From:	[redacted]
To:	A303Stonehenge@pins.gsi.gov.uk
Subject:	Clarification of Question Raised at Preliminary Hearing - A303 Scheme
Date:	03 April 2019 08:40:47
Attachments:	aqua_book_final_web.pdf review_of_qa_of_govt_analytical_models_final_report_040313.pdf R0003199E.pdf

Dear Sir or Madam,

At the Preliminary Hearing on 2nd April 2019, I raised a question along the the lines of "I understand the Office of National Statistics (ONS) recommends all models (eg computer and spreadsheet models) used by and on behalf of government to be appropriately verified and validated before use. Would it not be a good idea to have all such models (for noise, light pollution, hydrology, ecology") used in developing and underpinning the scheme to be confirmed as having appropriate verification and validation."

It seems my assertion that the recommendations came from ONS was incorrect and the correct source was, in fact, the Treasury. Two specific documents:

Review of quality assurance of Government analytical models

The Aqua Book:guidance on producing quality analysis for government

-provide the relevant information and these are attached below.

A third publication, which specifically deals with verification and validation:

Verification and Validation for the AQuA Book

- is also attached. Highways England, as a government owned company, and the Planning Inspectorate, as an executive agency, are both covered by the scope of these documents.

Please could you bring this correction to the attention of the Inspectors and pass the documents to Highways England who were tasked with investigating whether the models used were compliant with government policy.

Best Regards

Dr Andrew D Shuttleworth

[redacted]

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

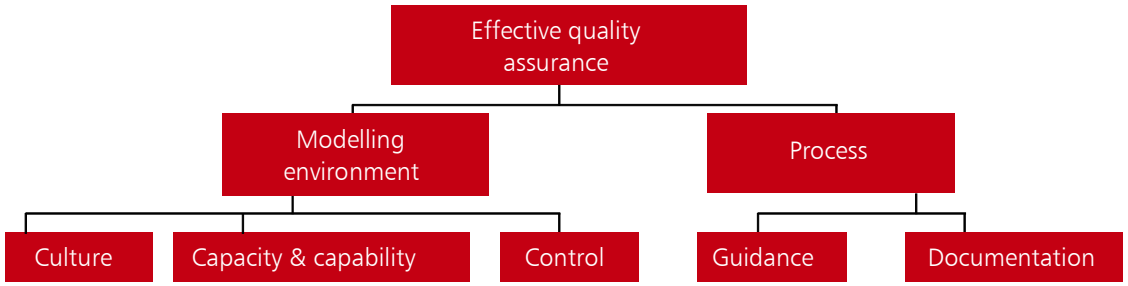
2.1 Quality assurance (QA) provides decision makers with key information about how a model works, and its risks and limitations. This is essential if a model's outputs are to be used with genuine understanding and confidence. As such, QA is a key aspect of the effective risk management of business critical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

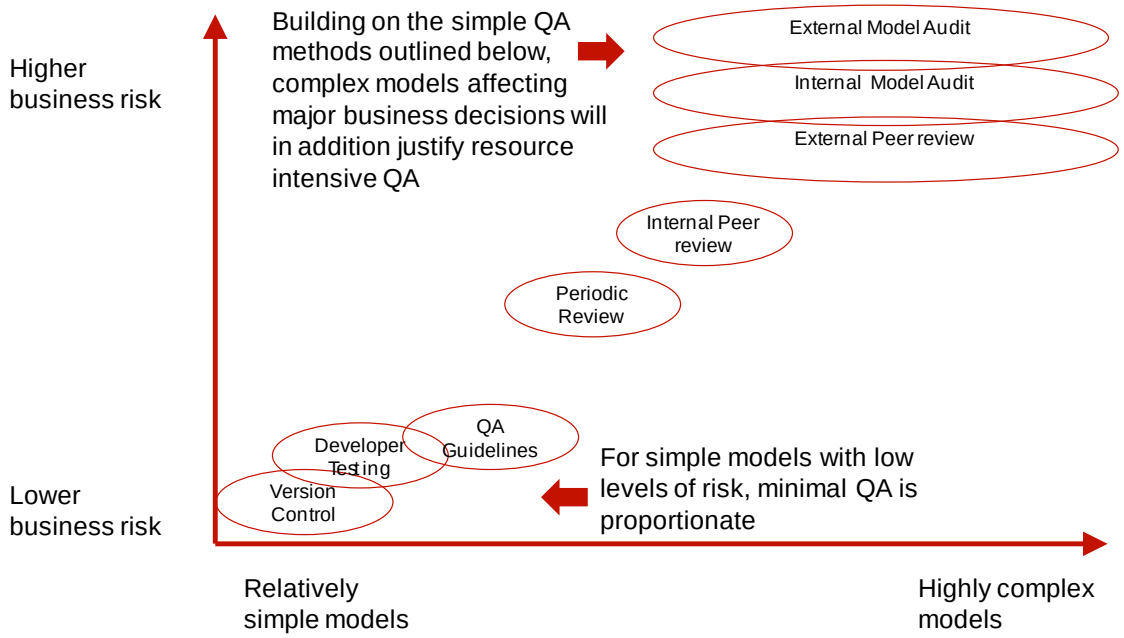
2.61 This illustrates the importance, at all stages of model development, that analysts and their customers take a conscious decision on the amount and type of QA that is appropriate. One way to achieve this is through a 'checklist' which some departments (HMRC for example) use to aid the QA process,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

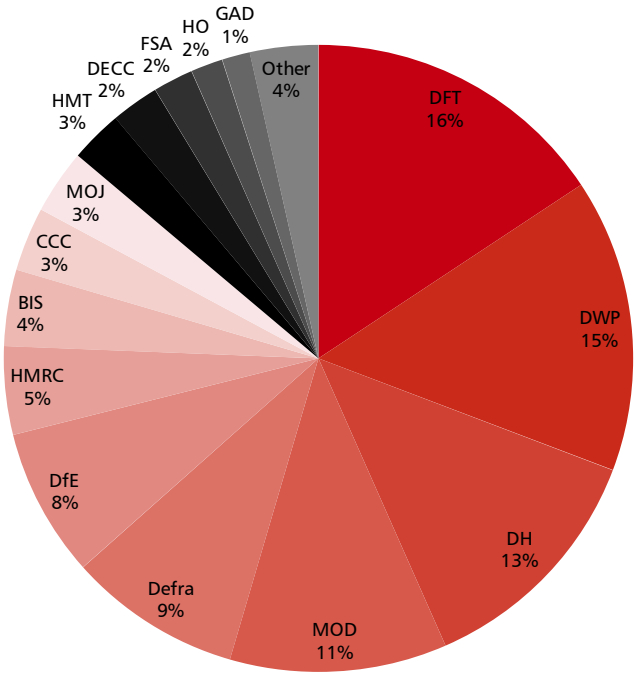
The extent and nature of government modelling

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.

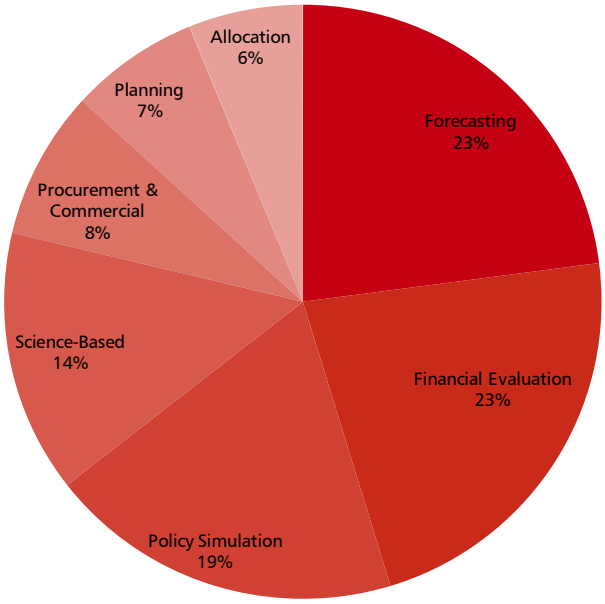
Types of business critical model

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.

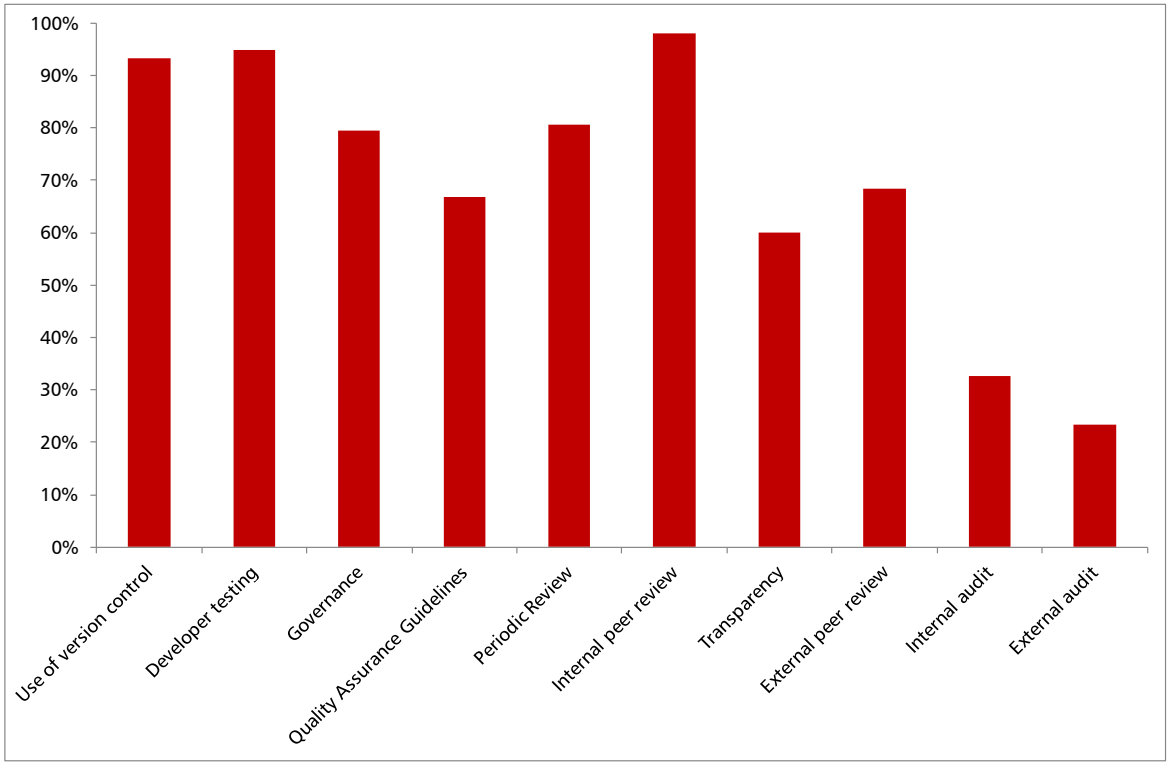
3.16 Around one third of models had some model audit – either internal or external. This is an often time-consuming and resource-intensive QA method, and it is to be expected that it would not be used across all models.

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.

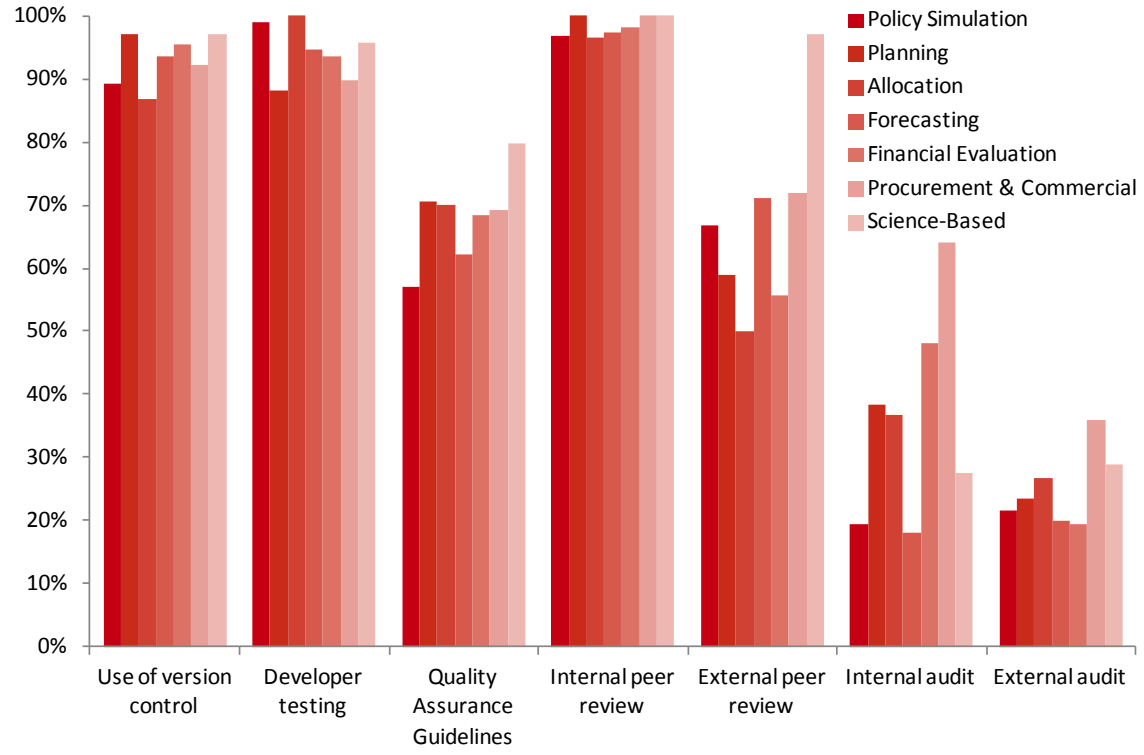
3.21 When considering the types of QA and the types of models Charts 3.D and 3.E below show that science-based models tend to have the most extensive types of quality assurance. This is especially the case with external peer review and transparency, reflecting a strong culture in the scientific community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

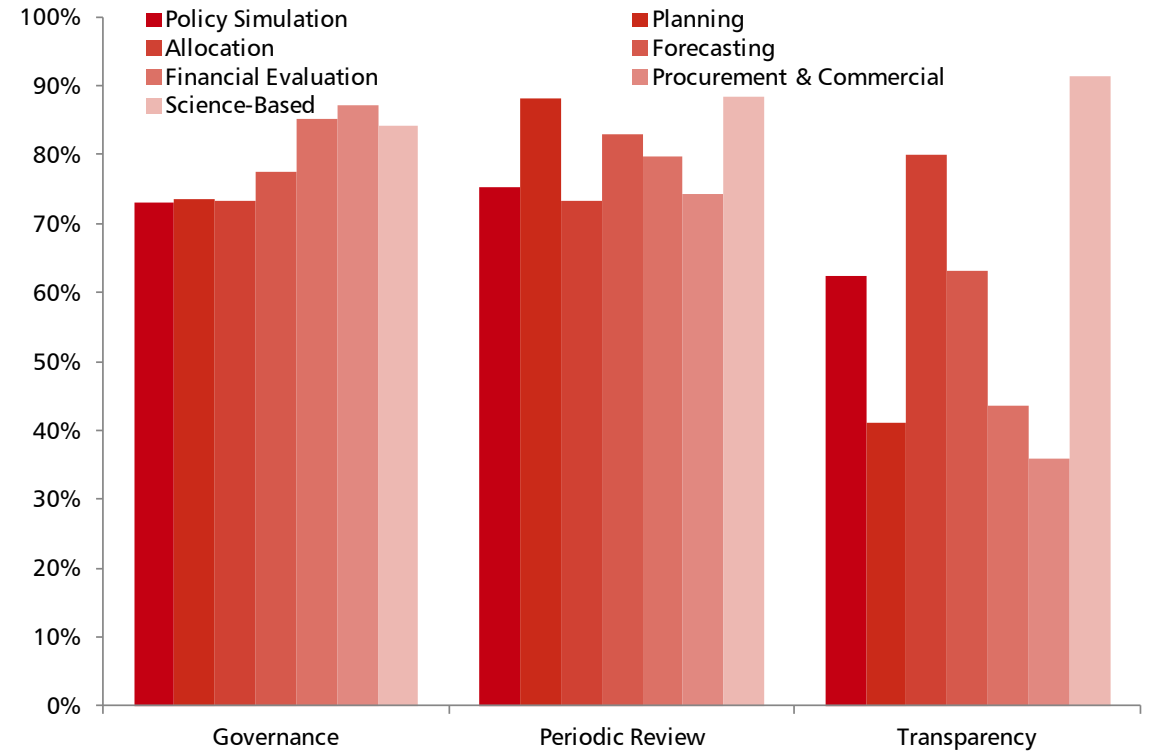
3.24 The degree of transparency tends to vary, with planning models understandably having low transparency as they are often modelling key government business. Procurement and commercial models might be expected to have a greater degree of transparency, owing to the open competition process and disclosure 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

HM Treasury

The Aqua Book:

guidance on producing quality analysis for government

March 2015

HM Treasury

The Aqua Book:

guidance on producing quality analysis for government

March 2015

This publication is available at www.gov.uk/government/publications

Any enquiries regarding this publication should be sent to us at public.enquiries@hmtreasury.gsi.gov.uk

ISBN 978-1-910337-67-7
PU1740

Contents

Foreword

Analysis and the use of evidence informs every decision we make as civil servants. It also underpins the civil service values of integrity, honesty, objectivity and impartiality by which we are held to account. The breadth of topics on which the civil service provides advice is extraordinary – from the health of the public finances to predicting the spread of pandemic flu, from forecasting our future energy needs to the complex commercial decisions that underpin our infrastructure ambitions. High quality analysis is therefore fundamental to the civil service's effectiveness.

I see high quality analysis all the time as I look at the complex and varied issues with which civil servants grapple. But since the difficulties with the Intercity West Coast franchise competition, the quality of our analysis has, rightly, been scrutinised.

In October 2012, I was commissioned to review the quality assurance of analytical models across government. The recommendations of that review were written with the aim to extend best practice across the whole of government. They focus on quality assurance, governance and accountability, culture, capacity, capability and control.

Continuous challenge and improvement is essential to ensure that the people we serve – ministers and, of course, the public – have trust in our analysis. Following the *Review of quality assurance of government analytical models*, a cross-departmental working group on analytical quality assurance was established. The Aqua Book is one of the products this group has developed. It outlines a sensible, achievable set of principles. These principles will help ensure that our work can be trusted to inform good decision making. I'm grateful to all those who contributed to it.

As the Aqua Book points out, we need to create an environment where the skills and time to deliver analysis is respected, and a culture that values it is encouraged. I commend it to you.

Nick Macpherson

Permanent Secretary to the Treasury

1 Introduction

Fit-for-purpose analysis in government

1.1 Analysis is vital to the success of policy development and the delivery of programmes, projects and operational services. Analysis helps to shape and appraise options, provides insight into how complex systems work and behave, measures system performance and improves efficiency.

1.2 However, if analysis and any supporting models, data and assumptions are not fit-for-purpose then the consequences can be severe ranging from financial loss through to reputational damage and legal challenge. In the most severe of consequences, lives and livelihoods can be affected.

1.3 The InterCity West Coast franchise competition of 2012 illustrated both the importance of analysis and modelling in delivering a major government project and the consequences when things go wrong. The subsequent *Review of quality assurance of government analytical models*[1] found significant variation in the type and nature of quality assurance used within, and between, departments. Much of this was to be expected given the differences in organisations' remits, and the levels of risk in question.

1.4 The review's work highlighted the benefits of creating a work environment that expects thorough quality assurance – including allocating clear responsibility for key models and how they are used, and giving specialist staff adequate time to manage quality assurance effectively. The review provided headline recommendations for departments and their arm's length bodies, including:

- 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;
- 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;
- There should be a single Senior Responsible Owner for each model ("a Model SRO") through its lifecycle, and clarification from the outset on how quality assurance is to be managed. Key submissions using results from the model should summarise the quality assurance that has been undertaken, including the extent of expert scrutiny and challenge. They should also confirm that the Model SRO is content that the quality assurance 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 output is appropriate.

