CONTENTS

Purpose 5
Scope 7
Executive Summary 8
Components of Validation 10
Introduction 10
Independence 10
Risk Coverage and Risk Indicators 10
Risk Ranking 11
The Validation Cycle 12

Validation by Risk Type 20
Data 20
Insurance Risk 20
Credit Risk 24
Market Risk 24
Operational Risk 25
Dependencies and Aggregation 25
One Year Risk and One Year SCR 26
Ultimate SCR 28

Appendices

1. Illustrative validation tests for specific risk types
2. Expert judgement and validation
3. References
1 PURPOSE

This document provides guidance to agents in respect of internal model validation requirements under Solvency II. It builds on the findings and observations of the Model Validation workstream of the Lloyd’s Solvency II Dry Run as well validation reviews carried out subsequent to completion of the Dry Run. Lloyd’s will issue Minimum Standards for validation of syndicate internal models in July 2014. This document is guidance.

Agents should note that all guidance issued is subject to ongoing discussion and change as the European Commission (EC), European Insurance and Occupational Pensions Authority (EIOPA) and PRA requirements become clearer.
2 SCOPE

This document offers guidance on validation of the methods, assumptions and expert judgement used in the internal model.

The following areas are not covered in detail in this document; additional guidance has been provided by Lloyd’s in the sources listed.

- Data (section 5.1 describes guidance currently available)
- Catastrophe risk (section 5.2.3 describes guidance currently available)
- Validation Policy (see section 10 of the Model Validation Evidence Template for a description of the requirements)
- Validation Report (see the Validation Report Guidance and the Model Validation Report Illustration for Reserving Risk on lloyds.com)
- Documentation (excluding documentation of expert judgement, which is covered in Appendix 2)
- Use test
3 EXECUTIVE SUMMARY

The overall purpose of the validation exercise is to ensure that the internal model provides a realistic and robust assessment of all the material risks faced by the syndicate. Validation is an iterative process of identification of model limitations and implementation of improvements. The most appropriate validation tools for achieving this task will vary by syndicate, depending on the size and complexity of its risks. This guidance is intended to assist agents with the selection and application of the validation tools most suited to their syndicates’ risk profiles.

The Lloyd’s Dry Run process, and more recently the use of internal models to set capital at Lloyd’s, has seen significant advancements in validation across the market. This has led to more confidence in the results of syndicate models.

Earlier in the process some of the key findings of validation were:

- Potential lack of independence or objective challenge
- Lack of clear validation test criteria and consequences of validation test fails
- Limitations around validation of expert judgement
- Communication of limitations of the model found in validation

In general, these areas have seen the most significant improvements.

Regulatory expectations have also been evolving. Whilst the fundamental objectives have not changed, there has been a degree of clarification provided. Two important and useful documents in this respect are:

- PRA Supervisory Statement SS5/14 (issued April 2014)
- EIOPA “Guidelines on the pre-application for internal models” (issued 27 September 2013)

Lloyd’s is advising agents to give special consideration to the following four areas that are given particular emphasis in the above documents:

- Documentation, communication and validation of expert judgement
- Identification and validation of parameter uncertainty
- Allowance and validation of calendar year effects
- Validation of emergence patterns

Where relevant, this Model Validation Guidance has been updated to be more closely aligned with the PRA Supervisory Statement and EIPOA Guidelines. Agents have received specific feedback on validation over recent years. Lloyd’s expects that outstanding items of material or critical feedback will be specifically addressed.

Therefore to ensure validation is completed to an appropriate level, questions agents should aim to address are:

- Does the validation address previous Lloyd’s feedback? What has been done to achieve this?
- Does the validation address the relevant key points raised in the PRA supervisory statement?

Has validation been able to assess the documentation, peer review and challenge of expert judgement in line with the EIPOA guidelines referenced above?
This is not an exhaustive list but is intended to ensure that, across the market, model validation is advanced from its current state as a well progressed and embedded process to closure of any potential “final gaps” and compliance with the Solvency II requirements in regard of validation.

As with all elements of Solvency II, there does remain some uncertainty in requirements. However, in terms of validation the fundamental position now should be considered relatively stable.
4 COMPONENTS OF VALIDATION

4.1 Introduction

It is essential for the Managing Agent, for Lloyd’s and for compliance with Solvency II that syndicate internal models are thoroughly validated. The requirements for model validation under Solvency II are extensive, and meeting them will bring many challenges for the industry. This is especially true for Lloyd’s, due to the unique and complex nature of the risks written in the market. Managing agents have invested significant time and effort over the last two years in order to meet the standards of Solvency II, including those for validation; progress has been promising.

The validation exercise consists of many components, which are covered in more detail in the following sections. It requires independence in order to ensure objective challenge. It must include a process to ensure that all material risks are covered in the model, and it should produce a risk ranking and a demonstration that the validation effort has been proportional to the materiality of the risk. The validation tools and tests should be well defined and appropriate for the risks being validated. Their outcomes should be clearly explained and justified, and the path from a “fail” outcome to escalation and model change should be clearly mapped. Finally, Solvency II is an evidentiary regime requiring documentation of the validation process and the reasons for the steps taken.

One finding of the recent reviews is that significant improvements to validation could be made by bringing greater clarity and integration to existing processes, as opposed to adding new ones. Examples of how this could be achieved include clearer links with the Validation Policy, more explicit risk ranking, better explanations of test outcomes, and more references to syndicate experience.

4.2 Independence

Validation requires objective challenge. Agents should be able to demonstrate that individuals responsible for validation have sufficient independence from the design, build, parameterisation and implementation of the model component being validated. They may make use of tests carried out by those responsible for the modelling, but not rely entirely on them. See the Validation Report Guidance for guidance around the independence requirements for the author of the Validation Report.

4.3 Risk Coverage and Risk Indicators

The objective of the validation process is to test that all material risks to the syndicate are adequately assessed in the internal model. The initial steps in the validation process should therefore be a gap analysis to test whether all material risks are indeed covered. A typical and acceptable way to do this is:

(1) an identification of risks to the business;
(2) an identification of which of these risks are not covered by the internal model; and
(3) an assessment of whether the risks not covered are material.

If material risks have been excluded, the model needs to be changed to include them.
The risk identification process should identify all sources of loss to which the syndicate could have non-trivial exposure. This process should not be restricted to insurance risks; it should for example include considerations such as the terms and conditions of the cover issued, data and operational systems, the current legal environment, recent market experience, and so on. It should take into account the possibility of new sources of loss not experienced by the syndicate or market in the past.

In recent years, many agents have maintained a risk register as a way of identifying risks faced by the syndicate. Lloyd’s considers this to be an appropriate approach under Solvency II. The format of the register should be the one most suitable to agents. The risk register may also be used to identify which risks are currently “in” or “out” of the internal model.

The risks not covered by the internal model should be assessed for materiality to the syndicate. This requires:

1. an explicit threshold for materiality; and
2. risk indicators for determining materiality.

It is agents’ responsibility to determine and describe the materiality threshold for their syndicates. The threshold should be closely linked to those for major and minor model changes; it should be less than that for a major model change.

Risk materiality can be assessed using risk indicators of materiality. These indicators can be approximate. Examples include: standard deviations; coefficients of variations (“CoVs”); allocated capital; and matrix methods for aggregation of risk. Both CEIOPS DOC 48/09 (5.3.4.2) and the 4 & 5 July 2011 Model Validation Workshops covered this topic adequately. Therefore, for further discussion of risk coverage and risk indicators of materiality please refer to these documents.

Historically, some of the largest losses to the market have come from sources of loss that were only partially modelled or non-modelled. Examples include asbestos and pollution, multi-year reserve deteriorations, the World Trade Center terrorist attacks, and the Japanese tsunami. For this reason, Lloyd’s considers a comprehensive risk identification process to be the basis for sound validation. Agents will be challenged on whether their model has captured all material risks.

Lloyd’s recommends that agents have a comprehensive process for identifying potential risks to the business and assessing their materiality. This should cover changes in risk profile, e.g. an increase in cyber risk, and re-affirmation that risks previously assessed as non-material remain so. The use of the outputs of this process within the validation exercise should be described in the Validation Policy.

