Actuaries Institute Australia

COMMUNICATION, MODELLING AND PROFESSIONALISM

MODULE 6: MODELLING

Education Program
ACTUARIES INSTITUTE

Education that Counts

Actuary Program

Module 6

MODELLING

Education that Counts

Table of contents

6. Modelling

This module covers the following learning objectives:

Item	Unit/Key Performance Objective/Learning Objective
2	**Construct an effective model**
2.1	**Design a model, considering its purpose, features required and restrictions**
2.1.1	Design a model flowchart
2.2	**Create a clear structure for a model**
2.2.1	Explain the key elements of a good model structure
2.3	**Check a model using a variety of techniques, such as spot and high-level checking, sensitivity testing and technical review**
2.4	**Document a model, including its purpose, data, key assumptions and methodology used, checks performed, outputs, sensitivities and limitations**
2.5	**Communicate model outcomes**
2.5.1	Identify key information required by the target audience
2.5.2	Explain a model's outcomes and their potential impact on the target audience
2.5.3	Communicate the uncertainty, reliances and limitations associated with the model's outcomes
2.6	**Explain the need to update a model over time**

6.1. Introduction

A model is a simplified mathematical representation of reality. A model is only a 'representation' of reality because no matter how complex it is, it will never perfectly mimic reality. This is why scientist George Box famously said:

> *Essentially, all models are wrong, but some are useful*[1].

Video 6.1 helps to demonstrate this concept of representing reality by discussing a model representation of Earth.

Video 6.1 – What is a model?

https://www.youtube.com/watch?v=OKA4_J5yeoU (5 mins)

Record your video notes here

As Video 6.1 highlights, there can be a temptation to make a model as complex as possible, to more closely represent the complexity that is reality. However, a more complex model can be more error prone and won't necessarily better mimic reality. Occam's razor, also known as the law of simplicity, dictates that a theory (or model) should provide the simplest possible (viable) explanation for a phenomenon. That is, adding more and more complexity to a model won't necessarily make it better.

[1] Box, G.E.P & Draper, N.R, (1987), Empirical Model-Building and Response Surfaces, John Wiley & Sons.

No matter how complex or simple a chosen model is, it needs to be calibrated to the real world, by comparing its outputs to those observed in reality. This concept will be explored further in Section 6.6.

So why do we need to build models? Models are used by actuaries to help decision makers in a range of ways such as by:

- testing a hypothesis (e.g. 'Only a small percentage of workers' compensation injuries happen on the way to work');
- predicting the future (such as the number and size of life insurance claims over the next 10 years); and
- understanding the impact of key variables on a modelled outcome (such as understanding how the liability valuation differs depending on the discount rate selected).

Some models are built for a specific, one-off task and have a very short shelf life. For example, an actuarial student may build a model to compare several potential employers they are considering working for. Once they have selected a position with their employer of choice, they may never need to use their model again.

Most models, however, are designed to be reused, sometimes over many years. For these models, continual updating and improvement of the model are important to ensure it remains relevant for its purpose and that it continues to provide a good representation of reality.

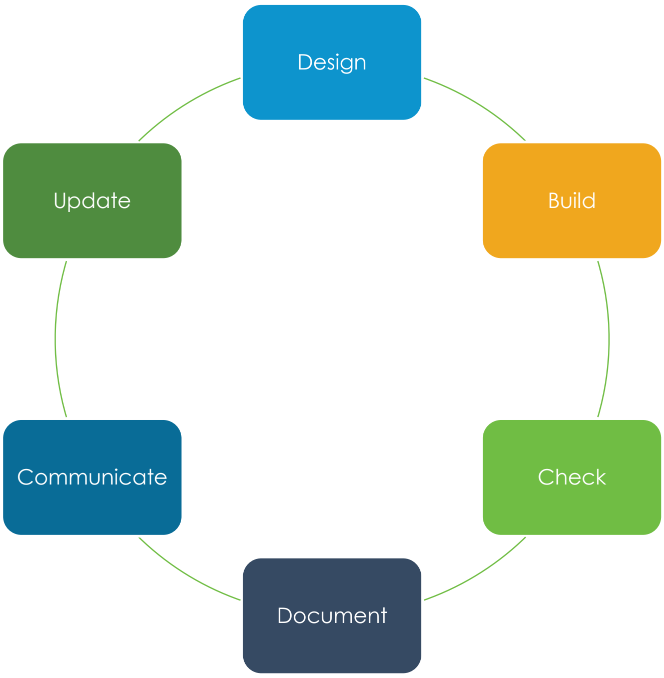
Different types of models are used by actuaries for different purposes and in different contexts. For example, the types of models used by actuaries working in life insurance, general insurance or retirement products can differ, as can those that are used by actuaries working in other fields such as banking, risk management or data analytics. Models unique to each of these fields of practice will be covered in relevant Fellowship subjects.

The principles introduced in this module are equally valid for all types of models.

6.2. Model construction process

The construction of a model is an iterative process involving the steps set out in Figure 6.1.

Figure 6.1 – Model construction process



Each of the steps shown in Figure 6.1 is discussed in separate sections of this module.

This module briefly revisits some of the modelling concepts introduced in the Actuarial Control Cycle subjects[2]. It then expands on these concepts. The topics of data collection, preparation and manipulation are covered in the Actuarial Control Cycle and Data Analytics Principles subjects. They are not covered in this subject. Model choice and assumption setting for different purposes will be covered in the relevant Fellowship subjects.

[2] See, for example, Bellis, C, Lyon, R, Klugman, S & Shepherd, J, (2010). 'Understanding Actuarial Management: the actuarial control cycle', Chapter 9, pp. 227–265, The Institute of Actuaries of Australia and the Society of Actuaries.

6.3. Model risk

Common modelling related problems that arise in business include:

- poor data entry controls, leading to low quality of data;
- inadequate access controls, resulting in unauthorised users accessing and making changes to data or models;
- inexperienced and/or overly confident model developers, leading to poor model design and modelling mistakes;
- lack of adequate processes for handling and storing sensitive data, resulting in privacy breaches;
- absence of back-up, leading to loss of work and unnecessary delays in rebuilding data sets or models;
- acceptance of results without understanding the underlying business rules and calculations, resulting in inappropriate conclusions being drawn from models;
- no or minimal model documentation, making it difficult to check, run or update models;
- minor tweaking of formulas is undertaken without knowing the flow-on effects and associated risks, leading to modelling errors when changes are introduced;
- poor or non-existent model maintenance and improvement processes, resulting in models quickly becoming out-of-date and out-of-sync with business requirements;
- ever-changing business drivers or requirements sometimes make models no longer fit for purpose but they are still used as there is no alternative; and
- no understanding of the business implications and impact of a model when there is a single person who has sole knowledge of how and why the model was developed, leading to unintended business consequences when models are updated.