1.5 The review found that the many components of best practice in quality assurance fall under 2 headings: the modelling environment, which can be generalised to cover the wider environment in which analysis takes place, and process.

1.6 The right modelling environment involves a culture where leaders value and recognise good quality assurance. It requires adequate capacity, including specialist skills and sufficient time to

[1] Review of quality assurance of government models, https://www.gov.uk/government/publications/review-of-quality-assurance-of-government-models, Accessed February 2015.

conduct quality assurance effectively. It also needs a set of controls, including a clear internal chain of responsibility and a route for challenge where analysts have concerns

1.7 The process side, on the other hand, is about a systematic approach to make quality assurance accessible, easy and comprehensive. It requires clear guidance on quality assurance and clear documentation for every model.

Quality principles

1.8 Following the review, the cross-government working group on analytical quality assurance was established to identify and share best practice across government. The Aqua Book is one of the products this group has developed. It draws together existing practice from departments and best practice from analysts across a variety of analytical professions within government. The Aqua Book builds upon the *Review of quality assurance of government analytical models* and expands the principles to cover all types of analysis as there is much common ground.

1.9 The Aqua Book sits underneath the *Review of quality assurance of government analytical models* as supporting material to help those implementing the review's recommendations and more generally to promote analytical quality. It should be read in this context: as providing further advice rather than making specific or binding recommendations.

1.10 No single piece of guidance can provide a route to a definitive assessment of whether a piece of analysis is of sufficient quality for an intended purpose. However, the Aqua Book sets out the following principles of analytical quality assurance that will help to support commissioning and delivery of fit-for-purpose analysis:

- Proportionality of response: The extent of the analytical quality assurance effort should be proportionate in response to the risks associated with the intended use of the analysis. These risks include financial, legal, operational and reputational impacts. In addition, analysis that is frequently used to support a decision-making process may require a more comprehensive analytical quality assurance response.
- Assurance throughout development: Quality assurance considerations should be taken into account throughout the life cycle of the analysis and not just at the end. Effective communication is crucial when understanding the problem, designing the analytical approach, conducting the analysis and relaying the outputs.
- Verification and validation: Analytical quality assurance is more than checking that the analysis is error-free and satisfies its specification (verification). It must also include checks that the analysis is appropriate, i.e. fit for the purpose for which it is being used (validation).
- Analysis with RIGOUR: Quality analysis needs to be **r**epeatable, **i**ndependent, **g**rounded in reality, **o**bjective, have understood and managed **u**ncertainty, and the results should address the initial question **r**obustly. In particular, it is important to accept that uncertainty is inherent within the inputs and outputs of any piece of analysis. It is important to establish how much we can rely upon the analysis for a given problem.

Accountability

1.11 Following the *Review of quality assurance of government analytical models,* all business critical government models should have a single *model* Senior Responsible Officer ("a Model SRO") through their lifecycle. 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.

1.12 This principle can be generalised to cover any business critical *analysis* and there may be a need for interaction between a model SRO and those responsible for a wider analytical project to ensure that quality assurance considerations are fit-for-purpose and are appropriately communicated.

1.13 Not all analysis will be business-critical or use business-critical models. However, it remains good practice to ensure that there is a single accountable individual with ultimate responsibility for the overall quality of the analysis, at all stages of the analytical cycle. For business-critical analysis, the single accountable individual should be of sufficient seniority for the associated risks.

Box 1.A: Accountability, responsibility and authority

It is important that departments and agencies have a cascade of accountability and responsibility from their senior management teams down throughout their organisation.

At the senior management level (e.g. the senior civil service), it is important that the accountability and responsibility for establishing the analytical quality assurance environment and processes are clearly defined. In addition, a mechanism should be established that determines the senior manager accountability for the analytical quality assurance activities supporting business critical analysis.

The senior accountable person for analytical quality assurance must ask the right questions and satisfy themselves that appropriate analytical quality assurance is being provided – but they do not need to be a specialist (or have an analytical background) 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 those providing analytical quality assurance are sufficiently senior and sufficiently experienced to take responsibility for the analysis in question.

Roles and responsibilities

1.14 To support those responsible for providing analytical quality assurance, and to deliver the quality principles set out above, it is useful to consider the wider activities that are involved:

- commissioning analysis
- providing analytical assurance
- delivering the analysis itself

1.15 Those accountable for the analysis may, or may not, be directly involved in the above activities, but they need to ensure that all 3 are being carried out to a suitable standard. In either case, governance arrangements for the wider programme requiring the analysis should consider analytical quality assurance needs.

Commissioning analysis

1.16 The person **commissioning analysis** must ensure that those doing the analysis understand the context of the question being asked so that they understand the likely risks and can determine what the appropriate analytical and quality assurance response should be. The commissioner has a role to ensure that there is sufficient time and resource for the required level of assurance to be delivered and that they understand the associated risks when time and resource pressures are unavoidable. When using the analysis, the commissioner must

understand the strengths, limitations, inherent uncertainty and the context of the analysis so that the results are interpreted correctly.

Analytical assurance

1.17 The person responsible for **providing analytical assurance** must ensure they receive evidence that appropriate analytical quality assurance activities have been conducted and that residual uncertainties and risks are understood and are communicated. Typically this would be done by a senior analyst or analytical project manager who is not one of the analysts delivering the analysis. This activity takes place throughout the life cycle of the analysis from understanding the problem, through designing the analytical approach, conducting the analysis and relaying the outputs. The analytical assurer is responsible for advising the commissioner on whether appropriate analytical quality assurance has taken place and advising them of any outstanding risks. The Aqua Book presents examples of the different type of assurance activity that can be undertaken including analyst-led testing, peer review and audits. For business critical analysis, more thorough assurance activities may be required.

Delivering analysis

1.18 The people responsible for **delivering the analysis** frequently assist the commissioner in structuring the question in order to ensure the appropriate analysis is performed. Some analysis may require external specialists and therefore analysts may also have responsibilities as part of the procurement process. Analysts, including those 3rd parties providing analysis, should also provide proportionate documentation that outlines the verification and validation activities undertaken and the associated conclusion. In addition, analysts should determine and communicate the uncertainty associated with the outputs of their analysis so that commissioners and users of analysis can make informed decisions.

Box 1.B: Local business practices and nomenclature

Each department and agency will require its own business processes and nomenclature to reflect their organisation's needs. Whilst the Aqua Book refers to commissioners, analysts and analytical assurers, it is the responsibilities identified that are important, not the name of the role. In addition, the Aqua Book makes no statement of the particular level of seniority or grade of each of the occupiers of the roles: this will vary from project to project and between departments and agencies.

At any time, analysts should refer to and operate in adherence to their local business processes and existing guidance on analytical quality assurance.

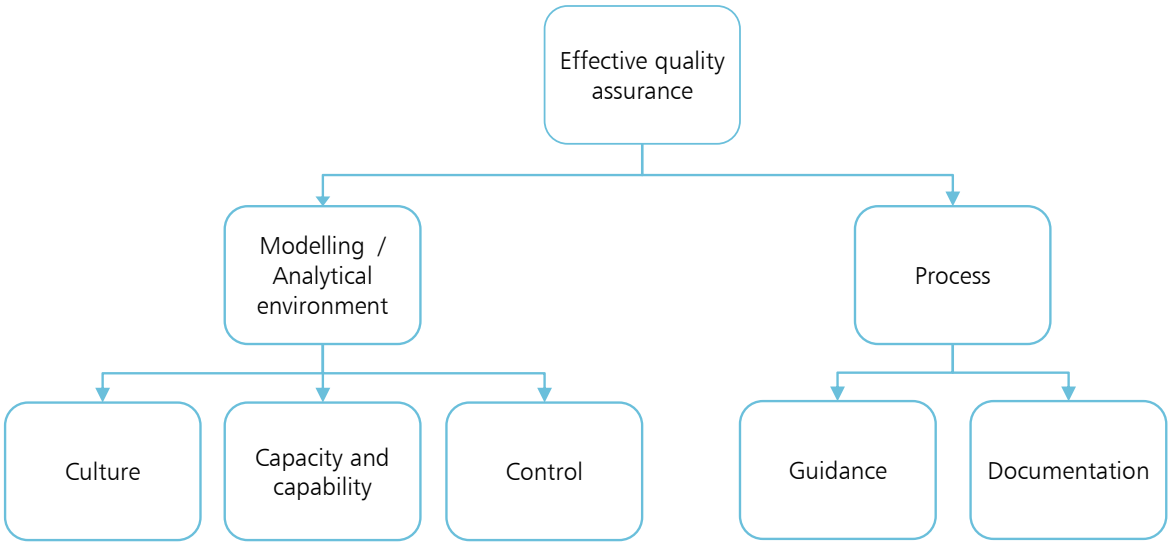
Quality assurance

1.19 Quality assurance is a key aspect of the effective risk management of analysis and the decisions it helps inform. Sitting above the many principles and techniques which ensure good quality assurance, there are 2 main requirements:

- **environment**: creating the conditions in which quality assurance processes can operate effectively, facilitated by a culture that values quality assurance and welcomes effective challenge, a well understood chain of responsibility and sufficient time for quality assurance; and

- **process**: establishing a clear process for every stage of the analytical life-cycle. This includes working alongside the commissioner and any other users to ensure there is a shared understanding about the purpose and any limitations of the analysis.

Figure 1.A: Effective quality assurance



Effective quality assurance is achieved by creating an environment that is conducive to quality assurance and embedding appropriate processes.

Source: Adapted from the Review of quality assurance of government analytical models

1.20 **Environmental** considerations includes creating:

- a culture where quality assurance is highly valued, and there are strong incentives to deliver appropriate quality assurance, backed by effective scrutiny of analysis, supporting models, assumptions and data
- capacity and capability where specialist staff have sufficient time built-in for quality assurance, and are able to draw on expertise and experience across government and beyond
- adequate controls, including a clear governance framework

1.21 An effective **process** involves on-going engagement between specialists and customers to ensure there is a shared understanding about the purpose and any limitations of the analysis. This includes understanding the consequences of sensitivity analysis, and the impact of the uncertainty associated with inputs, assumptions and outputs. This needs to be backed by:

- clear guidance that sets out the key considerations driving the approach to quality assurance
- clear documentation about the analysis and the quality assurance process

1.22 Further details can be found in the *Review of quality assurance of government analytical models*.

The Aqua Book and other resources

1.23 Combining the high-level principles of analytical quality assurance, together with clarified roles and responsibilities, the Aqua Book will help ensure that analysis is fit-for-purpose. For

those who commission or use analysis, a short overview provides sufficient depth to be able to ask the right questions about quality. Specific information targeted at the analyst and the analytical assurer covers more technical considerations.

1.24 The Aqua Book is a cornerstone of a suite of resources aimed at improving analytical quality assurance. A range of templates and guidance on specific analysis topics and analysis techniques that are frequently encountered within government analysis will supplement the resource. The aim of these resources is to:

- help departments and agencies embed an analytical environment that will assist the delivery of quality analysis by strengthening existing processes
- deliver greater consistency in the approach to analytical quality assurance processes across government
- ensure commissioners of analysis have greater confidence in analysis

1.25 The Aqua Book is divided into 3 parts:

- Part A is designed for the commissioners of analysis and those who have accountability for a programme in which analysis is used. It provides an overview of the each of the topics that are important for ensuring that analysis is fit-for-purpose.
- Part B is designed for the analytical assurer and the analyst and expands that material presented in Part A. It provides additional guidance on verification, validation and uncertainty.
- The Aqua Book draws together information that will be of benefit to all departments, agencies and analysts. However specific guidance covering verification and validation of particular types of analysis, as well as example templates and documentation, are also beneficial. Part C introduces the additional supporting resources that accompany the Aqua Book.

Part A

This part of the Aqua Book is written for commissioners of analysis and those that have accountability for programmes where analysis is important to successful delivery.

Chapter 2 explains how commissioning analysis relates to the problem under investigation and the decision the analysis is helping to inform. This chapter provides an overview of the types of analysis that may be commissioned in order to support the decision-maker. Analytical projects follow a similar cycle to other projects encountered in government, and a simplified analytical cycle is presented before 2 key themes are introduced: delivering quality analysis and working with uncertainty.

Chapter 3 provides an overview of how to commission analysis that will be fit-for-purpose. The commissioner of analysis plays an important role in the quality assurance of the analysis and this role throughout the analytical cycle is explained. To support the commissioner, the analytical assurer is there to provide the assurance that the analysis is fit-for-purpose.

Chapter 4 introduces the two main questions to be asked of any piece of analysis: is it the right analysis and has it been conducted correctly. This chapter provides an overview of the types of activity that can be undertaken to give the assurance required.

Chapter 5 explains that uncertainty is inherent in all analysis and needs to be acknowledged, understood, managed and clearly communicated. This chapter outlines why the commissioner of analysis should request an estimate of the uncertainty inherent within the analysis in order to make better informed decisions.

2 Decision-making and analysis

Box 2.A: Decision-making and analysis – key points

Decision-makers rely on various forms of analysis to inform their decisions.

Understanding the rationale for a decision enables the identification and appraisal of options. We can determine whether our desired outcome has been achieved by monitoring the impact of a decision and, if necessary, take further action as appropriate.

A variety of analytical approaches are used to build the evidence base that supports the decision-maker throughout the process. Different analysis may be used at different stages.

In order to ensure that the responsibilities required to deliver analysis that is fit-for-purpose, we identify a framework with 3 important roles: the commissioner of analysis, the analyst and the analytical assurer. Departments and agencies will wish to tailor these roles to meet their local business needs. Each has a role in ensuring the analysis is fit-for-purpose.

Some decisions are more important than others and the underlying analysis will require greater scrutiny. Departments and agencies may determine that a piece of analysis is business critical; in this situation, the governance arrangements and the seniority of the 3 responsibilities outlined in this framework – particular that of the analytical assurer – should be appropriate for the level of risk.

Analysis projects follow a similar life-cycle to that of other projects. Understanding the requirements enables planning and execution of the analysis. The analysis provides results that can be delivered and interpreted against the original requirements. In doing so, the requirements may evolve and further iterations of analysis may be commissioned.

Analysis should be fit-for-purpose. This is to say that there is sufficient confidence that the right analysis has been performed and appropriate analytical quality assurance activities have been conducted.

Even with the knowledge that the right analysis has been performed and appropriate analytical quality assurance activities have been completed, one artefact of analysis remains: the inherent uncertainty of the outcome of the decision. Uncertainty analysis helps to quantify and communicate the results of the analysis.

The analysis should be communicated to the commissioner with appropriate reference to the analytical quality assurance and statements of the residual uncertainty.

Introduction

2.1 This chapter provides an overview of the decision-making process and how it relates to analytical projects. It also includes an overview of the analytical cycle alongside an introduction to the topics of quality analysis and uncertainty.

Decision-making

2.2 Policy development and the delivery of programmes, projects and operational services frequently require identified issues to be overcome. The Green Book outlines the policy development cycle[1] – ROAMEF – which can be generalised to support any kind of decision-making and not just that associated with policy development. For any given decision, the **r**ationale and **o**bjectives must be understood. Following **a**ppraisal of the options and implementation of the decision the outcome should be **m**onitored, **e**valuated and the original rationale reconsidered with the completion of the **f**eedback.

2.3 Each part of the ROAMEF cycle involves gathering and appraising evidence. Different analytical approaches will supplement the evidence base at each stage of the cycle. Examples of different analysis are provided in section 2.4.

Types of analysis

2.4 Evidence bases contain a variety of facts, figures and analysis all of which must be correctly sourced, appraised and referenced. The analysis can vary widely between situations and many different types of analysis may be used to form the evidence base that supports the decision-making process. Examples of types of analysis that are frequently encountered in government are:

- actuarial analysis
- economic analysis
- financial analysis
- operational research / analysis
- scientific, technical and engineering research
- social research
- statistical analysis

2.5 Many branches of analysis make use of analytical models. Some analytical models can be used to inform several pieces of analysis and this flexibility and utility introduces the need for further quality considerations. Models are used for a variety of purposes including to:

- test systems under a variety of scenarios
- carry out investigations to understand a problem in more detail
- enable the monitoring of processes to facilitate risk management
- compare and appraise options
- understand past behaviour to better prepare for the future

2.6 Analytical tools – related to models – are frequently developed that allow the repetition of calculations to be performed in a more resource efficient manner.

[1] The Green Book: appraisal and evaluation in central government, https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent, Accessed February 2015.

Box 2.B: Types of model

Policy simulation: to better understand policy options that drive government decisions. Ministers make policy decisions based on assessments of the likely cost and impact of policy choices. For example, the Intra Government Tax Benefit Model is used to analyse the distributional impact of tax and benefit changes.

Forecasting: to predict the future and inform today's policy choices. For example, demographic projections are essential to understand future cost pressures for education and healthcare. Equally, DECC use the updated Energy and Emissions Model to forecast the energy demand and emissions by fuel type and business sector under existing firm and funded policies.

Financial evaluation: to better understand future liabilities or costs. For example, modelling to understand the future cost implications of current pension commitments or the future cost of decommissioning existing nuclear energy plants.

Procurement and commercial evaluation: for the letting and management of large contracts, and to ensure value for money – for example, where a key service is to be contracted out as in the case of railway franchises or where a major IT upgrades/ new system is being introduced.

Planning: to make workforce decisions which affect the delivery of future services. For example, these models may be used to assess the number of trainee teachers, doctors and nurses required to deliver education and healthcare into the future.

Science based: to better understand and simulate the physical environment, in order to evaluate possible solutions or to mitigate potentially devastating impacts – for example, climate change and flood risk.

Allocation of funds: to determine how funds allocated to departments are then distributed to, for example: local authorities, schools or across the health service. These models are essential to ensure funds are allocated properly across the country to underpin local service delivery.

Conceptual: to help understand the key influences that are important to a system being modelled. A variety of problem-structuring techniques are used to develop conceptual models.

Roles and responsibilities in analytical modelling projects

2.7 The *Review of quality assurance of government analytical models* introduced an important quality assurance role: all business critical models should have a single Senior Responsible Officer (a "Model SRO") throughout their development and application. The key prerequisites are that the Model SRO should be a named individual with sufficient seniority to take responsibility for the model throughout its life cycle and to provide sign-off that it is fit-for-purpose, prior to its use i.e. the Model SRO provides analytical assurance for the model and its outputs in each piece of analysis where that model is used.

2.8 Not all analysis will involve business-critical models so there may not be a formal requirement for a "Model SRO". However, it remains good practice to ensure that there is a single accountable individual with ultimate responsibility for the overall quality of the model, at all stages of the modelling cycle.

2.9 The concept of the Model SRO can be widened for any analytical project, whether it utilises modelling or other forms of analysis, that is used to draw conclusions that inform the decision making process. Good practice is to identify a single accountable individual with ultimate

responsibility for the overall quality of a piece of *analysis* that supports a specific decision making process (e.g. an *Analysis* SRO). For business critical analysis, the seniority of the person accountable for the analytical quality assurance should be determined based upon the risks associated with the analysis.

2.10 It is important that there is a clear cascade of accountability and responsibility from senior management teams down throughout the organisation so that the roles of a Model or Analysis SRO are understood within local governance arrangements.

Box 2.C: Accountability, responsibility and authority

Each department and agency will wish to review their local practices and guidance to ensure that there is an appropriate mechanism in place for determining the necessary accountabilities, responsibilities and authorities for the provision of analytical quality assurance. It is important that departments and agencies consider the complete cascade of accountability and responsibility from their senior management teams down throughout their organisation.

At the senior management level (e.g. the senior civil service), it is important that the accountability and responsibility for establishing the analytical quality assurance environment and processes are clearly defined. In addition, a mechanism should be established that determines the senior manager accountability for the analytical quality assurance activities supporting business critical analysis.

Departments and agencies will wish to tailor their accountability mechanism to accommodate local business practices and requirements, such as project risks, the need to work across organisational boundaries, and the interaction with other functions such as a *programme* Senior Responsible Owner and any existing analytical leadership structure.