4.4 Risk Ranking

The risk coverage exercise described above should confirm whether all material risks are covered in the model and should be proportional to their size and complexity.

The purpose of risk ranking in the context of validation is to identify those areas requiring the most extensive validation. Risk ranking should therefore be one of the first steps in the validation of the internal model.
Agents should be able to produce rankings consistent with the groupings used to manage the business (US medical malpractice, EUR-denominated corporate bonds, etc.). It is not necessary to rank every risk component of the internal model, although the methodology should have the capability to do so. The ranking methodology may be approximate, but should be broadly consistent across risk types.

There is overlap between risk ranking and the assessment of risk coverage described above. Both can be done using approximate metrics, such as risk indicators of materiality. It is preferable, however, that the risk ranking be derived from the outputs for the risk groupings used in the model.

Agents should view risk ranking as a tool for making the validation process more efficient. It will enable them to allocate resources to areas where model improvements will yield the greatest overall benefit.

Lloyd’s recommends that the Validation Policy include a description of how the outputs of the risk ranking exercise are used within the validation process.

4.5 The Validation Cycle

4.5.1 Overview of the validation cycle

The validation cycle is at the centre of the validation process. It is comprised of four steps:

1. the application of the validation test or tool;
2. the analysis of test results;
3. the escalation of test results to appropriate individuals in the business; and
4. the implementation of any changes necessitated by the validation test outcome.

In this section we provide an overview of the main features of each stage in the validation cycle. Guidance on the application of each stage to different risk types is provided under their respective sections (see section 5).

In the early stages of the Dry Run, a common finding was that agents provided relatively clear descriptions of step (1) of the cycle, but were less effective on step (2). For step (3), a process for escalation was often described but without sufficient practical detail. Step (4), the implementation of changes, was not usually covered. Individual feedback was provided and this is an area that has seen significant improvement.

With respect to step (2), all validation tests should result in a pass or fail. In practice, however, there will often be uncertainty around whether the result should be a pass or fail. One way to approach this problem is to group test outcomes into “buckets”:

1. a fail or rejection;
2. acceptance with minor or no modifications;
3. acceptance, but with recognition that there is uncertainty around the assumption and that an alternative could also have been reasonably selected.

In all cases, reasons should be given to support the test result.
In view of this, a reasonable approach to applying validation tests would be the following. First, apply one or more of the tests listed in section 4.5.2 other than sensitivity testing to determine which outcome “bucket” the method or assumption belongs in. Following from this, those assumptions in the third outcome bucket of acceptance, but with uncertainty, should be sensitivity tested. For the most material risks, these sensitivity tests should be based on plausible alternatives, rather than an across-the-board deterministic increase/decrease. The resulting changes in model output can provide an indication of potential limitations in the model arising from the uncertainty inherent in the assumptions.

As noted, two common findings from the Dry Run were that:

(1) agents often failed to clearly state what the test outcome was and why; and
(2) when the test outcome was stated, there was little or no assessment of potential limitations.

With respect to step (3), escalation, the validator must ensure that the implications of the validation results are considered appropriately. This could require model changes, further validation, or other actions. The validator may make specific recommendations.

With regards to step (4), the rationale and impact for changes to the model should be well documented. The changes should be classified as major or minor as specified in the Model Change Policy.

Lloyd’s recommends that each validation test shows evidence of the four stages of the validation cycle. In particular, it should be clear what the test outcome is and the reasons for that outcome.

## 4.5.2 Validation tools

### 4.5.2.1 Qualitative validation of methodology and assumptions

Article 121(2) requires that agents “be able to justify the assumptions underlying their internal model”.

The justification of methods and assumptions can be done both quantitatively and qualitatively. The other tests described in this section provide quantitative validation. However, qualitative validation can also play an important role, particularly if data is limited or where assumptions cannot be tested directly, as with an external model. Qualitative validation involves the same stages of the validation cycle as quantitative validation, with the analysis of results being based primarily on expert judgement.

Qualitative validation begins with a description of the most material assumptions of the method being used; it should identify where the assumptions may not be appropriate for the risk being modelled, and assess the limitations. An example would be the derivation of the premium risk volatility assumptions for a new Employers Liability portfolio from one with an established history, but with policyholders in a different industry. The agent could identify the main source(s) of potential mismatch in assumptions, such as the number of employees or claim frequency potential, and then accept/reject/modify the assumptions.

Lloyd’s recommends that agents be able to identify the most material assumptions of their methods and assess their limitations on a qualitative basis. It is not expected that agents should prepare an exhaustive list of all the assumptions relevant in any way to every risk. Qualitative validation should be more robust in cases where data is limited or assumptions cannot be tested directly (as with external models).
**4.5.2.2 Testing against experience**

Testing against experience includes tests of goodness of fit ("g.o.f") and comparisons between model outputs and historical results ("backtests").

In general, the claims history may be too limited to provide conclusive evidence for the optimum assumption or method. By definition, it will not include “Events Not in Data” (ENIDs). An obvious example is the selection of a large claims distribution, for which there may be a number of distributions with very different skewness, but which all provide close fit to the bulk of the claims. Despite this, the data will often be sufficient to indicate a test result in the first or third buckets (reject or accept with limitations).

In the Dry Run, there were examples of testing against a single historical value, such as the most recent year's loss ratio, with a fail occurring if the loss ratio was above/below the 99th/1st percentile. Such an approach may provide good grounds for rejecting a distribution, provided that the percentiles for a fail are appropriately high/low. The drawback is that it provides no information on the appropriateness of the shape of the distribution in the event of a pass. A test based on all relevant experience would be more informative.

In many instances, the portfolio will have changed over time. In these cases, there may be good reasons for excluding some parts of the history from the tests against experience. However, such exclusions should be based on objective reasons relating to unique characteristics of the risks, not simply on underwriting results. Conversely, it should also be recognised that a limited history may not capture the full tail risk of any portfolio; there may therefore be reasons based on expert judgement for model risk exceeding that indicated by the data. Lloyd’s will not accept agents taking a one-sided approach of excluding unfavourable history as being irrelevant, while not making allowance for tail risk that may not be reflected in the experience.

Finally, tests against experience can assist in communication of model outputs to management by linking them with recent results.

Lloyd’s recommends that comparisons to past experience are made wherever it is available. These comparisons should include clear explanations, based on expert judgement, of the relevance of this experience to the current risks. They should also describe how ENIDs have been taken into account (the PRA Supervisory Statement SS5/14 gives more information on ENIDs).

**4.5.2.3 Stress and scenario testing**

Stress and scenario tests are very valuable tools for validation of syndicate internal models. They rely heavily on expert judgement and are particularly useful where data is limited, such as in the tail of the distribution. They have the additional advantage of being readily understood by individuals across the business. Furthermore, despite often being viewed as relatively less sophisticated, the outcomes of these tests can be viewed as coherent risk measures. [6]

Stress and scenario tests must be based on realistic assumptions and extreme events in order to be credible. It is imperative that agents provide an explanation or narrative around their stress tests; otherwise the value of the exercise is diminished. It is also essential that the “stresses” are sufficiently severe.
The event severities and probabilities should be derived independently from the process used to derive the risk distributions in the model. In general, it will be more difficult to assess the probabilities than the severities. It may helpful in this regard to also consider scenarios at lower return periods, such as 20 to 50 years (twice or once in a career) and then extrapolate or build up to higher return period type events.

The evaluation of stress and scenario tests as validation tools requires the comparison of return periods from the tests with those from the model. As with other validation tests, agents should clearly state in which of the three test outcome buckets the test result lies. For example, if a scenario test indicates that there is a 1-in-40 year return period for the outcome of a reserve deterioration exceeding £35m and a cat loss exceeding £10m, while the model indicates that this would be a 1-in-100 year outcome, it should be clear whether this result is a pass/fail/accept with limitations and why.

4.5.2.4 Sensitivity testing

Sensitivity testing has broad application throughout the model. In general, there are two types of sensitivity testing. The first involves determining the sensitivity of model outputs to variations in key inputs or assumptions; the second involves testing the stability of outputs using fixed inputs while varying the random seed or number of simulations. We will refer to the first type as sensitivity testing and the second as stability testing.

