NOMs Methodology: Calibration, Testing and Validation Plan

1. The purpose of this document

The regulatory framework RIIO (Revenue = Incentives + Innovation + Outputs) introduces quantitative measurements on outputs which are defined in a portfolio sense instead of case by case. Before the starting of the RIIO period, a number of output delivery targets are prescribed in the license by the Authority. It is also stated in the license that the performance assessment methodologies need to be developed to enable the Authority to make decisions on the licensees' revenue (adjustment) at the end of the RIIO period.

For the transmission network owners (TOs), the NLR (non-load related) expenditure is allowed by the Authority to replace (and refurbish) the deteriorated assets in order to manage the condition related risks for the transmission assets. The specific output is termed "network replacement output" and "in assessing whether the licensee should be deemed to have delivered a particular Network Replacement Output under [paragraph 2M.3 of] this condition [i.e. the transmission license] the Authority will, amongst other things, take account of any trade-offs between asset categories which the licensee is able to demonstrate has or are likely to deliver an equivalent or better set of Network [Replacement] Outputs to those specified in [Table 1 of] this condition [i.e. the transmission license]" [1].

The modelling and data approach is extensively used during the development of the network output measures, with particular emphasis on the network condition measure, network risk measure, and network replacement output measure (including the trade-offs). It is therefore essential for such modelling and data driven approach be tested and validated before being implemented into the regulatory framework. While the testing, validation, and calibration of the models an enduring practice that implies continuous improvements, the Authority requested a clear testing, calibration, and validation plan, as well as models to facilitate this plan as part of the modified NOMs methodology[1]. Specifically, "the testing, validation, and calibration plans shall include: a) explanation of the approaches to testing, validation, and calibration; b) explanation of the data to be used, including any approach utilising data samples; c) any ongoing work necessary to refine and identify potential improvements to the Methodology; and d) timeframes for testing, validation, and calibration." At the same time, "outputs from the testing, validation, and calibration process shall include: a) identification of any points of weakness in the Model; b) calibrated input parameter values that achieve the requirements of [the Authority]; c) identification of common or licensee specific data gaps; and d) the Methodology must be designed to enable parameters to be easily adjusted to reflect results of the testing, validation, and calibration exercises" [2].

Figure 1 shows a logical flow chart describing the approaches and elements adopted in the modified NOMs methodology to provide a context. The testing, validation and calibration plan detailed in this

[1] In this document, the "current NOMs methodology" refers to the document that is published in Jan 2016 [3] while the "modified NOMs methodology" refers to the document (and supplemental materials) that is in development following Ofgem's direction [2]. Also, in this document the terms "Ofgem" and "Authority" may be used inter-exchangeably. Same applies to the terms "TOs" and "transmission licensees".

document is generally applicable to all the TOs regarding the specific items required in the Ofgem direction documents, e.g. health score, probability of failure, consequence of failure, monetised risk, etc. Any TO specific cases will be explained explicitly in this context. According to the modified NOMs methodology, the asset condition in Figure 1 is represented by the health score which takes into consideration condition information such as defect, fault, failure, condition assessment, age and duty. Those health scores are calculated based on the information which is unique to individual assets. This will be applied to the information that reflects a "representative asset" determined from the FMEA (failure mode and effect analysis) method. In the FMEA process, the failure modes, probabilities of failure, consequences of failure, and probability of each consequence are determined using engineering based workshops. The end-of-life failure mode analysis for the "representative asset" is then combined with asset specific information captured in the health score. The probability of failure, consequences of failure and probability of consequence are modified using the modifier technique to translate the health score to "equivalent age" of each asset. Those none end-of-life failure modes are also translated into asset specific information in a similar way using modifiers such as environment and duty. The consequences of failures can be further translated into cost of consequences using, for example, cost and impact information of historical events, as well as asset specific information such as criticalities of each asset. The probability of failure and cost of consequence will be combined using a risk model to calculate the monetised £risk, the details of which can be found in the modified NOMs methodology.

The purpose of this document is to provide a detailed plan with specific testing, validation, and calibration cases, against each relevant element in the risk monetisation model proposed by the TOs. The document also proposes in as much detail as possible, how the testing, validation, and calibration models work, together with modelling and data implications. The document is drafted such that for each element in the risk monetisation model, a context is provided regarding the functionality and rationales. A testing, validation, and calibration case is then proposed detailing: the logic behind each case; the data source and data collection/cleanse effort required; the previous experiences and overview of existing models and possible new models, together with insights for the possible outcomes of each case. It is aiming to list all the possible testing, validation and calibration cases that can be performed by the TOs. Not necessarily all the cases need to be carried out to demonstrate the compliance, but they should be done in the principle that sufficient confidence and assurance can be provided to the TOs and the Authority. For example, the calibration may be carried out as an enduring process, as "continuous improvement" both within each organisation and across the transmission licensees. For clarification purposes we also define testing, validation and calibration separately in the current context:

- **Calibration** is to develop planned cases /programmes such that the input parameters of the models in the modified NOMs methodology (that are tested and validated) can be derived and continuously improved from the data that is collected by each licensee. The calibration cases/programme shall be developed closely with the testing and validation cases to either revise the data inputs or the choice of models. Models that facilitate the calibration can be developed using data science technologies such as machine learning. For instances in which the information is too limited to provide statistical significance, subject matter experts should provide independent views on the data inputs in a consistent manner (facilitated by decision supporting tools for example). The TOs should also consider establishing data

collection plans and guidance such as structured workshops, stakeholder consultation, and R&D activities. A governance structure of the calibration process will also be created and agreed between the Authority and the affected network licensees.

- **Testing** aims to ensure the elements in the modified NOMs methodology meet the designed specification from the Authority while at the same time ensuring the models and processes capture the key underlying factors, and enables the models and processes to be interrogated using independent datasets and experts. The underlying assumptions, limitations and weaknesses will also be tested using different techniques such as logic test, functionality test, scenario test, sensitivity test, assumption test and extreme value test, etc.

- **Validation** is to provide confidence to the Authority and the TOs that the modified NOMs methodology provides appropriate outcomes for each element of the network output measures. It will examine the outputs using alternative approaches such as alternative modelling, envelope estimation, cross sector comparison, data mining, known risk assessment, etc. Any differences arrived from those approaches will be explained and justified. Failures in compliance will normally trigger the review of the model or the calibration of the model. For continuous improvement purposes, validation may be also performed by comparing the model predictions with future asset information (that will be collected as an enduring process).

A summary table of Testing, validation and calibration is shown below (Table 1):

Interrogation Type	**Model Elements**		**Follow on actions**
	Category	**Examples**	
Calibration	Input	• External input data • Derived coefficients • Licensee calculated input variable values	• Recalibrate input parameters • Seek alternative data sources
Testing	Process	• Function specifications • Assumptions • Correct functioning of the models	• Revise model/process/ function specification
Validation	Output	• Risk values • PoF values • CoF values	• Revise model/process • Recalibrate input parameters

The Authority has previously suggested that the testing, validation and calibration plan is constructed in four main parts: 1) the common methodology, 2) the trade-off model, 3), the specific appendix, and 4) the work plan and time scale, facilitated by an Excel spreadsheet summarising the testing, validation, and calibration cases. Given the fact the NOMs methodology is still under development as this document is currently being drafted, it is proposed that the current version of the document follows the logic order regarding different elements in the NOMs methodology and risk monetisation framework. For each element under consideration, a number of testing, validation, and calibration items are listed and the models for those items are also detailed to facilitate the

future implementation of the plan, including the validation, testing, and calibration cases applicable to the common methodology, specific appendix, and the trade-off model. Scenarios such as early life failure of an asset, scheme cancellation, intervention type change, intervention movement from load related to non-load related, etc. will be considered, and where possible real scenarios will also be considered. The time scale and resource requirement will become clearer at a later stage - due to the fact that the testing, validation, and the calibration time scale can be heavily affected by the complexity/quality of the model/framework being developed - if the model is well developed, the testing and validation cases will fall in place smoothly. The overall time scale and resource requirement will be finalised following discussions and agreements with all the TOs and the Authority. It is acknowledged that some of the elements (e.g. specific appendix) may need to be re-structured during the finalisation stage of the NOMs methodology, in order to be in compliance to the requirements set by the Authority. The current document also proposes a number of comparison studies between the TOs, for example, for the same asset type, similar service history and similar condition, it is expected the probability of failure of those assets should be comparable. Those studies are however subjecting to the confidentiality agreement between the TOs. Or alternatives can be arranged as part of the specific appendices (for example, comparison studies performed by the Authority who has got the visibility of the specific appendix from each TO.

The current document is mainly based on the methodology document that was issued in December, 2016 which may need further development in order to be compliant with the Authority's request. For example, one section not yet included, is the testing of the assumptions according to the assumption log in the common methodology. The current document is drafted in a way that any assumption that can be identified in the current version of the modified methodology will be discussed and a testing model is proposed. The Authority requested an assumption log during the course of the NOMs methodology modification, noting that "[the methodology] shall describe the rationale for any assumptions required for quantification purposes" [2]. It is however difficult to capture all the assumption without the assumption log, the testing cases against which will be detailed in a separate part of the document when the assumption log becomes apparent. It is also worth noting that the testing, validation, and calibration plan will be focusing on the elements in the common methodology that are closely related to the monetised risk modelling and risk trade-off (see Figure 1). Those rule based sections, such as high impact low probability (HILP) events[2], and the justification of over and under delivery, are considered as subject to agreement between TOs and the Authority, and will not be covered by the current document.

It is also worth stating some general principles regarding the data availability and data quality issue. Both the TOs and the Authority have recognised that data is one of the biggest challenges in the course of developing and implementing the NOMs methodology. Most datasets currently retained by the TOs are not gathered for the purpose of performing the statistics required in the NOMs methodology. In most cases, those datasets need to be cleansed, combined, and manipulated in order to perform any asset-related analytics. At the same time, different TOs might have different

[2] I the "direction" document [2], it is required that "[the testing, validation, and calibration plan] will provide sufficient confidence … the approach for assets requiring separate treatment (see paragraph 34) is appropriate, including demonstration of why normal treatment would lead to incorrect results". This justification of the separate treatment is detailed in the HILP section in the modified NOMs methodology under an HILP work stream and will be subjecting to the agreement from the Authority.

levels of data quality and data availability which will impact the ability of the TOs to carry out the testing, validation, and calibration cases. As a general principle, it is proposed to accept the data availability issue and maximise the utilisation of known information sources. This includes data from the system (e.g. fault failure, and defect database), from engineering knowledge (e.g. empirical or physics-based model), from public available information (e.g. cost of failure) and across the TOs and other sectors (e.g. DNOs). Data cleanse, combination and data manipulation effort will be essential to address the data availability issue by recycling otherwise unusable data at the same time any gaps and bias in the datasets can be identified for e.g. future data collection initiatives. The processed data, although (likely) still far from ideal, will provide a baseline to use other techniques such as expert judgement or simulations. The data gathered from the FMEA workshops can be also used to further fertilise the analyses and provide invaluable insights into the data gaps. The issue of different levels of data availability can be compensated by data sharing and comparison across the TOs, subject to confidentiality agreements. On the other hand, the existing data that is too spare to perform statistical analyses may be sufficient for the purpose of testing, validation, and calibration. For example, the cost of consequence may be difficult to obtain for each individual assets but the overall cost of failure in the whole system can be calculated by aggregating from the asset level (assume this information is already developed using the models described in the modified NOMs methodology document[4]) and compare with the (sparse, at the system level) cost data. The asset level cost can be then calibrated from a top-down fashion, providing the model in the methdology[4] can be verified to give the correct cost ratio between assets. In the current document, the data sources, data cleanse and manipulation effort and techniques, and any other implications are detailed and discussed specifically for each case where appropriate.