The senior accountable person for analytical quality assurance must ask the right questions and satisfy themselves that appropriate analytical quality assurance is being provided – but they do not need to be a specialist (or have an analytical background) 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 those providing analytical quality assurance are sufficiently senior and sufficiently experienced to take responsibility for the analysis in question.

To support the above, departments and agencies will wish to consider how those who manage, conduct and review business critical analysis should interact with the senior accountable person.

Roles and responsibilities in analytical projects

2.11 The Aqua Book presents a framework for addressing the responsibilities required to ensure analysis is fit-for-purpose and builds upon the role of the Model SRO that was outlined in the *Review of quality assurance of government analytical models*.

2.12 While analysts play a large role in the successful delivery of analytical projects, there are 2 other important responsibilities: commissioning analysis, noting that this responsibility may be on the behalf of the ultimate customer or decision-maker, and assuring analysis.

2.13 The commissioner must ensure that the analyst understands the context of the question being asked. This is so that the latter understands and can assess the likely risks and determine what the appropriate analytical and quality assurance response should be. The commissioner has a role to ensure that there is sufficient time and resource for the required level of assurance to

be delivered and that they understand the associated risks when time and resource pressures are unavoidable. When using the analysis, the commissioner must be confident in the quality of the outputs and understand the strengths, limitations and contexts of the analysis so that the results are correctly interpreted.

2.14 The **analytical assurer** – typically a senior analyst or analytical project manager, who is not one of the analysts delivering the analysis – must ensure they receive evidence of appropriate analytical quality assurance activity. These activities must take place throughout the life cycle of the analysis, from understanding the problem, through designing the analytical approach, conducting the analysis and relaying the outputs. They must ensure that the governance arrangements for the wider programme requiring the analysis consider analytical quality assurance needs. The analytical assurer is responsible for advising the commissioner on whether appropriate analytical quality assurance has taken place.

2.15 The analytical assurer may or may not be the same person as the Analysis or Model SRO if the analysis uses a business critical model.

2.16 The analytical assurer need not be an analyst. Projects that depend on highly complex and sophisticated analysis or modelling techniques may choose an analytical assurer with the ability to understand the technical or analytical considerations in order to "sense check" the outputs. Similarly, projects dependent on complex project or programme knowledge or theories may need an analytical assurer who can understand the sensitivities and uncertainties with the subject matter of the problem being addressed. The key requirement is that commissioners of analysis and analysts work closely together to ensure the analytical assurer is able to ask the right questions, fully understands the uses and limitations of the analysis and supporting model and is therefore able to sign-off to confirm it is fit-for-purpose.

2.17 Analysis is frequently used to inform business critical decisions. In such situations, it is important that the decision-maker has sufficient assurance from a senior level within the organisation before taking the decision. For business critical analysis and modelling, the commissioner should be satisfied with the seniority of the analytical assurer.

2.18 The analytical assurer sign-off provides confidence that:

- the quality assurance process used is compliant and appropriate
- analysis risks, limitations and major assumptions are understood by the users of the analysis
- the use of the analysis is appropriate

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

2.20 Reconfirmation of the sign-off would be required if the analysis or supporting 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 supporting model is being used for a new purpose or project, the analytical assurer should obtain confirmation from the Model SRO that the model is suitable for the proposed use.

2.21 If the analytical assurer cannot give their sign-off, this signals the analysis is not fit-for-purpose. In this case, the analysis should not be used until any specific issues are rectified. This may entail additional and / or alternative analysis or amending / re-developing any supporting model. In addition, further verification and validation checks may be required.

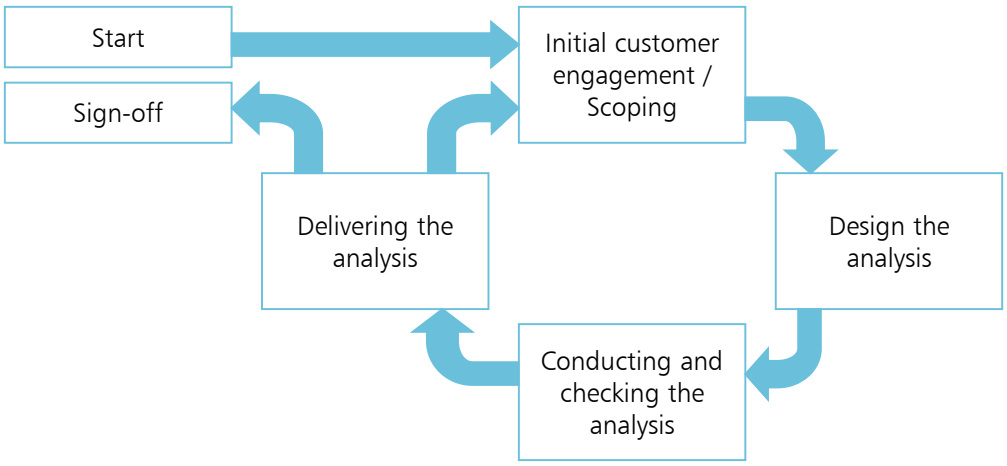
2.22 In addition to conducting the analysis, the **analyst** should also provide proportionate documentation that outlines the verification and validation activities that their analysis has been subjected to, in most cases by an analyst independent of the original analysis. In addition, the analyst should determine and communicate the uncertainty associated with the outputs of their analysis so that the commissioner can make informed decisions.

2.23 Departments and agencies will wish to tailor the presented framework to meet local business needs and the responsibilities may be distributed differently between organisations. It is the responsibilities that are important.

The analytical cycle

2.24 For analysis to be used to effectively inform the decision-making process, it should be fit-for-purpose. When commissioning a piece of analysis, a project is established and the typical project trade-offs between time, resources and quality must be made, recorded and communicated to the analyst. The analyst may also have a role in working with the commissioner to establish the optimal balance of these constraints. The analytical process, outlined in Figure 2.A, provides a simplified outline of the main steps in the delivery of an analytical project. It is worth noting that local business practices may require additional stages to be included, such as formal approval stages, or use different terminology. However the core stages below can be found in most versions of the analytical cycle.

Figure 2.A: The analytical cycle



The analytical cycle is often iterative as insight is gained and the original question refined. At each part of the cycle, analytical quality assurance activities take place to ensure the analysis is fit-for-purpose. While many checks take place at the point the analysis is conducted, it is not the only place where analytical quality considerations are made e.g. the customer's insight when inspecting the delivered analysis is an important part of the process.

2.25 The first stage of the analytical cycle presented above is **initial customer engagement and scoping**. The commissioner plays an important role in communicating the commission and working with the analyst to ensure that an appropriate understanding of the problem is captured. There should be a clear understanding of the requirements and scope between the commissioner and the analyst at the commissioning stage.

2.26 During the **design** phase, the analyst will convert the commission into an analytical plan and will consider the inputs, possible analytical methods, and the expected outputs that will be produced. A degree of iteration between commissioner and analyst is to be expected as the analytical solution develops over time. The analytical assurer should at this stage check that the proposed design meets the commissioner's requirements.

2.27 When the analyst is conducting their analysis, they will work through their analytical plan and will maintain a record of their analysis noting any deviations. In addition, they will be performing their own tests to check their analysis and they will commission other verification and validation activities as required.

2.28 During the delivery phase, the commissioner has additional important roles, both in providing feedback to assist in the correct interpretation of the results and to determine whether the analysis has addressed their commission. Analytical projects frequently require further iteration, as the original question is often refined in the light of initial results, and so the risk of further extensions to the analysis should be taken into account and managed constructively.

Delivering quality analysis

2.29 Before analytical output can be used to inform a decision, an appreciation of its fitness-for-purpose must be gained. This requires assuring that:

- the analysis undertaken aligns with its intended purpose and is relevant to the original problem
- the correct analysis has been performed
- the analysis has been conducted correctly and it is accurate
- the analytical output was provided in time to be useful and was presented in an accessible and clear manner
- the analysis is comparable and repeatable

2.30 By considering quality from the start of the analysis, the analysis is more likely to be right first time and thus save time and resources overall. However, quality management and control processes are also deployed to manage mistakes, handle changes to the analysis requirements and ensure the appropriate re-use of analysis for different purposes.

Working with uncertainty

2.31 Uncertainty is all around us and takes a variety of forms and so it should be acknowledged, understood and managed. All analysis contains inherent uncertainties and there are implications for the commissioner as well as the analyst.

2.32 Understanding the sources of uncertainty and the impact it has on the analysis will enable the decision-maker to apply appropriate weight to the results of the analysis.

2.33 Where practicable, uncertainty should be quantified. However, even where this is impossible or impracticable given time and resource constraints, a qualitative assessment of the uncertainty should be made. This must be communicated alongside the results of the analysis.

2.34 Further detail is set out in Chapter 5.

3 Commissioning analysis

Box 3.A: Commission analysis – key points

Commissioners play a vital role in assuring that analysis is fit-for-purpose.

At the initial engagement phase, the commissioner must share knowledge and contextual information that will support the analyst, and vice versa.

When interpreting the results of a piece of analysis, the commissioner provides constructive challenge and, with the analyst, explores whether further analysis is required.

When analysis is business critical or may become so in the future, the commissioner will wish to ensure that the analytical assurer is of an appropriate seniority.

Introduction

3.1 This chapter provides an overview of what commissioners of analysis will experience when working with an analyst to develop a commission, through to taking delivery of the analytical output.

Commissioners of analysis

3.2 People at all levels of seniority commission analysis. For large programmes, where business critical decisions may be taken, the commissioner may be a programme senior responsible owner or someone acting on their behalf. This chapter provides guidance for the commissioner of analysis that will assist in ensuring that they receive fit-for-purpose analysis.

Roles and responsibilities

3.3 During the **engagement** phase of the analytical cycle, see Figure 2.A on page 18, the commissioner and the analyst shape the analysis by developing a shared understanding of the problem and the context. The commissioner is responsible for ensuring that:

- key aspects of the problem, scope and complexities, including programme constraints, are captured and clearly communicated
- they are available to actively engage with the analysts in order to appropriately shape the work
- appropriate resources, including specialists, are commissioned for the analysis
- they work with the analyst to ensure that the scope of the project is proportionate to the issue being addressed and that the criticality of the analysis is understood by the analyst and the analytical assurer
- there is proportionate governance in place to support the analysis and its role in the wider project or programme – this is particularly important if the analysis is supporting business critical decisions

3.4 During the **design** and **conduct of analysis** phases, the commissioner may need to provide the analyst with important information for the analysis to proceed or be asked for other essential input. While it is not possible to provide an exhaustive list, the following are to be expected:

- identification of where the boundaries are between the problem in question and other topical issues or problem areas, including any dependencies
- agreement to the use of specific data and assumptions and provide agreement and sign-off to any assumptions that are developed as part of the project
- details of any changes to the scope and intended use of the analysis or change of importance of the analysis
- the level of precision and accuracy required

3.5 As part of the **delivery** phase, once the analysis has been interpreted and the results have been provided, the commissioner should:

- ensure that there is an assessment of the level of analytical quality assurance of the analysis, making note of where there has been a trade-off between time, resources and quality
- ensure that an assessment of uncertainty has been provided and that the implications of the uncertainties are understood
- provide constructive challenge and scrutinise the analysis against other evidence that has been gathered
- ensure that the views of the analytical assurer on the level of analytical quality assurance and the residual uncertainty are considered clearly and effectively when the analysis is used. In particular, the onward communication of the analysis should include the context of the problem being considered and include information on residual uncertainty, risks, limitations and constraints

3.6 The commissioner may also expect to take delivery of a variety of products that support and communicate the analysis. What is required will depend on the intended uses of the analysis.

4 Quality analysis and quality assurance

Box 4.A: Delivering quality analysis – key points

The process of ensuring that analysis is fit-for-purpose must be delivered in partnership across the suggested roles of commissioner, analyst, and analytical assurer. It is vital that the commissioner is confident that appropriate assurance is undertaken.

Effective communication and transparency are crucial throughout the whole analytical process: in understanding the problem, designing the analytical approach, conducting the analysis and relaying the outputs.

The scale and scope of the applied assurance needs to be proportionate to the purpose and constraints of the analysis – one size does not fit all analytical projects.

The need for assurance applies both to the whole analytical process and to smaller pieces of analysis or modelling that form part of a wider analytical programme.

Without a record of analytical assurance activities that have taken place, confidence in the analysis by the commissioner is reduced.

Introduction

4.1 This chapter provides a high level summary of what is meant by quality analysis and places it in the context of the project trade-off between quality, time and resources.

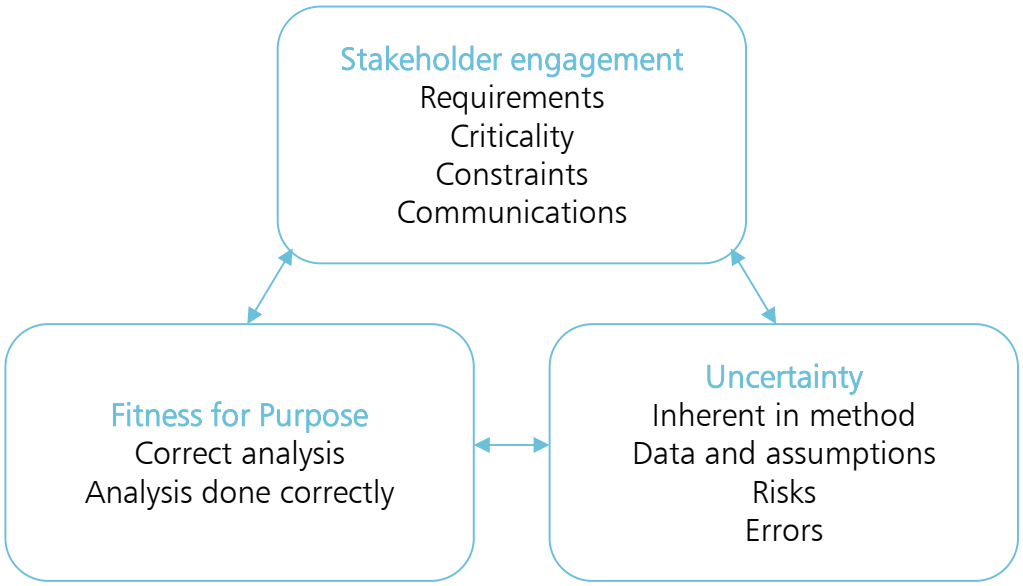
Requirements for quality analysis

4.2 To deliver quality analysis means the following have been addressed:

- the interactions between the analyst and the stakeholders have been effectively managed to ensure an proportionate amount of effort goes into an analytical project
- confidence has been provided that the analytical output is fit-for-purpose
- uncertainty and risks associated with the analysis have been quantified, where appropriate, and managed actively

4.3 The above are linked strongly. If there is an increase in the time available for the analysis project, then more time may be available for analytical quality assurance activities or for quantifying the residual uncertainty. However, in situations where the requirements of the commission increase but more time is not available, there is a pressure which may result in a less than ideal level of analytical quality assurance and/or an insufficient understanding of uncertainties.

Figure 4.A: Analytical project



There is usually a trade-off between the available resources and time for the project and the level of analytical quality assurance activities that can be completed. With any analytical project, the competing aspects of the project need to be considered.

Proportionate quality assurance

4.4 A wide range of factors will contribute to the overall quality of analysis. These include the skill and expertise of the analysts and users, the quality of the data and assumptions, the communication of the outputs and the understanding of the limitations and simplifications to the decision makers. All these factors play an important role in developing good quality analysis, and using it appropriately.

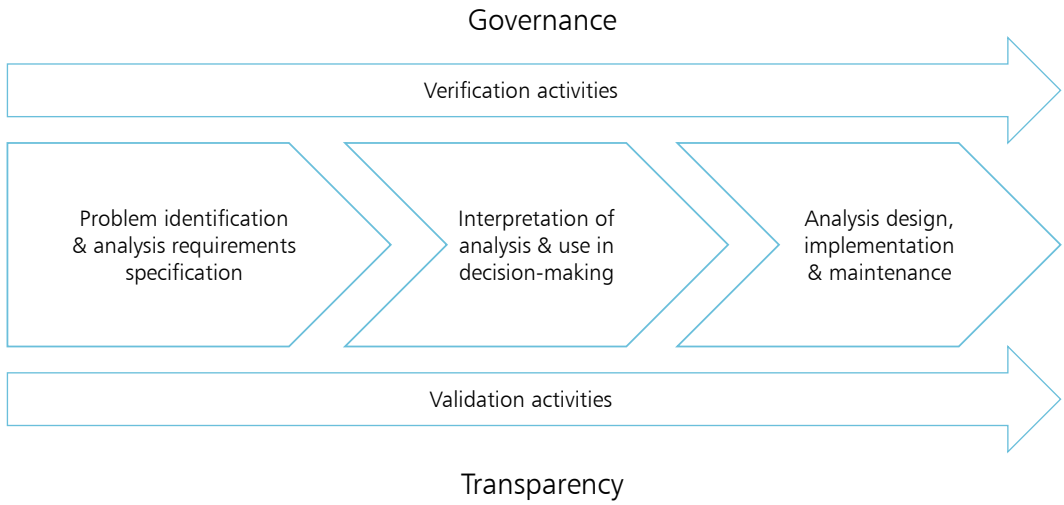
4.5 As the analysis progresses through the analytical cycle, there are various checks performed to ensure that the analysis is fit-for-purpose. Checks that confirm that the right analysis has been performed (known as validation) and that the analysis has been carried out correctly (known as verification) cover:

- the purpose for which analysis is developed
- the quality of any data inputs, and any assumptions that drive the analysis, including the estimation of parameters
- the use of the analysis' outputs
- the degree of risk and uncertainty in the analysis and it's outputs

4.6 To support these activities, appropriate governance and transparency between the analyst, the commissioner and the analytical assurer is required as is illustrated in Figure 4.B.

4.7 A governance framework should identify clear lines of responsibility and accountability, and transparency can help to ensure analysis benefits from external scrutiny. Effective governance and transparency can be particularly important where analysis is highly complex, and a level of expertise is required to understand the analysis and the risks associated with its complexity. An organisation's culture can also play a role in ensuring that appropriate quality assurance is highly valued and seen as fundamental to analysis, model development and use.

Figure 4.B: Quality assurance



Quality assurance combines both verification and validation activities throughout the life cycle of the analysis and are supported through appropriate governance and transparency between analyst, commissioner and the analytical assurer.

Source: Adapted from Review of quality assurance of government analytical models

4.8 There is no single "quality assurance" activity. Rather, quality assurance is delivered through a variety of different activities, each of which adds to the overall level of quality assurance. There can be no 'one size fits all' approach to determining what level of quality assurance is appropriate. There are good reasons why the approach to quality assurance will vary. These include the:

- type and complexity of the analysis
- novelty of the approach
- importance of the issue
- role and criticality of the analysis to the decision making process
- required precision of the analytical outputs
- amount of resource available for the analysis and the supporting assurance activities

4.9 This illustrates the importance, at all stages of model or analysis development, that analysts and their customers take a conscious decision on the amount and type of quality assurance that is appropriate. When there are time or resource constraints, analytical quality assurance activities should not be ignored. In such situations, the analyst should use a risk-based approach to highlight the areas of greatest potential error and focus assurance efforts on these areas. It is also important that the impact of any reduction in the thoroughness of analytical quality assurance activities is understood by the commissioner.