There are two approaches to sensitivity testing. One type (“ST-1”) involves deterministically varying a set of assumptions (such as loss ratio CoVs) by a given amount and measuring the effect on model outputs. This approach can be used to identify the relative materiality of different inputs; it can also test the mechanics of the model, in that if outputs do not move in the expected direction, it could be the result of a coding error, broken link, etc.

The second application (“ST-2”) involves varying the inputs, but using plausible alternative selections. The choice of plausible alternatives may be guided by a prior validation test, such as a test against experience. ST-2 will be less useful in determining the relative materiality of inputs, since the increase/decrease will vary for different inputs. However, ST-2 has the advantage of reflecting the uncertainty in different assumptions, and therefore in model outputs. As such, the results will be more informative to management. This approach also tests the mechanics of the model.

Lloyd’s prefers that both approaches are used. ST-1 has more relevance to risk ranking, whereas ST-2 is more useful for conveying the uncertainty in the internal model. ST-2 requires the additional effort of determining plausible alternatives, so it is more appropriate for larger risks. It should be noted that neither approach provides validation of the absolute value of the input being tested. For this reason, sensitivity testing on its own will generally not be sufficient for validation. Sensitivity testing will be difficult to isolate from other forms of validation, such as statistical quality testing and backtesting.

A finding from the Dry Run was that some agents relied almost exclusively on the first form of sensitivity testing. Furthermore, agents mostly restricted themselves to sensitivity testing parameters as opposed to distributions or other assumptions. This approach may have neglected key drivers of model outputs. For example, varying the input correlations for a Gaussian copula does not sensitivity test for the impact of alternative dependency structures. Another finding was that agents often defined a test as a “pass” if it resulted in a change to the SCR that was negligible or within a pre-defined range.
A sensitivity test on a material assumption will produce a large change in model outputs, regardless of whether it is correct. Conversely, a low sensitivity could be an indication that the risk has not been adequately accounted for.

With regards to stability testing, the principal challenges are:

1. the time and effort associated with re-running the model, and
2. defining appropriate criteria.

With regards to (1), it may be possible to obtain an (approximate) indication of the minimum number of simulations required for convergence by running separate simulations for the longest tailed risk in the model. For example, if US quake is the syndicate's longest tailed risk and the internal model input is a RiskLink ELT, the agent could separately simulate from the ELT to test the number of simulations required to obtain a satisfactory level of convergence. Some outputs may be amenable to analytic calculation, such as OEP in some cases, and comparing the simulated output with this accurate figure is a useful test.

It is the responsibility of agents to determine appropriate stability criteria; Lloyd's does however consider stability when assessing syndicate model results. EIOPA guidance also emphasises that models must be stable, and the validation should cover this point. The model should usually be run on a number of seeds and the results examined to ensure that the range is acceptable.

4.5.2.5 Reverse stress testing

Reverse stress testing is an essential validation test of the syndicate SCR. Management should have a view on the risks to solvency as well as the opportunities for growth and profit. Reverse stress testing provides the best opportunity to evidence that understanding.

Reverse stress testing begins with consideration of the events or combinations of events that could threaten the viability of the business. They should reflect the interaction of management (in) actions and external events. Historically, breaches of the ICA have resulted from items such as:

- Reserve deteriorations on multiple years of account of a casualty portfolio; or
- Large natural catastrophes, with a significant proportion of losses coming from perils not fully covered in the vendor models.

Lloyd's will take a particular interest in how agents have considered either or both of the above (as appropriate) in their reverse stress testing. Agents should not, however, confine their reverse stress testing to these scenarios.

Reverse stresses that result in a depletion of capital are by definition at return periods of 1-in-200 years or higher. These return periods are for the estimated capital requirement; that is, they are at the syndicate level. The stresses must therefore reflect the aggregation of risks across the syndicate, and not only the drivers of the insolvency. For example, a reverse stress could be based on the joint occurrence of a US windstorm and reserve deterioration on a casualty class, neither of which on its own would be sufficient to trigger an insolvency.

The reverse stress test outcome should be based on a comparison between the balance sheet distribution and the reverse stresses. The distribution in the neighbourhood of the 99.5th percentile should include aggregations of outcomes that are consistent with the reverse stresses. In this sense, the reverse stress tests can be thought of as a “reality check” on the balance sheet distribution.
4.5.2.6 P&L Attribution

Article 123 requires that undertakings review the causes and sources of profits and losses for each major business unit on at least an annual basis. It also requires undertakings to demonstrate how the categorisation of risk chosen in the internal model explains the causes and sources of profits and losses.

The basic concept is to ensure all material risk contributions are captured in the model as follows:

- Review the actual vs. expected outcome over the period
- Investigate the material sources of profit and loss
- Ensure that the model is sufficiently granular to capture the sources of profit and loss
- Reflect the results and conclusions in the model validation process

Lloyd’s would expect a “fail” if the profit/loss driver is not captured in the model (for example, tsunami losses), or a “pass” if the source is explicitly captured (for example, if IBNER releases on large motor claims are explicitly modelled).

From the reviews undertaken, Lloyd’s expects agents to:

- Maintain a clear, concise set of policies and procedures setting out the definition of P&L, granularity of business units, categories/line items to be tested and whether these are modelled (or not) and governance of the P&L process.
- Conduct, at least annually, a P&L attribution exercise for the year being modelled. This will include identifying the outcomes to be tested against the equivalent modelled probability distribution function.
- Incorporate the outcomes of the P&L exercise in the model validation process and include results and commentaries in the model validation report.

Given that Solvency II requires internal models to be constructed on an economic basis, P&L attribution should strictly speaking be done on an economic basis. Lloyd’s is aware that this will not be practicable for most agents at present. Agents should however make pragmatic use of available accounting information, with some qualitative consideration of how results of the test might differ on an economic basis.

4.5.2.7 Other tests

There are a number of other validation tests in addition to the above that may be appropriate.

Benchmarking can be both internal and external. An example of internal benchmarking would be comparison of a loss ratio distribution for a new or small class against those of more established classes. Conclusions could be made based on known differences between the classes, such as the length of payment patterns, size differences, market cycle, and so on.

External benchmarking may involve using market data to derive an alternative result. It may also include comparisons to market peer groups made by consultants or other third parties. Whilst such external benchmarking may add value, it should not form the basis of the validation process.
Validation should demonstrate that the internal model is appropriate for the syndicate’s own risks; external benchmarking cannot perform this task. When using benchmarks based on a peer group, agents should obtain specifics on the relevance of the peer group to their own business.

**Analysis of change** involves comparing the values of key inputs and outputs with those of the previous version of the model. Many agents made such comparisons with the ICA. The analysis should include an investigation of why the values have or have not changed, and reasons as to why the changes are or are not appropriate. The benefit of this exercise will depend on how well the previous version of the model was validated.

**Tests of model functioning** are designed to ensure that the model is functioning as intended. Examples include comparisons between the means and standard deviations of the input and output distributions and “as-if” calculations that push a single value of, for example, a cat event through the model and compare gross and net outcomes with those calculated manually. Tests of model functioning would normally be done most intensively during the model build stage.

### 4.5.3 Analysis of results

As noted in section 4.5.1, it should be possible to assign a test result to one of three outcome “buckets”:

1. fail;
2. acceptance with no or minor modifications;
3. acceptance, but with the recognition that there is uncertainty around the result, and that alternative assumptions could also yield an acceptable result.

The analysis of results should make clear which of the above categories the test outcome falls into. This does not require agents to follow the three categories above for their validation outcomes; for example, they may classify outcomes as a pass/fail, and explain any qualifications behind either result.

As noted, early in the Dry Run agents often applied a test without stating whether it was a pass or fail and why. This sometimes occurred with goodness of fit tests, where plots were presented, showing varying degrees of fit, without any supporting explanation. In other cases, usually with sensitivity tests, explicit criteria were specified in advance, without any justification of the criteria. (As discussed in section 4.5.2.4, pass/fail criteria for sensitivity tests may be inappropriate.) This would not be sufficient.

In cases where more than one validation test has been applied, there should be an overall conclusion based on assessment of the individual test results. This conclusion should be supported by explanations of the weight given to the different tests.

