer decides the policy need can be met by an existing model, the responsibility falls on them to ensure the existing model and QA processes are fit for the new purposes – in consultation with the model SRO. It is dangerously tempting to assume that because a model used to be appropriate in a similar area, it is just as appropriate in the new project. Subtle differences between business areas, as well as changes in assumptions over time, can affect a model's validity. The model customer must reassure themselves that the model they intend to use is appropriate to their needs.

2.59 Similar issues can arise where an existing model is to be re-used for the same purpose, as circumstances or assumptions may change with time. Again the model customer and the model SRO must reassure themselves that the model (including the data and assumptions as well as the model itself) is still appropriate to their needs.

Proportionality and ensuring levels of QA are appropriate

2.60 Even for highly business critical models, there can be no 'one size fits all' approach to determining what level of QA is appropriate. There are good reasons why the approach to quality assurance will vary between models. These include:

- the type and complexity of the model. Highly complex models require more QA;
- the novelty of the approach. Using a previously untried modelling technique requires more QA;
- the importance of the issue. Different issues will vary in their economic and social impact;
- the relevance of the model to the decision making process. When a model forms only one component of a broad evidence base, less QA is required than if the decision is heavily dependent on one model;
- the precision of the model outputs. Imprecise models can need different QA than precise models. This may be because of inherent limitations of the modelling technique, or a lack of data on model assumptions; and
- the amount of resource available for the modelling which includes QA. The value for money of any additional QA must be balanced alongside the benefits and the risk appetite that exists.

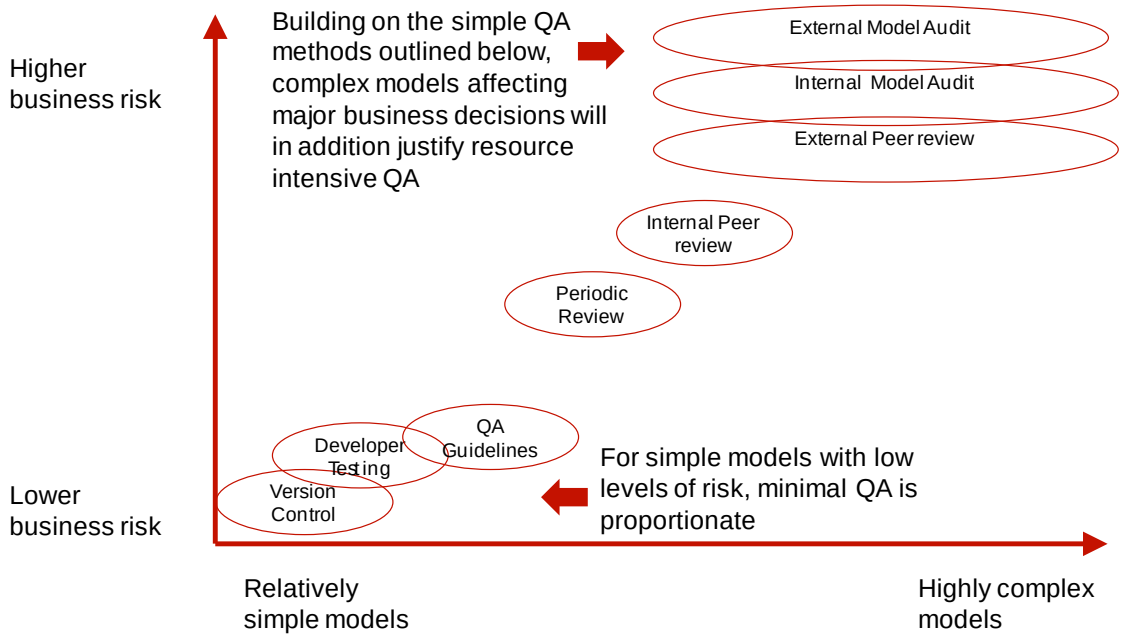
2.61 This illustrates the importance, at all stages of model development, that analysts and their customers take a conscious decision on the amount and type of QA that is appropriate. One way to achieve this is through a 'checklist' which some departments (HMRC for example) use to aid the QA process, and which enables the model SRO to sign-off that processes have been appropriate.

2.62 Unfortunately there is no shortcut or 'iron rule' which can define the ideal type of QA for a given model. Instead, model SROs should consider a range of QA measures, and when deciding whether they are appropriate, assess the risks and consequences of not undertaking them. If the

model SRO believes that exhaustive QA is not necessary to mitigate project risks sufficiently, this can be an appropriate approach to take.

2.63 Chart 2.C below illustrates some of the differences that might be expected in the approach to quality assurance, depending on the nature of the model, and variations in model complexity and business risk. This chart is indicative only and the detail of the various QA activities will vary depending on the model in question. Some methods, e.g. transparency, would be expected to apply across the piece, as well as identifying an SRO for all business critical models.

Chart 2.C: Schematic showing indicative types of QA that might be expected given different levels of risk



2.64 It is to be expected that most models will be subject to basic version control processes and developer testing, but that external model audit is appropriate primarily for the most complex models and/or those with high business risk. Circumstances when teams should particularly consider external model audit include higher levels of risk arising from influence on critical decisions, particularly complex models, where there is concern over possible "group-think" amongst those involved with the modelling, or where there have been recent changes in personnel, circumstances or model usage. Note also that the list of techniques is not exhaustive, nor are they mutually exclusive in any sense, for example a model being externally audited is likely also to have a number of other "lower level" techniques applied which may include internal auditing or peer review.

2.65 It is also worth emphasising that the nature and extent of each of these types of QA may vary depending on what is appropriate for each model. An important example of this is external model audit, where there is a clear distinction between:

- a comprehensive model-based audit which focuses on whether or not calculations are correct. This is likely to be resource-intensive but will probably only be needed once; and
- a less detailed results-oriented audit which focuses on whether or not the results are reasonable. This should be quicker but is likely to be required each time the model is used.

2.66 Box 2.F below outlines one government model, the Pandemic Model at the Department of Health, to which a range of QA measures apply.