Quality assurance activities

4.10 There are many different analytical assurance activities that can be undertaken. These include:

- analyst testing
- peer review
- analytical audits

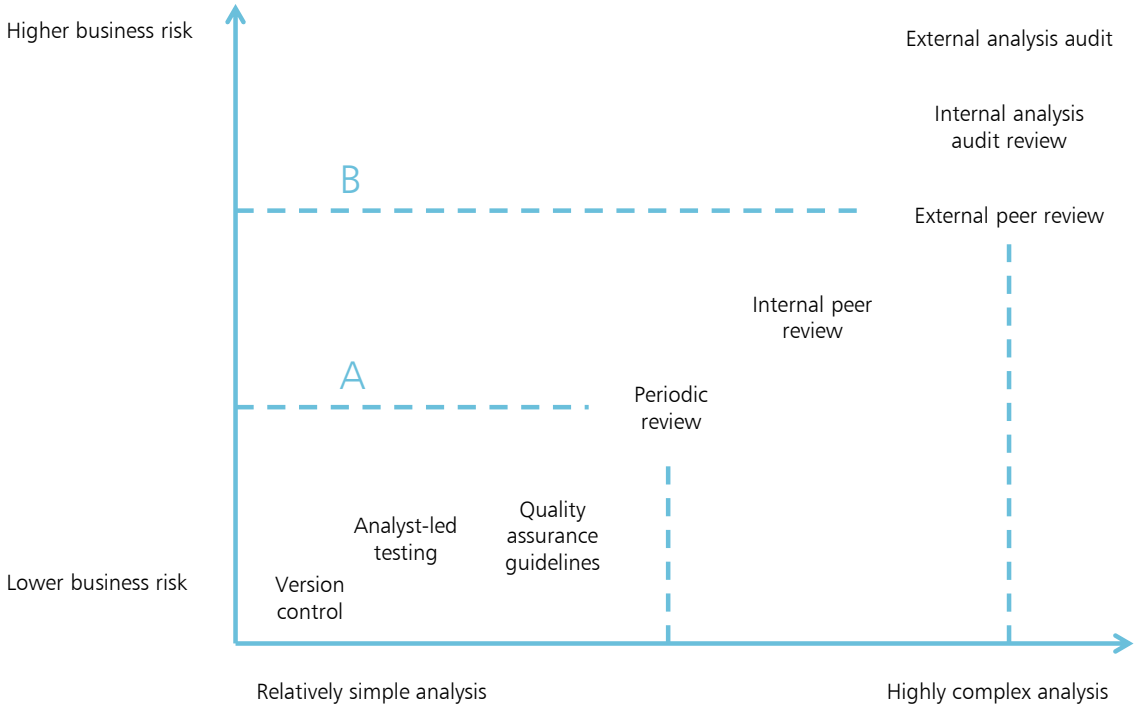
4.11 Analyst testing covers those activities where the analyst builds in checks and processes to ensure that their analysis is correct.

4.12 Peer reviews may use internal or external parties to provide critical challenge of the analysis. The peer reviewer might consider the entire analytical process from the user requirements through to the interpretation of the results, or focus on particular aspects of the project. Peer review provides constructive challenge and a fresh perspective on an analytical project. The peer reviewers should be unfamiliar with the detail of a piece of analysis.

4.13 An independent analytical audit focuses on the processes that were followed and the evidence gathered to provide the analytical assurance. The use of version controls and the adherence to guidance and checklists would form part of an audit.

4.14 In Figure 4.C, 2 projects (A and B) that carry different levels of risk are shown as an example of what types of quality assurance activity are required. It would be typical for all projects to include those activities to the left and below the most intensive quality assurance activity.

Figure 4.C: Types of quality assurance



The risk associated with a piece of analysis should influence the types of quality assurance activity that takes place. In addition to version control and analyst-led testing, Project B requires both internal and external peer reviews because it is riskier than Project A.

Source: Adapted from the Review of quality assurance of government analytical models

Box 4.B: Commissioning peer reviews and audits

In the cases of very complex analysis or analysis that drives a significant business critical decision, commissioners of analysis or analytical assurers may wish to request a formal peer review or analytical quality assurance audit of a piece of analysis. In doing so, consideration of the requirements should be carefully developed to ensure the appropriate assurance is obtained. If commissioning an external peer review or audit, it may be beneficial to obtain an internal peer review or audit first to establish the appropriate specification. A formal peer review or audit should also be accounted for in the time and resource needs of the analysis and the impact on the wider programme should be understood and managed.

Other supporting activities

4.15 Governance arrangements can be used to improve the analytical quality assurance activities. For example, formally agreeing assumptions will reduce the need for reworking of the analysis and provides greater time for assurance.

4.16 Greater transparency, e.g. through publication of inputs, assumptions and analysis, may enable wider engagement from experts who can provide external analytical assurance activities.

5 The importance and implications of uncertainty

Box 5.A: The importance and implications of uncertainty – key points

Analysis is used to inform decision-makers about which option to choose, often in unique situations. For each option, a range of real outcomes may occur – the actual outcome is uncertain. Uncertainty will always exist and is inherent in any analysis and real-world decision.

Decision-makers aim to achieve their desired outcome by adopting strategies which increase the chances of better outcomes occurring while decreasing the chances of less favourable outcomes occurring. This requires good information on uncertainty, such as the range of outcomes that may occur together with the likelihoods for each option they can choose. "Best estimates" are not usually enough.

In the proposed responsibilities framework, commissioners should always expect information on uncertainty from analysts, and challenge them when it is absent, inadequate or ambiguous.

Analysts often describe uncertainty in qualitative terms, e.g. moderate uncertainty, high confidence, etc., that do not express the range or likelihood of alternative outcomes. Commissioners should request further information, however project constraints and practicalities may limit what can be achieved.

If the uncertainties are too complex for analysts to quantify, even approximately, the analysts should say so in order that the commissioner can take this into account.

When communicating with decision-makers and stakeholders, commissioners of analysis need to describe the extent to which outcomes are uncertain and the reasons for this.

Introduction

5.1 This chapter discusses why understanding uncertainty is important and its implications for commissioning analysis and for communicating the analysis to decision-makers and stakeholders.

The importance of understanding uncertainty

5.2 Decision-makers aim to choose options that lead to good outcomes. It is for this reason that commissioners request analysis: to assess the prospective outcomes of alternative options.

5.3 However, the outcome of a decision is never known perfectly in advance. For each option, a range of real outcomes is possible: the outcome is uncertain.

5.4 Many different factors can contribute to the uncertainty of outcomes. These include uncertainty about the resources available to implement the decision, uncertainty about how the implementation will unfold, uncertainty about its immediate effects and uncertainty about its wider or longer-term consequences. They also include the potential for unexpected events, which should be considered as 'risks', or changes in the contextual environment.

5.5 Analysis seeks to take account of these factors, but is itself subject to uncertainties in the evidence, data and assumptions that are used and how they are combined. In addition, there

may be errors in analysis, though these should be mitigated by appropriate quality assurance. All of these factors contribute to the overall uncertainty about the decision outcome.

5.6 **"Best estimates" of outcomes are not enough**. For example, the best estimate of an outcome might be better for option A compared to option B, but if the uncertainty of A is greater it might carry an unacceptable chance of much worse outcomes.

5.7 To select the best option, decision-makers need to take account of the range of outcomes that may occur for each option and their relative likelihoods. In other words, they need information on uncertainty.

5.8 Information on the overall uncertainty is needed to inform choices between options. In addition, information on the main sources of uncertainty is useful for identifying risks that can be mitigated and areas where the data or analysis can be improved. This should be discussed between the commissioner and the analyst and should form part of the quality assurance process.

5.9 Uncertainty is generally increased in situations where there are time or resource pressures. This is accentuated in crisis situations, because they tend to involve new or unexpected problems and there is less time for investigation and analysis. Commissioners should expect to see greater levels of uncertainty in these situations.

Implications for commissioners

5.10 Commissioners should request and expect information on uncertainty from analysts, and challenge them when it is absent, inadequate or ambiguous. This information should go alongside details of the quality assurance effort that has been undertaken. In addition, commissioners of analysis may have identified sources of uncertainty as part of their wider considerations and should communicate these to the analyst.

5.11 Analysts may describe uncertainty in qualitative terms, e.g. moderate uncertainty, high confidence, etc. This does not express the range or likelihood of alternative outcomes, which need to be taken into account in decision-making. Furthermore, qualitative expressions are by their nature relatively ambiguous and subjective. The same expression, e.g. "moderate", is interpreted in different ways by different people, so commissioners and decision-makers may over- or under-interpret the degree of uncertainty and consequently take inappropriate decisions. When more quantitative expression of the range and likelihoods of alternative outcomes would be useful, commissioners should request it.

5.12 Uncertainty can be analysed at different levels of detail, with analytical methods of increasing sophistication, and requires time, resources and specialist expertise. It is therefore important to conduct analysis at the level required to support decision-making, and avoid spending resources on excessive detail and sophistication. If it is possible for the commissioner to indicate in advance the consequences for decision-making of different degrees of uncertainty, this may enable the analyst to conduct their analysis at a proportionate level. If this is not possible an efficient alternative is for the analyst to start by providing a simple analysis, which can then be refined over time until it provides a sufficient basis for decision-making.

Implications for decision-making

5.13 Information on uncertainty helps decision-makers take decisions that are more likely to achieve their desired outcome. They may wish to adopt a risk-averse or risk-taking strategy depending upon the assessment of uncertainty alongside other considerations such as legal, economic, social and political factors.

5.14 Commissioners of analysis should also consider whether it may be beneficial to commission more refined analysis of uncertainty, or evidence-gathering aimed at reducing uncertainty. Through discussion with the analyst, this can be targeted on those areas of the analysis where there is the best prospect of reducing uncertainty in a cost effective way. If there is a need for urgent action, such as a precaution against unacceptable but uncertain risks, the commissioner may request further analysis or evidence-gathering be commissioned in parallel to inform subsequent adjustments of the policy response when uncertainty is reduced.

5.15 If any source of uncertainty is so deep that nothing can be said about its impact on the policy outcome, this implies that the outcome could be anything, i.e. is totally uncertain. Deep uncertainties have major implications for decision-making: they are likely to require strategies that are precautionary and/or flexible, coupled with evidence-gathering and monitoring of emerging outcomes.

5.16 Decision-making should also take account of the quality of the analysis, which should be qualified alongside the analysis. When there have been time or resource constraints that have a negative impact on the quality of the analysis, the commissioner needs to consider, preferably in discussion with the analyst, whether this implies additional uncertainty about the outcomes, beyond that indicated by the analysis. If so, this additional uncertainty should be taken into account when considering the decision strategies outlined above.

Box 5.B: Further resource on risk

More detailed discussion of risk appetite and options for addressing risk is provided in the HM Treasury's Orange Book that is available at: www.gov.uk/government/publications/orange-book. Accessed February 2015.

Communicating uncertainty

5.17 Commissioners of analysis often have to further communicate the results of analysis to decision-makers or other stakeholders. In doing so, the commissioners should consider:

- the consequences of communicating certainty when there is an uncertainty, as credibility will be damaged if things turn out differently
- a balanced picture of the analysis covering what is known and what is uncertain
- describing the range of possible outcomes and their relative likelihoods
- identifying key risks or uncertainties that have a large impact on the predicted outcome, and explaining the reasons for the uncertainties and the circumstances in which the risks might be realised
- identifying options for managing risk and uncertainty
- being open about the existence of any deep uncertainties whose impact cannot be assessed, and explain how they are managing those uncertainties
- avoiding implying unwarranted confidence in particular outcomes. Focus instead on communicating the level of confidence in the appropriateness of the proposed decision, and explain how this is justified by the quality of the analysis and by the measures that can be taken to address risk and uncertainty

Part B

The Aqua Book clarifies the responsibilities required to ensure that analysis is fit-for-purpose: it presents a framework for addressing these responsibilities through the roles of the commissioner of analysis, the analytical assurer and the analysts throughout the life cycle of the analytical project. Departments and agencies will wish to ensure that responsibilities are met, though the roles may vary locally.

This part of the Aqua Book is written for analytical assurer and the analyst.

Chapter 6 provides a more detailed coverage of the principles of verification and validation and describes the activities that can provide analytical quality assurance. In addition, there is a helpful mnemonic to assist analysts in challenging their own analysis and that of others: RIGOUR. This chapter also outlines suggested activities and whether the commissioner, analyst or analytical assurer is best place to take ownership.

Chapter 7 outlines the responsibilities of the analytical assurer, who should provide the confidence to the commissioner that an appropriate amount of quality assurance has been performed on the analysis.

Chapter 8 discusses the approach to understanding uncertainty and how to go about quantifying, where possible, the uncertainty that inherently affects the output of any analysis. This chapter highlights the many different sources of uncertainty and offers approaches to help understand them.

6 Verification and validation

Box 6.A: Verification and validation – key points

Analytical quality assurance involves verifying and validating the analysis, i.e. that the analysis has been conducted as planned and that it is the right analysis, and it is delivered through a partnership of the commissioner, the analyst, and the analytical assurer.

Effective communication and transparency are crucial throughout the whole analytical process; in understanding the problem, designing the analytical approach, conducting the analysis and relaying the outputs.

The scale and scope of the verification and validation activities applied need to be proportionate to the purpose and constraints of the analysis – one size does not fit all analytical projects. Engagement between the commissioner, analyst and analytical assurer helps to identify the appropriate proportionate response.

While the results of the analysis is an important focus of any scrutiny, the need for verification and validation applies to the project as a whole as well as to the application of analytical techniques that deliver the analytical output.

Applying the principles of RIGOUR (repeatable, independent, grounded in reality, objective, uncertainty-managed, and robust) ensures that the key aspects of verification and validation are addressed.

Introduction

6.1 This chapter provides an account of the verification and validation measures that should be applied in order to provide analytical quality assurance. It describes approaches that are applicable to all types of analysis, although some may be more appropriate than others for a given piece of analysis and the available time and resources. This chapter focuses on those activities that are performed throughout the analytical cycle by the commissioner, analyst and analytical assurer, however an overview of additional assurance activities provided by third parties is presented for reference.

Quality analytical projects

6.2 Providing quality assured analysis means that the following must all be appropriately addressed:

- that the engagement between the analyst and the stakeholders have been effectively managed to ensure an proportionate amount of effort goes into an analytical project
- that confidence has been provided that the analytical output is fit-for-purpose and that there needs to be verification and validation of the analysis
- that the uncertainties and risks associated with the analysis have been understood, quantified where appropriate and managed actively

6.3 For analysis to be used to inform a decision it must be possible to assess its utility, reliability, and the degree of validation and verification to which it has been subjected:

- the utility of the analytical output can be assessed through gathering feedback from the end users about its practical benefits
- reliability refers to how consistent the analysis is in ensuring that it adds value to the commission and the intended purpose of the analysis
- verification activities ask whether the analysis has been conducted correctly
- validation activities ask whether the correct analysis has been performed

Verification and validation and the analytical cycle

6.4 To understand what utility, reliability, validity and verifiability mean in practice, it is worth considering the simplified analytical process presented in Figure 2.A on page 18 and expanding on each of the stages.

6.5 Analytical projects typically start with **customer engagement** although it is possible that other events trigger analytical projects. The commissioner plays an important role in communicating the questions to be addressed and working with the analyst to ensure an appropriate understanding of the problem is communicated. The commissioner is in fact verifying the understanding of the analyst, and the analyst is:

- validating that the analysis requested is actually what the commissioner requires in order to answer the problem
- identifying what form the findings need to be in to be of use to the commissioner

6.6 In this stage of the project it is important for there to be transparency of the understanding between the commissioner and the analyst. Best practice is to record the customer engagement process in scoping documentation, which forms a living document to record the details of the work requested originally plus any changes made to the commission later.

6.7 During the **design phase**, the analyst will convert the commission into an analytical plan. Key considerations include the inputs, possible methods of conducting the analysis and the outputs that will be produced. Throughout, it is important to validate that the analysis will deliver against the commission and to verify the conduct of the work against the aim. Some iteration between the commissioner and the analyst is to be expected as the analytical solution develops. The analytical assurance role assists in the identification and consideration of the validation evidence, as well as in reviewing the audit of the verification material. The commissioner also has an important role, since they may well be more an expert in the subject than the analyst. As such, their contribution towards the input assumptions, data requirements and the most effective way to present the outputs can prove invaluable.

6.8 The design phase requires that the conversion of the commission into an analytical plan is transparent. This transparency helps expose the assumptions that have been made and highlight any known limitations of the proposed approach. Best practice is to document the design process which should record how the proposed analytical process is intended to generate the requested insights. The concept of analysis should, if appropriate, be supported by design documentation.

6.9 When the analysts are undertaking their analysis, they will ensure that they are conducting the analysis correctly and will record any changes to their plan that they have had to make, such as if they encounter any difficulties or unexpected limitations. The analytical assurer can then comment on whether the analysis is still meeting the needs of the commission to ensure best use of the results.

6.10 Regular contact with the commissioner, for example through regular update reports on large projects, provides an opportunity for the commissioner to be able to advise on whether the analysis is still meeting their needs or whether there are any new requirements.

6.11 When conducting the analysis, it is important that it is transparent that the analytical plan has been followed and, if deviations have been necessary, any changes have been recorded. Best practice includes:

- maintaining a record of the work that has been done in a technical report
- logging the data and assumptions used in the analysis which should detail the source, ownership and a fitness-for-purpose risk assessment
- recording the verification and validation activities that have been undertaken, document any activities that are outstanding, and note what remedial action that has been taken and its impact on the analysis

6.12 During the **delivery phase**, the commissioner has another important role when they receive the interpretation of the results and determine whether it has addressed their request. However, as analytical projects frequently involve further iteration or extension this consideration may be the trigger for additional analysis.

6.13 Effective delivery requires a transparent translation from the results of the analysis to the conclusions presented to the commissioner. Best practice is to record this process in a customer report. The insights must also be presented in the most accessible form that can be achieved – the details of which should be determined on a case-by-case basis.

Box 6.B: RIGOUR of analysis

Throughout all the stages of an analytical project, the analyst should be asking questions of their own analysis. The helpful mnemonic "RIGOUR" may assist:

Repeatable

Independent

Grounded in reality

Objective

Uncertainty-managed

Robust

Repeatable: For an analytical process to be considered 'valid' it might reasonably be expected that for the "same" inputs and constraints the analysis produces the "same" outputs. It is important to note that different analysts will consider the analytical problem differently, potentially resulting in differing results, however if any one approach is repeated the results should be as expected.

Independent: To produce analysis that is free of prejudice or bias. In doing so, care should be taken to appropriately balance the views across all stakeholders and experts.

Grounded in reality: Quality analysis takes the commissioner and analyst on a journey as views and perceptions are challenged and connections are made between the analysis and its real consequences. Connecting with reality in this way guards against failing to properly grasp the context of the problem – which is being analysed.

Objective: Effective engagement and suitable challenge reduces potential bias and enables the commissioner and the analyst to be clear about the interpretation of the analytical results.

Uncertainty-managed: Uncertainties have been identified, managed and communicated throughout the analytical process.

Robust: Provide the analytical result in the context of residual uncertainty and limitations in order to ensure it is used appropriately.

Verification and validation roles and responsibilities

6.14 For each part of the analytical development cycle, there are multiple activities that help deliver quality analysis. Each analytical project will require more or less effort against each activity depending upon the complexity of the analysis and the familiarity of the analyst with the problem area.

6.15 The following tables list several verification and validation activities for each part of the analytical cycle. An indication is made as to whether the commissioner, analyst or analytical assurer has responsibility for that activity and which other roles may be involved (noting that departments and agencies may wish to tailor this framework to meet local business needs and processes). Additional activities may be required depending upon the specific analysis being covered. Within each table the term 'method' is used and should be interpreted to apply equally to analyses, tools, experiments or models.

Customer engagement

6.16 Successful engagement between the commissioner of analysis and the analyst will help identify the question that analysis can address and establish the context that the analyst needs to be aware of before they start the analysis.