The Validation Policy should define the next step in the validation cycle, depending on the outcome of the analysis of results. In cases where the outcome is “pass, but with limitations”, it is reasonable that for more material risks there should be some sensitivity testing of alternatives. Similarly, a “fail” for a material risk should lead to escalation, which is discussed below. For a practical example of these alternatives, see the Reserving Risk Illustration for the Validation Report under the Model Validation workstream section for Solvency II on lloyds.com.
4.5.4 Escalation

Escalation was discussed briefly in section 4.5.1. Elements of a robust escalation process include:

- A well-defined trigger in terms of validation test outcomes and risk materiality
- A description of the person(s) responsible and the tools they are expected to use
- A clear expectation of what is required for resolution of the validation escalation

A finding from the Dry Run was that agents often specified the escalation path and persons involved, without describing what these individuals would do should an escalation occur.

4.5.5 Changes to the internal model

The model may change either from validation findings or for other reasons. In either case, the changes must be validated, and the validation process must be completed before the changed model is used.

The wider model change process is covered in detail by Lloyd’s model change guidance and submissions. These can be found here: [Model Change](#).
5 VALIDATION BY RISK TYPE

Section 4 provided general guidance on the main components of validation. This section goes beyond providing guidance and offers some suggestions on how Agents can apply the validation cycle in practice to individual risk types.

Validation should be applied to each material method and assumption used in the analysis of each risk type. The methods and assumptions will vary by risk type, but will broadly include data selection and groupings, distribution selection and parameterisation, and inter-risk dependencies. These are discussed in more detail in the sections below.

Appendix 4 of the SCR Guidance includes a list of methods and assumptions for discussion in the SCR documentation; agents may find this useful when considering where to apply validation.

5.1 Data

Data is out of scope for this guidance document. Guidance has previously been issued to the market through the Data workshops of July 2011 and April 2012. Data Audit Report guidance was issued 31 March 2012 and all material is available on lloyds.com.

5.2 Insurance Risk

5.2.1 Reserve risk

Reserve risk is the largest risk for many syndicates. It should therefore be one of the main areas of focus for validation.

Aspects of the modelling requiring validation include:

- Use of paid or incurred data and how it is ensured that the CoVs are selected consistently with this. For example, model checking should ensure that the CoVs are calculated and applied consistently with the projection method and the best estimate reserves. In the Dry Run, there were examples of inconsistency here.

- Grouping of classes. More grouping means fewer assumptions, a less complex model and probably faster runs, but the groupings must be appropriately homogenous, including with respect of the effects of reinsurance and changes over time. Validation approaches can be qualitative (for example reviewing the changes in risks and reinsurance parameters over time) and quantitative (measuring the stability of more granular or grouped classes, and measuring the contributions of different granular classes to the grouped class over time)

- Whether and how attritional and large claims are modelled separately. Validation methods may include measuring how much of the overall volatility arises from large claims and measuring the dependency between large and attritional experience.

A significant validation challenge relates to the dependencies between accident years/years of account and the resulting skewness of the overall reserve risk distribution for a given class of business. (See section 5.6 for a discussion of other types of dependencies.) The majority of the market has relied in part or in total on triangle based methods for determining reserve risk – the Mack method and (more commonly) approaches that utilise bootstrapping. Agents often stated that the bootstrap implicitly accounts for dependencies between accident years.
It is true that the bootstrap is based on a sampling of the same set of residuals across accident years, and that this will introduce correlations between accident years. Similarly, in the Mack method, correlations are introduced through use of common development factors across accident years; this is explicitly allowed for in the formula for all years [4]. This correlation is however distinct from dependency between accident years’ process risk. In bootstrap approaches, the process risk distributions for individual accident year/development year cells are sampled from independently, consistent with the assumptions of the GLM on which the approach is based [1]. In the Mack method, independence is explicitly assumed. It is the process risk that is of primary interest in modelling reserve deteriorations; furthermore, as noted earlier, deteriorations across multiple accident years have been a frequent cause of ICA breaches within the market. Therefore, the assumption of independence between accident periods needs to be carefully addressed within the validation. To quote from a recent paper: “It should be emphasised that bootstrapping is a method rather than a model. Bootstrapping is useful only when the underlying model is correctly fitted to the data, and bootstrapping is applied to data which are required to be independently and identically distributed.” [3].

CEIOPS DOC 48/09 (5.245) described two approaches to the validation of dependencies that remain relevant:

(1) causal explanations based on expert judgement, and
(2) quantitative methods.

In terms of causal explanations, a number of considerations suggest that there will often be dependencies between accident years, particularly for long tailed classes of business. These include similar (or the same) risks, the underwriting cycle, claims inflation, and so on. Agents have however often included inter-class dependencies for which there is limited causal justification, but neglected intra-class dependencies between accident years.

In terms of quantitative tools, goodness-of-fit tests with residuals can provide evidence of calendar year trends that may invalidate the assumption of accident year independence. In the Dry Run, some agents stated that they “looked at the residuals”, but were unable to say how they would differentiate between outliers, which can be excluded vs. trends, which suggest a mismatch between the data and the model assumptions. Others expressed the view that trends or outliers in the residuals resulted in a more prudent estimate of reserve risk. The alternative view is that they signify a poorly fitting model whose limitations should be assessed.

If the analysis of validation results indicates that the assumption of independence between accident years has not been satisfied, there will not be a simple “one size fits all” solution. Whilst there are numerous alternatives to the bootstrap and Mack in the actuarial literature, few of these involve quantification of calendar year trends. Those that do will still require the modeller to take a view on whether/how the calendar year trend will continue. In some cases, adjustments to the data (for example capping claims) may yield a better agreement with the bootstrap assumptions. In general, Lloyd’s would encourage agents to make greater use of the detailed analysis done in the best estimate reserving process as opposed to relying on more sophisticated (but also more black box) statistical models. The reserving process will provide insights on the causes of past concurrent accident year deteriorations, and this could aid in, for example, the development of stress and scenario tests.
There are several other assumptions that may be made in conjunction with a bootstrap approach. These include the choice of stochastic model (over-dispersed Poisson, Mack, etc.), residual type and residual adjustments (scaled, bias-adjusted and zero-averaged). The materiality of the different options should be assessed qualitatively and/or quantitatively; the selections for the more material options should be validated. For example, zero-averaging the residuals will have varying impact depending on the difference of the mean of the residuals from zero and whether the difference is positive or negative. If the mean is positive, then zero-averaging will result in reduced simulated incremental claims; this in turn may impact the variability in reserves.

The reserve risk analysis will often result in distributions of reserves and ultimates with means different from those obtained from the best estimate reserving process. In these cases, the distributions should be scaled so that their means are consistent with the best estimates. Two possible approaches to scaling the simulated reserves are: (1) adding a constant, or (2) multiplying by a constant. The additive approach will change the mean of the reserves or ultimates but not their standard deviations. If the scaling constant is positive (which is usually the case), then the scaled distribution will have a lower COV than the unscaled. Agents should expect to be challenged by Lloyd’s on scaling that reduces the COV in this way. Model checking should be carried out to ensure any scaling does not distort the intended assumptions.

The methodology for determining reinsurance recoveries should be validated. Agents frequently use net-to-gross ratios (fixed or variable), which may not be accurate at high percentiles of gross claims, particularly for XOL programmes. The materiality of any approximations should be assessed, both in terms of recoveries and RI credit risk. There needs to be a clear justification for the allowance for reinsurance exhaustion.

This section has covered only two of the more prominent reserving risk validation issues identified over recent reviews. Other areas, such as latent claims/ENIDs, IBNER, legal rulings, and so on, will also need to be considered. Agents seeking further background may wish to refer to the topics listed in an appendix to the SCR Guidance.

See Appendix 1 for examples of validation tests for reserving risk.

5.2.2 **Premium risk excluding catastrophe**

The following discusses some of the more frequently raised validation issues for premium ex cat. In contrast to reserving risk, Lloyd’s did not identify one or two validation concerns of high materiality for the majority of the market.