Having effective procedures and policies in place for each step of the model construction process can avoid many of the risks outlined above. Entity-wide strategies that can be adopted to address such risks include:

- adopting better modelling standards (see Section 6.5);
- putting in place effective auditing and assurance arrangements (see Section 6.6);
- ensuring there are appropriate accountability and responsibility for the model (see Section 6.6); and
- having clear sight of how models are kept up-to-date (see Section 6.9).

It is important to ensure that such strategies are well communicated and understood throughout the entity. This will increase the likelihood that they will be adopted, as intended, by staff across all levels of the organisation.

6.4. Designing a model

Designing a model is a critical first step in constructing an effective model.

It can be tempting to skip the model design step and jump straight into building a model. However, this often leads to an unstructured model that is error prone, hard for someone other than the original builder to understand and unsuitable for its intended purpose.

For example, imagine your manager asks you to predict the number of new members joining your retirement scheme over the next three years. Eager to please her, you quickly build a spreadsheet that forecasts total member numbers for each of the next three financial years. After you show your manager your forecasts, she explains that she needs these numbers for the administration team to help them determine their seasonal staffing needs. They need to know the number of new members each month (rather than annually) and by various member categories. Frustrated, you rebuild your model to accommodate these requirements.

This illustrates the importance of considering the following when designing a model:

- purpose;
- features required; and
- restrictions.

6.4.1. Purpose

The purpose of a model will be determined by its business context, target audience and problems to be solved. In the retirement scheme example given above, the purpose of the model was to help the administration team (the target audience) determine their staffing needs. This purpose arose from the problem that staffing needs vary by season, making it hard for the administration manager to work out how many staff were needed at different times of the year.

When considering the purpose of the model, it is helpful to decide, up-front, how you will communicate the results of the model and to whom. This will help to inform the features required of the model, as discussed in the next section.

6.4.2. Features required

A model's purpose provides insights into the features required of the model, such as its:

- outputs;
- flexibility;
- transparency; and
- credibility.

There are many forms of actuarial model **outputs**. Choices include:

- summarised v detailed;
- grouped v individual policy or customer records;
- present value v cash flow projections;
- point estimates v probability distributions; and
- monthly v annual projections.

The form of output chosen will be influenced by both the purpose of the model and any restrictions on the model, such as available input, time or cost restrictions, as discussed below.

In the retirement scheme example from above, if the model's purpose had been to provide a high-level performance forecast for the scheme's Board of Trustees, a forecast of the total number of members likely to join each year could have been adequate. However, as the administration team required help with seasonal staffing needs, more detailed monthly projections by membership category were required.

If a model is to be used in an ongoing capacity, it may need to include **flexibility** to change its inputs, assumptions and outputs over time. For example, the retirement scheme's staffing model may need the flexibility to incorporate new membership categories as they are created. It is therefore important to discuss a model's flexibility (and other) requirements with the model's target audience.

The **transparency** and **credibility** requirements of a model are also influenced by its purpose. Some models are built as quick 'back of the envelope' calculations, for example to test an initial hypothesis before deciding whether to conduct further analysis. Such models might be very rough and not particularly transparent or detailed. This might be justified if they will be replaced by more transparent and credible models once a project is given the green light to proceed.

However, while you may not think that anyone will ever need to look at or rely on your model, it is important to make such limitations of your model clear in case someone does try to use it in the future. Documentation of models, including their limitations, is discussed in Section 6.7.

6.4.3. Restrictions

When designing a model, it is important to understand any restrictions that will be imposed on your model's development. These restrictions might include:

- financial budget;
- time;
- operational;
- available inputs; and
- existing models.

Model construction incurs a range of costs related to data collection, processing, storage, security and privacy, hardware and software, as well as resources required to design and build the model, estimate its parameters, test and validate model outputs and communicate its results. An understanding of the **financial budget** available for each of these construction components will influence the design of the model. For example, if the budget is very limited, the model is likely to take a much simpler form. A larger budget may allow complexities to be built into the model. However, even with an unlimited budget, the benefits of adding model complexities must be weighed against the risks of doing so (see the discussion about Occam's razor in Section 6.1).

The amount of **time** available to construct a model also needs to be considered when deciding on the complexity and features of the model. There are situations when a model needs to be constructed in less time than you would ideally like to spend in doing so. Sometimes you need to query these time restrictions and, at a minimum, clearly explain the limitations that your model will have because of these time restrictions.

Operational restrictions include the computer systems and software that are available and the expertise of staff who will need to access and possibly update the model.

The quantity, quality, detail and source of **available inputs** influence how to design a model to meet its purpose. If an abundance of high quality, easily accessible data is available, you will generally have more flexibility to design a model to best meet its intended purpose. However, in some actuarial work this may not be the case. You may be required to design and build a model with very limited and unreliable data, possibly resulting in a less detailed model, with outputs less credible than you would like. In this situation, it is important to test the impact of any model limitations, document the limitations and their impact and outline ways the model might be improved in the future.

In many types of actuarial work, you are likely to inherit **existing models** built either by yourself or others. Existing models include models that are typically used by others in your industry and adopted as standard practice. These existing models can also create restrictions on your model design.

If the business is very familiar and comfortable with existing models, it might be hard to convince people to accept a new type of model. A new model can initially be more time consuming and costly to build. This might lead users to continue to use an existing model because they can't afford the time to invest in building a new model. A new model also creates a new view of 'reality', which might be difficult to explain to others and have them accept the model's results.

For example, spreadsheet-based models are used in many actuarial organisations. While spreadsheet models can be very useful, there may be alternative models available that would better meet the model's purpose. However, organisations may be reluctant to move away from a spreadsheet model due to:

- difficulties in making system changes to accommodate new types of models;
- analytical staff not being willing or able to undertake the necessary training to learn a new software package or programming language; and
- senior management having more experience and comfort with spreadsheet models.

It is therefore often important to 'sell' the benefits of a new model to its users to overcome some of these issues in transitioning away from an existing model.

Despite the challenges listed above, you should not use an existing model simply because that is what everyone has always used in the past. As new software and computer technology emerge, it is particularly important for the actuarial profession to innovate by looking for ways to work more effectively and efficiently, even if this means having to overcome some of the challenges of moving away from existing models.

6.4.4. Designing a model flowchart

Once a good understanding of a model's purpose, requirements and restrictions has been obtained, the model can be designed.

It is possible to design a model using only text as a description of its features and processes. However, a model's design can often be communicated more effectively through a flowchart. A flowchart can better show a model's logic rather than just relying on a textual description of an algorithm.