Activity (to be tailored as appropriate to accommodate local practices)	**Suggested ownership (to be tailored as appropriate to accommodate local practices)**		
	Commissioner	**Analyst**	**Analytical assurer**
Ensure key aspects of the problem, scope and complexities are captured and clearly communicated.	Owner	Involved	Involved
Be available to engage with the analysts in order to appropriately shape the work.	Owner	Involved	
Clearly record the perceived purpose of the analysis and/ or modelling and the levels of quality and certainty that are required for this purpose.	Involved	Owner	Involved

Activity (to be tailored as appropriate to accommodate local practices)	Suggested ownership (to be tailored as appropriate to accommodate local practices)		
	Commissioner	Analyst	Analytical assurer
Challenge and test the understanding of the problem.	Involved	Involved	Owner
Ensure appropriate resources are commissioned for the analysis.	Owner	Involved	Involved
Ensure appropriate stakeholders have been identified so that the scope and boundaries of the problem can be appropriately explored.	Owner	Involved	Involved
Explore the requirements, boundaries, and scope with all of the stakeholders ensuring a wide range of perspectives are sought.	Involved	Owner	
Challenge the requirements, boundaries and scope and assess whether sufficient views have been considered.	Involved	Involved	Owner
Ensure expectations are managed to keep stakeholders expectations aligned with what can be delivered.	Involved	Owner	Involved

Designing the analysis

6.17 The analyst is responsible for planning their analysis including how they will demonstrate that they have considered analytical quality assurance throughout the 'conducting analysis' phase.

Activity (to be tailored as appropriate to accommodate local practices)	Suggested ownership (to be tailored as appropriate to accommodate local practices)		
	Commissioner	Analyst	Analytical assurer
Record and review the decision process from structuring the problem to developing the analytical plan. Check that the process reflects due RIGOUR.	Involved	Owner	Involved
Plan appropriate resources to deliver the analysis.	Involved	Owner	Involved
Capture the specification of any necessary methods. This must be adequate to allow subsequent verification testing / validation of the analysis. It should also specify what approaches will be used to identify, quantify and communicate uncertainty.		Owner	Involved
Produce appropriate design documentation. Best practice can include a concept of analysis, user requirements, design specification, functional specification, data dictionary, and test plan.		Owner	Involved

Activity (to be tailored as appropriate to accommodate local practices)	Suggested ownership (to be tailored as appropriate to accommodate local practices)		
	Commissioner	Analyst	Analytical assurer
Dry run the proposed approach to see if it delivers as intended. Then consider if the overall approach adequately addresses the complexities of the customer issue for this purpose. It is good practice to engage subject matter experts in this review.	Involved	Owner	Involved
Ensure the accuracy and limitations of the chosen methods are understood – and where appropriate tested (where possible baselining their response against independent reference cases).		Owner	Involved
Ensure the basis of the work is accurate, transparent (so that the basis of the findings can be understood) and well recorded.		Owner	Involved
Ensure the approach to the analysis is well-structured for the purpose, data driven, and reflects a robust overall design.		Owner	Involved
Ensure the level of quality checking of the analysis will be appropriate for the decision being supported.	Involved	Involved	Owner
Ensure that, if required, formal ethical approval is provided.	Involved	Owner	Involved

Conducting the analysis

6.18 As the analyst manages their analysis and follows their analytical plan, they are ensuring that the analytical assurer has sufficient evidence that proportionate quality assurance activities have taken place.

Activity (to be tailored as appropriate to accommodate local practices)	Suggested ownership (to be tailored as appropriate to accommodate local practices)		
	Commissioner	Analyst	Analytical assurer
Collect and manage data. Understand data accuracy and uncertainties. Capture, manage and understand implicit assumptions made.		Owner	Involved
Engage appropriate Subject Matter Experts, at the appropriate time, when collecting data. NB: The commissioner may be a subject matter expert.	Involved	Owner	Involved

Activity (to be tailored as appropriate to accommodate local practices)	Suggested ownership (to be tailored as appropriate to accommodate local practices)		
	Commissioner	Analyst	Analytical assurer
Record data and assumptions, including uncertainties and accuracy, in a master data and assumptions list to record the origin of all data used.		Owner	Involved
If applicable undertake parametric analysis to understand the consequences of missing or uncertain data and assumptions.		Owner	Involved
Ensure data formats, units, and context are properly understood and handled.		Owner	Involved
Ensure implications of any data dependencies or relationships to other analysis or methods are understood.		Owner	Involved
Ensure the level of quality checking of the analysis is appropriate for the decision being supported: All analysis requires some checks, at some level, by another competent person wherever and whenever practicable.	Involved	Involved	Owner

Delivery of the analysis

6.19 Communication of the results and the associated uncertainties and limitations is very important. The analytical assurer must be content that the analyst is presenting a true representation of the analysis that has been undertaken. The commissioner must be confident that the results of the analysis are fit-for-purpose in the context of the original question asked. They must also ensure they have sufficient statements of assurance and understand the uncertainty associated with the analysis.

Activity (to be tailored as appropriate to accommodate local practices)	Suggested ownership (to be tailored as appropriate to accommodate local practices)		
	Commissioner	Analyst	Analytical assurer
Ensure the results of the analysis are communicated clearly and effectively to the commission with statements of the degree of assurance associated with the analysis alongside a statement of the residual uncertainty.	Involved	Owner	Involved
Ensure that onward communication of the results beyond the commissioner are communicated clearly and effectively in the context of the problem being considered.	Owner	Involved	Involved

Activity (to be tailored as appropriate to accommodate local practices)	Suggested ownership (to be tailored as appropriate to accommodate local practices)		
	Commissioner	Analyst	Analytical assurer
Ensure uncertainty, risk, limitations, and constraints are communicated clearly, along with the results, to the study commissioner.	Involved	Owner	Involved
Ensure uncertainty, risk, limitations, and constraints are communicated clearly, along with the results, to the decision-makers and stakeholders.	Owner	Involved	Involved
Ensure an analytical record is provided to i) facilitate access to the analysis by broader stakeholders, ii) make the analysis exploitable for wider decisions, and iii) inform continual improvement.	Involved	Owner	Involved
Ensure a suitable audit trail is in place that clarifies the level of validation, scope, and risks associated with the analysis. Best practice includes the production of validation log books.		Involved	Owner
Undertake reflective learning to capture successes and difficulties and ensure these lessons are available to improve future analysis.	Involved	Owner	Involved

Documenting the assurance effort

6.20 Local business practices and processes will determine the appropriate form and format of those resources and templates that facilitate the verification and validation process supporting the delivery of quality analysis. The need for appropriate proportionality of analytical response and of the verification and validation efforts needs to be kept in mind.

6.21 The following templates should be considered:

- Specification documentation: this documentation captures the initial engagement with the commissioner and identifies the question, the context, and any boundaries of the analysis. It provides a definition of the scope and a mechanism for agreeing the project constraints, e.g. deadlines, available resources, etc., and capturing what level of assurance is required by the commissioner.
- Design documentation: this document outlines the design of the analysis, including conceptual models to illustrate the analytical problem, and forms an important tool for ensuring that the analytical assurer has the confidence that the analyst can deliver quality analysis.
- Assumptions and data log: a register of assumptions whether provided by the commissioner or derived by the analysis and data that have been risked assessed and signed-off by an appropriate governance group or stakeholder.
- Quality assurance plan: a detailed plan of what verification and validation activities are to be undertaken can also form the basis of a log for those analysts conducting

the verification and validation checks. Any additional verification and validation checks that have been performed should be recorded on the quality assurance plan.

- User / technical documentation: all analysis should have documentation for the user, even if that is just the analyst leading the analysis. This is to ensure that they have captured sufficient material to assist them if the analysis is revisited in due course. For analysis that is more likely to be revisited or updated in the future, documentation should be provided to assist a future analyst and should be more comprehensive. This documentation should include a summary of the analysis including the context to the question being asked, what analytical methods were considered, what analysis was planned and why, what challenges were encountered and how they were overcome and what verification and validation steps were performed. In addition, guidance on what should be considered if the analysis is to be revisited or updated is beneficial.
- Customer reports: a report that provides a detailed summary of the analysis outlining the commission through the design, conduct of analysis and delivery phases. The report should outline the decisions taken in order to complete the analysis and provide detailed descriptions of the analytical assumptions, levels of uncertainty etc.
- Assurance statement: a brief description of the analytical assurance efforts that have been performed to assure the analysis. The statement should make reference to known limitations and conditions associated with the analysis.

Box 6.C: Additional Aqua Book resources

Accompanying the Aqua Book is a suite of resources that include generic templates that can be adopted and tailored as required. For further information on templates, please see Part C. At any time, analysts should follow local guidance.

Proportionality

6.22 There is no "one size fits all" for the verification and validation process. The level of time and resource spent on demonstrating fitness-for-purpose should also be balanced against the criticality and urgency of the decision being supported. It is likely that more effort will be required in the following situations:

- when complex analytical techniques are used
- when a novel approach is adopted
- when the analysis is business critical or addresses a potentially controversial issue
- when there is limited evidence to provide challenge of the results of the analysis
- when the results are required to a high level of precision and accuracy

6.23 Quality assurance activities should be considered for every piece of analysis, including:

- analyst testing
- peer review
- analytical audits

6.24 **Analyst testing** covers those activities where the analyst builds in checks and processes to ensure that their analysis is correct. This may be supplemented by "dual running" where the analysis is repeated by a different analyst to ensure it is consistent and "sense checks" where alternative calculations are performed, often to a reduced accuracy for ease and speed, to improve confidence.

6.25 Activities that an analyst may perform under this strand of assurance checks include:

- adopting standards, e.g. spreadsheet or coding standards, to assist peer review checks through standardisation of approach
- comparison to other analysis, models or real events (e.g. historical data)
- applying "built in" checks to the analysis to highlight areas of concern
- test the analysis with alternative input data or assumptions to check the analysis behaves as expected
- run the analysis with extreme values to facilitate the identification of errors or unexpected behaviour

6.26 **Peer reviews** may use internal or external parties to provide critical challenge of the analysis. They might consider the entire analytical process from the user requirements through to the interpretation of the results, or focus on particular aspects of the project.

6.27 The views of others – whether in an expert capacity or not – can provide constructive challenge to the analytical approach, application of a methodology or interpretation of the analysis. Peer review enables:

- knowledge transfer
- "lessons learnt"
- a fresh perspective to test the logic and analytical approach adopted

6.28 An **analytical audit** focuses on the processes that were followed and the evidence gathered to provide the analytical assurance. The use of version control and the observations of guidance and checklists would form part of an audit. To assist an audit, recognised best practice is to:

- maintain a log of the analysis and the verification and validation checks that have been performed
- keep track of changes and modifications – and the reasons for doing so – alongside the analysis for ease of reference
- establish an appropriate version control system for both the analysis as a whole and for the supporting data and assumptions
- maintain a risk register and issues log for the analytical project

Quality under time and resource constraints

6.29 In some situations, such as for analysis that is needed under significant time constraints, it may not be possible to do as much work as is usual in order to understand and improve the verification or validation of the analysis as desired. In these situations:

- verification and validation efforts should focus on those areas of the analysis that are likely to have the largest impact on the analytical output and that are associated with the greatest risk

- the analysis should be communicated with appropriate caveats outlining what has and, importantly, has not been through verification or validation together with a practical interpretation of the associated risk
- when time allows, further assurance activities should be performed after the event so as to capture lessons learnt

7 Analytical assurance

Box 7.A: Analytical assurance – key points

In the proposed responsibilities framework, analytical assurance provides the sign-off that a piece of analysis is fit-for-purpose. Departments and agencies will each ensure that this is done, however the specific role and seniority may vary locally.

For analysis that is business critical, the seniority of the analytical assurer should be appropriate to the risks associated with the analysis and the wider project or programme it supports.

Introduction

7.1 This chapter provides an outline of the analytical assurance role and how it provides the commissioner with the required assurance, specifically that there has been proportionate and appropriate verification and validation of the analysis.

The analytical assurer

7.2 The concept of analytical assurance is not new: it is commonplace for someone with responsibility for the delivery of analysis to request that an independent analyst provides an appropriate review before the analysis is communicated.

7.3 However, this important responsibility requires greater acknowledgement to help create an appropriate environment for the quality assurance of analysis.

7.4 The analytical assurer should produce a report outlining their view of the level of analytical assurance checks that have been completed. The report should include a short assurance statement that can be carried alongside the output of the analysis for the benefit of the commissioner. An assurance statement could include a summary of the assurance activities that have been undertaken and any important risks that have been identified.

7.5 Departments and agencies may wish to consider formalising the reporting of such assurance statements and include them in formal ministerial or senior official submissions, board papers and reports.

Business critical analysis and the seniority of the analytical assurer

7.6 The proportionality of the quality assurance response, such as the number and type of verification and validation checks that are necessary to provide assurance, will vary by project. However, business critical analysis is more likely to require:

- verification and validation that are formally project managed with a formal governance arrangement
- external subject matter experts to validate the approach and assist with the interpretation of the findings such as through academic peer review
- verification checks performed by an analyst who has had no involvement in the design and execution of the analysis in order to ensure that a fresh pair of eyes considers the analysis and the adopted approach

7.7 To provide the appropriate assurance, the senior responsible officer must be satisfied that the analytical assurer has appropriate seniority for the risks associated with their wider project or programme.

7.8 The analytical assurer should expect to review evidence confirming the following have been completed over the life cycle of the analysis:

- evidence that the requirements have been captured, agreed through an appropriate governance process and shared with the commissioner, analysts and other stakeholders
- supplementary evidence of agreement and scrutiny of assumptions, analytical requirements and design considerations, including an analytical appraisal of options with considerations to uncertainty, limitations and weaknesses
- evidence that the commissioner and the analysts have discussed options and have agreed the verification and validation activities that are proportionate to the business critical risks
- supporting material confirming the suitability of the deployed analytical resources to both conduct the analysis and to carry out the verification and validation activities
- evidence that there has been sufficient networking between analysts as the providers of data and other evidence sources
- detailed descriptions of the verification and validation activities and their conclusions with supporting documentary evidence of the conduct of the work

7.9 The analytical assurer should provide a formal report that communicates the proportionality of the verification and validation checks. It is important that information covering the scope of the analysis, and relevant caveats, are communicated. In addition, an assessment of the residual uncertainty should be provided.

8 Analysing uncertainty

Box 8.A: Analysis of uncertainty – key points

Decision-makers need information on the uncertainty of decision outcomes, i.e. the range of outcomes that may occur and their relative likelihoods, in order to act appropriately and be confident with the messages they communicate.

To provide this, analysts need to:

- identify sources of uncertainty, i.e. the reasons why outcomes may differ
- assess the impact of the uncertainties, i.e. the range of outcomes they may cause and the relative likelihoods of those outcomes
- communicate the analysis of uncertainty clearly

Causes of uncertainty of the outcome include: uncertainties associated with data and assumptions used in the analysis; changes in the wider environment; and unexpected events or risks that may influence the outcome.

Uncertainty should be expressed as the range of possible outcomes and their likelihoods, as this is what commissioners need to communicate to the decision-maker. It is important to express this quantitatively if possible, even though it may be approximate.

A range of quantitative methods is available for analysing uncertainty, including sensitivity analysis to explore the range of possible outcomes and probabilistic modelling to estimate their likelihoods.

Analysts also need to use expert judgement to assess any additional uncertainties that are not quantified by modelling, because it is the overall uncertainty that is important for policy-making. In doing this, analysts need to guard against overconfidence and other biases that commonly affect such judgements.

Uncertainty analysis requires time and resource and should be proportionate to the commission.

Introduction

8.1 This chapter describes how analysts can identify sources of uncertainty and risk, and how they can assess the impact on the analysis that supports decision-making. The importance of communicating the uncertainty associated with the analysis is highlighted.

Objectives for uncertainty analysis

8.2 Analysis is used to inform decision-making by predicting and understanding the likely outcomes. For example, a decision may be required today in order to achieve a desired outcome in many years' time. The actual outcome of a decision is inherently uncertain: many other events could occur that will affect the actual outcome. Decisions-makers need information on the range of outcomes that may occur and their relative likelihoods.

8.3 To provide this, analysts need to:

- identify sources of uncertainty, i.e. the reasons why actual outcomes may differ from those presented

- **assess the impact of the uncertainty**, in terms of the range of outcomes they may cause and the relative likelihoods of those outcomes

Identifying sources of uncertainty

8.4 Many different factors can contribute to the uncertainty of outcomes including:

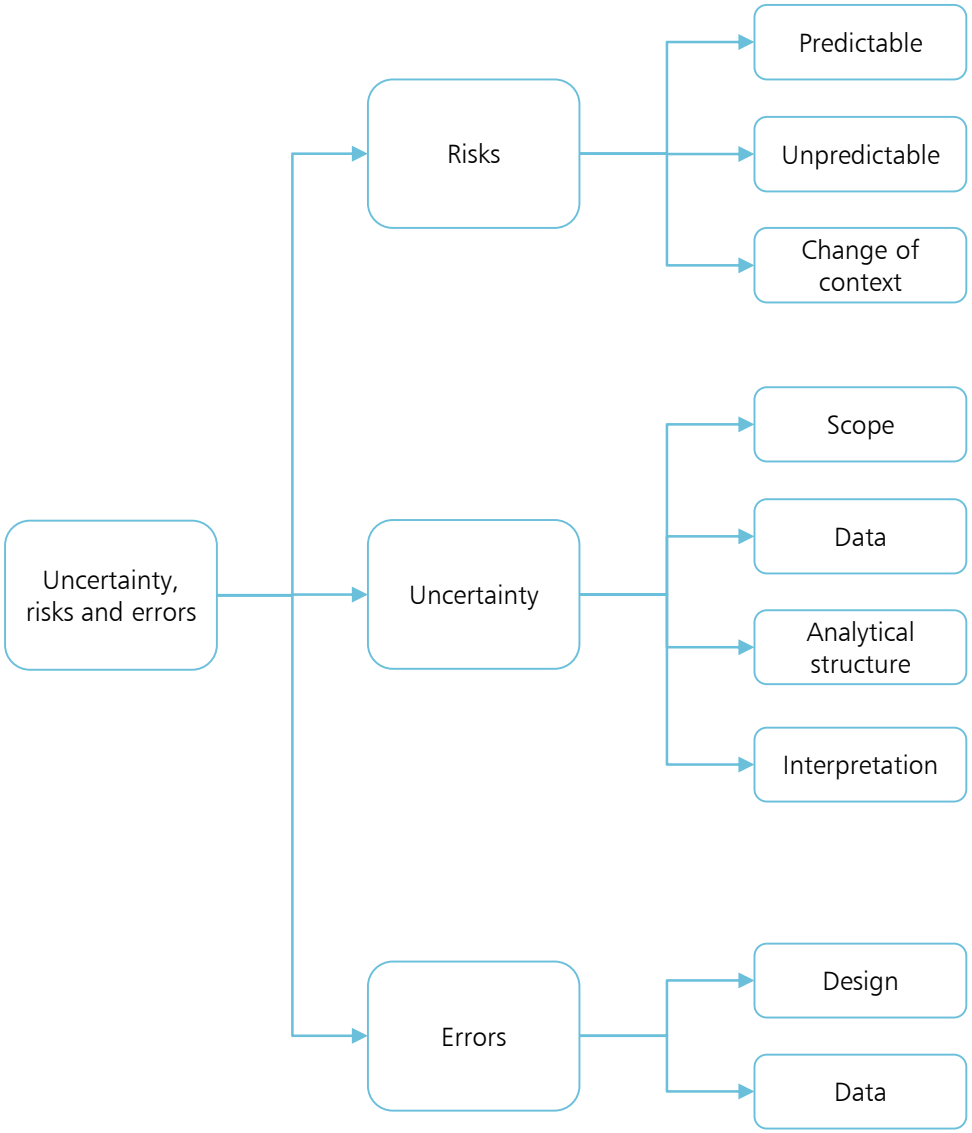
- the resources available to implement the decision
- how the decision will be implemented and its immediate effects
- the wider or longer-term consequences of a decision
- the potential for unexpected events, which might be considered as the realisation of 'risks', or changes in the contextual environment

8.5 Analysis seeks to take account of the above factors, but is itself subject to uncertainties, see Figure 8.A on page 51. Aspects of the analysis where uncertainty may arise include:

- inputs, including numerical data, evidence, intelligence and assumptions
- structural considerations, such as the logical flow and choice of analytical techniques for the problem in question, or the use of models beyond the defined scope
- external risks, influences and behaviours that may affect the outcome but are not taken into consideration within the analysis

8.6 In addition, there may be errors in analysis, though these should be limited by appropriate quality assurance. All of these factors contribute to overall uncertainty of knowing the outcome of a decision.