Box 2.F: Pandemic model – Department of Health

The Pandemic model is a suite of science-based models that model the impact of future pandemics. The model's rationale is to guide pandemic planning and preparedness plans, assist stockpile procurement and identify potential pressure points on the NHS and other aspects of national life (e.g. absenteeism), to cover infection rates and cost effectiveness of countermeasures. It enables real time modelling of a pandemic to inform Cabinet Office Briefing Rooms during a pandemic and supports World Health Organisation and European Centre for Disease Control processes. Overall ownership of outputs and advice is retained by the DH, but the model requires the input of ALBs and external parties.

QA robustness relies on multiple planks ranging from expert peer review, both internally and externally, through publication of results, to reliance on the professional and internal modelling standards of the various model development teams and parallel modelling streams to confirm a consensus view. There is a standing specialist governance group to oversee pandemic modelling. The results are either published in the scientific literature or are presented in the pandemic modelling summary on the DH website.

3 Current quality assurance in government

3.1 This chapter sets out the extent and nature of QA which currently applies to business critical models in government.

3.2 It is based on data returns from departments, and qualitative information about QA practices. To our knowledge, this is the first time data on business critical models and their quality assurance has been systematically collected. The purpose of doing so is twofold:

- to gain a picture, across government, of the nature and extent of modelling and any patterns or lessons emerging on quality assurance and through this;
- to give departments and their Arm's Length Bodies (ALBs) a point of comparison with models across government.

What the review asked of departments

3.3 At the start of the review, the team asked departments to detail any models used by themselves or their ALBs which qualified as 'business critical'. In assessing business criticality, the review asked departments to bear in mind the extent to which the model drives key financial and funding decisions, the extent to which it was essential to the achievement of their business plan, and the extent to which error could lead to serious financial, legal or reputational damage.

3.4 The review also asked departments to detail the QA processes that applied for each business critical model, for themselves and their ALBs. To assist in this, the team provided a list of elements of quality assurance. These included: developer testing, internal peer review, external peer review, use of version control, internal audit, QA guidelines and checklists, external audit, governance, transparency, and periodic review of model development over its lifetime. The team invited departments to add their own categories if they felt it appropriate. Chapters 1 and 2 of the review's interim report set out full details of the request to departments.

3.5 The review requested data to help build a picture of current business critical models and their QA. The team also met with individual departments to further understand the way quality assurance is conducted. The team then summarised this data to provide a snapshot of the different types of QA in use across government.

3.6 Two key caveats are important to bear in mind when considering the data analysis below:

- these statistics represent **a snapshot** of business critical models and QA status. They capture a point in time, late 2012, not including models in development and models that have been used in the past and that are not currently expected to be used again; and
- this analysis is necessarily descriptive, and **should not be used to form judgements**. As discussed already, the review would expect there to be a wide range in the approach to QA across different models. To be effective, and represent value-for-money, QA needs to be proportionate to the significance of the decision, the complexity of the model (including key inputs and assumptions) and the degree of risk and uncertainty.

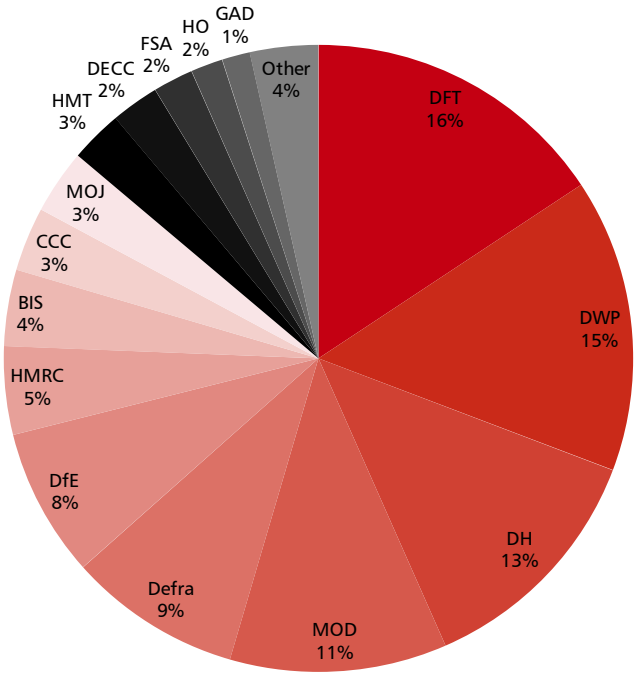
The extent and nature of government modelling

Numbers of business critical models and distribution by department

3.7 Part of the review's remit was to identify and map where the most significant models lie in government. The departments, and their ALBs, identified just fewer than 500 business critical models.

3.8 There is a large variation in the number of business critical models different departments' use, as would be expected given the range of functions departments fulfil. Smaller departments generally have fewer models – and indeed some small departments identified no business critical models, for example the Department for International Development. As Chart 3.A shows, the larger departments make up a large proportion of the models, with DfT, DWP, DH and MOD holding over 10 per cent of business critical models each and making up just over 50 per cent of the total number.

Chart 3.A: Chart showing distribution of models by department



Source: Analysis of departmental returns
NB: Departments with fewer than 6 models are included in the 'Other' category. FSA refers to the Food Standards Agency and CCC refers to the Committee on Climate Change.

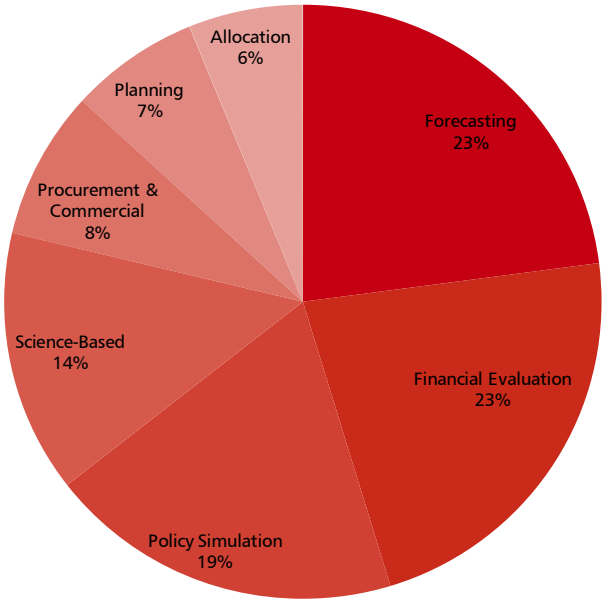
Types of business critical model

3.9 To give a better sense of how models are used in government, the review asked departments to classify them according to type.