A flowchart is a graphical representation of the sequences of steps and decisions required to perform a process. A flowchart uses defined diagram shapes that each represent a type of activity. Each step is contained within a single diagram shape. Directional lines capture the sequencing or logic of steps.

Many types of diagram shapes can be used in a flowchart. Table 6.1 lists five common flowchart symbols[3].

[3] There are commercially available programs that enable a user to create complicated flowcharts. This module only uses five common flowchart symbols available in the Microsoft Word package.

Table 6.1 – Flowchart symbols[4]

Symbol	Meaning
Terminator	This represents the start or end of a process.
Process	This represents an operation. The statement of the process may be at a high level in the flowchart. The detail behind the process will need to be written down elsewhere.
Data	This represents a flow of information into, or out of, the overall process.
Document	This indicates information that can be read by people, such as printed output.
Decision	This represents a decision point between two or more paths in a flowchart, often in the form of a true or false statement with two directional arrows.

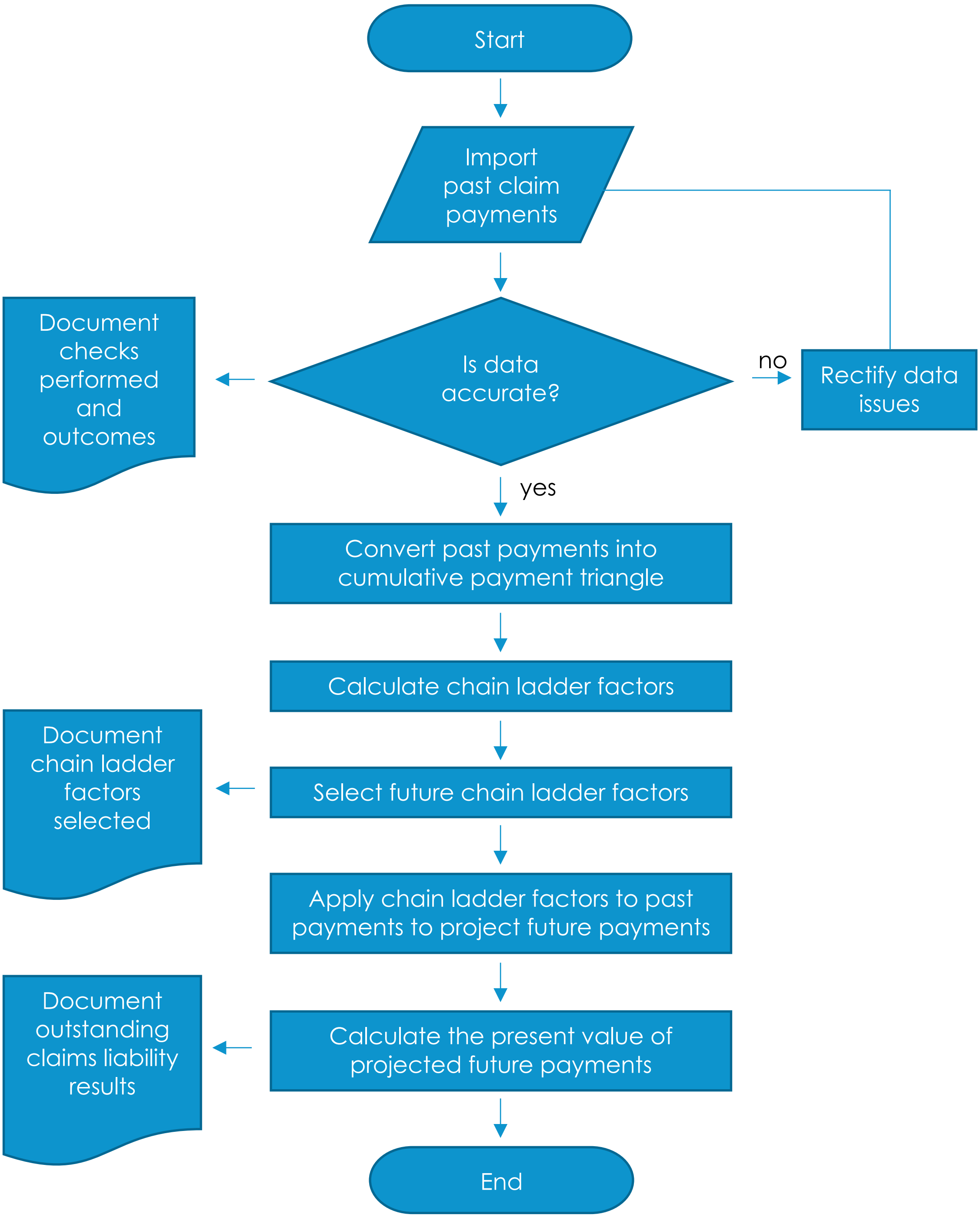
For example, suppose you have been asked to calculate the outstanding claims liability for a general insurance portfolio using a simple chain ladder model. Assume you have already determined the model's:

- purpose: such as understanding the types of cash flows to be discounted and what the outstanding claims liability will be used for (e.g. performance management v reporting);
- required features: such as the discount rate that should be used; and
- restrictions: such as the input available and how much time you have to build the model.

[4] Source: https://www.conceptdraw.com/How-To-Guide/flow-chart-symbols

Figure 6.2 shows a flowchart that depicts the liability model's proposed design.

Figure 6.2 – Example model flowchart



The flowchart in Figure 6.2 is included only as an example to show how a model's design can be described visually. In practice, this flowchart would be more complex[5], including steps such as inflating past and projected payments, comparing the chain ladder factors to those selected at the last valuation, checking the calculated liability for reasonableness and documenting not just the outcome, but also elements such as the modelling approach and assumptions adopted and the uncertainty inherent in the calculation.

Flowcharts are useful for ensuring a model is well structured. This is because you are forced to think through the structure of your model when designing the flowchart. The process of having to provide a solid description of a complex process, so that you can explain it to others, often helps to clarify your thinking about the model's design and identify any issues before moving to the model building step.

A visual diagram such as a flowchart is also usually an easier way to describe complex processes to others. This helps others understand the logic of your model and often leads to more fruitful discussions in relation to identifying and resolving issues with the model design.

Model flowcharts are also an effective tool to use when documenting the model for future reference (see Section 6.7).

[5] For some examples of different and more complex flowchart designs, see: https://www.gliffy.com/blog/common-flowchart-templates-flowcharting-examples-gliffy

6.5. Building a model

Model building (the second step in the model construction process) is itself an iterative process that involves the following stages:

- collect and analyse available data;
- form a hypothesis;
- build and test the model;
- refine, rebuild and retest the model; and
- finalise the model.