One validation issue (which has overlap with catastrophe risk) relates to the assumption that non-modelled cat perils are covered in the attribitional claims distributions. A common approach is to use the attribitional loss ratio as a balancing item to ensure that the attribitional, large and cat mean loss ratios sum to the business plan loss ratio, which in turn is meant to include allowance for non-modelled perils. This approach has a number of drawbacks, in particular that it is validated at the mean only. Agents using this approach should validate the entire attribitional distribution, bearing in mind that losses from non-modelled natural catastrophe perils may be highly skewed.

A second issue relates to the threshold for large losses. Agents often stated that this was done based on internal reporting considerations (and therefore supported the Use test). While this is a valid rationale, the implications should be assessed from a statistical standpoint. The threshold will affect the credibility of the data for making frequency and severity distribution selections.
In some instances, agents relied on prior loss experience unadjusted for inflation, IBNER or other changes as the basis for their premium risk distributions. In other cases adjustments were made, but the basis for the adjustments was not validated. An example is the revaluation of large claims to bring them to the current cost level. In some cases the revaluation rate was selected judgementally and not supported by expert judgement or quantitative tests. (See Appendix 1 for an example of the latter.) For short tailed classes, the selected revaluation rate may not be a material assumption, whereas for some long-tailed classes it could be very material. If the revaluation assumption is material, it should be validated.

A fourth issue relates to the reliance on underwriter judgement for parameter selection or validation. While many underwriters will have valuable insights gained from years of experience, their views should be supported by specific reasons. Another finding was that the validation questions posed to underwriters were sometimes open to more than one interpretation. An example was a request for an underwriter’s view on large claim “return periods”, when in fact the question related to the large claims severity distribution.

The use of an alternative model may be a useful validation tool for some classes. An example is a property class for which the risk profile has changed significantly, or where the sums insured of the historical claims are not known. In these cases an exposure based frequency-severity model could provide a valuable alternative indication to an experience based model.

Agents should also ensure that there is consistency between reserve risk and premium risk. It will generally be expected that the volatility for a given year of account should decrease with time. See Appendix 1 for examples of validation tests for premium risk.

5.2.3 Catastrophe risk

Detailed guidance on catastrophe risk is out of scope for this guidance document.

Claims and sources of loss related to catastrophe perils can be grouped into those modelled using external catastrophe models and those modelled using other methodologies.

External catastrophe models

During 2012 Lloyd’s undertook an extensive process of validation working with the LMA and closely with individual managing agents. If managing agents are looking to change their underlying catastrophe model or to change any adjustments to the existing model they must re-submit their validation documentation prior to using the model in the 2015 capital setting process.

Key outputs of this process which remain relevant for validation are located on the model validation section of lloyds.com.

Dr Mohammed Zolfaghari carried out a validation of the USGS earthquake catalogue for the Lloyd’s market which was presented in the Old Library and this report was sent round afterwards.

As we stress in the above documents it is critical for managing agents to focus on why the model is thought to be appropriate for their portfolio.

“Non-modelled” catastrophe risks

Syndicates may have exposure to sources of loss not covered (or only partially covered) by external catastrophe models. Internal models must include all material causes of risk so that the
derived SCR is complete and accurate. However the methods used will vary according to available information and materiality.

During 2013 Lloyd’s took part in a London market exercise convened by the ABI to consider good practices in this area. The outcome of the process was the production of a document which covers definitions, methods of identifying risks, monitoring exposures and modelling within an internal model.

The ABI document can be found here.

This document should be considered good practice and we note that the document keeps materiality at the forefront of the decision on which process to adopt from the many suggested alternatives.

5.3 Credit Risk

See Appendix 1 for examples of validation tests for credit risk.

5.3.1 Reinsurance

Most commonly, agents have relied on default probabilities published by the rating agencies. Qualitative validation should be based on an understanding of how they were produced and their potential limitations. Other common assumptions include an allowance for credit migration and a dependency between large catastrophes and default probabilities. Sensitivity testing may be an appropriate validation tool.

Reinsurance credit risk will of course be affected by how the reinsurance is modelled. Many agents use (fixed) net-to-gross ratios for reserves. The materiality of modelling recoveries using ratios vs. explicitly should be assessed for years with the potential to generate significant recoveries.

5.3.2 Other

Other sources of credit risk will vary by syndicate. Broker balances may be one. Generally these will be less material than reinsurance credit risk, so validation will be less extensive. Sensitivity and scenario tests may be of use.

5.4 Market Risk

For most syndicates, the main drivers of market risk will be foreign exchange risk and interest rate risk. Many agents have chosen to use vendor-produced Economic Scenario Generators to model these risks.

The CEIOPS DOC 48/09 provided advice on the validation of external models. Key elements of the validation include:

- Being able to demonstrate a detailed understanding of the model’s methodology and limitations
- Detailed validation of the external model output

One approach to demonstrating detailed understanding would be documentation of the knowledge gained from vendor publications and presentations and the agent’s own in-house validation. Validation of external model output could include sensitivity testing of key assumptions and comparison of (for example) interest rate movements with their own scenarios.
See Appendix 1 for examples of validation tests for market risk.

5.5 Operational Risk

For most agents, the modelling of operational risk will be based on limited data (for example, of “near misses”), supplemented by expert judgement. As with other cases in which expert judgement is relied on, a clear rationale should be provided for the assumptions made. This should be supported by sensitivity tests.

The comments on operational risk in the SCR guidance are relevant to validation.

5.6 Dependencies and Aggregation

The validation of dependencies is a broad and challenging subject. The discussion below highlights some of the key issues identified, many of the points raised here were previously discussed in the model validation workshops.

One useful distinction to make is that between top-down and bottom-up validation. Bottom-up validation begins with the component risks and considers whether the dependencies between them are appropriate. Top-down validation begins with a view on the aggregate distribution and seeks to ensure that the model dependencies produce outputs that are consistent with this view.

The table below summarises the main features of each.

<table>
<thead>
<tr>
<th>Dependencies: two approaches to validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>VALIDATION COMPONENT</td>
</tr>
<tr>
<td>Risk distributions</td>
</tr>
<tr>
<td>Inputs/Outputs</td>
</tr>
<tr>
<td>Sample metrics</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Key validation tools</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Both of these approaches have their advantages and limitations. The bottom-up approach explicitly addresses the appropriateness of the individual drivers of dependencies between pairs of risks; the limitation is that it can provide only indirect assurance on the aggregate result. Conversely, the top-down approach can provide comfort that the aggregate result is appropriate, but will be of less use in providing assurance on the appropriateness of the individual drivers of dependency.
The Dry Run confirmed that agents have relied primarily on bottom-up validation based on sensitivity testing of correlations. Whilst this approach can play an important role in the validation of dependencies, when used on its own it has some significant drawbacks. First, many agents relied on Gaussian copulas, which are not tail dependent; therefore, varying the correlations had little effect at the 99.5th percentile. A more robust sensitivity test would consider alternative dependency structures. Second, there was often only limited support (either qualitative or quantitative) provided for the choice of correlations or dependency structure. Third, there will frequently be drivers of dependency between pairs of risks in addition to the correlations and copula structure; the correlations-based bottom-up approach may not address these other drivers.

Given these limitations, top-down validation can provide a useful addition to the validation toolkit. At the highest level, it begins with the question “What are the aggregations that could (nearly) bankrupt the syndicate?” This question should be an essential part of the reverse stress testing process (see section 4.5.2.5); agents should be able to articulate an answer.

At a lower level, agents have made (for example) comparisons between historical YoA loss ratios aggregated across classes of business with the distribution of aggregated output from the model. Like the bottom-up validation, the top-down approach will not be sufficient on its own.

One of the primary areas of focus for Lloyd’s is the level of diversification, both within and between the different risk categories (reserving, market, etc.). The discussion of the bootstrap in section 5.2.1 provides an example of how a combination of granular risk distributions and independence can result in an overall distribution that lacks skewness. Similarly, the use of granular risk distributions elsewhere in the model, combined with non-tail dependent Gaussian copulas, will result in a lower 99.5th percentile, both at the risk category level, and for the overall SCR. Some internal models contain hundreds of pairs of correlations but still produce diversification benefits of 30% or more. Agents should address the level of the diversification benefit as part of their validation of dependencies.