2. Asset health index and health score

2.1 Introduction

For the modified NOMs methodology, a quantitative approach is required by the Authority while the qualitative elements in the current network condition measure (i.e. asset health) will be removed as much as possible (not that pure qualitative methods are used in the current NOMs methodology). This is explicitly specified as part of the "objectivity" principles: "The Methodology will be unambiguous and enable any two competent independent assessors (with access to the same input data) to arrive at the same view of licensees' performance (over-delivery, under-delivery, or on target delivery) and to identify and quantify the relevant factors contributing to performance [2]". The translation from qualitative elements to quantitative elements will inevitably lead to some inconsistencies between the asset health indices (AHIs) which are described in the current NOMs methodology and the health score in the modified NOMs methodology. This is in line with the Authority's note that "in designing the Methodology, licensees must not be constrained by trying to arrive at the same replacement priorities as indicated by Table 1 (Replacement Priority Outputs) of [the transmission license section] 2M" [2]. In the sense of portfolio risks, the modified health score methodology must ideally be compatible from the current target to ensure that "the Methodology shall enable the translation of existing RIIO-T1 (volume based) replacement priority targets to equivalent monetised (or alternative) output targets and shall provide the basis for setting targets in future price control periods. [2]". Such compatibility is also essential in terms of the past and future capital investment plan for T1, which is largely based on the asset health index [and criticality, which will be discussed in a later context] and the Replacement Priorities: "…the NOMs Methodology Objectives are … the assessment of historical and forecast network expenditure on the licensee's Transmission System …" [2].

In the modified NOMs methodology, a single health score will be assigned to each asset and is calculated as a function of the asset specific information such as condition data, performance, age and duty. It is defined as a numerical score, which is an integer rising from 0 representing good condition assets to a maximum value[3] representing the state requiring replacement, i.e. end-of-life. From the modelling point of view, the health score framework is a classification model that differentiates the assets using a large array of known information such as condition and duty. These data which normally comprise many different dimensions, need to be merged into a single health score measure using classification modelling techniques. On the other hand, the asset health index is also a classification model that categories the assets into a much smaller number of categories using a (largely qualitative) method base on the same amount of information (at the same point in time). The condition information can be further enriched in terms of volume and variety due to e.g. data collection exercises and innovative condition monitoring technologies. For example, as described in the modified NOMs methodology[4], the health score for a circuit breaker will be a function of anticipated asset life, age and duty, performance history, i.e. accumulative fault current, number of defects, and asset family specific information such as expected probability of failure (that is expressed as a probabilistic function). The relative importance or combined contributions of the factors to the health score are then modelled by assembling them using a formula (or a group of

[3] The (numeric) value of the maximum number may be different depending on the type of assets

[4] As per version 05 Dec. 2016 - there might be inconsistencies between this document to a later version of the NOMs methodology.

formulae) that applies to all the circuit breakers. The assembling formulae have got a series of variables the value of which are either asset specific information such as condition and duty or asset family specific information such as anticipated asset life. Apart from those variables that quantify the information needed for the health score, there are a large number of coefficients (e.g. relative weightings between two condition indictors) modelling how those variables can be combined together. Those variables, coefficients, and the constitutions of the formulae, (i.e. how the variables and coefficients are assembled together), will need further testing, validation and calibration to meet the criteria discussed in the preceding paragraph.

2.2 Calibration of health score

Assuming the formulae have got all the ingredients (represented by a number of variables discussed above) needed to calculate the health score in the correct form (this assumption will be discussed in-detail at a later context), the coefficients involved in the formulae can be calibrated using known constraints between the asset health and health score. Those coefficients can then be calculated by maximising/optimising the compatibility between the asset health index and health score. By definition, such compatibility requires a mapping scheme such as "when a health score is between a certain value and another certain value (i.e. thresholds), the asset health index will be a certain value between 1 and 4", i.e. a many-to-one mapping. More generally, such mapping scheme can also be probabilistic (many-to-many mapping) aiming to provide an equivalent portfolio of risks. (Precisely speaking, risk is subjected to different types of interventions and other elements such as the criticality and cost consequences. These are all constraints needed for an equivalent portfolio of risk and will be discussed in a later context). For most applications a deterministic and many-to-one mapping is adopted for simplicity reasons [5].

Calibrate the coefficients and the models

One calibration model can be designed using computer aided search in the parametric space of the coefficients. In the case of overhead line conductors, for example, the ingredients/factors contributing to the health score include (as detailed in the modified methodology document[4]): anticipated asset life (which includes the design, environmental variations), age, number of repairs, fittings defects, conductor sampling information (including corrosion, fatigue, grease level, broken strands, tensile tests and torsion test) , etc. Those ingredients/factors are quantified by the variables in the formulae, where the coefficients can be obtained using an optimisation method. The parametric space of the coefficients, the mapping rules (between asset health indices and the health score), and the aim is to optimise the compatibility between asset health indices and health scores, for known individual assets or for the whole risk portfolio. The performance of each combination of the parameters can be obtained, measured by the % matching between the asset health indices calculated from the health score. The best performing ones are extracted and reviewed by the subject matter experts, to avoid any over-fitting of the data.

Based on a similar principle but in a more generic framework, a machine learning approach can be adopted to optimise the compatibility between the health score and the asset health index (at the same point in time). Figure 2 [6] shows the performance of a generic machine learning model, in which the health score formulae are modelled by random decision trees (random forest method) instead of a group of formulae with given formats. The coefficient and the mapping rules are then

subjected to calibration by comparing the health score with the asset health indices for the whole asset portfolio. Many-to-many mappings between asset health indices and health scores are allowed. Two best performing models are shown in the figure for scenarios where conductor sampling information is available/unavailable. It can be clearly seen from the figure that 1) 100% compatibility between the asset health indices and the health scores is difficult if not impossible to achieve. This may be attributed to the inconsistencies when making those qualitative asset health indices decisions; 2) the relative importance[5] of the factors contributing to the health score/asset health indices may change according to whether the sample information is available; 3) the historical condition information (i.e. previous years' asset health) makes a contribution to the current asset condition information (which is not currently modelled in the health score framework); and 4) those information can be fed back (assessed or reviewed by) to subject matter experts to confirm that is how the decisions were made while the asset health indices were scored, as a validation to those machine learning models.

Calibration case 2.2.1: Use computer aided parametric search to calibrate the coefficients aiming to maximise the compatibility between the health score and the asset health index. The calibration results will need to be reviewed together with the model by subject matter experts.

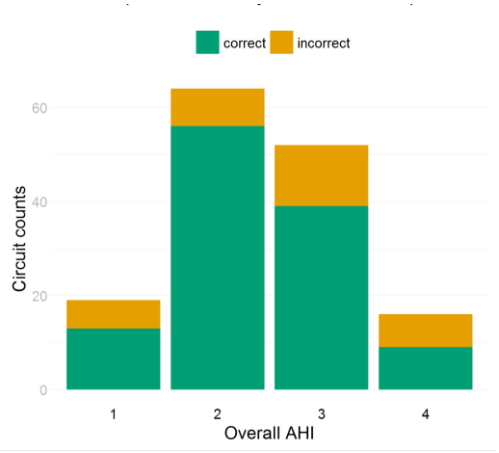
Calibrate anticipated asset life

An important input used in the modified methodology[4] to calculate the health score is the anticipated asset life, which can be either obtained from the asset policies of each TOs or calibrated from the asset end-of-life data (subject to data quality issues, etc.). One of the criteria of having the anticipated asset life in the asset policy is to ensure the asset end-of-life behaviours are truthfully and consistently reflected in the asset condition at the state requiring replacement. It is proposed that wherever possible, asset end-of-life data should be used to calibrate the asset life of the TOs. The advantage of using a data-driven approach is that it ensures the asset policy has got data to back it up. At the same time, it enables a continuous improvement when more data is available. Theoretically, the anticipated asset life can be obtained by the age that 50% of a large number of assets of the same type operate until the point of end-of-life failure. An assumption is normally made that the other half will fail following a certain probabilistic curve such as a Weibull distribution[6]. However, many asset types have not been through the half life cycle, and in the electricity transmission industry, the well-known asset management philosophy is to replace the assets before failure. This type of data can be enriched by forensic analysis after decommissioning to determining the remaining useful life and hence the anticipated asset life of this asset type. For many asset types however the post-decommission analysis is not always available for each TO. At such instances, international database and collective opinions from subject matter experts can be used to fulfil the gaps in the dataset. In order to calibrate the asset life, it is proposed to maintain the current asset lives that are specified in the transmission licensees' asset policy, unless strong data-driven evidence suggests otherwise. A comparison study among the transmission licensees' asset policies will be performed with the presumptive principle that two assets of the same asset type with similar operational regime and history, should have the same anticipated asset life. In case of inconsistencies, various optimisation techniques can be used to calibrate asset life that minimises

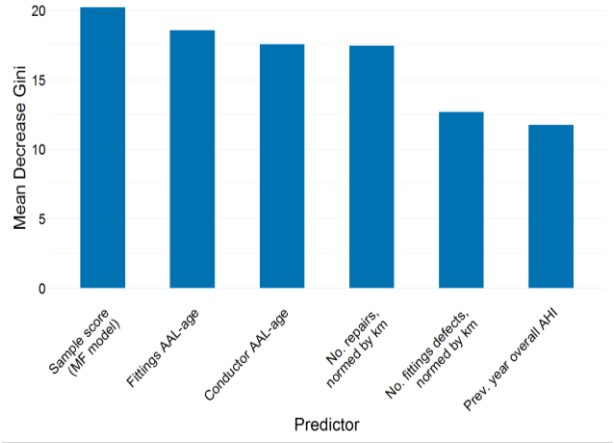
[5] The relative importance is measured by the Gini ratio the details can be found in http://www3.nccu.edu.tw/~jthuang/Gini.pdf

[6] The choice of the distribution will also be discussed in a later context of probability of failure

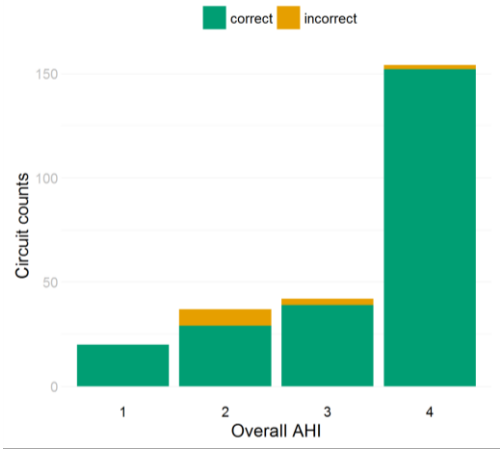
the impact on transmission licensees' capital programme while retaining an equivalent network risk (between the current methodology and the modified NOMs methodology).



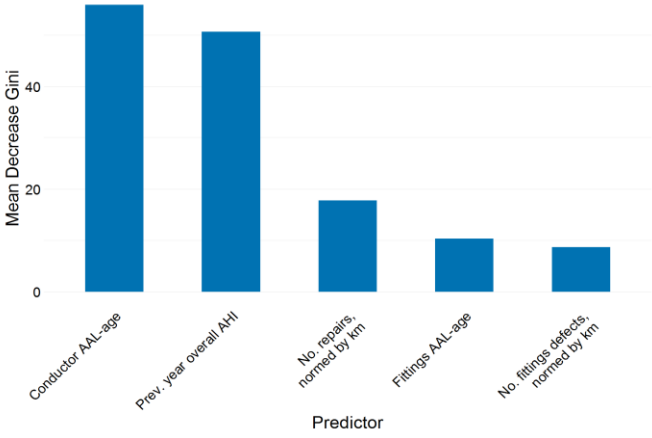
(a) Performance of model and calibration (conductor sample available)



(b) Relative predictor importance (conductor sample available, the relative importance is measured by the Gini ratio)



(c) Performance of model and calibration (conductor sample available)



(d) relative predictor importance (conductor sample not available, the relative importance is measured by the Gini ratio)

Figure 2 An illustration using machine learning method to carry out the modelling coefficients, by comparing the health score and the asset health index of overhead line conductors. (a) and (b): the accuracy performance and the predictors with relative importance, when conductor samples are available; (c) and (d): the accuracy performance and the predictors with relative importance, when conductor samples are not available.