The model building process is covered in the Actuarial Control Cycle and Data Analytics Principles subjects. Model building specific to different practice areas will also be taught throughout the Fellowship subjects.

This module will instead focus on ways to effectively structure or lay out a model.

6.5.1. Model structure

Having a clear structure for a model is important because it makes it easier:

- for someone else to understand what your model is doing;
- to check or update the model;
- to find the source of any errors in a model when it is not behaving as expected; and
- to document the model.

These benefits of having a clear structure for a model apply both in the workplace and throughout your actuarial studies. For example, when you are asked to build a model in an examination, the marker will be able to give you more marks if they can easily see and understand what each component of your model is doing.

Elements of good model structure include:

- simplicity;
- logical flow;
- clear labelling; and
- tidiness.

These elements, which are each discussed further below, apply equally to models built in a spreadsheet, programming language or elsewhere.

Simplicity

Many actuarial models are very complex. They might draw on a range of different data inputs, require a large number of assumptions, include many dependencies in the calculations and provide an output that is quite detailed.

However, no matter how complex the overall model is, it can usually be broken down into much simpler individual components. Using a model design flowchart, as described in Section 6.4.4, may help to determine what these individual components should be. For example, the chain ladder model described in Figure 6.2 may be broken down into the following key components:

- past claim payments;
- checks on past payments;
- assumptions (e.g. discount and inflation rates);
- cumulative payment triangle;
- chain ladder factors;
- projected future payments; and
- calculated liability.

Breaking a complex model down into simpler individual components has two key advantages: it makes it easier to check a model by reviewing each component separately and it allows individual components of a model to be reused either within the model or in another model.

In a spreadsheet, you can break a model down by having a separate colour-coded tab or worksheet for elements such as data, assumptions, calculations and outputs. Further, you might have a separate worksheet for each main step in the calculation or engine of the model. For example, in the chain ladder model above you could have separate worksheets for the cumulative payment triangle, chain ladder factors and projected future payments.

In a programming language, such as R or Python, you can break a complex model down by writing separate code for each component and then calling on each of these pieces of code in one summary code. Alternatively, you can use functions to perform certain processes in your model and call on those functions as you need them.

When using functions in code, each function should do only one thing and not a range of different things. This assists in making your model as simple as possible and easy to understand.

Another method for simplifying a model is to limit the number of arguments within a function. For example, suppose you are writing a function within a model to calculate the insurance premium payable by an individual. The following variables are needed to calculate the premium:

- date of birth;
- date policy was purchased;
- current date;
- gender;
- smoking status; and
- occupation.

A single function that calculates the individual's current age and policy duration and then uses these calculations and the other variables to look up the correct premium will be lengthy, complex and not particularly transparent or easy to follow. This will make the function difficult to check. Instead, it would be better to write one function that calculates the individual's age, another that calculates their policy duration and then a third function that combines these and the other factors to look up the correct premium.

Taking this example further, it would make sense to include each of the variables in the same row of a table that contains the calculated premium. This is preferred to having each input variable in separate places in the model, again making it more difficult to understand what the function is doing.

An instance of complexity arises when a spreadsheet model links together many different individual spreadsheets. This can cause errors if one of the spreadsheets changes name or is moved, or if the links in a spreadsheet are not automatically updated when a spreadsheet it depends on is updated. It is better to avoid having to link spreadsheets together if possible. If this cannot be avoided, the links should be as simple as possible, pointing to single cells in the related spreadsheet rather than performing complex operations on multiple cells in a corresponding spreadsheet.

Logical flow

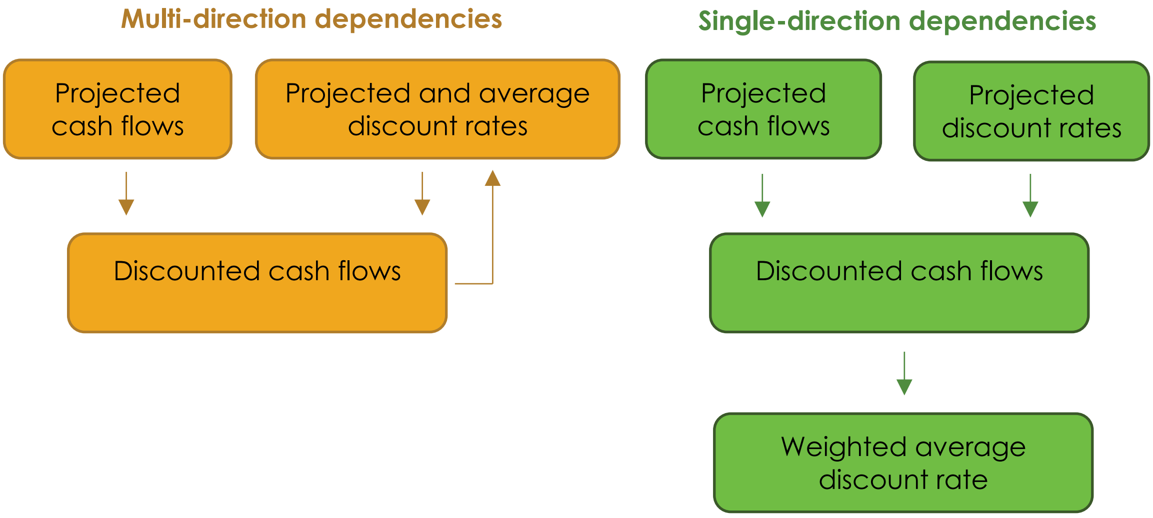
The logical flow of a model should be evident from the model's design flowchart. A model should be structured so that each calculation is set out in a logical order, as shown in the flowchart.

For example, a logical flow when using a spreadsheet would set out the spreadsheet tabs in the order in which they are used by the model. Tab 1 should do the first step in the calculation, such as importing data. Tab 2 should perform the next step, perhaps checking the data. Tab 3 should perform the next step such as manipulating the data and so on. In this way, as you work from top to bottom in the flowchart, you move from left to right in the spreadsheet.

This also applies to models built in a programming language. Someone reading the model's code should be able to start at the top of the code and read through to the bottom, with each step building on the step performed above it, rather than jumping back and forth between different parts of the code.

One feature of a model with poor logic flow is one that has multi-direction dependencies. For example, consider the orange multi-direction dependencies model depicted on the left of Figure 6.3. In this model, the 'Discounted cash flows' sub-model depends on the projected discount rates from the 'Projected and average discount rates' sub-model. The 'Projected and average discount rates' sub-model in turn depends on the 'Discounted cash flows' sub-model to work out the cashflows-weighted average discount rate. This two-way dependency adds unnecessary complexity to the model. Instead, a separate sub-model should be built to calculate the weighted average discount rate, as shown in the green model on the right of Figure 6.3, resulting in only single-direction dependencies.