3.10 As Chart 3.B below shows, around two-thirds of the business critical models in government are of the financial evaluation, planning, policy simulation or forecasting variety. It is to be expected that these are important areas for modelling. Government departments undertake a large amount of commercial and procurement activity and this tends to involve a suite of models which are often then applied to many competitions. This explains why the proportion of procurement and commercial models is less than the review would have expected; many

departments provided the overarching model and the QA process which would be the same for a variety of activities.

Chart 3.B: Chart showing types of model as a proportion of all business critical models in use in Government



Source: Analysis of departmental returns

Distribution of different model types by department

3.11 As would be expected, some departments have a higher prevalence of certain types of models than others.

3.12 For example, Defra has a high proportion of science models – in fact nearly half of Defra's models come under this category. Some departments have higher proportions of forecasting and policy simulation models; HMRC, DWP, BIS and DECC for example, whilst others have a high proportion of allocation models as at DfE. This variation in model type each department employs correlates well to the main functions and remits of the departments.

Quality assurance mechanisms applying to government models

3.13 Charts 3.C, 3.D and 3.E below set out key statistics on the QA mechanisms which apply to business critical government models. This information is descriptive only.

3.14 As discussed earlier in this chapter, the review would not expect all models to have all types of QA. This would be disproportionate, as the complexity of the model, risk involved and other factors will determine the appropriate QA process for each model. However, it is instructive to build a sense of the trends across government.

3.15 The charts show that nearly all models use developer testing, internal peer review and version control. This is to be expected given these are often relatively simple standard practices in any modelling. However, it is nonetheless encouraging to see that every business critical sent to us by departments either has developer testing or some form of peer review.

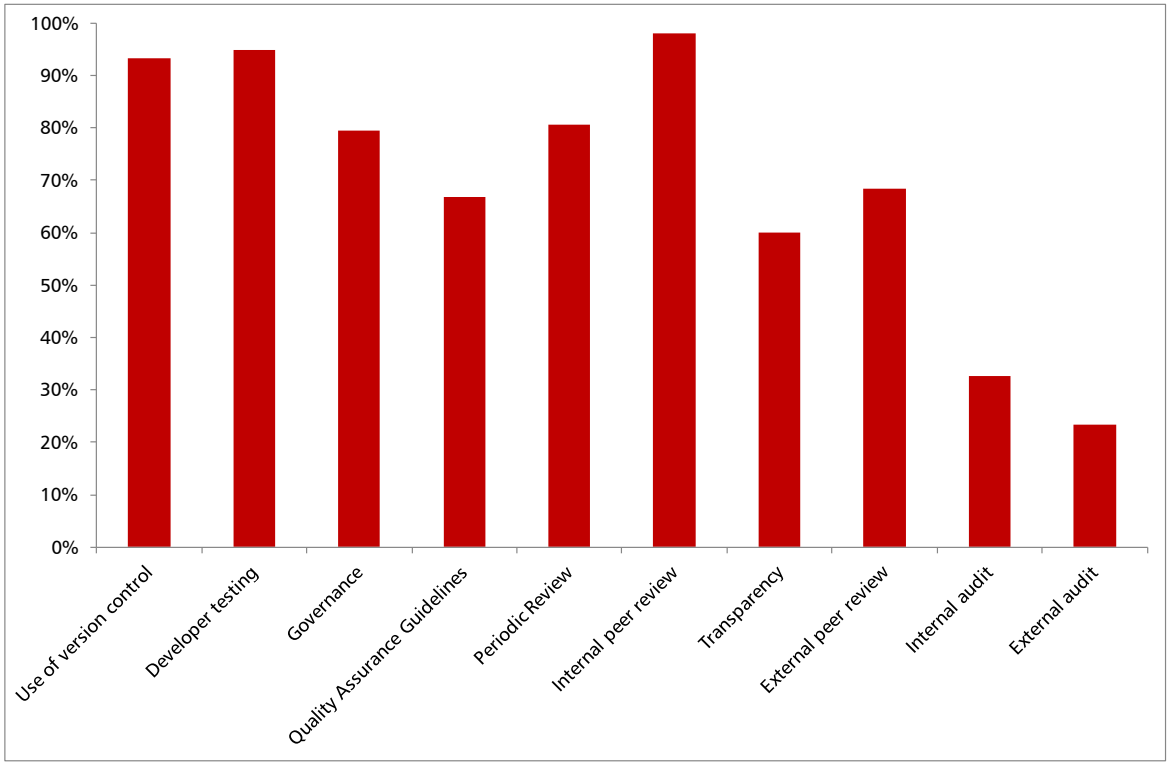
3.16 Around one third of models had some model audit – either internal or external. This is an often time-consuming and resource-intensive QA method, and it is to be expected that it would not be used across all models.

3.17 A fairly high proportion of models (around 50 per cent) had outputs that were available to external scrutiny and so are classified as 'transparent', though only a small proportion of these have the model itself in the public domain. Many of the decisions which are underpinned by business critical models are for internal government use only, though it is clear that where model outputs can be shared more widely this is often done.

3.18 In many cases, models are created and developed by external contractors such as accountancy and economic consultants. In these cases, it is to be expected that the contractor would often follow QA guidelines as per professional standards within the firm. In addition, if the model is maintained by an external firm, then the firm would use version control as dictated by the agreed contract. Similarly if the model is bought off-the-shelf it is reasonable to expect it would have been tested and internally peer reviewed. As a consequence, in either of these two cases any QA by the department would be classified as external review/audit, not internal.

3.19 The review team is aware that many departments are currently undertaking internal reviews of their own models and QA processes and as part of this they are producing and/or revising QA guidance. Therefore in many cases we expect that departments will have QA guidelines which have been worked up in parallel to this exercise.

Chart 3.C: Chart showing types of QA used in government models



Source: Analysis of departmental returns

3.20 Charts 3.D and 3.E below focus on the different elements of QA and whether they contribute to strengthening the wider model environment, or form part of putting effective process in place.