Calibration case 2.2.2: Calibrate asset life using data-driven fashion, where possible. The data quality issue can be mitigated using international databases and expert judgements, etc.

Calibration case 2.2.3: Calibrate asset life by comparing the same asset type across the TOs to enable a consistent approach.

2.3 Testing of the health score

The above calibration methodology assumes that the skeleton of the formulae that lead to the health score is correct, i.e. including the correct ingredients and the correct formulation. According to Table 1, the current testing exercise is used to make sure that the ingredients and the format, e.g. calculus and functionals[5], are correctly modelled, without knowing beforehand the values of those coefficients. It is therefore an iterative process in that the coefficients, calculus and functionals, are inter-dependent that when one is updated, it is essential the others be re-calibrated and re-tested.

Test the factors and ingredients

The starting point is to test whether the health score formulae include the right ingredients. This is a relatively straightforward task as those ingredients are the factors considered in the asset health review to determine or update asset health indices. A comparison study between the formulae in the modified methodology and the asset health review process in the existing NOMs methodology can identify whether the right ingredients are used in the formulae that generates the health score. A similar type of testing serving the same purpose is to use the machine learning method. As shown in Figure 2 (b) and (d), the machine mimics the logic that humans used to generate the asset health indices. The predictors identified from the machine learning exercise should be able to provide assurance that the health scores are calculated using the right ingredients.

Testing case 2.3.1: A comparison study between the current and modified NOMs methodology to ensure the current ingredients are used in the health score formulae.

Test the health score formulae

The format (e.g. calculus, functionals[7]) in the health score calculation is less straightforward to test. Unlike the many approaches used in natural science, where formulae are derived deductively from principles of science whereby the format of the functionals is automatically guaranteed, the modified NOMs methodology takes an inductive approach in which the formulae are constructed in a relatively empirical manner. Those empirical models, by definition, depend for their success and validity on how much they can be used to explain (all) the phenomena observed in the designed domain (i.e. the model applicability). On the other hand, the success of the empirical models is very much dependant on how much and to what extent the underlying mechanisms are captured in those analytically explicit formulae. Testing cases using the outputs from the health score models can be designed following a procedure that: Firstly, part of the asset health index data is used to calibrate the coefficients in the health score formula using calibration case 2.2.1; then, the health score formulae are used to calculate from the remaining part of the dataset, to calculate the asset health indices from the health scores (and the mapping rules); finally, the calculated asset health indices are compared with the existing asset health index in the dataset and the percentage accuracy can be measured to test the formulae in question.

[7] A functional can be understood as a generalised function and is used here to represent a group of functions or a function of functions

Testing case 2.3.2: Use part of the asset health index data to generate the coefficient and use the rest to test the formulae using the same set of coefficients.

Test the assumptions

Another important aspect of testing is to make sure the assumptions made while developing those empirical models are acceptable and largely representative of the reality. While a complete assumption log is not available at this point of time this document is drafted, examples are used in this document, where appropriate, to provide insights to the models, and to demonstrate how assumptions can be tested using different techniques. The first example is the health score formula of the circuit breaker (see the modified methodology document[4] for details). It has got an adjustment term to take consideration of the defects using the "number of defects raised against a specific asset" divided by "average number of defects for that asset's family". This treatment implicitly assumes that the number of defect is the leading factor for defect measurement. The extent, type, and severity of those defects are relatively uniform or distributed in a way that an average value can be representative to the whole distribution. Such an assumption can be easily tested using some deep-dive analyses using real defect data. Depending on how structured the defect data is recorded for each TO (that is, whether the defects are recorded in the database with extent, type, and severity clearly categorised or simply a free text of description), those assumptions can either be tested directly (for structured data) or some data mining (e.g. text matching or manual data cleanse + machine learning) techniques is required before testing this assumption.

Testing case 2.3.3: Test the assumptions (this is only an example of the assumptions) of the defects in the health score formula of a circuit breaker, i.e., the number of defects is representative to the impact of the defect to the asset condition. Defect database can be used in this testing case and some data cleanse effort may be required depending on how defects are recorded for each TO.

Another assumption[8] is the general format of the formulae adopted in the health score calculation. By examining the health score methodology, the functionals of those formulae can be generally attributed to three types: (1) $\max\ (c_i x_i + c_j x_j + \ldots + c_n x_n)$; (2) $c_i x_i + c_j x_j + \ldots + c_n x_n$; and (3) $(1 - \bar{x}_i)(1 - \bar{x}_j) \ldots (1 - \bar{x}_n)$, in which i, j, …, n are indices representing different condition indicators or factors that are related to conditions; x is the variable value of each factor, the "bar" on top of a symbol represents normalisation against its maximum possible value (i.e. percentage))and c the coefficients of each corresponding variable. From the informatics point of view, a valid health score model should take all information from all the condition (related) factors while is able to differentiate asset conditions (and risks) for different assets, at least from the probability of failure prospective (which will become relevant in a later context). Qualitatively speaking, formula (1) indicates that all the condition (related) factors are either correlated at least to one other factor while there must be one factor contains all information that is given by other factors, or, if uncorrelated, they must be critical to the asset conditions therefore a poor factor will automatically lead to poor overall condition and other information does not really matter. Formula (2) indicates that all the condition (related) factors are uncorrelated and the only way to have a full picture is to sum those (weighted) factors. Formula (3) takes an approach similar to the reliability of series system, suggesting that those condition (related) factors are uncorrelated, but any one of those

[8] Use NGET model as an example here while in the final version of the methodology all those assumptions should be recorded in the assumption log and tested accordingly.

factor being poor will lead to an overall poor condition while more than one factor being poor can lead to an even worse (poor) condition. Those three functionals are used separately or jointly for different asset types according to different design, operation, and degradation characteristics. The assumptions listed above can be tested against the condition (related) factor data. This includes (1) correlation study of data between different condition (related) factors; (2) correlation study between each condition indictor with the asset health indices (which are already established in the current NOMs methodology); and (3) any other asset specific features in terms of the overall condition and whether those features can be represented through the formulae. Another possible method of testing those formulae is by guided interview/workshops/questionnaires with independent engineering and subject matter experts, for example, giving extreme values of the factors, or using known cases of outliers to test the capacity and limit of the health score methodology. Finally, the classification modelling of condition information for specific types of transmission assets is an active research area with a number of published international standards and literatures. References can be made during the testing exercise.

Testing case 2.3.4: Test the assumptions of the formulae, using correlation analysis of the condition data where possible. Justifications from subject matter experts, international standards, and published references can be also used to test/justify the assumptions.

2.4 Validation of health score

General validation and verification

Given a set of inputs and the formulae, the outputs of the health score models can be calculated for each asset. Those health scores should be largely consistent with the asset health indices described in the current NOMs methodology, as guaranteed through the testing and calibration methods described above. In the validation section, it is essential to firstly have assurance that the model is developed according to the designed specification (verification). The direction document from the Authority provides a list of specifications that can be verified against. The outputs from the model can be also compared with the existing asset health indices to provide further assurance but it should not be used as the sole validation case to avoid self-compliance. Particularly, those asset health indices that can't be accurately predicted by the health score methodology need to be reviewed on a case by case basis, together with the subject matter experts. While some inconsistencies may be generated due to the inconsistencies between decision makings, this type of inconsistency should be largely acceptable as the inconsistencies are expected to cancel (at least in a certain degree) in a portfolio sense. Should some inconsistencies be regarded as due to systematic misalignments between the quantitative method (i.e. health score) and the qualitative method (i.e. asset health index), further development will be required to identify those groups of assets for which special treatments may be arranged. The testing and calibration cases detailed in earlier sections will also need to be reviewed for those asset groups.

Validation case 2.4.1: A general validation study to verify that the inputs and the model are constructed according to designed specifications (e.g. the direction document [2]) and, at the same time to review and understand the inconsistencies between health score and the asset health index.

The DNOs validation methodology

The DNOs have published a brief case study to demonstrate the validation methodology for the monetised risk outputs [5]. In the validation methodology, the DNOs validate the NOMs methodology against known risks in the (existing) risk assessments and also different asset types are compared in order to validate that the outputs are as desired. It is essentially a process of back-tracking those factors and to evaluate their impacts on the outputs through the analytically explicitly formulae. The DNOs' validation adopts techniques such as visualisation, root cause analysis, comparison study, and (partial) scenario tests. It is performed on a population basis and can be applied for all aspects of the NOMs methodology. In the current section we only focus on the validation cases for the health scores and the similar methods will be adopted on other elements of the modified methodology in the later contexts.

Firstly, the key factors/ingredients of the asset type under consideration are listed, including values of all the coefficients required in the health score calculation. Then the health score for each asset is plotted and visualised, showing the volume of assets in each asset health category or in each health score (as illustrated in Figure 3). It shows the condition profile (the risk profile will be completed when taking the criticality or monetised consequence into consideration) in a similar way that is used in the current NOMs methodology. A forecast can be also visualised to show the (year-on-year) future projection of the condition profile. These risk profiles will be reviewed, compared with the condition profiles generated using the current methodology, as well as with known risk assessments to the TOs. For example, if it is known that a particular type of assets are more vulnerable to, using cables as an example, leakage, it is expected that the leakage condition information will be correctly reflected in the health score calculation and drive a poor condition profile for this type of assets. Finally scenario tests can be performed for a) breakdown the asset portfolio in various ways, for example, age group, to understand the interactions between the condition (related) factors (here we consider age is a condition related factor); and b) forecast using different investment scenarios to understand and validate whether the underlying investment scenario gives the condition profiles as expected.

Validation case 2.4.2: Carry out a similar validation study following the DNOs' validation methodology. This includes visualisation, root cause analysis, comparisons against known risk assessments, as well as scenario analyses.

<u>Sensitivity analysis</u>

Another validation method is sensitivity analysis[9], which is defined as the study of how the uncertainty in the output of a mathematical model or system (numerical or otherwise) can be apportioned to different sources of uncertainty in its inputs [7]. The Authority requested that the uncertainties need to be sufficiently considered and that "the Common Methodology shall: a) explain where relevant how uncertainty has been accounted for and resultant confidence intervals around the main output parameters; b) explain any adjustments or allowances necessary for varying levels of uncertainty in input data, e.g. due to differences in regularity of inspection and condition assessment, and c) explain how licensees can provide assurance that despite the levels of uncertainty …". [2]. At the current stage (when this paper is drafted), the uncertainties around various elements in the health score are not yet clearly articulated in the modified methodology

[9] Strictly speaking, a full sensitivity analysis covers the calibration, testing, and validation while in this paper we put all the sensitivity related analyses in the validation section for a clearer structure of the document.

document[4] but some level of uncertainty will be assumed around each of the asset condition (related) factors in the current context.