Figure 6.3 – Multi- v single-direction dependencies



Clear labelling

Clear labelling is another feature of models that are well structured and easy to follow. Clear labelling involves:

- using meaningful variable names and headings;
- using consistent definitions;
- including lots of comments; and
- using standard file naming conventions.

Using meaningful variable names and headings makes it easy for someone to look at your model and immediately work out what it is doing, even without reference to your model's design flowchart, comments or documentation. To make variable names and headings more meaningful, avoid the use of abbreviations. For example, if a column in your spreadsheet model includes data about a worker's occupation type, a meaningful heading for that column would be 'occupation type'. An abbreviated heading such as 'occ_typ' is not meaningful as users of your model may not instantly work out what such an abbreviation stands for.

It is also important to **use consistent definitions** throughout your model. This includes only using variable names or headings for one meaning throughout the model. For example, if you are calculating the average age of males in one part of your code and you assign the variable the name 'average_age', you should not then use this same variable name in a later section of the code that calculates the average age of females. Instead, one variable should be called, for example, 'average_age_male' and the other 'average_age_female'.

Including a lot of comments in a model also makes it easier to follow. Generally, comments written as text in a separate cell in a spreadsheet, next to the part of the model being described, will be more noticeable than text in a hidden comment attached to a cell. An abundance of comments should also be included in coding. This helps someone reviewing your code understand what each section is doing, as well as providing a visible break between different sections of code.

Comments or documentation within a model must be in a place where people will read them. For example, comments that are hidden at the very bottom of a worksheet, or the very bottom or to the side of a piece of code, are unlikely to be seen by the model's users. You should therefore carefully consider where and how such commentary should be incorporated into the model.

Using standard file naming conventions helps when automating tasks, avoiding and spotting typographical errors in code and when trying to find the files you are looking for. Standard file naming conventions might include a relevant and meaningful title for the file, the date it was created and a version control number.

Tidiness

In addition to adopting the other elements of well-structured models (simplicity, logical flow and clear labelling), a tidy model is one that has no redundant or unnecessary sections ('zombie code').

When building a model in a spreadsheet, using code or otherwise, you can sometimes build a section and then later decide that it is no longer needed or there is a better way to build it. This occurs a lot more frequently when the model is built without first properly considering its design (see Section 6.4).

It can be tempting to leave this zombie code within the model, perhaps greying it out in a spreadsheet or commenting it out in code, in case you think you might need to refer to it for another purpose or because you don't have time to go back and remove it. However, leaving the zombie code within your model adds to its complexity and makes it more difficult to follow and review.

It is better practice to either delete these redundant sections or save them in a separate spreadsheet or piece of code.

Video 6.2 discusses zombie code along with six other tips for giving your code good structure. Many of these tips also apply to models built in spreadsheets or elsewhere.

Video 6.2 – Seven tips to improve your programming skills

https://www.youtube.com/watch?v=vfqvzgkYYF0	(12 mins)

Record your video notes here

A tidy model also avoids using too many columns in a spreadsheet or too many characters in each line of code. Too many columns in a spreadsheet might mean that your sub-model is becoming too complex and could be simplified by breaking it down into smaller components. Too many characters in a line of code might indicate that your coding language could be simplified or more elegantly expressed. Some programmers contend that a line of code should not exceed 80 characters[6]. However, the ideal maximum number of characters in a line of code is a source of some debate[7].

It is useful to check whether the organisation you work for has modelling templates or guidelines for spreadsheets or code that specify rules for spreadsheet or code structure, formats and styles, naming principles, labelling and other elements of a model's structure. If such templates or guidelines don't already exist, it might be useful to develop one yourself and share it with your colleagues[8].

[6] See, for example, https://nickjanetakis.com/blog/80-characters-per-line-is-a-standard-worth-sticking-to-even-today and https://katafrakt.me/2017/09/16/80-characters-line-length-limit/.
[7] See the comments section in https://katafrakt.me/2017/09/16/80-characters-line-length-limit/.
[8] An example of best practice modelling guidelines that you could use as a starting point for developing your own guidelines can be found here: https://www.modano.com/guidelines/modano.

Exercise 6.1

Review the two spreadsheets provided in the learning management system for the module. One is labelled '… calcs.xls', the other is labelled '… projected mortality rates Dec 19.xls'.

How do these two spreadsheets compare? What features make '… projected mortality rates Dec 19.xls' more structured and easier to follow than '… calcs.xls'?

6.6. Checking a model

Errors can arise in a model from sources such as:

- errors or omissions in the raw data;
- errors that are introduced when importing or manipulating the data;
- lack of business knowledge or technical expertise of the person building the model;
- mistakes in the selection or application of assumptions;
- errors in formulas;
- coding errors (e.g. mistyped words or wrong syntax);
- mistakes introduced when updating a model; and
- computer glitches, such as when a link in a spreadsheet doesn't properly update.

It is therefore critical that adequate audit and quality assurance processes are in place to review a model before relying on its results or output. At a high level, this means that models used in a business should be subject to appropriate sign-off processes. These should involve having the model developer sign-off that they are comfortable with the accuracy of the model, as well as requiring sign-off by an independent technical reviewer of the model.

Such a sign-off process ensures that each model has been independently reviewed, but it also ensures that the model developer accepts responsibility and accountability for their model. While this may seem obvious, it is not always automatically the case. In business, it is common for actuaries to inherit models that others have built before them. In such situations, it is not reasonable for the inheriting actuary to assume the model is working or to assume that responsibility and accountability for the model still lie with the original builder of the model. Instead, if you have inherited a model from someone else, you need to take on responsibility and accountability for the model, such that you are comfortable to sign-off that it is working as expected and that its outputs can be relied on.