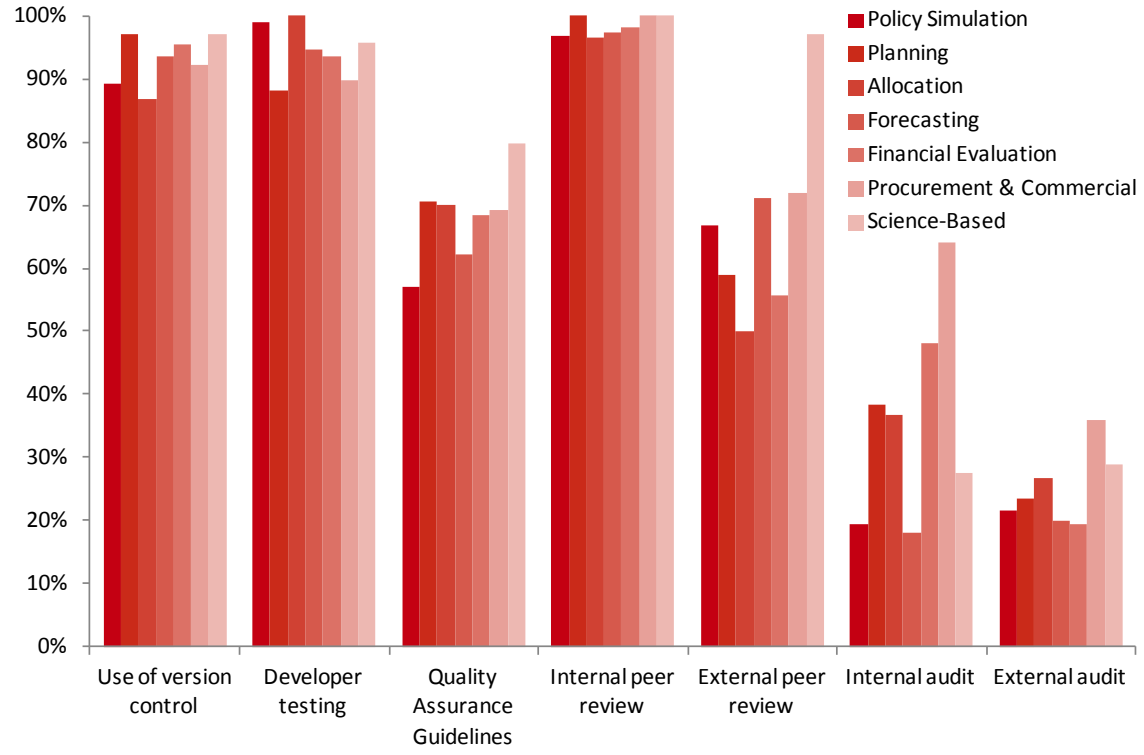
3.21 When considering the types of QA and the types of models Charts 3.D and 3.E below show that science-based models tend to have the most extensive types of quality assurance. This is especially the case with external peer review and transparency, reflecting a strong culture in the scientific community of peer review before publication.

3.22 To assess if audit was included in the QA processes the team asked departments and their ALBs to classify this as present only when there was evidence that the model had been checked by professional model auditors. In cases where teams had involved others in model audit-type activities, departments and their ALBs classed it as peer review. So it is not surprising that model audit, both internal and external, is rarely used and limited to a subset of models – most often in science-based, procurement and commercial and financial evaluation models.

3.23 In terms of the environment-based QA there are similar amounts of governance and periodic review across all model types. Periodic review is an assessment of whether the model is fit-for-purpose when a model is being used on an ongoing basis or after a period of time has lapsed for a different use to that originally intended. It therefore makes sense that periodic review is not present in all models and is spread across all model types as, in each model type, there will be some models which require this review and some that are one-off models so do not.

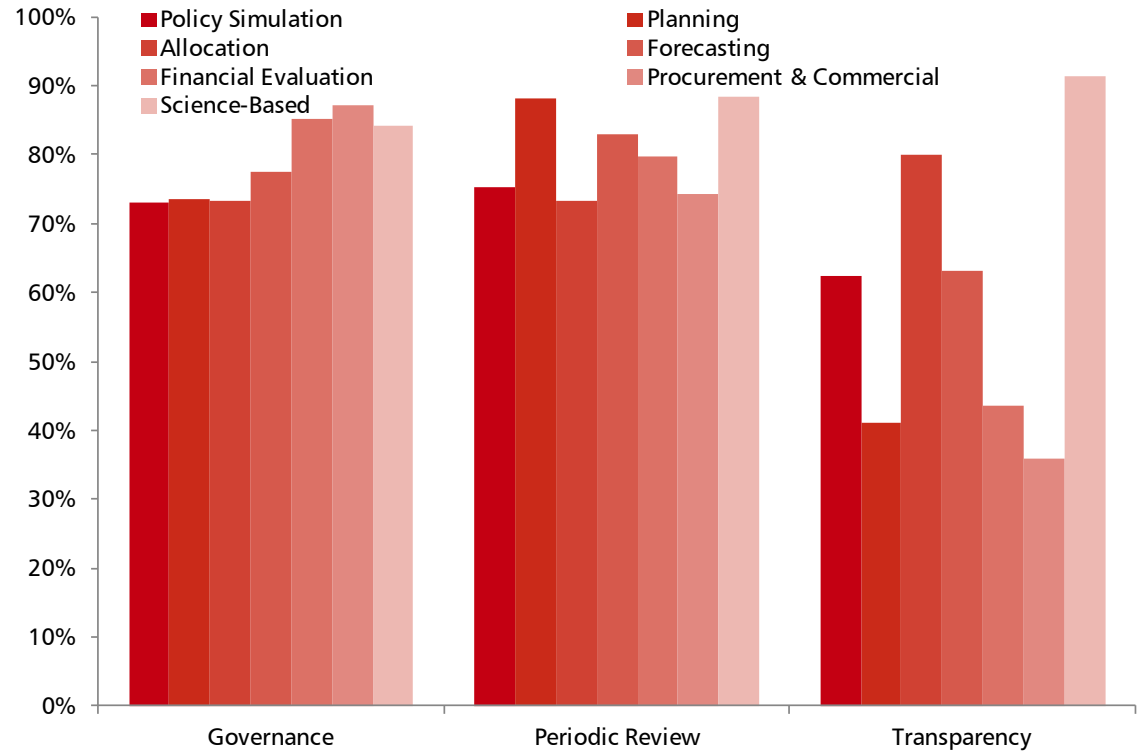
3.24 The degree of transparency tends to vary, with planning models understandably having low transparency as they are often modelling key government business. Procurement and commercial models might be expected to have a greater degree of transparency, owing to the open competition process and disclosure required by the EU procurement law. However, as these models may contain other sensitive information which is not required to be disclosed, the publication of the model might compromise the department's commercial position and therefore the models are retained for internal use only.