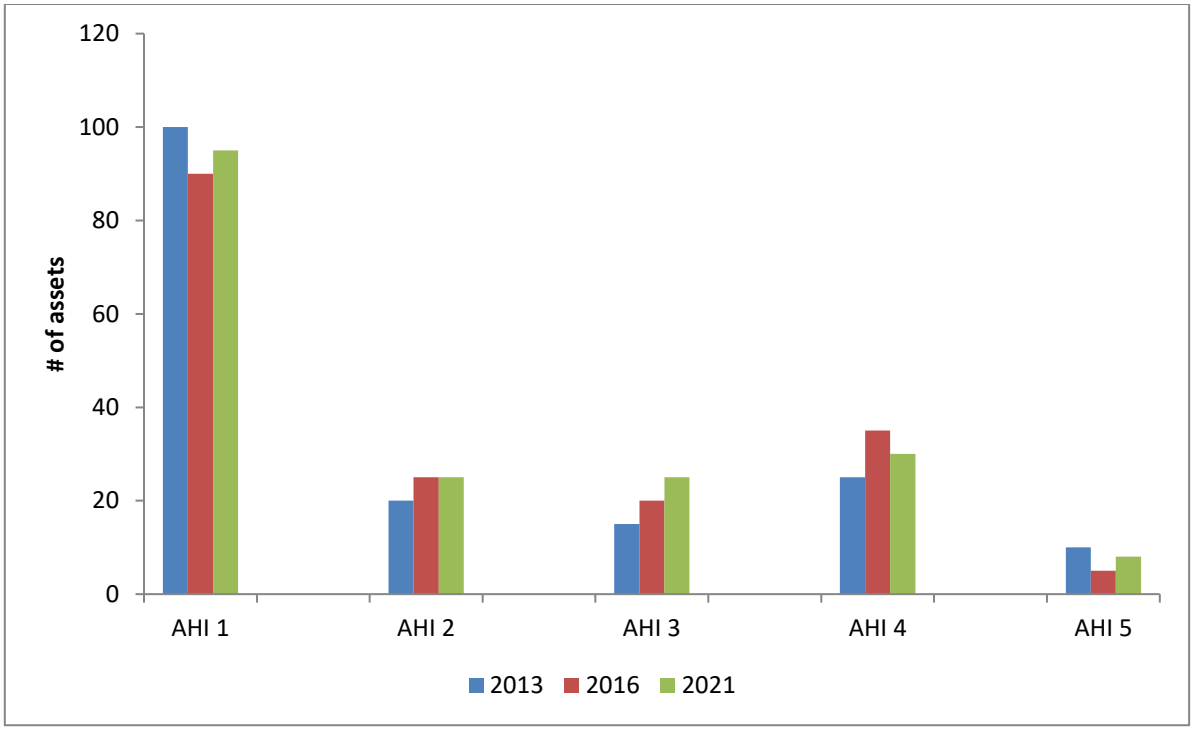


Figure 3 Visualisation of the volume of assets in each asset health category for a number of different years, according to the health score methodology. The risk profiles across the years can be explained by known risk assessments and risk mitigation activities such as deterioration and asset replacements. Reproduced in a similar way that is used in the DNOs validation document [5].

Uncertainty from inputs

The first uncertainty analysis to be performed is to assign a level of uncertainty to each of the input factor in order to evaluate the overall effect of uncertainty on the health score. Different factors will have different uncertainties, which can be measured statistically (what statistical method is used will depend on the availability, format, quality and nature of the data). A simple example is to assume the condition measure is not 100% accurate and a small standard deviation is assigned. Such standard deviation can be further modelled as a function of time since last inspection was performed (i.e. the condition will become more uncertain with time until a new condition assessment). The impact of the uncertainties on the overall health score can be calculated using Monte Carlo analysis, in which each of the uncertainties is a (assumed to be uncorrelated) random value following the standard deviation (following a normal distribution, the variance of which can be drawn from data, or, if data is unavailable, can be treated as a variable subject to parametric study), which will lead to different values of the health score after a large number of Monte Carlo simulations (rule of large numbers). The distribution of those Monte Carlo outputs will represent the overall uncertainties of the health score. The Monte Carlo technique can be also applied to the whole asset population to evaluate the portfolio uncertainties and will be used later for evaluating the uncertainties of other elements in the modified NOMs methodology (e.g. monetised risk).

Validation case 2.4.3: Apply Monte Carlo simulation to evaluate the uncertainties of health score.

Material impact of uncertainties

It is not straightforward (at least judging from the current version of the modified NOMs methodology) to answer the question: "explain any adjustments or allowances necessary for varying levels of uncertainty in input data" [2]. If the uncertainties are associated with financial costs (material impact), then the value of the associated uncertainty need to be calculated as part of the modified NOMs methodology. It is not the objective of the testing and validation plan to come up with a methodology to assign a cost to the uncertainty. However it is regarded that those costs can be validated against the cost of reducing the uncertainties – the cost of uncertainty must be somehow balanced with the cost of collecting ad hoc condition information to remove those uncertainties. Those cost data is generally well recorded in the work orders and is relatively easy to obtain using traditional costing models (men hours, access cost, material cost, outage cost if applicable, etc.)

Validation case 2.4.4: Evaluate the uncertainty cost proposed in the modified NOMs methodology with the historical cost data that is used to collect condition information.

The uncertainty from the model itself

The final source of the uncertainty is the uncertainty of the model itself, including the functionals and the coefficients. According to a recent study performed against a collection of widely used risk based models, the uncertainty of the model itself can be significant when the input data are moving away from the normal operational regime, i.e. to a stressed operational environment [8]. It is proposed that a stress test is applied to the model to test its performance and uncertainty where the inputs are artificially stressed. Each input can be stressed in an isolated manner for its stress-response from the overall health score. Also it might be reasonable to assume that the network is able to sustain if the condition factors are poorer overall than expected (e.g. 10%). Techniques detailed in case 2.4.2 can be used to understand different scenarios. The stress test of the model can be also used in combination with case 2.4.4 to see whether poorer condition (related) factors will lead to a higher or lower uncertainty level for the overall condition profile.

Validation case 2.4.5: Stress test the health score model(s) using isolated tests for each input, scenario tests, and uncertainty test for the portfolio condition profile, where appropriate.

3. The failure mode and probability of failure

3.1 The failure mode and effect analysis

A transmission asset, as the unit for which the majority of asset management activities are planned and carried out is comprised of an assembly of 'items' which perform a defined function. A failure mode, in the current context, refers to a transmission asset 'item' that is no longer able to fulfil its designed functionality. Each failure mode can be attributed to several different underlying causes and lead to several different failure effects. For the purpose of risk monetisation, the modified NOMs methodology has been focusing on modelling the failure effects (instead of cause) using the technique of failure mode and effect analysis (FMEA). In the National Grid method, an FMEA exercise involves reviewing as many components, assemblies and subsystem as possible to identify failure modes and effects [9]. For each component, the failure modes and their consequences, as well as resulting material impact (on the system, safety, or environment) are recorded in specific FMEA worksheets. For the Scottish TO's method the failure rates associated with an asset's failure consequence are determined from a combination of worldwide and specific TO asset operators historic failure and/or forensic data and knowledge networks rather than attempt to quantify every failure mode for every component as it is recognised that these assets are rarely allowed to operate until failure occurs. Criticality of individual assets can be also included in the analysis (FMECA, failure mode, effect, and criticality analysis) to differentiate assets of the same type but, for instance, in different locations (physically or in the system topology). In this document, unless necessary, we do not differentiate FMEA and FMECA. More details about the FMEA can be found in the standards SAE-ARP 5580, IEC 60812, BS5760-5, and MIL-STD-1629A.

The procedure for how an FMEA exercise is carried out for transmission assets is described in the modified NOMs methodology[4]. Each asset type is broken down into component level.

For NGET, known failure modes documented in existing TOs' knowledge database (including both data in the system and expert opinions) are used to provide a basic list of failure modes associated with each component. A failure mode describes an instance that a component fails to operate. All failure modes identified for the asset type are then assembled together to reflect the consequence of the failures (trip, disruptive failure, etc.).

For SPT & SHET the failure rates associated with an asset's failure mode are determined from a combination of worldwide and specific TO asset operators historic failure and/or forensic data and knowledge networks. Rather than attempt to quantify every failure mode for every component, the effort to establish a credible failure rate is aimed at the consequence of the failures (trip, disruptive failure, etc.).

The effects of those consequences are also recorded in the FMEA from financial, system, safety, and environmental perspectives. It is worth noting that the terminology can be slightly different from the current NOMs methodology and the modified NOMs methodology and can also differ between the individual approaches adopted by the TOs. In the FMEA approach used by NGET, the failure modes describe the components which upon ceasing to operate, lead to the consequence representing different failure types (trip, damage component, disruptive failure etc.), which further lead to materialised consequences (injury of personnel, pollution, loss of supply, etc., which will be discussed in the cost of consequence work stream instead of the FMEA). Comparing with the current

NOMs methodology [3], the FMEA approach determines the failure mode at the component level, which is the level at which different types of intervention are applied. To avoid confusion, in this document the consequence used in FMEA (trip, disruptive failure etc.) is termed as an "event" while the consequence referred in the current NOMs methodology (injury of personnel, pollution, loss of supply, etc.) is termed as "material consequence" of the "event". An illustration of the relationship of the terminology in different contexts is shown in Figure 4, in which the "multiply symbol" represent the monetised risk calculation (£risk = probability of failure × £consequence).

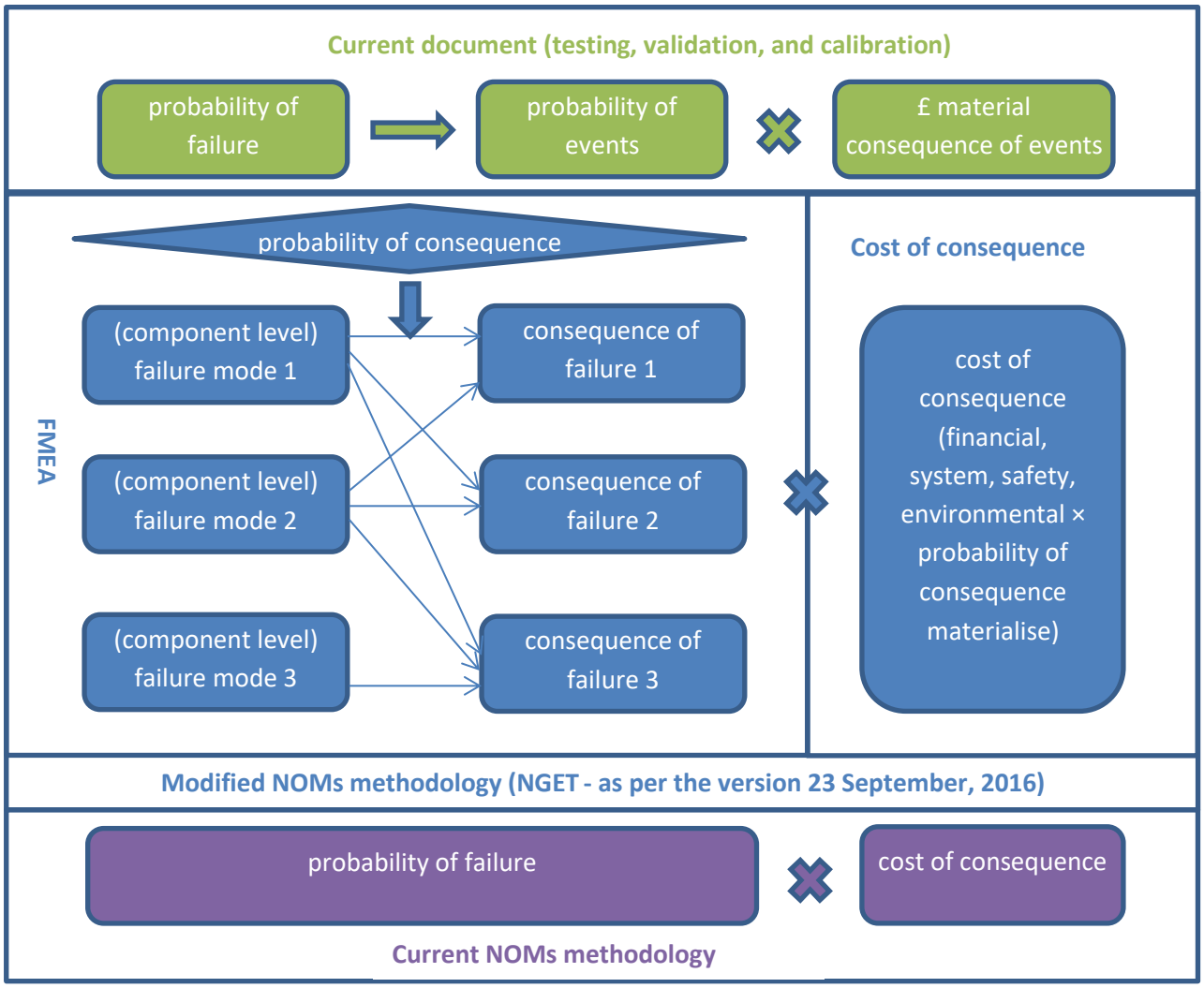


Figure 4 Illustration of the terminologies and definitions in the monetised risk modelling in different contexts. The vertical alignment between terminologies represents equivalent definitions. The "multiply" symbol represents the calculation of monetised risk: (£risk = probability of failure × £consequence).