Finally, agents should also consider whether all of the dependencies in their model are warranted. CEIOPS DOC 48/09 (5.245) stated that dependencies should be supportable by expert judgement on causal relations or quantitative evidence (or both). If such justification cannot be found, this should be stated as a finding of the validation process. An example could be dependencies between attritional and large claim premium risk distributions. Agents often stated that they chose their attritional/large threshold so that their attritional loss ratios were stable over time. If this is the case, adding dependencies between attritional and large may not make any material contribution to the final result (regardless of the dependency structure). Agents should avoid adding dependencies unless there is a quantitative or qualitative case for doing so.

See Appendix 1 for examples of validation tests for dependencies.

### 5.7 One Year Risk and One Year SCR

Syndicates are required to provide the one year 99.5th VaR for the individual risk categories and a one year SCR for the syndicate as a whole. The validation of one year risk presents a number of unique challenges.

Broadly speaking, there are two approaches to calculating one year risk:

1. a direct approach, and
2. an indirect approach that derives one year risk from ultimate risk.
Currently, the most common approach is to use some form of indirect approach (often called “recognition patterns”). Whichever approach is chosen, however, the validation task is to demonstrate that the level of risk over a one year horizon is appropriate. The validation tools will be largely the same.

Benchmarking against ultimate risk is a fundamental validation test. Lloyd’s requires that insurance risk be modelled to ultimate; for these risks (reserve, premium and cat), one year risk should not normally exceed ultimate. Lloyd’s takes this position based on the following principles.

- Reserving risk emerges over an extended period, and this leads to the volatility on a book of earned reserves to be lower over the next year compared to ultimate. Historical experience also indicates that “bad years get worse” (i.e. there is no real evidence of reversionary effects on reserves).

- The exposure to premium (or underwriting) risk is typically 18 months on an ultimate basis; “problems” (especially attritional issues) can take an extended time to emerge. These two effects will make premium risk more volatile on an ultimate basis.

- As with premium risk, cat risk will, on average, be covering 18 months exposure on an ultimate basis. This will result in more volatile results at a 1-in-200 on an ultimate vs. one year horizon.

See the 9 & 10 May 2011 Model Validation Workshop section on Calibration for a discussion of one year vs. ultimate risk.

With regards to reserve risk, a test against experience for one year risk should involve a comparison of historical one year movements in ultimate as a percentage of opening reserves. Two difficulties that may arise in applying such a test are:

1. a lack of credible historical data; and
2. past ultimates may not have been set on a true best estimate basis, as Solvency II requires.

Despite these difficulties, Lloyd’s considers the historical experience to be an essential point of reference. Where margins/deficits have existed, agents should attempt to explain the reasons and the impact on one year emergence. It should not be necessary to carry out detailed tests against experience for all classes; these could be done for a few of the more material classes, and smaller classes could be benchmarked against these, taking into account considerations such as whether the class is longer or shorter tailed.

A number of methods exist for calculating one year reserve risk emergence, sometimes referred to as the variability in the claims development result (“CDR”). These approaches range in complexity from the QIS 5 proportional method to multi-period simulation approaches that include recalculation of the risk margin. A recent paper [5] on risk margins found that an approach based on factoring ultimate risk into annual time periods was nearly as accurate and much easier to implement than the “full blown” multi-period approach. This result was specific to the data and model used, and may not be valid in general. Agents should provide qualitative validation on the assumptions and limitations of any method used to determine or validate one year reserve risk.

Validation of the syndicate one year SCR should include similar validation checks to those used for the ultimate SCR. These include stress and scenario testing and reverse stress testing. Historical profit & loss at the syndicate level, relative to some exposure measure, may provide an indication of whether the SCR will be breached. It is generally expected that the one year SCR will be less than the ultimate, since the 99.5th VaRs for the dominant risk categories are expected to be lower at one year than at ultimate.
There are other considerations however that could result in the one year SCR being (somewhat) higher. On an ultimate basis, both the SCR and the risk margin are available to protect against insolvency; in effect, the risk margin is an offset to the ultimate SCR. Conversely, on a one year horizon, the risk margin must be available to transfer the liabilities at year end. If one year risk is large relative to ultimate (as for say a syndicate writing mainly property), and depending on how the risk margin has been determined, the one year SCR could exceed ultimate. Another possible reason is that risks may aggregate differently at one year vs. ultimate. Lloyd’s will expect explicit justification when the one year SCR exceeds the ultimate.

The PRA are clear that in “actuary in the box” methods an unadjusted basic chain ladder is unlikely to be acceptable, and Lloyd’s will expect validations to include a test of the reserve simulation.

Similarly the use of paid or incurred patterns to represent emergence patterns are not acceptable.

5.8 Ultimate SCR

The validation of the ultimate SCR is based on the validation of the underlying components. Agents should also carry out validation tests at the syndicate level, including reverse stress testing, P&L attribution and an analysis of change against the previous SCR. These tests were discussed in section 4.5.2
# APPENDIX 1: ILLUSTRATIVE VALIDATION TESTS FOR SPECIFIC RISK TYPES

## Reserving risk

<table>
<thead>
<tr>
<th>TEST TYPE</th>
<th>DESCRIPTION &amp; CRITERIA</th>
<th>PURPOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative (data)</td>
<td>Assessment of appropriateness of claims history for predicting future emergence; criteria can relate to changes types of risks, size of data set, policy terms, claims reserving practices, etc.</td>
<td>Identification of potential biases to the reserve risk estimate due to limitations of claims history</td>
</tr>
<tr>
<td>Qualitative (methodology &amp; assumptions)</td>
<td>Identification of desirable criteria of a reserving risk model, including types of risk covered (parameter, inflation, etc.), allowance for dependence between accident years; differentiation between types of emergence (IBNR, IBNER); accuracy of RI recovery modelling; user-friendliness and ease of explanation; etc.</td>
<td>Qualitative assessment of strengths and weaknesses of approach; Demonstration of understanding of methodology</td>
</tr>
<tr>
<td>Sensitivity test (risk groupings)</td>
<td>Sensitivity of reserve risk estimate to alternative data splits (all claims vs. large and attritional; different class groupings, etc.)</td>
<td>Identification of risk groupings as a potential source of reserve risk estimation error</td>
</tr>
<tr>
<td>Goodness-of-fit (bootstrap residuals)</td>
<td>Analysis of residual plots by AY, DY and CY for non-normality; criteria defining outliers and trends (e.g. number of standard deviations for outliers; number of sequential residuals moving up/down for trend)</td>
<td>Assessment of appropriateness of bootstrap assumptions for data set</td>
</tr>
<tr>
<td>Backtest (reserve movements)</td>
<td>Historical accident/underwriting year reserve movements (adjusted for paid) as a percentage of opening reserve over 1, 2, etc. years vs. modelled distribution percentiles</td>
<td>Test the consistency of model distribution with historical experience; Provides link between history and model for senior management</td>
</tr>
<tr>
<td>Qualitative (AY dependencies)</td>
<td>Assess historical likelihood of more than one accident year having a large actual vs. expected in a given calendar year</td>
<td>Qualitative evaluation of accident year dependencies</td>
</tr>
<tr>
<td>Stress &amp; scenario test (large claims)</td>
<td>Stress test case reserves on largest claims &amp; assess impact on reserves; compare to reserve risk distribution percentiles</td>
<td>Assess appropriateness of reserve risk distribution based on information on known claims</td>
</tr>
</tbody>
</table>
## Premium risk excluding catastrophe