Chart 3.D: Chart showing QA by model type – process-based QA



Source: Analysis of departmental returns

Chart 3.E: Chart showing QA by model type – environment-based QA



Source: Analysis of departmental returns

Qualitative messages from departments

3.25 In addition to the quantitative data described above, the team also gathered qualitative insights from departments about QA, and its current strengths and weaknesses. The team also asked to see existing guidance which the department made available to its analysts.

3.26 The key messages from this broader, qualitative work were as follows:

- a **wide variety of guidance on QA** exists already within government. It is encouraging that a number of principles are common to this work. For example, in discussion almost all of the organisations highlighted the value of using someone independent from the project team to review the model and provide effective challenge, and almost all the processes require a formal review of the model by someone who has not been directly involved with its development. There is, however, a wide variation in the scope and format of these documents. For example, some but not all provide criteria to help decide on the extent of QA that should be undertaken. Some make the distinction between verification (the process through which the model is reviewed to ensure it is error free and satisfies its specification) and validation (a wider review to ensure that the model is fit for the purpose it is being used for), while others do not include this detail;
- **lack of both time and resource** can make good quality assurance challenging. This becomes a particular risk if caveats are not appropriately communicated to policymakers. Some aspects, including model documentation, can suffer when time is short;
- there are challenges in preserving good quality assurance when a model's **scope and purpose shifts** in response to often sudden change in policy and priorities;

- some departments have a very **clear governance structure** for models, with an SRO assigned from the start. However, this is not the case everywhere;
- **machinery of government change** can lead to legacy issues with models that started in one department, and subsequently end up owned by another. It can be challenging to track the development of these models and update them;
- departments with the most developed quality assurance processes appeared to have sufficient **specialist and experienced staff**, but not all felt they had the staff with the right skills in place to match the demands on them. Retaining specialist staff and providing career progression for experts was highlighted as a challenge;
- some departments have a strong **culture of openness and discussing mistakes**, but this is not uniform across government; and
- in general, and in summary, departments felt that there is a lot of **good practice** in government, but this is not always standard across or within departments.

Conclusion – quality assurance across Government

3.27 The data returns and work with departments show significant variation in the type and nature of QA used within, and between, departments and their ALBs. Much of this variation is to be expected, and is a natural function of the varying business critical models that different parts of a department and its ALBs will use – and of diverse departmental remits.

3.28 There are good signs of an effective baseline or 'minimum standard' for QA across government. These include the broad spread across departments of key basic techniques like internal peer review and the extent of internal QA guidance.

3.29 The conversations with departments indicate some challenges, including ensuring they have the right skills and capacity, and dealing with time pressure and sudden changes in scope.

3.30 There is therefore a need to define how best practice in QA can be systematised and extended across the whole of government. It is to this that the next chapter turns.

4 Conclusion and recommendations

4.1 In light of the experience on the Inter-City West Coast franchise competition, the Cabinet Secretary and Head of the Civil Service commissioned this review to identify best practice in quality assurance (QA) of business critical models, and recommend improvements.

4.2 Models are used extensively across government to make strategic investment decisions, ensure key services are properly planned and better understand future risks and challenges. It is essential to sound decision making that they are robust.

4.3 This review has collected information on around 500 business critical models and their QA frameworks. These models drive a range of vital outputs which affect the wellbeing of this country. As part of the review, the list of business critical models identified and the quality assurance procedures that apply to them is being published at Annex D of this report.

4.4 Models influence many billions of pounds' worth of government expenditure, as well as other significant decisions which cannot easily be quantified. In many cases, the models and those who produce them must respond, at pace, to a fast-changing policy environment. Recent high profile cases should not obscure the fact that much government modelling achieves its task quietly, yet effectively.

4.5 The review found many examples of good practice within government. Some departments and their arm's length bodies (ALBs) have a clear and structured approach to quality assurance and a well-defined governance framework. There is much that can be learnt from this. Equally, almost all models use developer testing and internal peer review, demonstrating there is a basic application of quality assurance across the board. A significant proportion had key elements of the model in the public domain, enabling external scrutiny. Similarly, the review found an appetite for continuous improvement across government, with many departments and their ALBs assessing their internal processes alongside the work of the review.

Learning from stakeholders inside and outside of government

4.6 Recent events highlight what can go wrong when complex models are used to tight timeframes, and without a clear and robust governance framework. While much effective QA is undertaken, there is scope to sharpen it and ensure it extends universally across government

4.7 Stakeholders from a wide range of backgrounds highlighted the foundations of good practice, which should be embedded across government. In particular, they pointed to strong leadership from the top that values and expects effective challenge, a clear governance framework, and adequate time to allow expert and experienced staff to carry out quality assurance. They emphasised that policy-makers should understand the limitations and risks of a model and take these into account in deciding the best way forward. Together these factors can create an environment where quality assurance is seen as a central plank of risk management and effective government.

4.8 Openness about key elements of a model can reinforce these foundations by allowing external experts to engage effectively, and can also help to spread knowledge and understanding about best practice.

4.9 Stakeholders also stressed the importance of process, including clear guidance setting out the overall approach to quality assurance, and model specific documentation. There is a significant range in the level of detail of existing guidance, and different parts of government should to decide how best to meet their specific needs. However, succinct and generic guidance is needed that is relevant across government.