NGET uses a workshop based approach is used to understand and record answers to questions such as [10]: "how can each component conceivably fail?"; "what could the effects be if the failure did

occur?"; "is the failure in the safe or unsafe direction?"; "how the failure is detected?", and so forth. The key outputs from the FMEA workshops include: failure mode, consequences (referred to as events, see Figure 4), the probability (likelihood) of each consequence (event) happening, detection method of failure (for each failure mode, where possible), the probability of that detection being effective, earliest onset, anticipated, and latest onset time to failure (for each failure mode), as well as any risk reduction before options (e.g. inspections) and after failure (e.g. replacement). The rationale behind the FMEA workshop is that the TOs won't run the transmission assets until failure therefore it is impractical to use failure data for all failure modes. It is however important to look into the available failure data to provide validations and confidence, during the cause of testing, validation, and calibration of the methodology.

Recall that the definition of testing (Table 1) is focused on the process between input and output: thus it might be difficult to test the workshop process directly. However, a comparison study may be performed between the worksheets used in the TOs FMEA exercise and those recommended from the literatures/international standards. The differences should be fairly justifiable while if significant deviation found some workshops need to be re-held. However it is regarded as a minimal risk as long as the outcomes from the workshop can be readily validated. No testing case will be performed directly against the process of the FMEA workshops except any assumption declared in the assumption log (due to the unavailability of which only examples will be discussed in the current document).

<u>Validate against defect, fault, and failure data</u>

The most obvious validation is that the outputs from the workshop could be validated against known datasets such as defect, fault, and failure record in 10 years' history (as per the "direction" from the Authority [2]). Firstly, the instances in the defect, fault and failure database needs to be cleansed and mapped against asset information such as type and age. Depending on the details of the database, the defect, fault and failure information can be validated against the failure modes and/or events. For each line (assuming one line record per one instance of defect, fault, or failure) in the defect, fault, and failure database, a failure mode and/or an event type needs to be assigned to enable further statistical analysis. It can be time consuming depending on the quality of the source data. Some techniques may be used to help in cleansing/combining/categorising the large database after certain amount of manual work is done (to train the algorithm). A statistical analysis can be performed on the cleansed database to validate (a) whether the failure mode and event types list identified from the workshops is complete; (b) give an indicative understanding of the failure rate and/or frequency of the events (e.g. order of magnitude, upper or/and lower limit), that can be further used to validate the probability of failure/probability of event (to be discussed in the next section) for each failure mode and event type (subject to data availability); and (c) provide reference figures for the means of detection of failure, as well the probability of detection (see section 3.4).

Validation case 3.1.1: Where possible**,** use historic defect, fault and failure data to validate where captured the failure modes, event types, probability of failure, probability of events, as well as providing further understandings and assurances to the outcomes of the FEMA workshops. This exercise requires data cleanse and data pre-processing which can be time consuming and labour intensive.

Data quality and calibration using defect, fault, and failure data

Note that the data might be biased due to decommissioned and newly commissioned assets over the past 10 years (i.e. non-stationary asset population base). The data is expected to have a substantial data quality issue due to the way those defect, fault, failure data has been historically recorded - they are recorded mainly for the purpose of raising work orders and normally consisted of a large amount of free text and are not necessarily consistent over the 10 years. It is advised that any differences obtained from the workshop and the data triggers a re-calibration exercise to run through the results with subject matter experts to understand the differences and re-calibrate the necessary parts of the outputs from the FEMA exercises.

Calibration case 3.1.2: Differences from validation case 3.1.1 will trigger a review guided by subject matter experts and possibly re-calibration, where captured, of the failure modes, event types, and corresponding probabilities.

Earliest onset, anticipated, latest onset time to failure and end-of-life failure mode

The failure modes can be categorised as end-of-life type of failure which normally triggers the de-commissioning of the whole asset and none end-of-life type of failure which can be rectified by repairs or partial replacement. The end-of-life failure mode have been documented relatively well in each TO's policy statements and is also used in the current NOMs methodology (shown in Figure 5). A validation case can be performed by comparing the current NOMs methodology with the modified NOMs methodology. Any inconsistencies between the FMEA workshop and the current NOMs methodology will trigger a re-calibration exercise to align the FMEA and the TOs policy statements. It might be more difficult to estimate the end-of-life curve (shown in Figure 5) directly from the failure data as the TOs do not run the asset to failure. For some asset types, the failure data, de-commission data, as well as the forensic data (that estimated the remaining useful life at the time of de-commission) can be used together to perform a statistical study to validate the earliest onset, anticipated, and latest onset of the asset end-of-life (for example, NGET has detailed transformer data). For those asset types for which de-commissioned without forensic analysis, the statistical analysis is likely to give a lower bound of the asset life (under-estimate). Note that there might be statistical bias due to the censored datasets (e.g. a large amount of assets are neither failed nor de-commissioned), in which case statistical techniques such as Kaplan-Meier estimator can be used to correct some bias [9].

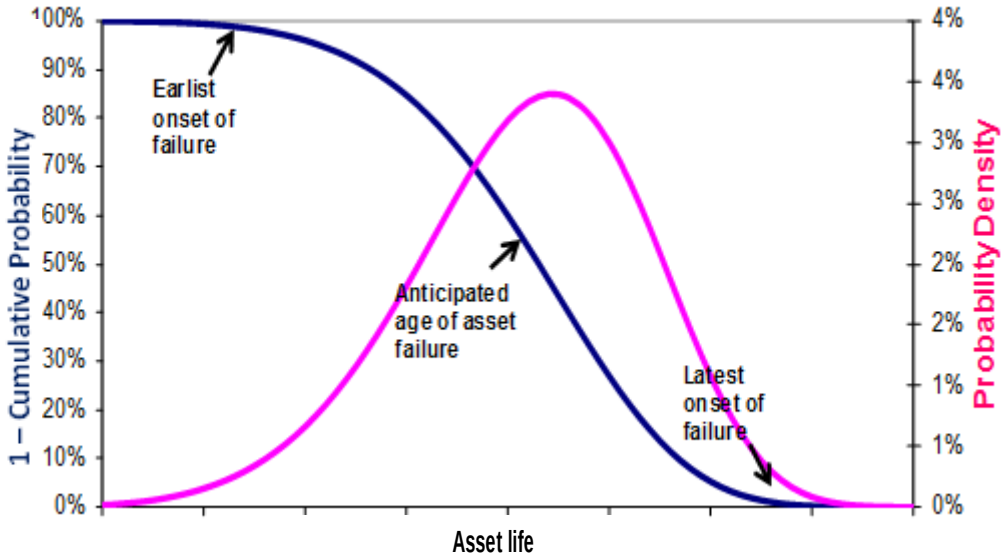


Figure 5: Earliest onset, anticipated, and latest onset of the asset life, represented by a probability density function and a cumulative probability function. Reproduced from the current NOMs methodology [3]

Validation case 3.1.3: Compare the end-of-life failure mode between the current NOMs methodology and the modified NOMs methodology to validate those end-of-life curves obtained from the FMEA.

Validation case 3.1.4: Combine the failure database and the replacement database and possibly forensic database to calculate probability of failure curve or its lower bound. Estimators can be used to correct some statistical bias.

Calibration case 3.1.5: Any large discrepancies will trigger a review and a re-calibration exercise to align the FMEA and the TOs policy statements, as well as the failure database.

Disruptive failures

In the FMEA exercise, the disruptive failure is modelled as an event (consequence) instead of a failure mode[4]. A number of failure modes might lead to the disruptive failure, which is associated with its own cost of (material) consequence. The disruptive failures are well documented and reported over the past 10 years and the failure rate, i.e. the probability of this type of events is relatively easy to calculate. The results can be directly compared with those from the FMEA workshops for validation and calibration purposes. The disruptive and non-disruptive but end-of-lie failures will have a combined effect leading to the asset replacement stated in the above section (i.e. the current NOMs methodology and TOs' policy statements). This however, is based on the assumption that the TOs' done necessary replacements to avoid the disruptive failure and if not, the disruptive failure probability will become strongly correlated with ageing although there is very little data on the earliest onset, anticipated, and latest onset time to failure. If such probability is required for modelling purpose (e.g. it is required by Ofgem at the point of RIIO-T1 submission to model the "without intervention" scenario) to calculate the risk avoidance, some assumptions need to be made

to calculate the implied probability of failure (e.g. without intervention). For example, it is reasonable to make an assumption that if the asset is not replaced when it reaches a state requiring replacements, disruptive failure may occur within in the next several years, as illustrated in Figure 6.

Validation case 3.1.6: Disruptive failure – the disruptive failure probability can be directly calculated from the failure database and compared with FMEA (calculated from the probability of failure and the probability of events).

Calibration case 3.1.7: Disruptive failure – if the implied disruptive failure probability is needed (i.e. the replacements are not done), there will be no data to calibrate the probabilities of this type of failure. Assumptions will need to be made to calibrate the disruptive failure probability (it is unclear yet whether these data is recorded in FMEA workshop, if yes, a comparison study can be carried out to increase the validity of the assumption).

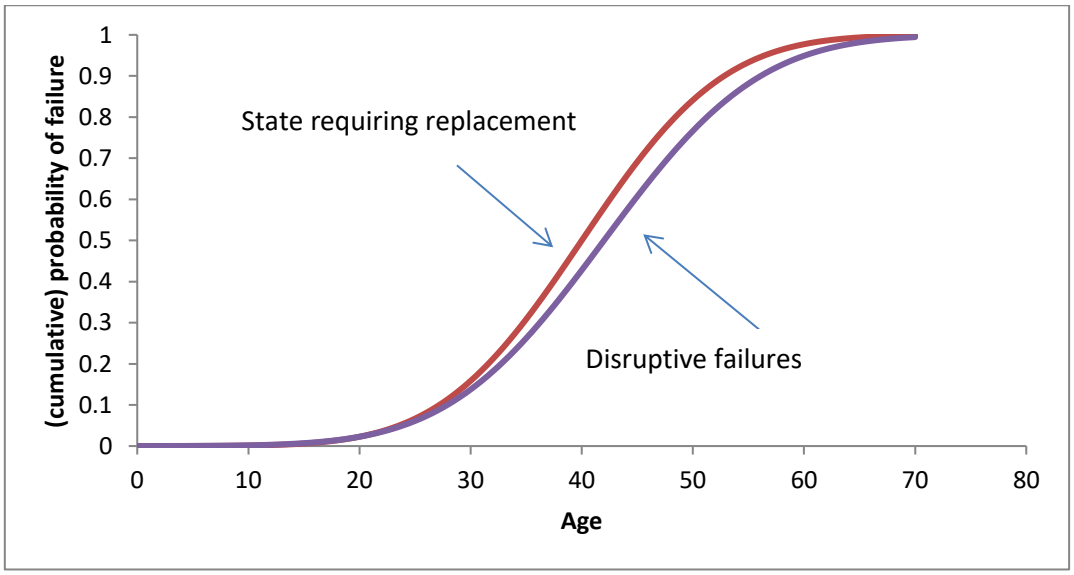


Figure 6 A proposed way of calculating the disruptive failure probability without intervention, by assuming disruptive failure will happen shortly after assets' end-of-life.