<table>
<thead>
<tr>
<th>TEST TYPE</th>
<th>DESCRIPTION &amp; CRITERIA</th>
<th>PURPOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative (data)</td>
<td>Assessment of credibility of experience for modelling premium risk; criteria based on considerations such as volume, differences between past policy T&amp;C and risk types vs. next year of account</td>
<td>Identification of potential biases to the premium risk estimate due to limitations of claims history</td>
</tr>
<tr>
<td>Qualitative (methodology &amp; assumptions)</td>
<td>Identification of desirable criteria of a premium risk model, including types of risk covered (parameter, inflation, underwriting cycle, etc.); differentiation between types of claims emergence (attritional, large claims, clash, non-modelled perils); accuracy of RI recovery modelling; user-friendliness and ease of explanation; etc.</td>
<td>Qualitative assessment of strengths and weaknesses of approach; Demonstration of understanding of methodology</td>
</tr>
<tr>
<td>Sensitivity (data adjustments)</td>
<td>Graphical evaluation of re-valued/on-level historical large claim frequencies, large claim severities and ULRs; criteria should relate to absence of trends; result should be random variation about long-term average</td>
<td>Indirect test that adjustments for rate changes, inflation, IBNER, etc. are reasonable</td>
</tr>
<tr>
<td>Goodness-of-fit tests (large claims)</td>
<td>Standard statistical tests such as Q-Q plots; criteria should involve consideration of tail</td>
<td>Test consistency between parametric distributions and claims history</td>
</tr>
<tr>
<td>Benchmarking (CoB comparisons)</td>
<td>Comparison of relative riskiness of different CoB as indicated by quantitative methodology; criteria based on underwriting knowledge</td>
<td>Test consistency between premium risk models for different classes</td>
</tr>
<tr>
<td>Scenarios (large claim)</td>
<td>Scenarios based on extreme large claims; criteria could involve comparison against policy limits</td>
<td>Assess appropriateness of premium risk distribution against actual exposure</td>
</tr>
</tbody>
</table>
## Credit risk

<table>
<thead>
<tr>
<th>TEST TYPE</th>
<th>DESCRIPTION &amp; CRITERIA</th>
<th>PURPOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative (methodology &amp; assumptions)</td>
<td>Assessment of methodology used to assess probabilities of default, recovery rates, transition probabilities, dependencies between creditors, allowance for credit enhancements (funds withheld, letters of credit, etc.) and a stressed environment; criteria could relate to considerations such as size of data set relied on, whether it reflects recent trends, any testing against experience or other validation carried out (if obtained from an external provider)</td>
<td>Qualitative assessment of strengths and weaknesses of approach; Demonstration of understanding of methodology</td>
</tr>
<tr>
<td>Stress test (r/i recoveries)</td>
<td>“As-if” scenario based on stressed environment following large cat; calculate r/i recoveries after default of one or more r/i after allowance for any mitigation (funds withheld, etc.); criteria could be based on difference with model output for similar scenario</td>
<td>Assess appropriateness of default severities</td>
</tr>
</tbody>
</table>
# Market risk

<table>
<thead>
<tr>
<th>TEST TYPE</th>
<th>DESCRIPTION &amp; CRITERIA</th>
<th>PURPOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative (ESG methodology &amp; assumptions)</td>
<td>Descriptions of the methodology and assumptions used, with emphasis on those parts most material to the syndicate (interest rate risk and foreign exchange risk) and any proprietary assumptions relating to dependencies; criteria could be based in part on any validation done by the vendor, and comparisons with other ESGs</td>
<td>Qualitative assessment of strengths and weaknesses of the ESG; Demonstration of understanding of methodology</td>
</tr>
<tr>
<td>Backtest (market data)</td>
<td>Model outputs are compared to historical market data, such as UK gilt or US Treasury rates; criteria could relate to differences between historical and model volatilities or extreme movements</td>
<td>High-level check on credibility of ESG output</td>
</tr>
<tr>
<td>Stress test (past crises)</td>
<td>Stress tests based on past crises e.g. 2008; criteria could be based on probabilities of similar movements in model output</td>
<td>Tests coverage of extreme market events in internal model</td>
</tr>
<tr>
<td>Sensitivity test (shift in key variables)</td>
<td>Sensitivity test value of asset portfolio for different movements in interest rates or exchange rates; criteria could be based on probabilities of similar movements in model output</td>
<td>Determine materiality of exposure to key financial variables; Aids communication with management</td>
</tr>
<tr>
<td>Stress test (future crises)</td>
<td>Stress tests based on understanding of current threats to the financial markets, e.g. insolvency of European banks or governments; criteria could be based on probabilities of similar events in model output</td>
<td>Tests coverage of extreme market events in internal model</td>
</tr>
</tbody>
</table>
## Dependencies

<table>
<thead>
<tr>
<th>TEST TYPE</th>
<th>DESCRIPTION &amp; CRITERIA</th>
<th>PURPOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative (driver identification)</td>
<td>Qualitative assessment of drivers of dependency (inflation, cats, shared UW cycle, common policyholders)</td>
<td>Assess appropriateness and comprehensiveness of approach to dependencies in model; Demonstrate understanding of sources of dependency</td>
</tr>
<tr>
<td>Benchmarking (aggregated vs. total)</td>
<td>Compare risk distribution obtained by aggregating marginal distributions with that obtained from the aggregate data (e.g. CoB bootstrap reserve risk distributions aggregated and compared to bootstrap on aggregated CoB data)</td>
<td>Top-down test on dependencies</td>
</tr>
<tr>
<td>Backtest (aggregated history vs. model)</td>
<td>Aggregate historical experience across risk categories (such as CoB loss ratios) and compare to aggregated output from model</td>
<td>Top-down test on dependencies</td>
</tr>
<tr>
<td>Backtest (joint exceedance probability)</td>
<td>Historical probabilities of pairs of distributions both exceeding a given percentile (e.g. two CoB premium risk distributions both exceeding their 75th percentile)</td>
<td>Provides indication of tail dependencies</td>
</tr>
<tr>
<td>Reverse stress test</td>
<td>Identify aggregations of events that could (nearly) bankrupt the syndicate; compare to SCR distribution</td>
<td>Top-down test on dependencies and SCR</td>
</tr>
</tbody>
</table>
APPENDIX 2: EXPERT JUDGEMENT AND VALIDATION

Expert judgement is relevant to all aspects of the internal modelling process, including assumption setting, governance, documentation and validation. With regards to validation, we will distinguish between validation by expert judgement as opposed to validation of expert judgement. The former relates to validation tools such as stress and scenario tests that rely primarily on expert judgement; the latter relates to the justification provided for expert judgements used to set assumptions and parameters. An analogy can be made with data, which is the basis for validation that tests against experience, but which itself must also be validated for completeness, accuracy, and so on. Validation tools relying on expert judgement are covered in 4.5.2; this appendix will deal mainly with the how expert judgement itself should be validated.

EIOPA has recently issued further “Guidelines” on key considerations for national regulatory authorities when forming a view on how prepared an “insurance or reinsurance undertaking is to submit an application for the use under Solvency II of an internal model for the calculation of the Solvency Capital Requirement” (1.8). Chapter 4 of this document is devoted to “Assumption setting and expert judgement”, and covers how expert judgements are set; the governance of their use; how they are communicated and documented; and how they are validated. The Guidelines are the most up-to-date guidance from EIOPA on what is expected of agents on expert judgement. This appendix will not summarise the entire contents of the Chapter 4, which agents are strongly encouraged to read for themselves. It will instead provide practical suggestions on compliance in the areas that have proved most challenging for agents to date.

In addition, the PRA has released Supervisory Statement SS5/14, which explains its expectations on how undertakings address certain critical areas of their internal models. The relevance of expert judgement to some of these areas is also discussed below. It is the responsibility of all agents to read SS5/14 and ensure that their internal models meet its requirements.

The first requirement is that the use of expert judgement and its validation are documented (1.66). In particular, the documentation must cover “the rationale for the opinion, including the information basis used, with the level of detail necessary to make transparent both the assumptions and the process and decision-making criteria used for the selection of the assumptions and disregarding other alternatives” (1.63). Lloyd’s has also emphasised from the start of the Dry Run the importance of providing a rationale for expert judgements.

Expert judgements are most often relied on when there is a lack of relevant or credible information. Two strategies that may facilitate the development of rationales in these circumstances are: (1) limiting the expert judgements to a few distinct alternatives; and (2) defining the alternatives in qualitative terms, which can then be quantified. The first strategy is reasonable since a high degree of precision is not warranted when data is lacking. Validation should test whether the alternatives represent the full range realistic assumptions. The second separates the rationales, which may be very plausible, from their quantification, which may involve a high degree of uncertainty. It also allows different experts to focus on their area of specialty, whether it is qualitative or quantitative.