Figure 8.A: Sources of uncertainty



An overview of the risks, errors and uncertainties that may contribute to the overall uncertainty within a piece of analysis.

8.7 Analysts should examine their analysis systematically for all possible sources and types of uncertainty, to minimise the chance of missing any that might be important. Tips that may help with this include:

- Develop a list of types of uncertainty encountered related to the field of work, or adopt a general list. This list can be used as a checklist when searching for uncertainties affecting a piece of analysis.
- Make a list or table of all the input data, evidence and intelligence used in the analysis and then consider each type of uncertainty that could affect it. It may be useful to maintain an assumptions and data log for this purpose.
- Write down each step of the analysis and/or draw diagrams that represent the structure of the analysis. For each step, consider what additional factors might act at that point and affect the analysis outcome.
- Identify any external risks and potential changes in the wider policy landscape that might affect the outcomes of interest.

- Review the structure of the analysis as a whole and consider carefully whether there are any other ways in which it could have been approached.

8.8 Document all the sources of uncertainty and risk that are identified, including any considered negligible and/or tolerable. This provides a transparent record of what has been considered and contributes to the credibility of the analysis.

Expressing the impact of uncertainty

8.9 It is not sufficient to identify and describe risks and uncertainties. It is essential also to assess their impact on the outcome of the analysis and their contribution to the range and likelihoods of possible outcomes.

8.10 If this is not done explicitly as part of the analysis, it will be done implicitly when decisions are made. For example, the decision may be based on the best estimate, which would imply that the combined impact of all the uncertainties is assumed to be negligible.

8.11 If the analyst does not evaluate the impact of the uncertainties, it will be left to the commissioner or decision-maker to do this. This is undesirable, because they will generally be less familiar with the details of the analysis and therefore less able to evaluate the impact of the uncertainties.

8.12 Qualitative expressions of uncertainty, e.g. 'moderate' or 'high', are ambiguous and mean different things to different people, so the degree of uncertainty may be misunderstood by others and decision-makers may over- or under-interpret the degree of uncertainty and consequently take poorly informed decisions. In addition they do not describe either the range or likelihood of alternative outcomes.

8.13 Therefore, it is recommended to express the impact of uncertainty quantitatively when possible in terms of the range of outcomes and their likelihoods, even if this is approximate and/or subjective.

Quantifying uncertainty about the structure of the analysis

8.14 Sometimes there is uncertainty about the structure of the analysis, such as the equations or logic used to combine the analysis inputs. The impact of this may be quantified by implementing different versions of the analysis, for example using alternative formulations and examining how this changes the outputs. This can be applied both to uncertainty about specific steps in a single piece of analysis and to different approaches to the problem, which require wholly different analysis.

8.15 This process will result in one set of alternative outputs for each alternative piece of analysis. If relative likelihoods can be assigned to the different versions of the analysis, and hence to their outputs, this will provide better information for decision-making.

Quantifying uncertainty about additional factors, risks and external changes

8.16 If it is uncertain whether a particular factor should be included in the analysis, then **repeating the analysis** with and without it will show the range of alternative outcomes. Once key factors have been identified, their uncertainty can be assessed.

8.17 A similar approach using **alternative scenarios** can be used to quantify the impacts of external risks and potential changes in the policy landscape. Repeating the analysis with and without a risk event, or with and without an external change, will show their impact on the range of alternative outcomes.

8.18 If relative likelihoods can be assigned to the alternative scenarios and their impacts with and without the additional factor, risk or policy change, then this will provide better information for decision-making.

8.19 If it is not possible to add a potentially relevant factor, risk or external change into the analysis, this must be made clear. Approaches for addressing uncertainties that are not included in the analysis, or which cannot be quantified, are discussed below.

Quantifying uncertainty associated with the use of data in the analysis

8.20 When an input or parameter is estimated from data, this will often rely on **statistical methods** and theory to identify the most appropriate estimate. Further statistical methods may be used to directly quantify the uncertainty in the estimate. Whilst this output may provide an estimate of uncertainty within the context of the statistical method being used, it is important to realise that the underlying statistical methods are in themselves a form of model and may introduce further uncertainty.

Box 8.B: Uncertainty associated with statistical techniques

Where parametric techniques are used, for example by assuming a statistical distribution, there may be uncertainty about the applicability of this assumption or about the value of the parameters of the distribution.

Where non-parametric methods are used, for example using bootstrapping techniques to re-sample from the data, the resulting estimates of variance will themselves be uncertain, especially if the quantity of data is limited.

8.21 It is important to consider how well available data meet the needs of the analysis. Often, no data are available that are directly and precisely relevant to the parameter and conditions of interest. In such cases, it is often possible to use **surrogate data**. These are measurements of another parameter, or of the parameter of interest under different conditions, that are related to the parameter and conditions of interest. This implies an **extrapolation** between parameters, or between conditions for the same parameter, which introduces further uncertainty, additional to that associated with the data themselves. It may be possible to quantify this additional uncertainty using expert knowledge of the relationship between the surrogate and the parameter of interest.

Box 8.C: The limitations of data

It is rare to have the perfect dataset for an analytical commission. Reasons for this include:

- the data is not available in the time frame required for the ideal analysis
- the data definition does not perfectly align with the commission
- there are data or coverage gaps
- the data may be experimental or there are other reasons why it is not 'mature'

As a consequence it may be necessary for an alternative dataset to be used as a proxy and further uncertainty must, unfortunately, be introduced into the analysis. The impact of using a proxy dataset should be explored and, if the uncertainty associated with the dataset has a large impact on the analysis, its appropriateness should be revisited. This exploration, and the decision to use a particular dataset or input, should be recorded for the benefit of the analytical assurer.

Quantifying uncertainty with expert knowledge

8.22 When neither direct nor indirect data are available for a parameter needed in the analysis, **expert judgements** about that parameter may be sought from people with relevant knowledge.

8.23 Expert judgement may be used to estimate uncertainties associated with data, as well as in the absence of data. Examples might include estimating the size of adjustment needed for extrapolating from surrogate data to a parameter of interest, or to correct for biases in sampling or measurement.

8.24 When using expert knowledge it is important not to rely on the expert's 'best estimate', as this gives no indication of the generally large uncertainty involved. Rather, experts should be asked to provide a range or a range plus a central estimate, or a distribution representing both the range of alternative values and their relative likelihoods.

8.25 There are various formal methodologies for eliciting knowledge from experts such as the Cook, Sheffield or Delphi methods. These are designed to reduce the influence of cognitive biases that affect expert judgements, including over-confidence which can give too narrow a range or distribution, anchoring in which one fixes too strongly to an initial estimate and 'group think' whereby there is a lack of critical challenge in order to quickly reach a consensus. They also include standardised procedures for selecting the experts, training them in the elicitation process, and combining judgements from multiple experts.

8.26 Judgements of different experts will differ to some degree. This is another source of uncertainty, which needs to be taken into account. If there is considerable disagreement among experts, one option is to repeat the analysis using alternative expert judgements to show how much they affect the outcome.

8.27 Formal expert elicitation is costly in time and resource. It is therefore efficient to use simpler, less formal expert judgements to provide initial, approximate estimates and use these to identify the more critical parameters or uncertainties that may merit the expense of more formal elicitation. However, less formal methods should still guard against cognitive biases, such as by requiring experts to review their initial ranges and ask themselves why they could not be wider, to guard against over-confidence.

Quantifying the impact of uncertainties on the analysis outcome

8.28 It is not sufficient to quantify the uncertainty of individual components within an analysis. It is necessary to also quantify their impact on the overall outcome, because this is what matters for decision-making.

8.29 The impact of individual uncertainties on the analysis outcome may be quantified by sensitivity analysis or probabilistic modelling:

- Ranges or alternative point estimates representing a range of alternative values or scenarios may be propagated by simply repeating the calculation with each estimate in turn. This is a simple form of **sensitivity analysis**.
- Distributions may be propagated by repeating the calculation many times, sampling different values from the distributions each time: often referred to as stochastic or **probabilistic modelling**. This is often done by Monte Carlo simulation.
- If the form of a parameter distribution is uncertain, the impact of this can be quantified either by repeating the analysis with alternative distributions or by using imprecise probability approaches such as probability boxes, which envelope sets of potentially relevant distributions.

8.30 When quantifying the impact of multiple uncertainties, it is important to take account of potential dependencies between them. In sensitivity analysis, this means excluding implausible or impossible combinations of input values. In probabilistic modelling, it means specifying correlations between input distributions so as to take account of how the likelihoods of combinations of values deviate from the products of their individual probabilities. This is a critical challenge because wrongly assuming independence or mis-specifying dependencies can lead to highly misleading results, for example combinations of extreme values that rarely or never occur together in practice. Uncertainty about the form and magnitude of dependencies needs to be captured either within the model, such as using imprecise probability methods, or outside it.

8.31 Once the impact of uncertainties on the analysis output has been established, their relative contributions to overall uncertainty can be quantified, such as the percentage of overall variance. This is particularly valuable if time and resources allow further work to be done to reduce the most important uncertainties, but it can also provide guidance for future research.

Evaluating uncertainties that are not quantified by modelling

8.32 Even when many uncertainties are quantified by modelling, there will always be other uncertainties that affect the analysis outcome but are not included in the analysis. These must also be evaluated, because it is the overall uncertainty that is important for decision-making.

8.33 The basic requirements for evaluating uncertainties outside the analysis are the same as apply for quantifying uncertainties within the analysis. First it is necessary to evaluate the individual sources of uncertainty, and then it is necessary to evaluate their combined impact on the analysis outcome.

8.34 This can be undertaken in a stepwise manner:

1. List all identified sources of uncertainty in a table, including uncertainties relating to the structure of the analysis; uncertainties associated with data, evidence and intelligence used in the analysis; additional factors; external risks; and potential changes in the policy landscape.

2 Evaluate the impact of each uncertainty on the analysis outcome, when considered in isolation. For uncertainties affecting input data, it may be helpful first to evaluate the uncertainty of the data themselves, and then consider the consequences of this for the analysis outcome.

3 Evaluate the combined impact of all the uncertainties on the analysis outcome, when considered together. In this step it is important to consider potential dependencies between the different sources of uncertainty, just as is necessary when uncertainties are quantified within the model.

8.35 The evaluations at steps 2 and 3 should be expressed in quantitative rather than qualitative terms if possible, to avoid the ambiguity associated with qualitative expressions. Ideally one would use formal expert judgement to obtain these quantitative estimates, but generally there will be too many uncertainties for this to be practical. Instead, it is efficient to start by performing the evaluation using less formal expert judgements. If decision-makers require a more refined analysis of the uncertainty, formal expert judgement could be introduced progressively, such as first assessing the combined uncertainty, step 3, and then, if needed, for the individual impacts of the most important individual uncertainties.

Dealing with deep uncertainties that cannot be quantified

8.36 Quantifying uncertainty is difficult. Quantifying it statistically from data requires significant statistical expertise, and experts often find it very challenging to express subjective judgements in a quantitative form. If an uncertainty really cannot be quantified it is vital to communicate this to commissioners and decision-makers, as such **'deep' uncertainties** have special implications for the interpretation of analysis outputs and for decision-making.

8.37 If even one source of uncertainty is so deep that nothing can be said about its impact on the decision outcome, this implies that the outcome could be anything, i.e. is totally uncertain. This will have major implications for decision-making. Therefore, before concluding that an uncertainty is really unquantifiable, it is recommended to consider very carefully whether absolutely nothing can be said about it. For example, it is rare that a parameter could literally take any value from plus infinity to minus infinity, which is implied if one states it is unquantifiable.

8.38 If it is too difficult to express an uncertainty as a distribution, try to identify an approximate minimum and maximum or a range of alternative scenarios. Even when this results in a very wide range of estimates, it is still useful for decision-makers because it is essential for them to understand the magnitude of the uncertainty involved.

8.39 If even an approximate range cannot be given, bring the deep uncertainty to the attention of decision-makers at an early stage and describe its nature and causes as fully as possible.

Making the uncertainty analysis proportionate to the problem

8.40 Uncertainty can be analysed at different levels of detail with analytical methods of increasing sophistication, and requires time, resources and specialist expertise. It is therefore important to conduct analysis at the level required to support decision-making, and avoid spending resources on excessive detail and sophistication.

8.41 If it is possible for the commissioner to indicate in advance the consequences for decision-making of different degrees of uncertainty, this may enable the analyst to conduct their analysis at a proportionate level.

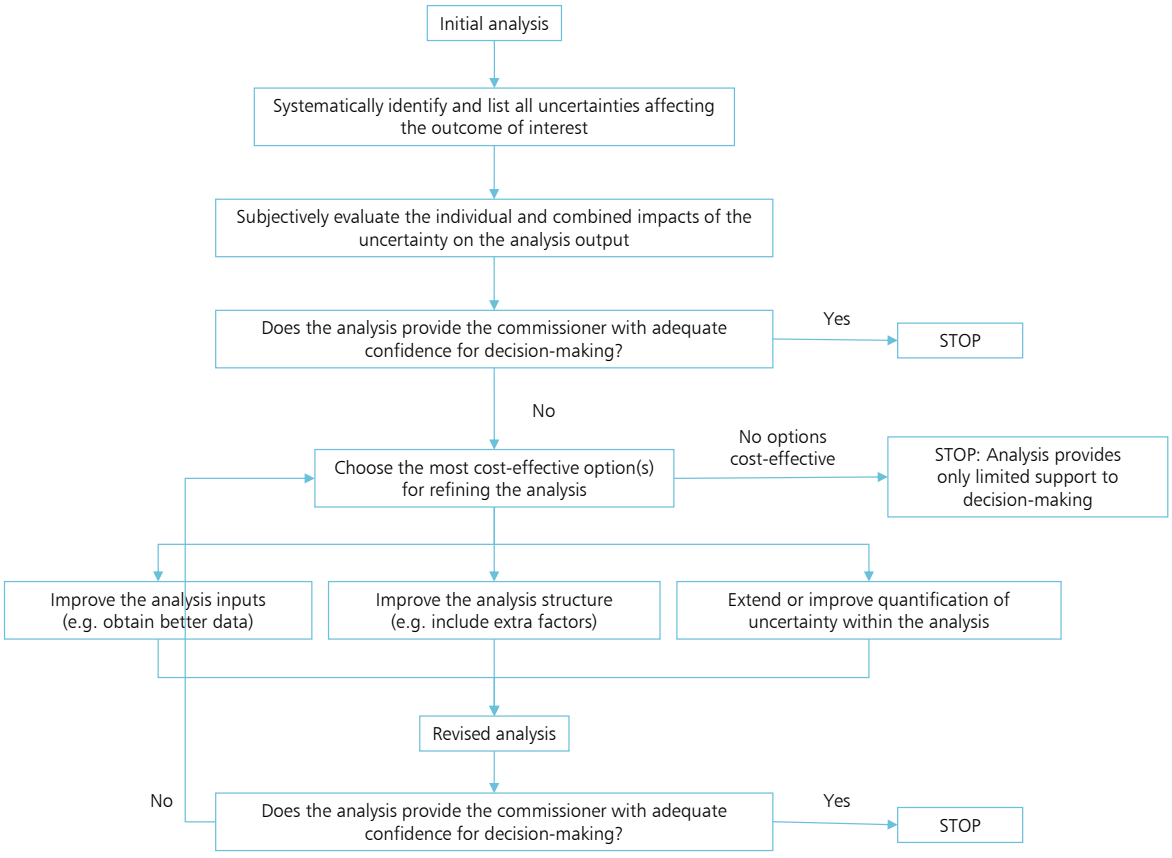
8.42 If this is not possible then an efficient alternative is for the analyst to start with simple and approximate methods and progress gradually to more refined approaches only when this is needed for decision-making, see Figure 8.B below:

1 Initially, do not attempt to quantify any uncertainties by modelling. Instead, systematically identify all the uncertainties, evaluate their combined impact on the analysis outcome by informal expert judgement, and communicate the result to the customer, making clear its approximate and subjective nature.

2 If the initial evaluation of uncertainty is not sufficient for the decision-maker to reach a decision, use the approximate initial evaluation to identify the most important sources of uncertainty, and consider quantifying them more rigorously. This can either be by using more formal methods of expert judgement, or by quantifying them by sensitivity analysis or probabilistic modelling. Revise the analysis and communicate it to the customer.

3 If the customer requires still more refinement of the analysis, repeat step 2 iteratively, progressively extending more rigorous methods to more and more of the uncertainties in order of their importance. Continue this until the customer is able to reach a decision.

Figure 8.B: Refining uncertainty analysis



An iterative approach for refining uncertainty analysis to the level required to support decision-making.

Communicating the outcome of the uncertainty analysis

8.43 The principal output of uncertainty analysis is information on the range of alternative outcomes and their likelihoods. It is important to communicate this in ways that are accessible

to commissioners and decision-makers and minimise the risk of misunderstanding. The analytical assurer may wish to insist this information goes alongside any communication of the analytical assurance, see paragraph 7.5.

8.44 A tiered approach to communication may be helpful; a concise headline conclusion in narrative form, accompanied by one or more levels of more detailed information which the recipient may consult if they wish.

8.45 If the likelihoods are quantified, such as by probabilistic modelling or expert judgement, the range of outcomes and their likelihoods can be presented in tabular form or as a probability distribution or bar graph.

8.46 It is important to provide an overall characterisation of the uncertainty, combining any that have been quantified by modelling with the contribution of any additional uncertainties that were evaluated outside the model. This may be achieved by first giving the estimates provided by modelling, and then indicating how much this is increased by uncertainties evaluated outside the model.

8.47 Depending on the complexity of the analysis, communication of results can be a challenging and may require separate approaches for different audiences. This will have to be determined on a case-by-case basis and if appropriate, seek assistance from communication specialists.

Part C

This part of the Aqua Book provides an overview of the accompanying resources that will help turn high-level guidance and principles into embedded practice.

Chapter 9 introduces the templates and checklists that have been developed. They can be tailored to complement existing business processes or to act as a starting point for the development of new business processes. Supplementary guidance on particular analysis problems will complement the Aqua Book.

9 Aqua Book resources

Box 9.A: Aqua Book resources – key points

The Aqua Book forms part of a suite of resources that will help analysts deliver quality analysis with analytical quality assurance. Additional resources include templates, checklists and specific guidance for specific analytical techniques and common areas of analysis.

Templates and checklists can be adapted to meet local business processes or specific analytical project needs.

Introduction

9.1 To support the Aqua Book, a series of templates, checklists and supporting guidance and more have been created and can be found alongside the Aqua Book on the civil service 'collaborate' workspace on analytical quality assurance.

Quality assurance resources

9.2 Templates and checklists are often used to facilitate the adherence to business processes. A suite of templates and checklists have been developed that cover topics such as scoping out analysis and recording assumptions.

9.3 Spreadsheets are commonly used for a variety of analysis. A spreadsheet template has been developed that helps structure project and analytical quality assurance information alongside clearly structured calculations.

9.4 A series of supporting resources delve deeper into the concepts of analytical quality assurance and uncertainty and what they mean in principle when implementing specific analytical techniques or addressing common analytical problems. The supporting resources are not intended to provide a detailed explanation of the theory and methodology behind each of the techniques, for which there will be numerous textbooks and training courses available. Instead they will provide a point of reference for analysts setting out on a project using the technique, providing guidance on:

- analytical quality assurance considerations
- common pitfalls to avoid
- project considerations and what additional information needs to be communicated to the analytical assurer and to the commissioner

10 Acknowledgements

10.1 The government's chief analysts would like to thank the cross-government quality-assurance working-group – chaired by Mike Marriott – for steering the development of the Aqua Book. In particular, thanks go to the chapter leads, contributors and editors:

- Jason Dewhurst
- Nick Harris
- Paul Glover
- Andy Hart
- Willem Roelofs
- James Palmer
- Ed Heardman
- Ian Rogers
- John Walls
- Alec Waterhouse
- Emma Frost
- Jon Tecwyn
- Andy Jones
- Hiroko Plant
- Adrian Jonas
- Kevin Reid

10.2 In addition, the chief analysts would like to thank those analysts from across government and from across the analytical professions who have provided input into the project.