None end-of-life failure modes

For those none end-of-life failure modes, there is a much smaller amount of data can be used directly for validation and calibration purposes. Routine maintenance and ad hoc repairs are the common asset management practices (intervention types) to address those types of failures however the effect and consequence of those failures, as well as the time to failures, are less well recorded in the database. Some none time-dependent data, such as defects vs. failure modes, can be obtained using a similar approach introduced in the validation case 3.1.4 to give an indication of the failure rate. However a time dependence curve from the data alone is regarded as difficult because the database is biased due to the asset management activities such as repair, replace, and maintenance, at the same time the nature of defect recording (e.g. defects observed vs. defects not observed). It might be instructive to validate the none end-of-life failure modes by making some

assumptions regarding the TOs' asset management activities, that they must be essential activities and cost efficient to mitigate those none end-of-life failure modes. For example, the routine maintenance is able to address a large number of failure modes (for example, lubrication will address the driver train malfunction of a circuit breaker), the maintenance cycle is considered to balance between the cost of maintenance and the cost of failure if a maintenance were not carried out, in a portfolio sense. Life cycle costing models can be built to take into consideration the outputs from the FMEA workshops as well as other information, namely, probability of failure, probability of events, cost of (material) consequence, cost of maintenance, across all asset types and for a period of time (say, 5 maintenance cycles, but before the end-of-life failure mode start to be significant). The maintenance cycle currently adopted by the TOs should minimise (or close to the optimal position) the equivalent annual cost (for which time value of money can be also incorporated). This is a powerful validation case as it validates a group of elements included in the risk monetisation. It interrogates the combined effect of probability of failure, probability of events, and cost of (material) consequence for a large number of failure modes. The details regarding each individual failure modes cannot be validated however, the risk monetisation model is only replying the combined failure modes instead of individual ones. It is noted that this exercise is based on the assumption that the maintenance is relatively optimised through the operational experiences from the TOs. Large discrepancies do not necessarily mean the none end-of-life failure modes are incorrect but some review need to be established to understand the underlying reasons. Data wise, the only additional data required in the exercise is the TOs' maintenance cost, which is relatively easy to obtain. The probability of failure, probability of events, cost of (material) consequence are assumed to be already developed in the modified NOMs methodology.

Validation case 3.1.8: Minimise the equivalent annual cost over a long term cost for maintenance and risk cost and compare with the current maintenance cycles adopted by the TOs.

Calibration case 3.1.9: Any large discrepancies found through the above validation case should trigger a review of cause and (possibly) a re-calibration exercise to pass the validation case.

<u>Comparison between TOs</u>

The final validation and calibration case will come from the comparison study across TOs. In fact most areas in the NOMs methodology can be tested and validated using the comparison across TOs (e.g. calibration case 2.2.2). The area of the none end-of-life failure modes is particularly relevant because the TOs developed this area separately while the data set is relatively smaller and of poorer quality than other areas in the NOMs methodology. The principle will be the same asset type should have the same list of failure modes with similar failure rates, probability of failures, probability of events, as well as (material) consequences of failures (subjecting to criticality).

Validation case 3.1.10: Comparison studies between TOs' none end-of-life failure modes. Note that translation between definitions from different TOs might be essential to enable a like-for-like comparison.

NB. Some stakeholders commented that cases 3.1.8-3.1.10 are scenario planning cases and may be substituted by appropriate cases or move to other sections. Some stakeholders also commented that they won't have data to perform those exercises. The three cases are left as they were for completeness purposes regarding those non end-of-life failure modes and may change in the future.

3.2 Modelling the failure modes and the probability of failure

The assumption of independence of the models (NGET specific)

In the modified NOMs methodology[4], the NGET specific sections, the failure modes and events are modelled in such a way that the independence between each other is assumed to enable the direct multiply and summation across the failure modes and events. For example, monetised risk = summation across (the probability of consequence × cost of consequence)[4]. It is a strong assumption that intuitively the failure modes and events are generally correlated with each other rather than independent. It is however also valid for a weaker assumption, that after pre-processing of the data/inputs, the failure modes and events are independent to allow those model to be valid. For failure modes that are determined from the component level, while it might be easier to justify those components are independent in terms of failing to function, the failure modes associated with the same components may be due to common cause (hence inter-dependent). Once those common cause failure modes are separated from the failure modes associated with each component (which are further used as data and inputs to model the probability of failure), the assumption of independence will become valid and the modelling proposed by the modified NOMs methodology can be carried out.

For some types of events, the assumption of independence is generally applicable if the events are sequential. For example, as the shown in Figure 7(a), for the event of transformer fire, it is always associated with trip but an event of trip does not necessarily lead to fire. The event data can be pre-processed to separate events as "trip but nothing happens further" and "trip and fire". In such cases, only the probability of failure for each failure mode, in this example, "trip", and "fire" is needed although they are not independent before the pre-processing. Some of the inter-dependences between events are more difficult to model if they are not in a "sequential relationship". As shown in Figure 7(b), two events have got some overlapping, i.e. one event happening will increase or decrease the probability of another event happening. In this type of situation, more information than only two failure modes is required such that "the probability of event A happens", "the probability of event B happens", and "what the probability of events A and B happen together". Note that events happen together including happen concurrently as independent events also happen due to e.g. common cause. The assumption test for the failure mode and events modelling can be performed to investigate whether the data and model work together to sufficiently address those inter-dependence and where independence is assumed, expertise opinion and data from the defect, fault and failure database need to be examined (e.g. correlation analysis) to justify the assumptions. Latest updates (Dec 2016): it has notified that in the FMEA workshop "the probabilities of damaging other assets" are identified when a failure mode happens. This is related to the "inter-dependence" discussion in this section however it is unclear so far how this is incorporated in the models.

Testing case 3.2.1: Assumption test on independence to investigate: 1) whether the failure modes and events are independent or "independent after processing", 2) whether the data is sufficient to model the "independence after processing", and 3) if independence is assumed in the modelling, expert reviews and data analysis are required to justify the assumptions used.

Calibration case 3.2.2: Where data is not available, further data collection (e.g. FMEA workshop) is required to calibrate the data required for the inter-dependent failure modes and events.

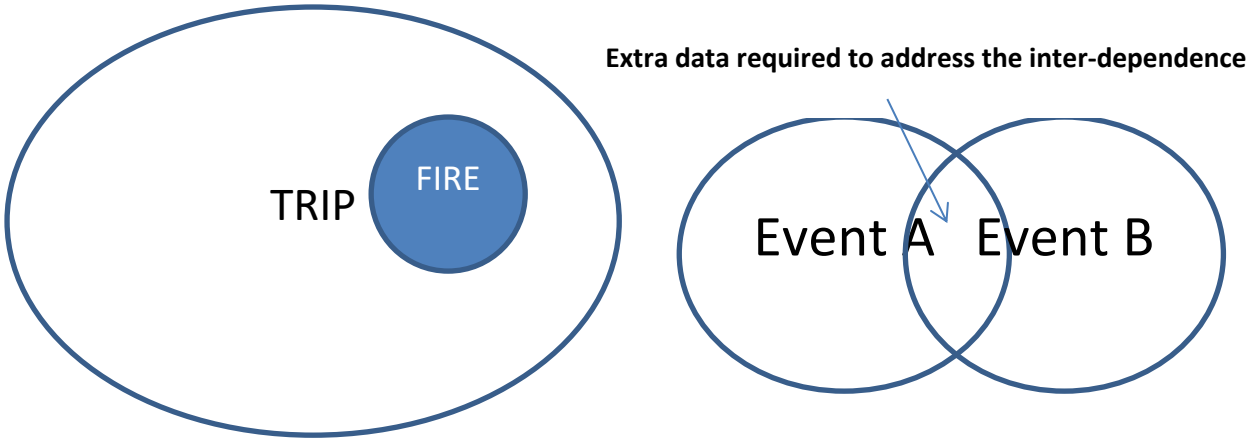


Figure 7 Illustrations of the inter-dependence of failure modes and events and how the independence can be modelled by "pre-processing".

The choice probability curves (NGET specific)

In the NOMs methodology, the probability of failure is modelled as "the most appropriate underlying density function (Weibull, bi-normal)"[4]. To justify a chosen distribution is the "most appropriate" is however not a straightforward task. The Weilbull distribution is chosen to model the ageing process in a large number of engineering-related subjects. It represents a mono-increasing hazard rate/conditional probability of failure that qualitatively mimics the increased underlying source of failures with the elapse of time. In the Scottish TOs' approach (similar to the DNOs NOMs methodology) approach, on the other hand, the conditional probability of failure is modelled directly using a polynomial expansion of an exponential function to mimic the mono-increased failure rate [11]. None of the above models takes into consideration the "burn-in" phase which might need to be considered in some of the failure modes. This is however justified by the fact that the NOMs methodology is focusing on the "… asset failures where condition is an underlying cause"[2], while most of the "burn-in" cases are (theoretically) due to none condition related causes such as human errors, lack of experiences, installation defects, etc.

The maximum likelihood type of statistical study is regarded as impractical to evaluate which model is the best suited as it requires a relatively complete dataset in which the relationship between the failures and age can be drawn for each failure mode, and each asset type. For some special cases with known data availability such as transformer end-of-life failure mode, this test might be performed to demonstrate the differences between each models and the principle might be something like "using the same model (e.g. 2 parameter Weibull, 3 parameter Weibull, the Scottish TOs' approach, bi-normal, etc.) as much as possible for simplicity and audit purposes. It is also noted changing of the probability of failure model can be a way of performing calibration if any of the validation cases (that is related to probability of failure) fail to give a satisfactory outcome. On the other hand, the success of the validation cases (that is related to probability of failure) will provide a level of assurance that an appropriate curve is chosen.

Validation case 3.2.3: Use known dataset to perform a case study to demonstrate the impact and the rationale of choosing a certain type of probability of failure model (for NGET, the transformers end-of-life failure mode can be used).

Calibration case 3.2.4: Upon failure to conform the probability of failure related validation cases, calibration might also be performed by using alternative underlying models for the probability of failure.

NB. The SPT/SHE Transmission implementation utilises a predetermined generic and common PoF (probability of failure) curve to define the relationship between HIs (health indices) and PoF. This relationship has been developed based on a combination of worldwide and specific TO asset operators' historic failure and/or forensic data and knowledge networks. The parameters of this curve are the C-Value that defines the shape of the curve and the K-Value that scales the PoF to a failure rate. The HI-PoF relationship is validated and calibrated in accordance with cases 3.1.1 and 3.1.2. In addition the HI model is also validated and calibrated in accordance with section 2.

Intervention vs. probability of failure modelling

An assumption made in the modified NOMs methodology[4] is that after intervention, the probability of failure (for the affected failure modes) is reset to zero or the value of the starting point of the probability of failure curve. As long as the intervention is deemed to be effective to address a certain failure mode, it is assumed to be perfect and complete. Imperfect intervention is not considered. Intuitively, this assumption is generally valid for the replacement (either the whole asset or components) related intervention while it may be difficult to identify and quantify an imperfect repair (e.g. joining broken conductor strands, or maintenance induced failures). One of the more important imperfect intervention types is the targeted approach on overhead line fittings to only replace the poor condition fittings instead of replacing the whole fittings system. After the intervention, the probability of failure, for the fittings system as a whole, may not be reset to zero due to some old fittings (that are regarded as fit for purpose but not necessarily in perfect condition) are still in service. The extent and frequency of such imperfect intervention need to be examined to justify the assumption of "perfect intervention". Another requirement related to intervention is that, after a replacement related intervention, the probability of failure model may need to change to reflect the newly replaced assets and components (which might be different from the previous ones).

Testing case 3.2.5: Test the assumption of perfect intervention, using techniques such as expertise survey, data deep dive, or scenario test (to compare the outcomes of perfect and imperfect interventions) in order to understand the importance of the imperfect intervention, i.e., whether needed to be modelled in the monetised risk.

Calibration case 3.2.6: If the above is deemed as important, new models need to be derived and calibrated to incorporate the imperfect intervention.

Testing case 3.2.7: Test the model to see whether address appropriately, after intervention, the probability of failure for newly replaced assets or components may be different. Justifications are required for any none compliance of the above specification.