To take an example, agents will often specify (say) five alternative dependency structures based on a single copula and five different correlation assumptions (high, medium, and low, plus nil and 100%). Underwriters are often asked to select the correlations. This approach incorporates the first strategy of limiting alternatives, provided that care is taken to ensure that the copula and correlations cover a sufficient range of dependencies. The drawback is that it is difficult to provide a rationale for a correlation (a statistic) except by evidence from data.
An application of the second strategy could begin by defining dependencies in the qualitative terms of strength and shape. The shape alternatives could be defined as stronger in extreme scenarios vs. stronger in non-extreme scenarios. (Guidance would need to be given on what constitutes “extreme”.) The strength alternatives could be defined as no more likely than chance, somewhat more likely than chance, much more likely than chance, and always occurring together. The supporting rationales could be based on predefined considerations. In the case of premium risk dependencies between different classes, these could include common policyholders, a shared underwriting cycle, exposure to the same cat events, etc. The different permutations of strength and shape could be mapped by the actuaries to several joint exceedance probability distributions. This separation between qualitative judgements and their quantification makes the rationales more transparent, while pre-empting the need for underwriters to speculate on the quantitative impact of their judgements.

Another frequently encountered case in which these strategies could be applied is that of modelling reserve risk for a class of business with only a few years of history. The reserve risk emergence could be defined in terms of the speed of run-off (fast, medium or slow) and type of emergence (primarily attritional vs. primarily large or a mixture of the two). Underwriters or claims staff could justify their choices by comparison with classes with longer histories that exemplify the different alternatives. Actuaries could then map the qualitative judgements on speed of run-off and type of emergence to a limited number of distributions. This approach should help to make the rationales more transparent, as opposed to (for example) selecting a COV of a distribution of ultimate claims and trying to justify it with qualitative reasons.

A second requirement from the Guidelines is that material assumptions or expert judgements are updated in response to new information (1.67). Expert judgements must incorporate new information as it becomes available if they are to improve over time. A useful analogy can be made with Bayesian analysis, in which an initial “prior distribution” is selected for a parameter, which is later updated to a “posterior distribution” that gives weight to additional observations that have become available. A prior based on purely on judgement is sometimes referred to as a “weak” or “non-informative” prior; over time, however, the posterior distribution will become more closely weighted to the data. Its variability will also decrease, reflecting the reduced uncertainty resulting from the additional observations. This analogy is not intended to suggest that a Bayesian framework is necessary or even optimal for setting expert judgements; the principle of revising assumptions in light of new information is however the same.

The task of ensuring that expert judgements respond to new information is a non-trivial one, in part because “new information” may consist of only one additional observation a year. For example, in our earlier example of premium risk dependencies between attritional and large claims within a class, each subsequent underwriting year will provide only one new observation for comparison against the selected joint exceedance probability distribution. While this additional observation should be given some weight when updating the expert judgements on which the dependency structure is based, one observation is not a credible basis for materially changing the assumed structure. Moreover, a single observation cannot be used to define the circumstances under which a joint distribution assumption would be considered false (more on this requirement below).

It may be possible to augment the value of the new information by taking advantage of the limited number of alternatives used in the expert judgement setting process, if that strategy (as described above) was adopted. In the example of dependencies between attritional and large claims, the choice was between five alternative dependency structures based on the considerations of shape and strength. Suppose that the same structure was selected for four classes.
The four observation pairs (one for attritional and one for large from each class) in the subsequent year could be treated as independent samples from the same joint exceedance probability distribution for attritional and large claims. (This assumption would need to be revised if dependencies existed between classes for premium risk.) A “success” could be defined as both attritional and large claims in a given class exceeding a certain percentile (say the 75th) of their respective distributions in the next year. All four observations could then be used to test the appropriateness of the expert judgement selected joint exceedance probability distribution. For example, suppose that the selected distribution gives a probability of 10% for both attritional and large claims exceeding their 75th percentiles. The probability of more than one success in four would be 5%. If the test criterion were set at 5%, then observing more than one success would require the original expert judgement to be revised in order to result in a distribution more consistent with the new information. Clearly this approach would not be appropriate if credible data for setting dependencies were available for each class individually. In our example, however, that information is not available. Grouping the observations could provide more insight than treating each observation as relevant only to the class from which it arises.

A third requirement relates to how agents use validation tools such as stress testing and sensitivity testing (1.68). As noted at the beginning of this appendix, these tests are covered elsewhere in the Guidance. They remain central to how expert judgement is communicated and applied in the validation process.

Fourth, the Guidelines require justification of the internal or external expertise used to make the expert judgements (1.69). One pitfall in this regard that should be highlighted is the risk of misunderstanding or miscommunication (1.58) that arises when experts from different disciplines are required to jointly form a view on a particular assumption. In practice this is most likely to arise when underwriters are asked to translate their underwriting experience into a statistic, such as a correlation or a COV. As discussed above, one way to avoid this pitfall is to separate the provision of the qualitative rationales for the expert judgements from their quantification.

A final requirement relates to the specification of occurrences under which an assumption or expert judgement would be considered false (1.70). Satisfying this requirement depends critically on the ability to provide rationales and incorporate new information. If the basis for the expert judgement is unknown, then it cannot be said under what circumstances it would be invalid. Similarly, if new information is not incorporated into the expert judgement setting process, then new information that contradicts current judgements may not be acted on.

Conversely, if these requirements are met, then deciding what conditions would falsify an expert judgement should reduce to a relatively straightforward (but subjective) judgement about the tolerance for error. An example was provided above with dependencies, in which the selected dependency structure would be considered “false” in the event of more than one joint exceedance at the 75th percentile. In other cases the criteria might not be so conveniently quantifiable. In the example of reserve risk for a new class, the two considerations were speed of emergence and proportion of attritional vs. large claim sizes. If the amount of emergence in the subsequent year was in line with expectations, but the proportion of attritional vs. large was not, then a reasoned explanation would need to be given as to why this occurrence did or did not falsify the original expert judgement. The Guidelines require that undertakings be able to “detect” the occurrence of circumstances under which the original assumption was false. Lloyd’s does not interpret this as a requirement to specify all “fail” criteria in advance; it does however necessitate rationales that are sufficiently detailed and explicit to differentiate between a reasonable and unreasonable outcome. “Looks reasonable” or “within expectations” do not meet this test.
The PRA has recently released Supervisory Statement SS5/14, which provides several important messages on the calculation of technical provisions and the use of internal models. We will comment on the parts of the Statement that have direct relevance to the use of expert judgement, in light of the EIOPA Guidelines.

The Statement sets out the expectation that “where there is significant uncertainty around a sensitive parameter…firms should seek to reflect the parameter uncertainty in their choice of parameter value unless they have otherwise quantified and allowed for this estimation in the model” (3.10). In Lloyd’s view, the first step of identifying sensitive parameters with significant uncertainty should be a structured process of sensitivity testing. Note that reflecting parameter uncertainty does not imply adopting the worst case assumptions. This process should identify the reasons for the uncertainty and provide objective criteria (either in absolute terms or relative to the risk type, SCR, etc.) for what is “significant uncertainty”. It should be obvious that any parameter derived primarily from expert judgement (as in the dependencies and reserve risk examples provided above) will have a significant degree of uncertainty. The second step of reflecting the parameter uncertainty in the parameter values is most often done by allowing the parameter values to take on values from a distribution, which may be determined by expert judgement. A commonly used example is that of a Poisson frequency distribution with a mean parameter that is itself a random variable following a gamma distribution. The gamma distribution could be parameterised based on expert judgement on the level of uncertainty on the true value of the mean. The Statement does not require a distribution based approach, and it appears to allow for less explicit ways of capturing parameter uncertainty. Regardless of the approach used, agents should be able to demonstrate that they have adequately accounted for the impact of parameter uncertainty on the SCR. Expert judgement will be relevant both in identifying where parameter uncertainty exists, and how to quantify it.

Firms are also advised not to assume that “parameterising the internal model using only historical data will take into account all quantifiable risks” (3.2). Clearly, expert judgement will be required in cases where historical data is not sufficient. Agents should expect to be challenged on how they have determined whether the historical data is sufficient, and to support any expert judgements that make allowance for its insufficiency.

The above sections of the Statement are most relevant to expert judgement. The Statement also includes discussion of other areas relevant to validation, specifically 3.24 – 3.28. Agents are expected to act on these areas as well.
APPENDIX 3: REFERENCES