Both the model developer and the technical reviewer should review the model by performing high-level checks on the outputs as well as more detailed checking of the working of the model. These checks should include:

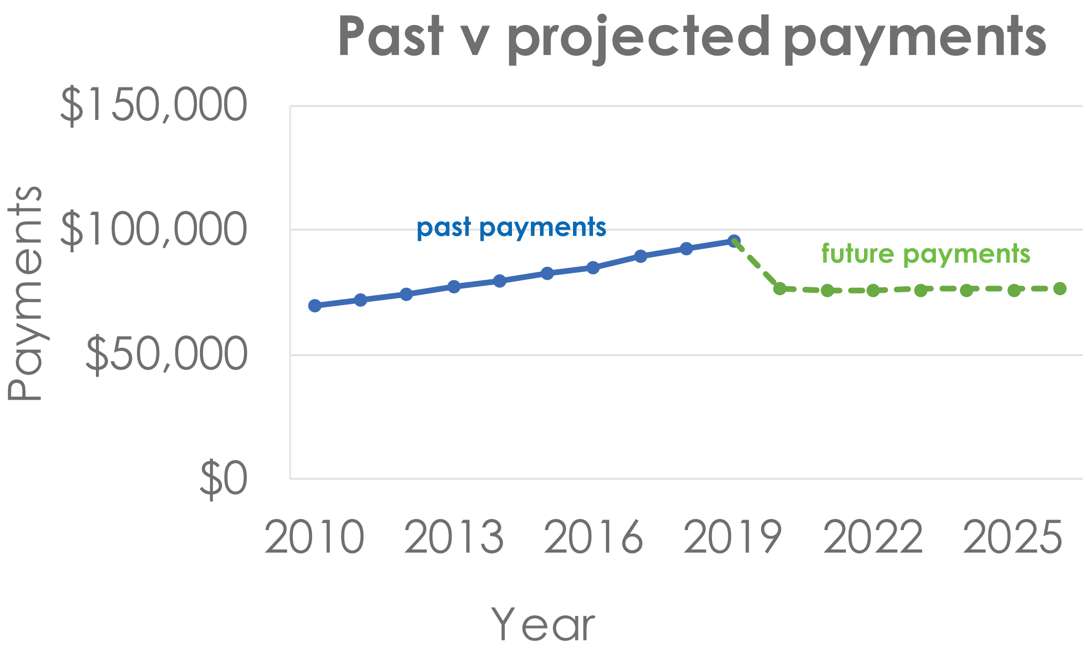
- sense checking the reasonableness of the results;
- comparisons of past and projected outcomes;
- back testing the model against historical data;
- out-of-sample testing;
- comparisons to previous model results (if available);
- comparisons to other sources;
- spot checking;
- checking outputs from each sub-component of a model; and
- sensitivity testing.

High-level sense checking involves, as the name suggests, performing an assessment of whether, at a high level, the results of the model 'makes sense'. It is the most important check to perform. A model might pass all the other checks applied to it, but if its results fail the sense check, then it is likely that your other checks missed something important. It is essential that you are comfortable with the reasonableness of your model's output before you can rely on it or communicate it to others.

High-level sense checking naturally involves the application of judgement, given the business context for the model and your experience in building and using similar models. For example, if you have built a model to estimate discount rates that should be applied in a discounted cash flow calculation, the resulting rates should be compared to those you have used in other similar exercises. If the model suggests rates of around 10%, but you know from experience that current discount rates used for similar purposes are around 2%–4%, this is likely to indicate that there is something wrong with your model.

If the model provides a projection of future outcomes (such as claim numbers, sales volumes or payments), a high-level check should include a visual **comparison of past outcomes against those projected**, such as is shown in Figure 6.4.

Figure 6.4 – Example of high-level sense checking



Education that Counts

The past payments in Figure 6.4 are shown by the solid blue line. They have been trending upwards over the last nine years. The future payments, shown by the dotted green line, are considerably lower than the most recent past payments and are relatively flat. At face value, this comparison suggests that the model results may not be reasonable[9]. In this case, detailed checking must be performed on the model to understand why the projected payments are so much lower (and flatter) than the past payments.

Historical data can also be used to **back test the model**. This involves running past data through the model, holding back the last few years of data. Projected outcomes from the model can then be compared to the known withheld recent data to check for the goodness of fit against recent outcomes. Alternatively, **out-of-sample testing** can similarly be used to test a model by withholding a random sample of the input data and comparing the model outcomes against the known out-of-sample data.

While comparisons of outcomes to past experience, back testing and out-of-sample testing can provide useful checks on a model, it is important to avoid over-fitting a model to historical data. This can result in a model being very good at replicating past outcomes but poor at predicting unknown future outcomes. This concept is explained in more detail in the Data Analytics Principles subject[10].

High-level checking should also include **making comparisons of model results to those from previous models**, where these are available. For example, if a model is estimating premiums for different policyholder cohorts, how do these estimated premiums compare to ones produced by the model at the same time last year, both at an aggregate level and for individual cohorts? Differences from the previous model may be reasonable, as experience may have been different from that previously anticipated. However, such differences need to be both understood (to check the model is working as expected) and explained to interested stakeholders, such as the marketing team in this example (see Section 6.8).

---

[9] The model may have considered other information not evident in the past data. For example, if the past payments represent insurance claim payments, upcoming legislative change, alterations to policy conditions or expected new business sales may provide legitimate reasons as to why future payments will be at a different level to past payments.
[10] The article found here: https://elitedatascience.com/overfitting-in-machine-learning also provides a very useful introduction to the concept of over-fitting.

Model results should also be checked for reasonableness by **comparing them to other sources**. These sources might be the company's accounts or annual report, or other sources such as the company's website. For example, if you have built a model to perform an expense allocation to determine the bank fees to charge for different products, total expenses coming out of the model should be compared to those in the bank's annual accounts. As another example, if you have built a model that forecasts membership numbers for a sporting organisation, the forecasts should be compared to the membership numbers shown in the organisation's annual report. Again, any differences need to be understood and explained to key stakeholders.

Spot checking involves randomly checking the working within each individual section of a model. In a spreadsheet, this might involve randomly checking a number of individual cells within each part of the model. This can be a useful checking technique for a model built in a spreadsheet where the same formula is copied down a column for many rows, making it inefficient to check every cell in the spreadsheet.

Outputs from each sub-component of a model should also be checked for high-level reasonableness. These sub-components can be checked using the techniques described above for checking the overall results of a model. This checking process is made much easier when a model is well structured, containing simple, clearly separated individual sections of a whole model.

Sensitivity testing is another useful technique. This involves checking whether the model responds sensibly to changes in assumptions or retesting the model to check how stable its outcomes are when using different data inputs.

Checking a model is a critical step in the construction process. It should not be left until the very end, when there may be limited time remaining to perform all the required checks or to rectify any issues uncovered during the checking process. Instead, a well-structured model that is built using logical, simplified sub-components should have interim checks performed along the way as it is being built.

To this end, you may sometimes face requests or demands from others to give them your model results before the model has been adequately checked. However, since you are responsible and accountable for your model's results, you need to be prepared to push back on releasing the results of your model before it is ready. This can feel difficult to do in practice, but is an important component of meeting your Competence and Care obligations under the Actuaries Institute's Code of Conduct (see Module 8 – Professionalism in practice).

A final important check on your model is whether it is useful. Business contexts and problems are ever evolving. It is important to keep up with these changes and consider whether your model continues to be useful for the context in which it is to be used. This will be discussed further in Section 6.9.