3.3 Probability of events

The probability of event is the concept introduced in the modified NOMs methodology to enable the monetised risk modelling. Instead of modelling the (unconditional) probability of events directly, the methodology uses the probability of failure (for each failure modes) and the (conditional) probability of the events when failures happen (which is a constant value for each failure mode) to give a better understanding the root cause of those events, as well as enable appropriate types of intervention. As detailed in the validation case 3.1.1, the (unconditional) probability of events can be inferred from the defect, fault and failure database if the source data is categorised appropriately. What is more, if both failure mode and event type are assigned to each line of the database, the (conditional) probability of events upon each failure mode can be also calculated using statistical analysis (already covered in the validation case 3.1.1).

Mutual exclusivity (NGET specific)

The modified NOMs methodology[4] has proposed a model to calculate the probability of event that forms the basis of the risk monetisation. As discussed in section 3.1, the FMEA exercise will determine the probability of events when a failure mode happens. Such (conditional) probability of events is then used together with the probability of failure to derive the (unconditional) probability of events[10]:

$$P(C_j) = 1 - \prod_{i=1}^{n} (1 - P(F_i) \times P(C_j | F_i)), \quad (1)$$

in which i is the failure mode count (from 1 to n), and j is the event (consequence) count, C and F represents event (consequence) and failure mode. Eqn. (1) also assumes there is no detection of failure which will be discussed in the next section. As shown in Eqn. (1), the (unconditional) probability of an event type is related to all the failure modes that are associated with this event type, through the probability of failure (conditional on the failure mode does not happen at the point of observation), and the probability of the event (conditional on the failure occurs). Eqn. (1) also indicates that the events and failure modes are independent (see discussions in section 3.2), at the same time, all different failure modes can happen concurrently (but independently), i.e. there are no mutually exclusive failure modes (see ref. [9] for more details of this type of model)[11]. This is

[10] Reproduced from the modified NOMs methodology document (version 05 Dec 2016)

[11] Looking at Eqn. (1), for one failure mode, one event (consequence) $P(C_1)$= P($C_1|F_1$)$P(F_1)$, this requires C_1 and F_1 being independent; for two failure modes, one event, apart from the above independence, we have $P(C_1)$=1-(1- P($C_1|F_1$)$P(F_1)$)× (1- P($C_1|F_2$)$P(F_2)$), = 1- (1- P(C_1)$_{mode\ 1}$× (1- P(C_1)$_{mode\ 2}$), in which 1- P(C_1)$_{mode\ 1}$ models mode 1 does not happen, 1- P(C_1)$_{mode\ 2}$ models mode 2 does not happen, their product means both mode 1 and mode 2 does not happen by assuming 1 and 2 are independent, and the overall equation models the cases EXCEPT the case that both mode 1 and mode 2 does not happen, which including the following cases: mode 1 happens while mode 2 not happen; mode 2 happens while mode 1 does not happen; mode 1 and mode 2 both happen. The above three cases can be also shown in the way that $P(C_1)$=1- (1- P($C_1|F_1$)$P(F_1)$)× (1- P($C_1|F_2$)$P(F_2)$), = 1- (1- P(C_1)$_{mode\ 1}$× (1- P(C_1)$_{mode\ 2}$) = P(C_1)$_{mode\ 1}$ +P(C_1)$_{mode\ 2}$+P(C_1)$_{mode\ 1}$ ×P(C_1)$_{mode\ 2}$, which the three terms represent the three cases. For three failure modes 1, 2, 3, the equation models the case EXCEPT mode 1, 2, 3 not happen at the same time, which cover the cases: mode 1 happens others not, mode 1 and mode 2 happens and mode 3 not, mode 1, 2, 3, all happen, etc.

also a strong assumption that needs to be tested using the detailed data/outputs from the FMEA exercise. Peer reviews from subject matter experts and collective opinions from workshops are needed to confirm all the failure modes for the same asset type associated with the same event type are not mutually exclusive.

Testing case 3.3.1: Use peer reviews from subject matter experts and collective opinion from workshops to confirm that the failure modes identified in the FEMA are not mutually exclusive.

Calibration case 3.3.2: If the above can't be met, the model (e.g. Eqn. (1)) and relevant parts in the risk monetisation model needs to be re-visited and re-derived.

Probability of failure (Scottish TOs) vs. probability of events (FMEA)

The validation case 3.1.8 will provide some assurance on the probability of event (a combined effect of probability of failure, probability of event, material consequences of failure, etc.) that is used in the monetised risk modelling. On the other hand, the probability of some event types should be directly comparable with the probability of failure derived from the Scottish TOs' modelling, for example, the disruptive failure, for the same asset type across the TOs. The independent development of the probability of consequence models ensures such comparison meets the Authority's requirement of comparability across TOs [2].

Validation case 3.3.3: Where possible, perform comparison studies between the probability of events (NGET) and the probability failure (Scottish TOs).

3.4 Probability of detection (and acting accordingly)

The cost of consequence for detection (and acting accordingly)

The modified NOMs methodology also introduces the concept of "detection" for which, if "acting before consequence materialises", the event can be avoided. The equation for the probability of the event (Eqn. (1)) becomes[4,10]:

$$P(C_j) = 1 - \prod_{i=1}^{n} (1 - P(F_i) \times P(C_j | F_i) \times (1 - P(D_i)) , \qquad (2)$$

in which D is the detection and the additional term to Eqn. (1) represents the probability of (imperfect) detection and the risk will be reduced if acted on accordingly. Due to to Eqn. (2), the detection is associated some failure modes. If a failure mode i can be detected (and acted on before the event j happens), the failure can be avoided and the failure mode i will have zero monetised risk regarding event j. The cost of "acting" (e.g. repair, replacement) is not considered in this equation and will be modelled separately to give a monetised risk[12]. The probability of detection is modelled

[12] It is not clear yet how this is modelled in the modified NOMs methodology according to the document (version 05 Dec 2016).

as a fixed number and will be set to zero if inspection is not performed or its validity period has expired.

Testing case 3.4.1: Make sure the probability of detection addresses the cost of "acting accordingly" at the same time the assumptions of fixed probability of detection is a valid assumption.

Validate the probability of detection

The logic of introducing the probability of detection (and acting accordingly) is to highlight the effectiveness of inspection as well as other asset management activities from the monetised risk point of view. According to Eqn. (2), it can be premised that the effectiveness of detection is to reduce the risk, if combined with follow-up activities. Eqn. (2) therefore does not cover the general inspections that are used by asset managers to understand the asset condition while not necessarily trigger any risk reduction actions. The data regarding probability of detection and the validity period is recorded from the FMEA workshops.

To validate the probability of detection values from the FMEA workshops is not a straightforward task. The defect, fault and failure database does not always contain the information on detection. How effective the inspection method can be inferred qualitatively/indicatively from the workshops but a quantified value may be difficult, especially if it is shown from Eqn. (1) that the monetised risk is very sensitive to the values of the probability of detection (and acting accordingly). The inspection data, the work orders ("acting accordingly" data), and the defect, fault, failure data need to be combined together to give any useful conclusions. This exercise is however requiring high data quality and data completeness, facilitated by huge data mining/collection/cleanse effort. It might be productive to draw some insights from the validation case 3.1.8, that the balance between risk and cost to validate backwards on the risk inputs (e.g. probability of failure, probability of detection). For the inspection activities, the cost of inspecting the whole asset population on a periodical basis must be justified by avoiding certain types of events, which would happen to a small number of assets if inspections were not performed. The balance between the inspection cost and the risk avoidance will therefore provide some assurance to the probability of inspection (and act accordingly) values that are obtained from the FMEA workshops.

Validation case 3.4.2: Use the balance between cost and risk to validate the probability of detection and the validity period of inspection.

Calibration case 3.4.3: If large discrepancies are found in the above exercise, the values obtained from the FMEA workshops need to be re-calibrated.

NB. Cases 3.3.1-3.3.2, and 3.4.1-3.4.3 are NGET specific due to the nature of the approach adopted by SPT/SHE Transmission which considers only those failure modes with material effects, thus avoiding any unnecessary analysis of failure modes that do not have material effects. In addition detection is built into the calculation which determines the HI and subsequently used to determine the PoF.

3.5 Modifiers and equivalent age (or equivalent time since intervention)

The rationale of introducing modifiers in the modified NOMs methodology is to map the time/age based probability of failure model to condition based probability of failure and monetised condition risk whilst incorporating asset specific information (namely, the health score, duty, etc.) in the probability of failure and monetised risk. There are a number of modifiers proposed in the modified NOMs methodology and the health score introduced in section 2 is used here as an example for the end-of-life modifier (in the current section the concept of health score and end-of-life modifier are used inter-exchangeably). The testing, validation, and calibration plan proposed here is however regarded as generic for all types of modifiers[13] as the probability of failure for all the failure modes are based on a similar type of time/age based model. For the end-of-life modifier, it is proposed in the modified NOMs methodology that the linkage between the health score and the probability of failure is established by constructing a health score - age curve, i.e. equivalent age. It is then used together with the probability of failure - age model to identify the probability of failure for each asset.

The relationship between age and health score

As shown in Figure 8[14], the key to the equivalent age methodology is to establish the function between age and health score. The health score of an asset is compared with a "typical" asset ageing curve which plots the health scores against age. The equivalent age can be obtained by knowing what the age would be for a "typical" asset with the same health score to the asset under consideration. It is then used for the probability of failure curve (see section 3.2) to obtain the probability of failure for the particular asset. Conceptually this is a straightforward approach assuming the "typical asset" has got a "typical" ageing curve, and at the same time "typical" probability of failure curve. It is necessary to test that the ageing curve and the probability of failure curve is compatible through the models, i.e. for the same "typical asset". By knowing the two curves (probability of failure - age curve, and health score - age curve), a curve can be plotted for the probability of failure as a function of health score, for the "typical asset". This curve needs to be reviewed and tested for each asset type. Similar to validation case 3.1.1, the defect, fault, and failure database can be used by assigning the health score to the underlying assets, for each line in the database. The failure rates/probability of failure can be calculated for each health score and compared with the probability-health score curve. Due to scarcity of the dataset, it might be that on occasion subject matter experts' or Industry normalised review of those curves is needed. For the same asset type, it is expected that the probability of failure - health score curve should be comparable between the Scottish TOs' approach and the FMEA+ equivalent age (NGET) approach. At this stage, we assume the health score, probability of failure is already validated and calibrated through section 2 and section 3.2 therefore the major calibration effort is on the equivalent age modelling itself, i.e. what the "typical" ageing curve (see figure 8) looks like[15].

Testing case 3.5.1: Construct the probability of failure as a function of health score (and other modifiers), plot the curves for each asset type (and/or each failure modes). Those curves need to be compared with known database (e.g. known failure rates for assets with certain health scores) and

[13] It is not yet clear in the methodology (version 05 Dec 2016) how many modifiers will be used in the final version of methodology.

[14] Reproduced from the modified NOMs methodology document (version 05 Dec 2016).

[15] For none end-of-life modifiers, the modifiers parallel to the health score is not clear yet in the methodology document (05 Dec 2016) and they need to be tested and calibrated accordingly.

peer reviewed by subject matter experts or industrial normalised for the cases where data is not available.

Validation case 3.5.2: Compare the probability of failure - health score curve between the NGET approach and the Scottish TOs' approach, for the same asset type and same failure mode or event, where possible.

Calibration case 3.5.3: If large discrepancies cannot be resolved, calibration exercises will be needed to align the models and/or align the models with data. Calibration will be focusing on equivalent age model itself providing the probability of failure and health score are sufficiently tested and validated.