10.3 The authors wish to thank the external reviewers from the private sector and academia for providing comments on this resource as it was developed. In addition, the authors would like to thank James Richardson for his invaluable insight.

HM Treasury contacts

This document can be downloaded from www.gov.uk

If you require this information in an alternative 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

Email: public.enquiries@hmtreasury.gsi.gov.uk

Verification and Validation for the AQuA Book

Paul Glover

DSTL/TR81015
01 September 2014

Dstl
Policy and Capability Studies
Portsdown West
Portsdown Hill Road
Fareham
P017 6AD

Release Conditions

This document has been prepared so as to openly share the Verification and Validation (V&V) considerations that underpin the V&V advice presented in the government Analytical Quality Assurance (AQuA) book, which is hosted on the Treasury web site.

Executive summary

This guidance is concerned with the Verification and Validation (V&V) that is necessary to deliver appropriate analytical quality in support of successful programmes and their constituent projects and activities. This guidance has been written to underpin the higher-level guidance presented in the Analytical Quality Assurance (AQuA) Book, which is hosted on the Treasury web site. The concept of V&V presented here is based upon an assessment of 'fitness for purpose' rather than seeking to accredit analytical approaches for a stated range of purposes, which is both bureaucratic in ethos and against the advice presented in the literature.

Key definitions:

- Validation – literally meaning to make valid, through the agreement of those judged competent to take such views. The central question that validation raises is the extent to which the right work is being engaged in, given the purpose and constraints placed upon that work. The key output from the validation process is a judgment, based on evidence, concerning the extent to which the work is 'fit for purpose';

- Verification – is concerned with the extent to which the work that has been agreed to is being done in the 'right' or 'accepted' way, given the 'art of the possible'. The key output from the verification process is a judgment, based on evidence, concerning the extent to which the agreed work has been conducted appropriately;

- Programme – a governance structure designed to co-ordinate, organise, direct and implement a portfolio of projects and activities that together achieve outcomes and realise benefits that are of strategic importance;

- Project – a governance structure created for the purpose of delivering one or more business products against an agreed business purpose;

- Activity – a specific piece of work that has been tasked in order to make an identified contribution to an agreed business purpose;

- Commissioner role – the person who commissions analysis for the purposes of a Programme, Project or Activity;

- Analyst role – a person tasked to conduct analysis on behalf of the commissioner;

- Analytical assurer role – a person tasked to provide analytical assurance of the work conducted by an analyst. For small rapid projects the person working as the analyst could also be fulfilling the analytical assurance role, although it is advisable to always have an independent person to provide the analytical assurance check upon the work.

The purpose of this guidance is to inform all members of a broader programme about what constitutes appropriate V&V for the work conducted. It is envisaged that this document will be of particular interest to:

- The commissioner of the analysis concerned with what constitutes appropriate V&V;
- The Senior Responsible Owner (SRO) tasked to produce an assurance statement for the work that has been undertaken;
- The design authority providing strategic level governance to a programme.
- The person working in the programme assurance role;
- The project manager responsible for day to day delivery;
- The person working in the in the project assurance role;
- The person working in the analytical assurer role;
- People working in the analyst role, concerning what is expected of them, with respect to the conduct of V&V in their work.

This advice:

- Considers what constitutes analytical quality and broadly sets out how analytical quality is achieved (see section 2);
- Presents a four-stage model for the conduct of V&V in analysis activities (see section 3);
- Identifies the V&V activity to be conducted in each stage of the work (see section 4);
- And closes with a consideration of common analytical pitfalls, the key responsibilities of each of the three main roles identified by the AQuA book (commissioner, analyst and analytical assurer) in overcoming them; and it also briefly examines the limits of what is knowable (see section 5).

Table of contents

List of figures

1 Introduction

This guidance is concerned with the Verification and Validation (V&V) that is necessary to deliver appropriate analytical quality in support of successful programmes and their constituent projects and activities. This guidance has been written to underpin the higher-level guidance presented in the Analytical Quality Assurance (AQuA) Book, which is hosted on the Treasury web site. The concept of V&V presented here is based upon an assessment of 'fitness for purpose' rather than seeking to accredit analytical approaches for a stated range of purposes, which is both bureaucratic in ethos and against the advice presented in the literature.

The purpose of this guidance is to inform all members of a broader programme about what constitutes appropriate V&V for the work conducted. A glossary of key terms is provided at the end of this paper. Terms in the following list that have not been included in the glossary have been drawn from managing successful programmes (MSP). It is envisaged that this document will be of particular interest to:

- The commissioner of the analysis concerned with what constitutes appropriate V&V;
- The Senior Responsible Owner (SRO) tasked to produce an assurance statement for the work that has been undertaken;
- The design authority providing strategic level governance to a programme;
- The person working in the programme assurance role;
- The project manager responsible for day to day delivery;
- The person working in the project assurance role;
- The person working in the analytical assurer role;
- People working in the analyst role, concerning what is expected of them with respect to the conduct of V&V in their work.

This advice:

- Considers what constitutes analytical quality and broadly sets out how analytical quality is achieved (see section 2);
- Presents a four-stage model for the conduct of V&V in analysis activities (see section 3);
- Identifies the V&V activity to be conducted in each stage of the work (see section 4);
- And closes with a consideration of common analytical pitfalls, the key responsibilities of each of the three main roles identified by the AQuA book

(commissioner, analyst and analytical assurer) in overcoming them; and it also briefly examines the limits of what is knowable (see section 5).

2 Purpose

The purpose of V&V is to ensure the quality of content of the analysis, the quality of the process by which it is produced and the quality of outcome that is achieved (Robinson (2002)). These aims can only be achieved if our concept of analytical quality embraces the whole span of the analysis process from the inception of the work through to initial delivery and then formal publication.

These three aspects of the quality of the work are delivered through the analyst, analytical assurance and commissioner roles working together in partnership. The reason for this is rooted in the nature of validation, which fundamentally is a shared collective judgment amongst key parties concerning what is jointly understood to be 'fit for purpose' in a given circumstance (for more information see Kleindorfer et al (1998)).

It is the responsibility of the analyst to focus upon quality of content, supported by the analytical assurer role. It is the responsibility of the analytical assurer role to deliver quality of process, supported by the analyst role. It is the responsibility of the commissioner of the analysis to deliver quality of outcome, supported by both the analyst and analytical assurer roles. Advice on the quality assurance issues to be managed and who should lead on them is presented in part B of the AQuA book, chapter 5, Verification and Validation (The AQuA Book (2014)).

The key V&V issue that the analyst should focus on initially is the conceptualisation of the work. This conceptualisation then frames the expectations against which the quality of their work can be assessed. While the detailed planning and conduct of that work clearly matter, errors are often traceable to the way in which the work was conceptualised. Errors in conceptualisation can be trapped through:

- Taking care to understand the benefits that are sought from the work that has been tasked and then explicitly considering how the work conducted is contributing to the intended goal at each stage of the analysis process;
- Periodically reflecting on how the framing for the analysis has been selected and thus the starting point from which the research was launched, its boundaries and the structural weaknesses of that perspective (Jackson (2003)). In particular:
    - Identify the foundational narrative upon which the research stands and consider if any of the claims made by that narrative appear extraordinary in the light of the emerging findings from the work. Claims that fundamentally shape the research and as a result of the work begin to appear extraordinary, on further investigation often are. Unpicking the beliefs which are causing the customer for the research to founder and thus request analytical support can be one of the most useful insights that an analyst can raise. A means of unpicking such beliefs, when the emerging evidence makes this appear to be necessary is to:

- Characterise the claims implicit in the narrative upon which the tasking for the analysis is founded;

- Take a Popperian approach to assessing these claims against the data (Popper (1972, 1979)), particularly with respect to perceived outliers in the data. The reason for 'attacking the outliers' in this way is that the reasons that the outliers do not fit the pattern implicit in the framing narrative helps to identify the causes of misalignment between the narrative and the issues that the customer is seeking to engage with;

- Formally identify the gap between the accepted narrative and the contradictory evidence;

- Posit less extraordinary explanations that embrace the previously excluded data;

- Submit these candidate explanations to progressively more demanding examination and thorough ongoing review until sufficient clarity emerges such that it is possible to propose a new framing for the research upon an evidential basis.

- Consider scoping and boundary issues and their consequences (the boundaries are there to make the work tenable given the time and cost available and to ensure that we minimise 'lost opportunity' from the deployment of staff and other resources). There follows a means of reflecting upon such boundary issues:

  - In order to reflect on the degree to which the defined breadth of the research is appropriate, more broadly frame a conceptual model and then consider what the implications are, including sensitivity analysis that could be conducted in order to work towards an understanding of the degree to which we need be concerned;

  - In order to reflect upon the degree to which the depth of the work is appropriate, consider the extent to which the findings from the components of the analysis align with the evidence for the current baseline of the system being examined;

  - In order to reflect upon the degree to which the granularity of the research is appropriate, consider the degree to which aspects of the emerging findings align with more detailed studies in the area.

The key V&V issue that the analytical assurer role should focus on is the credibility of the work. While the quality of the content of the work clearly matters, it is the quality of the process which is fundamental to its credibility. Robinson (2002) identified that 75% of the concerns of the recipients of analytical work typically relate to quality of process issues, while 25% of their concerns typically relate to quality of content

issues. This view was founded in Robinson's own research experience and for the purposes of the advice presented here and in the AQuA book was cross checked through a 'crowd sourcing' exercise across experienced analysts in the Civil Service. This process identified that 80% of the errors that occur in analysis are typically with respect to quality of process issues and 20% are typically with respect to quality of content issues. The specific errors identified through this process are addressed in part B of the AQuA book, chapter 5, Verification and Validation, in terms of the responsibility of each of the roles (analyst, analytical assurance and commissioner) to guard against these faults.

The key V&V issue that the commissioner role should focus on is the acceptability of the work. In particular, it is the responsibility of the commissioner role to understand the benefits that the analysis is seeking to support and in consequence who should be 'bought-in' concerning the process of V&V that is put in place. In doing this it is useful to remember that "essentially all models are wrong but some are useful" (Box and Draper (1987)). As a consequence of this, the commissioner of the analysis has a key role in shaping the work and in assisting in the interpretation of the results, from what is necessarily the simplified context of the research into the 'real world' context in which the intended benefits are sought (more detail concerning these responsibilities is presented in part B of the AQuA book, chapter 5, Verification and Validation).

The inter-relationship between these key aspects of quality is illustrated below in Figure 1.

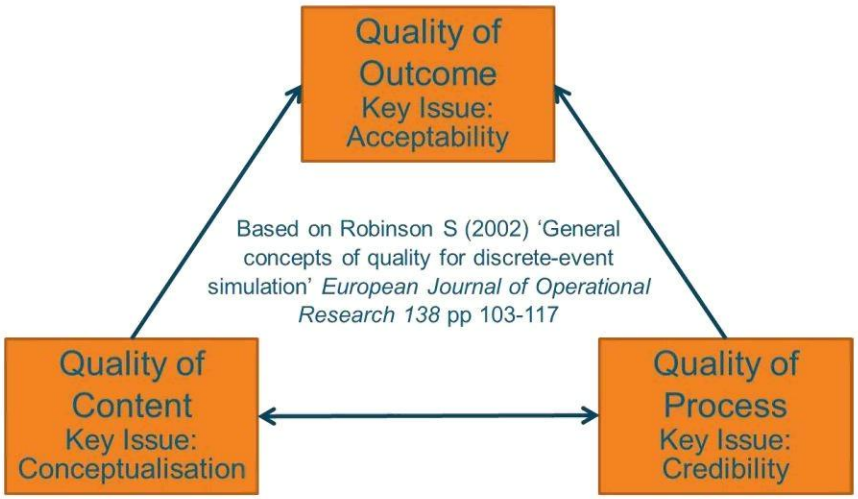


Figure 1: The quality triangle

3 Transparency of process

The AQuA book identifies four stages to the analysis process. Each stage in this process needs to be captured so as to enable the analysis to be delivered with appropriate transparency. Four forms of transparency have been identified in the construction of this advice, one for each stage of the work. The stages of the work are illustrated in Figure *2* and further described below:

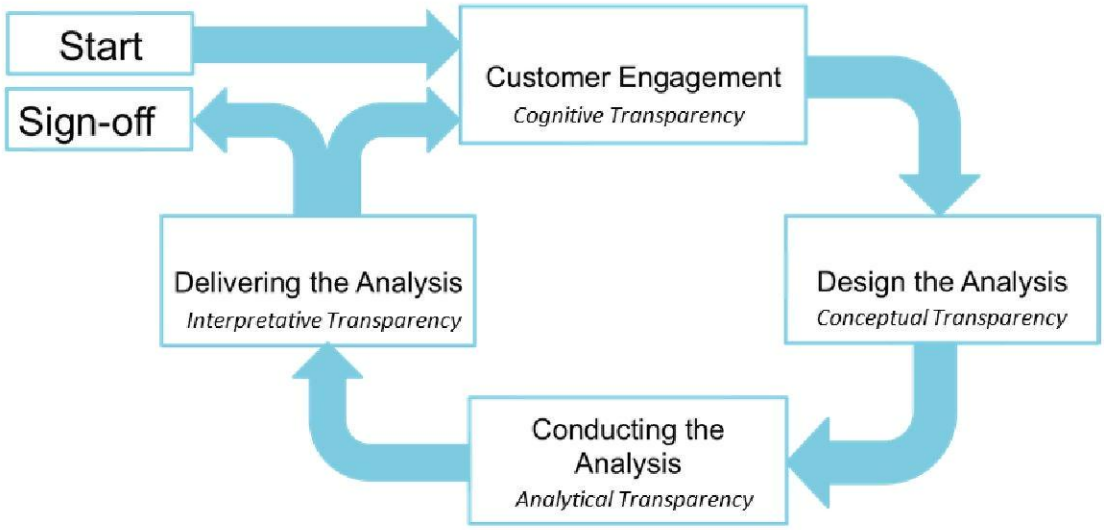


Figure 2: The stages of the work

- **Customer engagement:** The aim of customer engagement is to surface the purpose or purposes of the work and identify the benefits that the analysis is seeking to contribute to. In consequence it seeks to identify the breadth and depth of enquiry that is needed and the range of perspectives that are to be taken into account through an open dialogue which seeks to agree an appropriate balance between analytical tractability and appropriate constraints. Customer engagement is facilitated through cognitive transparency:

    - *Cognitive Transparency:* The purpose of cognitive transparency is to be clear about the benefits that the analysis is being commissioned to support, the assumptions upon which the analysis shall stand and the reasons for the selection of this view of the world. It is recommended that this understanding is formally recorded in an analytical estimate, although for very fast turnaround work it is acknowledged that it may be necessary to do this after the event. The analytical estimate is a living document, the purpose of which is to act as a vehicle to record the understanding derived from an open engagement with the commissioner of the research, in order to reach a view on the analytical tractability of the work, noting the limitations of the constraints within which that work would be required to proceed. The idea behind this approach is that reflecting back the analyst's

understanding to the commissioner role serves to identify and clear up any misunderstandings that may occur during the tasking process, producing a shared understanding of 'the art of the possible' given cost and time constraints. Since the analytical estimate seeks to surface the assumptions upon which the analysis is founded it also serves as a vehicle by which the evidence arising from the analysis can be considered, allowing the basis upon which the work has been constructed to be questioned if one or more key assumptions no longer appear to hold. The analytical estimate thus serves to ensure that as far as possible there are no surprises arising from the conduct of the work, but where such surprises do arise it provides a point of reference to help identify their source and provide a basis from which they can be managed.

- **Design the Analysis:** The design of the analysis should be firmly based upon the mutually agreed requirements identified through customer engagement. For this reason it has been recommended that the results of customer engagement are formally captured through an analytical estimate which provides a comparator against which the 'fitness for purpose' of the design can be assessed. Such assessment both enables the design to be tested for completeness and the coverage of the customer engagement to be reflected upon in the light of issues raised by the production of the design. The design of the analysis is facilitated through conceptual transparency:

    - *Conceptual Transparency:* The purpose of conceptual transparency is to be clear about the process by which the benefits for which the analysis that has been commissioned shall be obtained. It is recommended that this process is formally recorded, both in a concept of analysis and as appropriate in the documentation of the methods, techniques or models used or specifically developed for the work. The concept of analysis should clearly set out the issues to be addressed by the analysis and the means by which it is proposed that the intended work shall be achieved. Care should be taken to ensure that the concept of analysis addresses all of the requirements agreed within the analytical estimate. It is recommended that the concept of analysis is supported by documentation setting out the verification and validation status of all methods, techniques and models to be used. A way of providing this information is through a validation logbook or for smaller methods, techniques or models a validation logsheet. This verification and validation documentation should set out:
        - The purpose or known competence of the method, technique or model;
        - The history of usage in the relevant field of application;
        - Those currently known to competently deliver this capability (in full or in part with any known limitations on their current competence identified);
        - Evidence or reference to evidence concerning the verification of the approach against the requirements that it seeks to address;

        - Evidence of validation of the approach through inspection both of the approach and its results by relevant experts in the field and through comparison to evidence previously reported; and
        - Any known limitations of the approach.

    - As with the analytical estimate, the aim of the concept of analysis and the documentation for methods, techniques or models used (including formally recording their verification and validation status), is to ensure that there are no surprises arising from the conduct of the work as far as possible, but where such surprises do arise it provides a point of reference to help identify their source and provide a basis from which they can be managed.

- **Conducting the Analysis:** As far as possible the conduct of the analysis should follow the design set out in the concept of analysis. None the less it is recognised that information sources may have previously unrecognised limitations which will need to be managed in order to ensure proper delivery of the work. This will require consideration of uncertainties and dependencies in the data, with appropriate use of parametric variation in order to help bound the problem space. Thus the role of the analyst is: to ensure that the right inputs are introduced into the agreed analysis process in the right way; that any input error that does occur is sifted from the results set through validation checks; and that the recognised results conform to the process that was intended. The enactment of the analysis is facilitated through analytical transparency:

    - *Analytical Transparency:* The purpose of analytical transparency is to be clear about the process that it actually proved possible to conduct, given emergent limitations of process and technique. Central to this are considerations concerning the extent to which the research is: Repeatable; Independent; Grounded in Reality; Uncertainty Managed; and Robust (captured in the acronym RIGOUR) outlined in Chapter 5 of the AQuA book. It is recommended that the data used in the analysis is captured in a master data and assumptions list which is subject to validation through peer and expert review. The work itself should be captured in technical reporting which includes: the purpose of the work; the method; the results; any significant limitations or caveats associated with the conduct of the work; preliminary interpretation setting the results in the context of previous research and elucidating the mechanisms which both lead to conformity with previous research and those mechanisms which drive key differences.

- **Delivering the Analysis:** As identified in Chapter 5 of the AQuA book, commissioner, analyst and analytical assurer all have key roles in the delivery of the analysis. The reason for this is that the translation from the simplified representation of the world reflected in the analysis to the complexities and perturbations introduced by a more complex reality require a measure of interpretation in order to map the results to the context of their intended application. This translation process needs to be jointly owned, such that: the

analyst is content that the interpretation is a fair reflection of the meaning of the results from the process that was run; the analytical assurer is content in terms of the mapping to the extant literature (both that conformity has been achieved where that is appropriate and that new insights stand given proper considerations of process (structures that effect change, their levers and the functioning of their mechanisms (Harré R (1970)); and that the commissioner is satisfied with the work that has been engaged in, the process by which it was produced and how it relates to the context in which advice is required. Furthermore, once satisfied, it is the responsibility of the commissioner role to ensure that the analysis has the impact that it deserves (see Collins (2001)). The interpretation of the analysis is facilitated through interpretative transparency:

- *Interpretative Transparency:* The key to interpretative transparency is to be clear about the reasons for selecting the assumptions which were adopted and the alternatives that were considered in making the interpretation of the analysis that is offered. This enables the commissioner of the research and the customers to judge for themselves the extent to which they accept the findings of the work. It is imperative that the customer reporting of the analysis is tailored for each intended audience. It is recommended that the key findings are presented first along with any significant limitations or caveats associated with the work in order to assist the busy reader. Other information to include is the relationship between this analysis and previously reported findings (including appropriate referencing), the reasons both for similarities and key differences and the recommendations that in consequence are being made.