6.7. Documenting a model

Documentation of a model is another important step in the model construction process. It is a step that is sometimes overlooked either because time has run out or its importance is not well understand. However, documentation of a model is important for a number of reasons.

Documenting a model can help you check its logic. Often issues in a model's logic only become clear once you try to explain how it works to someone else. Model documentation can be used by the technical reviewer to check whether the model is doing what it was intended to do. Documentation also provides an audit trail for a model, making sure there is adequate information so that the model can stand up to scrutiny by others such as a peer reviewer or auditor. Finally, documentation of a model will help you and others know how to run the model and update it in the future.

A model's documentation should include features such as the[11]:

- purpose of the model;
- types, sources, volume and treatment of data;
- methodology adopted;
- key assumptions used;
- results;
- checks performed;
- skill sets of model creators and experts providing input;
- degree of confidence or uncertainty around the results, including sensitivities; and
- limitations of the model and how these influence the usefulness of the model.

When documenting a model's methodology and assumptions, justification should be provided for the method and assumptions used. This justification should include any options considered. It should also include any differences to past methods and assumptions adopted, including reasons for these differences.

The theoretically 'best approach' should be discussed, including how this compares to the actual approach adopted. For example, a client or regulator may require a certain approach to be adopted, even if it is not theoretically the 'best approach'. In this situation, the model's documentation should discuss how the use of this prescribed approach has impacted the model's results.

Ideally, a model should be documented such that an independent reviewer can understand the basis of the model and make a conclusion about its appropriateness and reliability for the purpose it is being used for.

[11] This list and subsequent commentary are adapted from the UK Institute and Faculty of Actuaries 'Risk alert: Disclosure of information relating to models': https://www.actuaries.org.uk/system/files/field/document/Risk%20alert-%20Disclosure%20of%20information%20relating%20to%20models.pdf and the UK Financial Reporting Council's Technical Actuarial Standard 100: Principles for Technical Actuarial Work: https://www.frc.org.uk/getattachment/b8d05ac7-2953-4248-90ae-685f9bcd95bd/TAS-100-Principles-for-Technical-Actuarial-Work-Dec-2016.pdf.

Actuarial techniques are often complex and can be accompanied by specific, technical language. However, documentation of a model should use plain language wherever possible. This makes your documentation readable for many different audiences. Use of plain language can also be a good test of how well you understand the model. If you can't explain the model in plain language, there is a good chance that you do not fully understand it yourself.

6.8. Communicating model results

The documentation for a model should contain a lot of information about the model and its results. However, the audience for such documentation might be restricted, for example, to the model developer and reviewer.

Other audiences, such as senior management in a company, will be interested in the model's results but will not necessarily be interested in looking through the detailed model documentation.

When communicating the results of a model, the Communication principle under the Actuaries Institute's Code of Conduct (see Module 8) requires you to have regard to the:

- intended audience;
- purpose of the communication;
- significance of the communication to its intended audience;
- potential need of the audience for further explanation; and
- capacity in which you are acting.

These requirements are discussed more generally in Module 8 (Professionalism 2).

The need to consider both the purpose of the communication and its intended audience was discussed in detail in Module 2 (Effective communication). These considerations will help to answer the following questions in relation to communicating a model's results:

- What decisions might the audience be relying on the model's results for?
- What key information is required that will influence the audience's decisions?
- What outcome is the audience expecting?
- How are your model outcomes different from those expected and why?

Thinking about the **decisions your audience might make** based on the model's results will help you determine what key information is appropriate to include in a communication. It will also help you avoid adding unnecessary complexity that will detract from the key information your audience is interested in.

The **outcomes the audience is expecting** might be based on previous model results, business targets or draft model results that you previously communicated to them. Understanding an audience's expectations will enable you to anticipate questions they might have about the model's results. This will allow you to respond to such questions in your communication, providing reassurance for the audience that the results are reasonable.

The amount of detail to be included should differ depending on the audience for the communication. For example, if a model's results are being presented to the company board, the audience will likely be interested in a higher-level summary of the results. On the other hand, if the audience is your technical reviewer, they may want you to talk through the model and its results in detail before they start checking it.

Depending on the amount of detail to be included in the communication, the following basic elements of your model should be communicated:

- purpose;
- approach;
- results;
- interpretation/conclusions;
- reliances and limitations; and
- next steps.

The UK Financial Reporting Council's Technical Actuarial Standard 100: Principles for Technical Actuarial Work outlines a number of specific provisions that should be followed when communicating the results of a model[12]. These provisions include:

- making communications clear, comprehensive and comprehensible, allowing users to make informed decisions;
- stating a model's users, scope and purpose and who commissioned the work;
- ensuring a communication's style, structure and content is suited to the skills, understanding and technical knowledge of the audience;
- providing any material information in a permanent form (rather than just orally);
- comparing the model results with the results of the previous modelling, together with an explanation of any differences (if they exist);
- outlining any material uncertainty in the model's results;
- stating the nature and significance of each material risk or uncertainty faced in relation to the modelled results;
- indicating any material changes or events that have occurred since the model results became available;
- providing clarification or information if the communication appears to have been misunderstood by its audience; and
- excluding immaterial information if it obscures material information in the communication.

[12] https://www.frc.org.uk/getattachment/b8d05ac7-2953-4248-90ae-685f9bcd95bd/TAS-100-Principles-for-Technical-Actuarial-Work-Dec-2016.pdf. Note that the provisions in this technical actuarial standard have been reworded in this section of the module as the technical standard relates to actuarial advice more generally, not just specifically to modelling.

6.8.1. Communicating uncertainty

When communicating the uncertainty of a model's results, it is particularly important to use plain language your audience will understand. For example, when communicating the results of a liability valuation, rather than saying 'the provision has a 75% probability of sufficiency', it might be more effective to show the range of possible outcomes on a graph. You could then say that liabilities at different levels give a different level of confidence that you will have enough money set aside to meet your company's future insurance obligations. You could then point to different levels of liability in the graph and explain the estimated confidence you have that each level will be enough to meet your company's future obligations.

Video 6.3 discusses some simple, practical tips for communicating uncertainty to an audience.

Video 6.3 – Communicating uncertainty

https://www.youtube.com/watch?v=fm3GJvbl8iQ	(2 mins)

Record your video notes here

While Video 6.3 focuses on uncertainty in weather forecasting, these principles also apply when communicating the uncertainty inherent in actuarial modelling. These principles include:

- using simple, consistent terminology;
- painting pictures with your words;
- using analogies; and
- reiterating key messages in several different ways.