Delivering best practice across government

4.10 The review has identified two cornerstones of effective QA: appropriate, well-defined processes, and an environment conducive to QA – encompassing culture, capacity and capability, and control. On both counts, process and environment, more can be done. There is scope to strengthen and share skills across government, to ensure clear governance and leadership, and to develop effective processes and guidance across the board. More can be done to develop effective challenge, allowing modellers to raise concerns at a senior level, and to create a culture which discusses and learns from mistakes.

4.11 An effective **environment** includes creating:

- a culture where QA is highly valued, and there are strong incentives to deliver appropriate QA, backed by effective scrutiny of key models;
- capacity and capability where specialist staff have sufficient time built-in for QA, and are able to draw on expertise and experience across government and beyond; and
- adequate control, including a clear governance framework.

4.12 An effective **process** involves ongoing engagement between specialist and policy staff to ensure there is a shared understanding about the purpose and any limitations of a model. This should include sensitivity analysis, and the degree of uncertainty about model inputs, assumptions and outputs. This needs to be backed by:

- clear guidance that sets out the key considerations driving the approach to QA; and
- clear documentation about the model and QA process.

4.13 This review comes at a time of considerable scrutiny by departments and their ALBs of their own internal procedures on QA. The review has benefited enormously from their openness about the challenges they face and desire to bring about further improvements.

4.14 The recommendations below therefore aim to support departments by setting out the key elements needed for good environment and process. The review recommends departments and their ALBs should develop plans for both the above elements in a way that fits with their remit, and is proportionate to risk. The inputs specified below are not exhaustive, but rather give a minimum guideline as to what organisations should address in QA plans going forward.

4.15 The review also sets out recommendations to create incentives for continued good practice. Central to this is embedding Board level responsibility for ensuring an appropriate quality assurance framework is in place and backed by clear process.

4.16 The recommendations below relate to business critical models, which by their nature require greater consideration of QA. It is for departments to determine the extent to which they may also apply these recommendations to non business critical models in their remit.

Recommendations for government departments and their ALBs

Recommendation 1: All business critical models in government should have appropriate quality assurance of their inputs, methodology and outputs in the context of the risks their use

represents. If unavoidable time constraints prevent this happening then this should be explicitly acknowledged and reported.

Recommendation 2: All business critical models in government should be managed within a framework that ensures appropriately specialist staff are responsible for developing and using the models as well as quality assurance.

Recommendation 3: There should be a single Senior Responsible Owner for each model ("Model SRO") through its lifecycle, and clarity from the outset on how QA is to be managed. Key submissions using results from the model should summarise the QA that has been undertaken, including the extent of expert scrutiny and challenge. They should also confirm that the Model SRO is content that the QA process is compliant and appropriate, model risks, limitations and major assumptions are understood by users of the model, and the use of the model outputs are appropriate.

Recommendation 4: The Accounting Officer's governance statement within the annual report should include confirmation that an appropriate QA framework is in place and is used for all business critical models. As part of this process, and to provide effective risk management, the Accounting Officer may wish to confirm that there is an up-to-date list of business critical models and that this is publicly available. This recommendation applies to Accounting Officers for Arm's Length Bodies, as well as to departments.

Recommendation 5: All departments and their Arm's Length Bodies should have in place, by the end of June 2013, a plan for how they will create the right environment for QA, including how they will address the issues of culture, capacity and capability, and control. These plans will be expected to include consideration of the aspects identified in Box 4.A below.

Box 4.A: Modelling environment

1 There should be visible leadership at the top of the organisation – backed by incentives – to create a culture that expects high quality QA, including by:

   a senior staff, including the Accounting Officer, demonstrating in practice the importance they attach to appropriate QA;

   b creating opportunities for non-specialist senior staff to better understand key aspects of quality assurance, either as part of ongoing management training, or through seminars. This could be formally recognised in performance objectives;

   c valuing effective challenge. Internal steering groups or project boards should routinely invite critical challenge from experts both within and outside the organisation;

   d being open – where possible – about key elements of a model, and with a view to engage with external experts and invite effective scrutiny. This can help to drive forward best practice;

   e recognising the importance of QA for specialist staff through personal objectives, development plans and performance management systems; and

   f expecting policy staff to have a good understanding of a model's purpose and limitations as well as the risk and uncertainty surrounding the inputs and outputs. This should be reflected in the process supporting model use and development.

2 There should be appropriate capacity and capability where specialist staff have sufficient time built-in for QA, and are able to draw on expertise and experience across government and beyond, including by:

   a ensuring access to appropriate and specialist staff with the necessary skills to conduct QA, reflecting the needs and risk tolerance of each organisation and the required types of QA;

   b recognising the critical role of expert and experienced staff, including the importance of clear communication, through the performance appraisal process and in the options for career progression;

   c recognising that specialist skills are important at a senior level if there is to be effective challenge for how key models are used and interpreted;

   d harnessing the expertise and experience that exists across government and beyond. For example, using professional networks to identify staff with expertise in particular types of modelling and the appropriate quality assurance framework; use of short-term secondments within government and outside to build expertise, and help to embed best practice; and

   e using project and programme management techniques to ensure sufficient time for QA is built in from the outset, and ensure analysts are empowered to highlight the substantial risks where they have significant concerns about the robustness of the work. To support this, the governance framework should include a specific route for effective challenge.

3 There should be adequate controls in place, including a clear governance framework. Key elements are:

a a clear chain of responsibility within a robust governance framework. There should be a single Senior Responsible Owner for each model ("Model SRO"), for each model through its lifecycle, and clarity from the outset about how QA is to be managed. The SRO should determine the optimal approach to QA with reference to value for money, including whether the best value for money will come from outside government;

b a senior QA champion with responsibility for ensuring there is an appropriate QA framework in place that is understood and used across the organisation; and

c a route for effective challenge where analysts have strong concerns. This could be through Heads of Profession, a QA champion or other senior staff member with clear responsibility for this role.

Recommendation 6: All departments and their Arm's Length Bodies should have in place, by the end of June 2013, a plan for how they will ensure they have effective processes – including guidance and model documentation – to underpin appropriate QA across their organisation. These plans will be expected to include consideration of the aspects identified in Box 4.B on page 38. To support this recommendation, succinct guidance setting out the key, generic issues that drive effective quality assurance will be added to "Managing Public Money", which offers guidance on how to handle public funds properly.