4 Stages of the work

The key purpose of the approach to verification and validation that is presented here is to enable the analyst to:

- Orient onto the current stage of the work and reflect on the purpose it is seeking to fulfil;
- Consider that purpose in light of the current plans and the assumptions and beliefs in which they have been founded;
- Provide guidance on the key considerations to resolve when engaging with the phenomena central to this stage of the work;
- Verify that they have done what they intended, capturing evidence to show the extent to which this has been achieved; and,
- Validate the product that they have generated against the requirements placed on them.

This process is illustrated below in Figure 3.

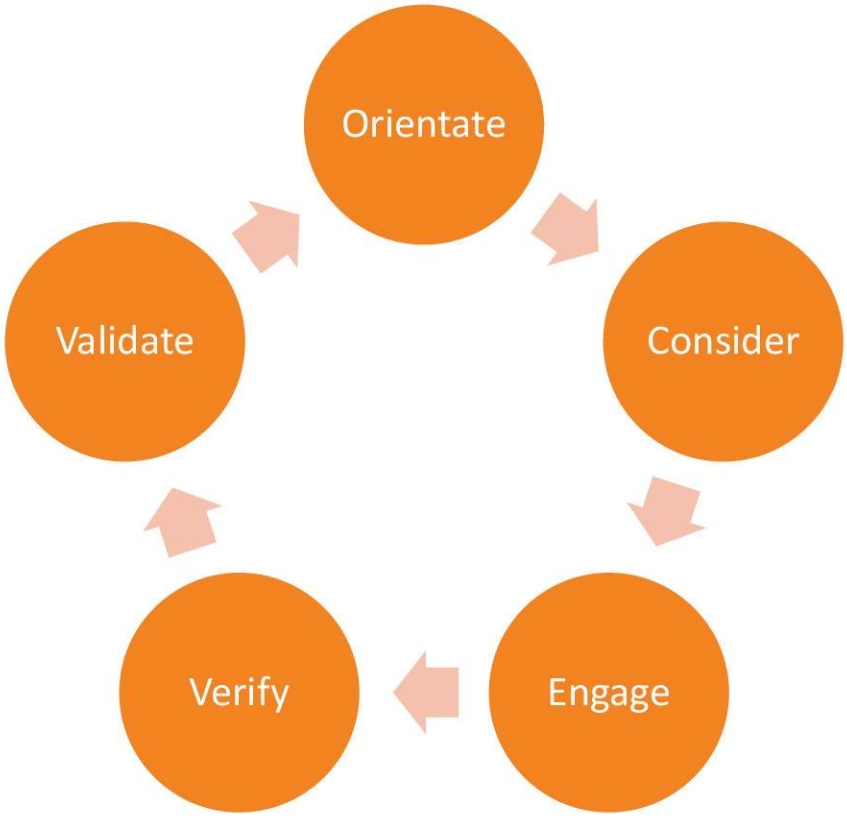


Figure 3: The verification and validation process

Orientate: The purpose of this first step is to clarify what has been asked of the analyst at this stage in the work and the role that this has in producing the benefits

that are sought. Pausing in this way serves to guard against rushing into the work, without first understanding the success criteria. The aim is to ensure that what the analyst sets out to do in 'best faith' is what the commissioner of the work would most wish of them, so as to produce work that can in principle be considered to be reliable. The evidence captured in the orientate step serves to frame the considerations made in the validate step (see below).

Consider: The second step is to pause again and consider if the plans that frame the envisaged approach to this phase of the work align so as to enable the realisation of the success criteria, identified in the orientate step (above). In considering these plans it is useful to reflect upon the extent to which the assumptions and beliefs upon which the current plans were founded continue to hold, given what has been learnt in the orientate step. Should potential merit in changing the plan be identified then this should be made known to the quality assurer role for validation of this view and with their agreement made known to the commissioner role with a view to authorising a variation to the plan. The evidence captured in the consider step serves to frame the assessments made in the engage and verify steps (below).

Engage: The third step is to engage with the phenomena which are the subject of the analysis at this stage. When engaging with the appropriate phenomena there are five validation criteria that need to be considered. Again there should be a pause before proceeding, to do a final conceptual check that the work as planned (captured in the consider step (above)), makes sense in terms of the questions that these criteria raise. Adequate transparency (for the purposes of the work being conducted), concerning how these criteria were addressed, provides a key part of the evidence needed for consideration in the validation step below. The criteria are as follows:

- *Reliability*: which considers the degree of alignment between what is studied in the analysis (in terms of breadth and depth), given the constraints upon the work, and the benefits that it is designed to provide;
- *Face Validity*: which considers the degree to which the stakeholders who prove key, consider there to be an adequate alignment between the characterisation of the issues being examined in the analysis and their understanding of the 'problem space'. Lack of alignment between key stakeholders' expectations concerning what is examined and the detail of the work leads to lack of confidence in the product of that work;
- *Criterion Validity*: which considers the detailed engagement with the phenomena being examined in the analysis and the extent to which the work actually engages with the phenomena that it claims to;
- *Construct Validity*: which considers the adequacy (for the purposes of this analysis) of the representation of how the phenomena being examined are structured, the key factors to which they respond and the mechanisms by which they do this;
- *Content Validity*: which considers the interpretative weight that the work proposed can bear, as a result of its breadth, depth and granularity. The aim is

to match the interpretative value that can be derived from the analysis to the need to produce insight in order to enable the delivery of benefits.

Verify: The fourth step is to verify the extent to which the work conducted aligns with the work that was planned and to capture evidence as to the extent to which this is the case. The evidence from the verify step is combined with the evidence from the previous steps in order to produce the evidence that is considered at the validate step (below). Reasons for variation from the plan can include 'human error', lack of some of the required data within the required timescales, or a need to make an ad hoc change to the analysis plan due to the emergence of additional constraints that were not anticipated at the time the analysis plan was constructed. Where variation from the plan has occurred, the analytical assurer role needs to consider if aspects of the work need to be repeated or if the differences can be adequately controlled and understood through appropriate parametric variation, which is then tasked to the analyst.

Validate: The validate step considers all of the evidence gathered in the previous steps conducted at this stage in the analysis process. It is recommended that this validation step is conducted as a gated review with one of four possible outcomes:

- The work is accepted and permission is given to proceed to the next phase (from customer engagement to designing the analysis; from designing the analysis to enacting the analysis; from enacting the analysis to delivering the analysis; and from delivering the analysis to formal publication);
- The work is accepted, but noted as having limited utility with respect to the benefits that were sought, with particular reservations noted. A decision is then made concerning how to proceed (the work could be re-shaped through returning to an earlier stage in the analytical process, the work could continue with noted caveats, or the work could be put on hold or stopped);
- The work is noted as still having potential, but only if identified issues can be resolved, with a decision made on granting permission to examine these issues further;
- The work is rejected, with reservations noted and a decision made concerning how these issues should be actioned.

5 Common Analytical Pitfalls

There are three main levels of potential pitfall that those involved in the analytical process need to be aware of:

- The framing of the analysis;
- The choice of approach within the selected frame;
- The engagement with the detail of the work.

It follows that each of the three analysis roles (commissioner, analyst and analytical assurer) has a role to play in mitigating and managing the potential pitfalls that may become apparent during the course of the work. In particular the commissioner role should lead on the framing of the analysis, the analytical assurer role should lead on the choice of approach within the agreed frame and the analyst role should lead on engagement with the work, using the approach that has been chosen. The conduct of work at each level is subject to review by the person responsible for the shaping decision at the level above.

It should further be noted that an apparent oddity identified at one level in this hierarchy can point to a problem in the level above. As such, those responsible for each level of potential pitfall should remain open to evidence both from the level above and the level below of a need to reconsider their approach. Each level of pitfall is further discussed below.

Framing: The fundamental decision made with respect to any analysis is how it is framed in order to give access to the benefits sought by the commissioner role. It is for this reason that the commissioner of the analysis owns this decision, but is also the reason why the commissioner needs to remain open to evidence from the conduct of the work that a different framing could realise these intended benefits more easily.

The following mistakes concerning the framing of the work have been identified by Salt (2008); while this paper specifically focuses upon simulation its findings can be generalised to other forms of analysis. The mistakes that have been identified are:

- Assuming that as a perfectly accurate external observer the commissioner role can accurately specify the programme of research that is required. Reasons why this is not the case are expanded upon in Jackson (2003). Instead the commissioner role owns the need for the research, expressed through the framing in order to realise the benefits that are sought in consequence of the analysis. It is the analytical assurer role that owns the programme of work to realise this aim and the analyst role that owns the work to realise the programme that has been agreed;
- Assuming that additional detail in the research necessarily delivers additional benefit. It can simply cause the work to take longer with no appreciable improvement in the quality of outcome to be derived from the

work, although considerations of face validity (see section 4 above) can sometimes over rule this;

- Assuming that using more frames of reference for the analysis necessarily produces a more useful understanding. While the use of different frames can help to 'triangulate' onto an understanding of a phenomenon where there are concerns about the veracity of any given method the use of additional frames should not be engaged in needlessly;

- Assuming that dynamic systems can adequately be analysed through static models. Senge P (1990) identifies that: "The real leverage in most management situations lies in understanding dynamic complexity, not detail complexity". This insight is discussed in greater detail in Georgiou (2007);

- Assuming that since analysis has been conducted the findings can simply be accepted as a 'fact'. It is for this reason that the commissioner of the research has an important role in the interpretation and promulgation of the work. Furthermore, it is important that the reporting of the analysis reveals not only what is known but also areas of uncertainty, Jackson (2003);

- Assuming that analysis can accurately predict. While analysis can be indicative, given stated uncertainty bounds, such as the weather forecast, there are generally too many factors in operation for the future state of a system to be predicted with certainty.

Choice of approach: The choice of approach to the analysis is dependent on the frame, including issues of timeliness and allowable resource. It is the responsibility of the analytical assurer role to assess the 'fitness for purpose' of the work against the requirements of the frame within which that work has been set and to consider if the frame chosen is the most effective way to unlock the intended benefits of the work. If the analytical assurer believes that the intended benefits the commissioner seeks could better be reached through a different framing of the problem space then they should make this known along with the evidence which leads them to believe this. The key pitfall with respect to choice of approach is:

- Assuming that since an analytical approach has been successfully used in a similar context in the past that such an approach should be mandated. It is the responsibility of the analytical assurer role to advise on the choice of analytical approach or approaches, given the benefits that are sought and the limitations of cost and resource that are available.

Engagement with the work: While those commissioning analysis and working in the analytical assurer role are drawing on their experience in order to shape the work, it is the analyst who is engaging with the details of the evidence that indicates how well those shaping decisions have been made. The following types of pitfall have been identified:

- A key pitfall in conducting analysis is the degree of self-awareness of the analyst with respect to the extent to which their prior beliefs may be driving the way in which they conceive, conceptualise, analyse and subsequently

interpret their work. The commissioner and analytical assurer both have roles in helping the analyst to overcome any such prior views. In particular the following mechanisms assist the analyst to address such limitations: the framing of the analytical question; the identification of the stakeholder group for the analyst to engage with; and the verification and validation processes (described above in section 4) assisted through transparency of process (described above in section 3). Other approaches that can help to ensure due RIGOUR in the conduct of the work (see section 3 above) are described in Petty et al (2012). One of these approaches is the use of a reflexive journal in which the analyst takes time to consider the shaping effect of their beliefs upon the analytical process, the extent to which emergent evidence may indicate a need to modify such beliefs and the means by which this emergent view could be verified and validated. Georgiou I (2007) identifies that it is not possible to simultaneously conduct analysis and consider how the emergent meaning of the research may suggest a need to re-frame the work, either in terms of the choice of approach or the overall framing of the work. Instead specific timeouts are required in which to use the reflexive journal. An example of the use of a reflexive journal while conducting analysis is provided by Boulton (2011). The analytical assurer role is there to provide the analyst with a sounding board for emergent views deriving from the use of the reflexive journal.

- There are also three well-known philosophical problems which have the potential to limit insight delivered by analysis (DiFate (2007)), in particular:
    - *The Ravens Paradox* which illustrates the limitations of 'bootstrapping' through logical argument, since almost anything can be asserted to be evidence of a pre-held view through such argument DiFate (2007). The use of *criterion validity* (see section 4) protects against the vices of the Raven's Paradox;
    - *The Grue Paradox* which seeks to assert either that the nature of the way things interact has changed or is about to change and hence seeks to shape the analysis through greatly limiting the data set that is drawn upon or demands that the analysis is founded upon a theoretical posit that reflects the assertion which has been made DiFate (2007). The use of *construct validity* (see section 4) protects against the vices of the Grue Paradox, particularly if an analytical approach which allows the examination of dynamic complexity is used (Harré (1970), Senge (1990) and Georgiou I (2007));
    - *Underdetermination of Theory by Evidence* which leaves the analyst with no sound way of distinguishing between the veracity of a set of different hypotheses concerning a phenomenon, all of which could be valid given the evidence currently available DiFate (2007). Such problems have for example bedevilled archaeological interpretation (Hodder (1986, 1991, 2003)). Where this problem occurs it is vital to be clear about the range of possible interpretations that can be placed upon the evidence available (Jackson (2003)).

Bibliography

- *'The AQuA Book: Analytical quality assurance'* June 2014
- Boulton J G (2011) *'The Complexity Turn: Narrative, Science and Utility'* A thesis submitted for the degree of Master of Philosophy University of Bath School of Management October 2011
- Box G Draper N (1987) '*Empirical Model Building and Response Surfaces*' John Wiley & Sons Chichester
- Collins J C (2001) 'Good to Great: Why some companies make the leap … and others don't' William Collins
- DiFate V (2007) 'Evidence' *Internet encyclopaedia of philosophy: A Peer-Reviewed Academic Resource* http://www.iep.utm.edu/evidence accessed 28th May 2014
- Georgiou I (2007) '*Thinking Through Systems Thinking*' Routledge London
- Harré R (1970) '*The Principles of Scientific Thinking*' MacMillan & Co Ltd London
- Hodder I (1986) '*Reading the Past. Current approaches to interpretation in archaeology'* (revised 1991 and, with Scott Huston, 2003) Cambridge University Press
- Jackson M C (2003) *'Systems thinking: Creative holism for managers'* Wiley Chichester
- Kleindorfer G O'Neil L Ganeshan R (1998) 'Validation in Simulation: Various Positions in the Philosophy of Science' *Management Science* Vol. 44, No. 8, dated August 1998 pp 1087-1099
- Petty J N Thomson O P Stew G (2012) 'Ready for a Paradigm shift? Part 2: Introducing qualitative research methodologies and methods' *Manual Therapy* 17 (2012) pp 378-384
- Popper K (1972) '*Objective Knowledge: An Evolutionary Approach*' Revised (1979) Oxford Clarendon Press
- Robinson S (2002) 'General concepts of quality for Discrete Event simulation' *European Journal of Operational Research* 138 pp 103-117
- Salt J D (2008) 'The seven habits of highly defective simulation projects' *Journal of Simulation* 2 pp 155-161 November 2008
- Senge P M (1990) *'The fifth discipline'* Doubleday

Glossary

Activity

An activity is a specific piece of work that has been tasked in order to make an identified contribution to an agreed business purpose.

Analyst role

The analyst is a person tasked to conduct analysis on behalf of the commissioner.

Analytical assurer role

The person tasked to provide analytical assurance of the work conducted by an analyst. For small rapid projects the person working as the analyst could also be fulfilling the analytical assurance role, although it is advisable to always have an independent person to provide the analytical assurance check upon the work.

Commissioner role

The commissioner is the person who commissions analysis for the purposes of a Programme, Project or Activity.

Programme

A programme is a governance structure designed to co-ordinate, organise, direct and implement a portfolio of projects and activities that together achieve outcomes and realise benefits that are of strategic importance.

Project

A project is a governance structure created for the purpose of delivering one or more business products against an agreed business purpose.

Validation

Validation literally means to make valid, through the agreement of those judged competent to take such views. The central question that validation raises is the extent to which the right work is being engaged in, given the purpose and constraints placed upon that work. The key output from the validation process is a judgment, based on evidence, concerning the extent to which the work is 'fit for purpose'.

Verification

Verification is concerned with the extent to which the work that has been agreed to is being done in the 'right' or 'accepted' way, given the 'art of the possible'. The key output from the verification process is a judgment, based on evidence, concerning the extent to which the agreed work has been conducted appropriately.

List of abbreviations

AQuA	Analytical Quality Assurance
RIGOUR	Repeatable; Independent; Grounded in Reality; Uncertainty Managed; and Robust.
SRO	Senior Responsible Owner
V&V	Verification and Validation

Initial distribution

1.	KIS	Dstl	Electronic

Report documentation page v4.0

* Denotes a mandatory field

1a.	**Report number:** *	DSTL/TR81015	1b.	**Version number:**	1.1
2.	**Date of publication:**	01 September 2014	3.	**Number of pages:**	i + 20
4a.	**Report UK protective marking:** *	OFFICIAL			
4b.	**Report national caveats:** *	NONE			
4c.	**Report descriptor:** *	NONE			
5a.	**Title:** * Verification and Validation for the AQuA Book				
5b.	**Title UK protective marking:** *	OFFICIAL			
5c.	**Title national caveats:** *	NONE			
5d.	**Title descriptor:** *	NONE			
6a.	**Alternative title:**				
6b.	**Alternative title UK protective marking:**				
6c.	**Alternative title national caveats:**				
6d.	**Alternative title descriptor:**	NONE			
7.	**Authors:** * Paul Glover				
8.	**Name and address of publisher:** * Dstl Policy and Capability Studies Portsdown West Portsdown Hill Road Fareham P017 6AD		9.	**Name and address of sponsor:**	
10.	**Sponsor contract:**	CSA Research Programme			
11.	**Dstl project number:**	707214			
12.	**Work package**	0004			
13.	**Other report numbers:**				
14a.	**Contract start date:**	01 Apr 2014	14b.	**Contract end date:**	31 May 2018
15a.	**IPR:** *	CROWN COPYRIGHT			
15b.	**Patents:**	NO			
15c.	**Application number:**				

Please note: Unclassified, Restricted and Confidential markings can only be used where the report is prepared on behalf of an international defence organisation and the appropriate prefix (e.g. NATO) included in the marking.

* Denotes a mandatory field

16a.	**Abstract:** * This report supplements the Analytical Quality Assurance (AQuA) book hosted on the Treasury web site, specifically providing additional advice concerning: what constitutes analytical quality; a four stage model for the conduct of V&V in support of analysis activities; the verification and validation activity to be conducted in each stage of the work; and a consideration of common analytical pitfalls.	
16b.	**Abstract UK protective marking:** *	OFFICIAL
16c.	**Abstract national caveats:** *	NONE
16d.	**Abstract descriptor:** *	NONE
17.	**Keywords: Advice, Analytical Quality, AQuA book, V&V Verification, Validation**	

18.	**Report announcement and availability** *		
	Title and abstract	Full document	
18a.	☒	☒	UK MOD has unlimited distribution rights
18b.	☐	☐	UK MOD has no rights of distribution
18c.	☐	☐	Can be distributed to UK MOD and its agencies
18d.	☐	☐	Can be distributed to all UK government departments
18e.	☐	☐	Can be distributed to all UK defence contractors
18f.	☐	☐	Can be distributed to all foreign government departments
18g.	**Additional announcement:**		
18h.	**Additional availability:**		
18i.	**Release authority role:** *		

Please note: Unclassified, Restricted and Confidential markings can only be used where the report is prepared on behalf of an international defence organisation and the appropriate prefix (e.g. NATO) included in the marking.

THIS PAGE INTENTIONALLY LEFT BLANK


Ministry
of Defence