When communicating uncertainty to your audience, it can be useful to incorporate different visual aids to relate this message. This could include the use of graphs, pictures, videos or other visual aids. The use of visual aids in communicating a message is discussed in Module 4 (Presentation).

Scenario testing is another useful way to communicate the uncertainty in a model's results. This should include scenarios that are meaningful to the audience. For example, rather than showing a scenario of a '10% increase in expense assumptions', show a scenario of 'expenses increasing by $5m due to a replacement of the current IT system' or 'expenses dropping due to reducing the support staff head count by 20 people'. These latter two scenarios use language that means something to its audience, making the scenarios more relatable for them. This allows the audience to form a judgement about how likely each scenario is to occur.

6.9. Updating a model

Most models are designed to be reused. At the same time, business contexts and problems continue to evolve over time. This, therefore, means that models that are not kept up-to-date can quickly become redundant.

Models need to be updated to reflect:

- changes in the data inputs to the model;
- assumptions that are more realistic in the current environment;
- changes in the business requirements, such as needing more detail or longer projection periods; or
- emerging methods or techniques that are expected to give more reliable modelling results.

Before updating a model, you need to refer to the model's documentation so that you fully understand what each aspect of the model is doing. This will help avoid unintended consequences during the update.

Section 6.6 talked about the importance of taking responsibility and ownership for a model. This also applies when you are updating a model that was created by someone else. For example, it is not professionally acceptable to change one or two assumptions in a model and not check the full impact of those assumption changes or be able to explain how the result has been impacted by the changes. When you make such updates to an existing model, *you* are responsible for the accuracy and reliability of the model's new results, not the original model builder.

6.10. Key learning points

- A model is a simplified mathematical representation of reality.
- Models are used by actuaries to help decision makers test hypotheses, predict the future and understand the impact of key variables on an outcome.
- Model construction steps include designing, building, checking, documenting, communicating and updating the model.
- A model's design is influenced by its purpose, required features and restrictions.
- Model flowcharts help to summarise and effectively communicate a model's design.
- Having a clear structure for a model makes it easier for someone else to understand the model, check or update it, find the source of any errors and document the model.
- Features of a clear model structure include simplicity, logical flow, clear labelling and tidiness.
- Errors can arise in a model from a number of sources. It is therefore critical that a model is appropriately checked and signed-off before relying on its results.
- It is important that a model's developer, and anyone who subsequently updates the model, takes responsibility and is accountable for the model.
- Clear documentation of a model is important to check its logic and provide information to the model's reviewer, as well as an audit trail and information for people who may need to update the model in the future.
- The communication of a model's results needs to consider both the purpose of the communication and the intended audience. This includes considering the reliance that an audience might place on the model's results and the results the audience might expect from the model.
- Communication of a model's results should be clear, comprehensive and comprehensible for the intended audience.
- Models need to be updated to reflect changing business and environmental circumstances.
- Care must be taken not to cause any unintended consequences when updating a model.

6.11. Answers to exercises

The following table compares each spreadsheet according to the four elements of a structured spreadsheet outlined in Section 6.5.1.

Feature	'… calcs.xls'	'… projected mortality rates Dec 19.xls'
Simplicity	• contains all the data and calculations in one tab, making it look like a very complicated model • the factors used in the projection of mortality rates (e.g. the current mortality rates and future improvement factors) are in columns that are not close to the actual calculation, making it harder to check these inputs to the calculation	• the data, assumptions, assumption selections and calculations have been broken up into smaller, simpler steps within the model, with each step appearing in a separate tab of the worksheet • the projection of mortality rates is done in a separate tab for males and females. Each of these tabs brings in the necessary inputs to the calculation (e.g. the current mortality rates and the future improvement factors)
Logical flow	• there is a logical flow to this model, although it is difficult to see the flow due to the lack of clear labelling and because the entire model is included in one worksheet	• it is much easier to see the logical flow in this spreadsheet due to the model information set out in the Notes tab, the clear labelling of different sections of the model and the use of different tabs for different elements of the model, including useful section dividers such as 'Output ==>', 'Data ==>' and 'Future Improvements ==>' • the 'Output ==>' tab comes before the 'Data ==>' and 'Future Improvements ==>' tabs. Although the model flowchart would show a different order (with the 'Output' coming last), it can be useful to have model output available for users towards the front of a spreadsheet so that it can be found easily

Feature	'… calcs.xls'	'… projected mortality rates Dec 19.xls'
Clear labelling	• there is very little labelling in this spreadsheet • the tab is just called 'Sheet 1' • the headings of each column are difficult to interpret (e.g. 'x', 'M', '15:14') • there are no comments in the spreadsheet explaining what the various model components are doing • the file name '…calcs.xls' doesn't tell you anything about the model • there is no summary information about the purpose of the model or when it was created and reviewed • the graphs in the sheet are not labelled, so it is difficult to work out what they are showing	• each tab in the spreadsheet has a relevant name such as 'life tables', 'projection – M' • the headings of each column are meaningful (e.g. 'Year 0', 'q_x from Standard Australian Life Tables (VIC)') • there are a lot of worksheet headings and comments explaining what the various model components are doing • the file name '…projected mortality rates Dec 19.xls' relates to the purpose of the model • the 'notes' and 'review' tabs give you background information about the model, including that the review of the model found a material issue with the data source being out-of-date • the graphs have been clearly labelled, making it easier to understand what they were used for
Tidiness	• the numbers in the worksheet are formatted inconsistently, with different numbers of decimals used in different cells • different fonts are used for different parts of the model • all calculations are in one worksheet, resulting in too many columns being used in one worksheet • there appears to be some zombie code in rows 105:210. Without any labelling or commentary, it is difficult to tell whether this section of the worksheet is still needed	• the numbers are all formatted consistently • the same font is used throughout • there aren't too many columns used in each tab: all the columns fit within a normal computer screen width

About the Actuaries Institute

The Actuaries Institute is the sole professional body for actuaries in Australia. The Institute provides expert comment on public policy issues where there is uncertainty of future financial outcomes. Actuaries have a reputation for a high level of technical financial skills and integrity. They apply their risk management expertise to allocate capital efficiently, identify and mitigate emerging risks and to help maintain system integrity across multiple segments of the financial and other sectors. This expertise enables the profession to comment on a wide range of issues including life insurance, health insurance, general insurance, climate change, retirement income policy, enterprise risk and prudential regulation, finance and investment and health financing.

Published December 2019

Institute of Actuaries of Australia

ABN 69 000 423 656
Level 2, 50 Carrington Street,
Sydney NSW 2000, Australia
t +61 (0) 2 9239 6100
f +61 (0) 2 9239 6170
actuaries@actuaries.asn.au
www.actuaries.asn.au