Recommendation 7: To support the implementation of these recommendations, the review recommends the establishment of an expert departmental working group to continue to share best practice experience and to help embed this across government.

Recommendation 8: Organisations' progress against these recommendations should be assessed 12 months after this review is published. HMT will organise the assessment, possibly with support from another department.

Links with civil service reform

4.17 The themes of professionalism, openness and accountability do not just apply to quality assurance. The review's recommendations link to several key themes of civil service reform, as outlined in the Government's Civil Service Reform Plan[1]:

- strengthening professions – central guidance and proposed network on QA;
- open policy making – publication of models/results, culture of raising concerns;
- of raising concerns;
- sharpening accountability – clear SROs, governance statement on QA; and
- policy skills and expertise – appropriate expertise for modelling QA.

[1] http://resources.civilservice.gov.uk/wp-content/uploads/2012/06/Civil-Service-Reform-Plan-acc-final.pdf

Box 4.B: Process

1 Each department and ALB should have clear **guidance** setting out their approach to QA. For relatively simple models with a low level of risk a comparatively light QA framework may be appropriate. Where there is a higher level of risk – for example for more complex models or those influencing particularly critical decisions, where there is a concern over possible "group-think", or where there have been recent changes in personnel, circumstances or model usage – a more extensive approach may be required. For the latter group a key judgement will be how to engage sufficiently expert review. In addition, consideration should be given to the degree of independence of the reviewers that is appropriate.

2 Each business critical model should have clear **documentation** that sets out the following. In line with good practice in managing other types of complex business critical systems, the above could take the form of a "checklist" to ensure all these points are covered and formally signed-off as the model is developed and used:

   a the model's scope and specification;

   b the purpose, limitations and risks;

   c the quality assurance undertaken;

   d the identity of an appropriately senior model SRO with overall responsibility to ensure the model is "fit-for-purpose", who will confirm the QA process is compliant and appropriate; that the model risks, limitations and major assumptions are understood by model users; and that the use for the model outputs are appropriate; and

   e that the model customer has understood the outputs and any major uncertainties, including the results of any sensitivity analysis.

3 There should be a clear process for **handover** of responsibility where the model SRO needs to change for any reason.

A Terms of reference

Background

A.1 In light of the experience on the InterCity West Coast franchise competition, and given the Civil Service's commitment to better policy making, the Cabinet Secretary and the Head of the Civil Service have commissioned a review of the quality assurance of analytical models that are used to inform policy decisions.

Purpose

A.2 Government departments are responsible for the analytical models they use to inform decision-making in the policy areas on which they lead. This review will consider the quality assurance mechanisms that central government departments have in place to scrutinise the robustness of analytical models and will make recommendations for improvement.

A.3 In doing so, the review will:

- ask departments to identify existing Government models that are business critical, as well as identifying and justifying the existing quality assurance systems, processes and methods in place that apply to those models;
- identify best practice on model development, operation and quality assurance both within Government analytical models and in non-Government analytical models; and
- make recommendations for improvements.

Governance

A.4 The review will be led by Sir Nick Macpherson, Permanent Secretary to the Treasury and chair of the Whitehall Heads of Analysis Group. The review will be supported by a multi-disciplinary team, including economists, actuaries, statisticians and scientists. The Whitehall Heads of Analysis Group, enlarged to provide for external challenge and support, will act as Steering Group for the review.

Evidence

A.5 The review will gather evidence from a wide range of stakeholders, including government departments, the private sector, public policy organisations in the UK and overseas; as well as the academic community.

Reporting

A.6 The review will provide an interim report by end November 2012 which will identify the business critical models identified across Government and map the quality assurance mechanisms that apply to those models, and a final report to the Cabinet Secretary and Head of the Civil Service by end January 2013 setting out lessons from best practice and recommendations for improvement.

Contact

A.7 For more information, please contact HM Treasury public enquiries at public.enquiries@hm-treasury.gov.uk.

Notes for editors

A.8 The Steering Group will be as follows:

- Sir Nick Macpherson (Chair);
- Tera Allas – Deputy Head of the Government Economic Service;
- Richard Bartholomew – Joint Head of Government Social Research;
- Sir John Beddington – Chief Scientific Adviser;
- Ian Davis – Non Executive Director, Cabinet Office;
- Jenny Dibden – Joint Head of Government Social Research;
- Richard Douglas – Head of the Government Finance Profession;
- Trevor Llanwarne – Government Actuary;
- Jil Matheson – National Statistician;
- Tony O'Connor – Head of Government Operational Research Service;
- Dave Ramsden – Chief Economic Adviser; and
- Chris Wormald – Head of Government's policy profession.

A.9 Richard Brown, former Chief Executive Officer of Eurostar International Ltd is also leading a related review into the InterCity West Coast franchise competition.

B Organisations participating

B.1 In addition to government departments and their ALBs who contributed returns, the review team would like to thank the following organisations who gave their time and expertise:

- Aetha Consulting;
- BAE systems;
- Centre for Science and Policy;
- Chartered Institute of Management Accountants (CIMA);
- Deloitte;
- Ernst & Young;
- Financial Reporting Council;
- Financial Services Authority;
- Institute for Fiscal Studies (IFS);
- International Monetary Fund;
- Institute & Faculty of Actuaries;
- KPMG;
- London School of Economics;
- Bank of America Merrill Lynch;
- Met Office;
- Milliman;
- National Audit Office;
- National Institute of Economic and Social Research;
- Organisation for Economic Co-Operation and Development (OECD);
- Office for National Statistics;
- Oxford-Man Institute;
- Prudential;
- PwC; and
- University College London.

C Model process – issues to consider

C.1 This annex sets out examples of issues and types of QA which should be considered at different stages of the model development process. It is intended to complement the process sections of the Chapter 2 of this report, on best practice in QA.

C.2 The examples given in the tables below are illustrative and in no way exhaustive. The QA needs of a specific model will depend on a range of factors such as its complexity, the associated risk, its range of application, potential users etc which are likely to be model dependent. The issues outlined below are therefore meant to be a guideline only, and departments will want to provide more specific information in their own guidance.

Table C.1: Table showing issues to consider at Scope stage of model development

When to consider	Description
Should always be considered	Business Criticality – an understanding of the level of business risk (this could be financial, reputational or business continuity) in the decision the modelling output will be used to support; Availability and Credibility of Input – a high level specification of the model inputs, whether they are available and the level of confidence in their accuracy; Complexity/Novelty – an understanding of the likely complexity and novelty involved in the modelling and an understanding of the associated risk; Level of Resources – an understanding of the amount of time and expertise required to develop the model and whether sufficient time and expertise is available; and Governance, ownership and QA – A suitable governance structure is identified (including model ownership) and an appropriate level of Quality Assurance needed to ensure the model is fit-for-purpose.
Should often be considered	Range of Use – Whether the model will be used to answer a single policy question or be used for a number of different questions, possibly across policy areas or over an extended period of time. End-User – definition of the envisaged user of the model, their expertise and the level of training that may be required.
Should be considered if appropriate	Management of changes – How any requests to change the requirements will be managed during development. Maintenance – If required for multiple or continued use, how the model and supporting data will be maintained to ensure it remains fit-for-purpose.

Table C.2: Table showing issues to consider at Specification stage of model development

When to consider	Description
Should always be considered	Model description – a clear, agreed definition of what the model will do and the main assumptions it will contain. Risks – a clear characterisation of the risks associated with the model, any mitigation strategies and any residual risk. Inputs/Outputs – a comprehensive list of the expected model output and the required inputs, including a list of likely sources for input data. Level of QA – the agreed amount and type of QA that is necessary to ensure the model meets the requirements and is fit-for-purpose Model use – a description of how the model will be used and by whom – e.g. will it be used only by an expert or non-expert user.
Should often be considered	Sign-off procedure – the process by which the model will be accepted as meeting the requirements and being fit-for-purpose
Should be considered if appropriate	Training – Any requirements for end-user training, including any necessary training material. Maintenance – Any requirements for model maintenance (updating model inputs, assumptions etc) including timescale and estimates of resource required.

Table C.3: Table showing examples of appropriate QA at the Model Build stage

When to consider	Description
Should always be considered	Version control – systems in place to manage the development of the model and ensure any changes are captured; Unit testing – individual testing of components of a model to ensure they are correctly coded and give the right result; Logic testing – the logic flow within the model follows that defined at the model design stage, (at the level of individual units, multiple units or the complete code); Internal code review – independent review of model coding may be worthwhile to ensure it meets the specification and is as free from errors as possible. This should be conducted by someone who is not part of the development team; and Internal test review – independent review of the verification testing results to ensure results are consistent with the model design specification. This should be conducted by someone who is not part of the development team.
Should be considered for more complex/ high-risk models	External code review –peer-review of model logic, assumptions and coding to ensure the model meets the specification and is as free from errors as possible. This will generally be conducted by someone external to the organisation; Test review – independent review of the verification testing results to ensure results are consistent with the model design specification. This will generally be conducted by someone external to the organisation; and Parallel builds – for complex, high-risk models there may be value in developing parallel builds to ensure cross-checking of results

Table C.4: Table showing examples of appropriate QA at the Model Test stage

When to consider	Description
Should always be considered	Checking against data – checking model outputs against available data, for example recreating historical datasets; Reviewing assumptions – checking that assumptions remain valid e.g. circumstances haven't changed since the assumptions were originally set; Limit testing – sample testing of the range of validity of all input variables – this may not be possible for complex models, but parameter ranges of key variables should be tested. Input values outside the accepted ranges should also be included to test any exception and error handling within the model; Cross checking – checking model output with similar independent models where available; Internal independent testing – independent testing of the full system may be advisable at this stage; Reviewing outputs – checking that outputs are sufficient for the purpose of the decisions being taken, including assessment of limitations, alternative scenarios, etc; and Transparency – publication of the model itself, or the test schedule and results, may to provide additional external review if appropriate.
Should be considered for more high-risk/ complex models	External independent testing – external peer-review of the full system; Internal audit – a formal audit conducted within the organisation. This would need to be supported by full model specification and test documentation; and External audit – a formal external audit. A comprehensive model-based audit would need to be supported by full model specification and test documentation, although a results-oriented audit might be a better alternative in a number of circumstances, particularly where there is regular updating and usage and "lower level" checks such as internal peer review are already in place.

Table C.5: Table showing ongoing QA considerations

When to consider	Description
Should always be considered	Periodic review – to ensure the model is fit for its current and upcoming uses.

D Departmental returns

D.1 This annex explains the data that was requested from departments and their Arm's Length Bodies (ALBs).

D.2 The review team wrote to all government departments asking for information on their business critical models. This data formed the basis of the descriptive analysis in Chapter 3 of this report.

D.3 Part of the remit of the review was to "identify the business critical models identified across government and map the quality assurance mechanisms that apply to those models" and this map is presented in the table published alongside this report. The table provides the name, description and QA summary of the business critical models for each department and their arm's length bodies. The review does not include information for organisations that sit independently of government including the Office for National Statistics and the economic regulators.

D.4 The review requested data from departments to build a picture of current business critical models and their QA. The team also met with individual departments to further understand the way quality assurance is conducted. The team then summarised this data to provide a snapshot of the different types of QA in use across government.

D.5 Two key caveats are important to bear in mind when considering the data:

- these statistics represent **a snapshot** of business critical models and QA status. They capture a point in time, late 2012, not including models in development and models that have been used in the past and that are not currently expected to be used again; and
- this analysis is necessarily descriptive, and **should not be used to form judgements**. As discussed already, the review would expect there to be a wide range in the approach to QA across different models. To be effective, and represent value-for-money, QA needs to be proportionate to the significance of the decision, the complexity of the model (including key inputs and assumptions) and the degree of risk and uncertainty.

D.6 The table of published returns is published alongside this report on the Treasury website.

HM Treasury contacts

This document can be found in full on our website: http://www.hm-treasury.gov.uk

If you require this information in another language, format or have general enquiries about HM Treasury and its work, contact:

Correspondence Team
HM Treasury
1 Horse Guards Road
London
SW1A 2HQ

Tel: 020 7270 5000

E-mail: public.enquiries@hm-treasury.gov.uk

ISBN 978-1-909096-53-0