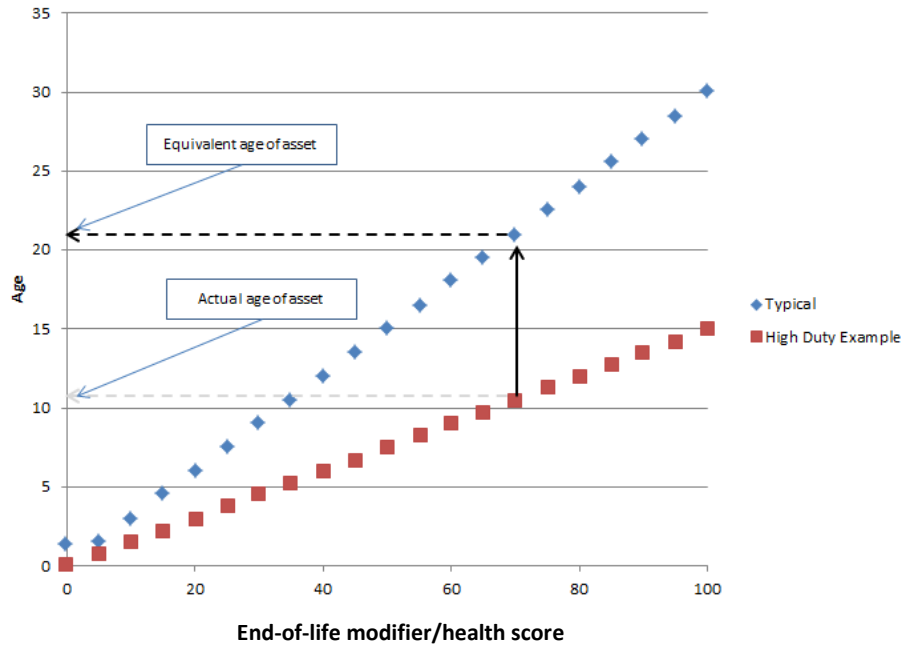


Figure 8 An illustration of the equivalent age methodology by comparing a health score curve of a "typical" asset with the health score of the asset under consideration. Reproduced from the modified NOMs methodology[4] document.

Probabilistic modelling of health score as a function of age

Another way of looking at Figure 8 is the uncertainty around the health score and age. An asset's health score deviates from the "typical curve". An old asset might have a low health score (good condition) while a new asset might have a high health score (poor condition). By plotting the portfolio of asset of the same type, using the health score - age plotting for the available data from current and historical asset inventory, the uncertainties around the deterioration can be observed. Such uncertainty can be modelled using a probabilistic plot which will form the basis of forecasting

the future deterioration of the asset health score. It might also worth noticing that this plot might be subject to bias from de-commissioned assets which may or may not have a health score recorded at the time of de-commission. For NGET's dataset, good records have been made for the decommissioned transformers therefore such plot is available. This health score - time plot can be super imposed with the asset health - time plot that is currently used for the forecast of the future asset health and deterioration [3] (see Figure 9 below). This is a valid comparison by recalling that any forecast methodology needs to have some elements of age in it and recalling that there are certain type of mapping between the health score and asset health (although not 100%, see section 2 for details). This comparison study provides further assurance of the end-of-life modifier (i.e. health score) as well as its forecast methodology.

Validation case 3.5.4: Compare the deterioration model in the current NOMs methodology [3] with the health score - age plotting, to validate the equivalent age methodology and the health score forecast methodology. Note that it may subject to the bias of data due to de-commissioned assets. This validation case is only applicable for the end-of-life modifier.

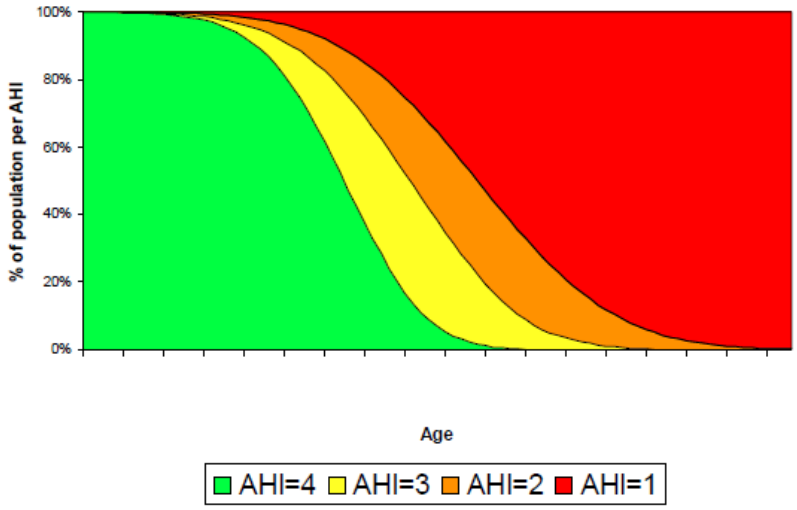


Figure 9 The deterioration model that probabilistically models the asset health index as a function of age. Reproduced from NGET's asset policy statement.

Extreme value testing

By examining Figure 8 closely, we can conclude that the equivalent age methodology assumes that the equivalent age is only related to the current health score of the asset instead of the trajectory of the health score (i.e. the health score history does not matter in terms of equivalent age modelling). This assumption is widely used across the energy network asset owners and for most asset management frameworks and no further testing of assumption is proposed here. One thing worth noticing is the details of the equivalent age model. It is currently unclear[4] (for NGET) how much the element of real asset age will be included when translating health score to the equivalent age. Where this is the case, special caution need to be taken to justify that the age element in the equivalent age model is not a duplication with the age element in the health score model (see section 2). An extreme value test is proposed for the end-of-life modifier (i.e. health score) all asset

types, that an old good condition asset will not be prioritised against a young poor condition asset, subjecting to uncertainties around the condition information, etc.

Testing case 3.5.5: Extreme value tests designed to test the model to the limit that for the rare cases (e.g. good condition old asset) the model is still valid and fit for purpose.

Validate the health score forecast modelling

The final validation case to the forecast methodology of the health score is to compare the historical, year one year health score profile with the forecast results assuming the starting position of a historical year. Specifically, the TOs have got the asset health indices since 2010 for all the lead assets. Those asset health indices need to be translated to health score using the same information that arrives at the asset health at the time (in fact such translation is essential to translate the current target of network replacement outputs). The deterioration model can be then applied to the 2010 health score and forecast the health score in a probabilistic manner (or deterministic in the Scottish TOs' approach). The forecast also needs to take into consideration the known interventions that have been applied to the assets (by definition only refurbishment and replacement will impact the health score). Those forecast results can be compared on a year-on-year basis with the health scores in 2011-2016, in order to provide validation to the deterioration forecast. Any large discrepancies require a re-calibration of the deterioration model, recalling that the data may be biased due to the de-commissioned assets (which might not be recorded at the time of de-commissioning)

Validation case 3.5.6: Use historical year-on-year health score profile to validate the deterioration forecast model, by comparing the forecast results using the starting point 2010 health score and known interventions (refurbishment and replacement) between 2010 and 2016.

Calibration case 3.5.7: Any large discrepancies may lead to re-calibration of the deterioration model. Note that the difference between the deterioration model and the health score - age plot may be considered and justified (for example, justification can be made due to data bias)

3.6 Testing, validation and calibration of the probability of events

The probability of events is the final outcome from the activities and models described in the current section. For each asset, there should be a list of probabilities of events, e.g. trip, component damage, partial replace, fully replace, disruptive failure, etc. Those probabilities will be used together with the cost of material consequences to calculate the monetised risk for each asset. It is worth pointing out that a large proportion of data used in the monetised risk modelling is generated from the FMEA workshops instead of data recorded from the databases. At the same time most of the models proposed in the modified NOMs methodology are empirical based. The testing, validation and calibration presented in the current section (section 3) is focusing on finding independent information sources that are able to provide like-for-like comparisons for instance the assurance of the model and data. Ideally speaking, if all the testing, validation and calibration cases are performed in a satisfactory manner, the probability of events, as an output based on validated

inputs and model, should be automatically validated. Due to the importance of this element, some top-down type of validation approaches are proposed here.

Against total number of events on the network

Firstly, the probability of events for each asset, when aggregated at the network level, should be comparable to the total number of events recorded by the network owners. While the probabilities of events are defined over a year and are broken down into event types, the fault, failure and defect database needs to be categorised accordingly, and on a year by year basis. Note that case 3.1.1 proposes the exercise for probability of failure, the exercise for probability of events is using the same database, following the same logic, but with different data manipulation target. In fact, this exercise is also used in the DNOs methodology [11] in order to validate, and calibrate the probability of failure and its progression with time. The data cleanse and manipulation effort will be similar to case 3.1.1 and this exercise also implicitly validates the model that aggregates the probability of failure to the probability of events (Eqn. (1)), as well as some associated inputs. One of advantage of this exercise is that, even with limited data availability, some probabilities associated with certain types of events are known from pervious exercises performed by the TOs. For example, the TRIP database, and disruptive failure database are well populated and widely used by the TOs for various reporting and benchmarking purposes.

Validation case 3.6.0: where possible, use the fault, failure, and defect database to validate the probability of event, on a like-for-like basis. Note that considerable data manipulation effort may be required to map the database to the event types and asset types. Similar exercise is also performed by the DNOs methodology as one of the major means for validation and calibration.

Against known risks and established models

The probability of events can be compared with known risk assessments to the TOs. This is similar in principle with the DNO's validation methodology described in section 2.4.From a portfolio point of view, the higher probability of events will lead to a higher priority of intervention, for assets with the same criticality. A validation case can be therefore designed such that for each asset type, and each criticality level, the asset intervention list obtained from the current capital plan and maintenance plan should be given higher priority according to the probability of events. For end-of-life related events, the sum of the probability of those events for each asset, when ranked from high to low, should be comparable with the replacement list in the capital plan. The sum of all assets' probability of end-of-life related events, should be comparable with the established modelling approaches (before the criticalities are used in the journey of asset management maturity), for example, Ofgem's survivor modelling [12]. Note that some categorisations and treatments may be required to enable a like-for-like comparison, for example, the survivor model only models the ageing related asset replacements therefore only ageing related end-of-life events will be used for the validation case.

Validation case 3.6.1: Use known risk assessment results for the assets to evaluate the prioritisations of intervention. Some categorisation may be needed to enable a like-for-like comparison. For example, the end-of-life related events should be known in the replacement intervention list, especially for those assets with high criticalities; the random failures related probabilities might be comparable with strategic spare holdings; etc.

Validation case 3.6.2: Use previous modelling techniques such as Ofgem's survivor model to validate the volume of intervention needed, in a portfolio sense. Note that the models may only address one type of intervention. For example the survivor model only addressed ageing related end-of-life therefore should be compared with the relevant probability of events.

Against deferred maintenance risk assessments

For maintenance related events, the above validation cases may be more difficult as the maintenance is applied on the failure modes and not directly identifiable from the event types. The validation case 3.1.8 provides a good validation for the probabilities of those maintenance related events. Some more insights may be obtained from the historical records of the deferred maintenance. The probabilities of each event can be calculated for each asset, both before and after the maintenance is deferred. It can be immediately observed (aided by visualisation tools) that the deferred maintenance will lead to a tolerable increase probability of events. The TOs are routinely doing the risk assessment on the deferred maintenance assessment (not based on the probability of events) and it would be instructive to compare the historical deferred maintenance risk assessment reports with the risk assessment using the probability of event methodology.

Validation case 3.6.3: Calculate the change probability of events by deferring maintenance and compare those results with known/historical risk assessments of deferred maintenance reports.

NB. Some stakeholders suggest that SPT & SHET do not use probability of consequence in their models like this.

4. The cost of (material) consequences

4.1 An overview

According to the modified NOMs methdology[4], the (material) consequence is associated with a £cost value of each event type. The modified NOMs methdology[4] also categorises the cost of (material) consequence into financial, system, safety and environmental aspects, as defined in Table 2 (note that the financial consequence is the cost of intervention and other cost to recover the service, instead of all financial related consequence, e.g. the financial penalty related to loss of supply is attributed to the system consequence). Comparing with the current NOMs methodology [3], two additional areas are introduced (in the modified NOMs methodology): 1) the probability of the consequence (events) materialisation is taken into consideration explicitly, as per requirement from the Authority that "[the methodology] realistically quantifies the probability and monetised consequence of all material consequence condition related failures" [2]; and 2) different types of failure modes and events will lead to different costs of (material) consequence, due to the failure mode analysis discussed in the preceding section.

Table 2 definitions of different (material) consequences that are associated with the costs, reproduced from the modified NOMs methodology document[4].

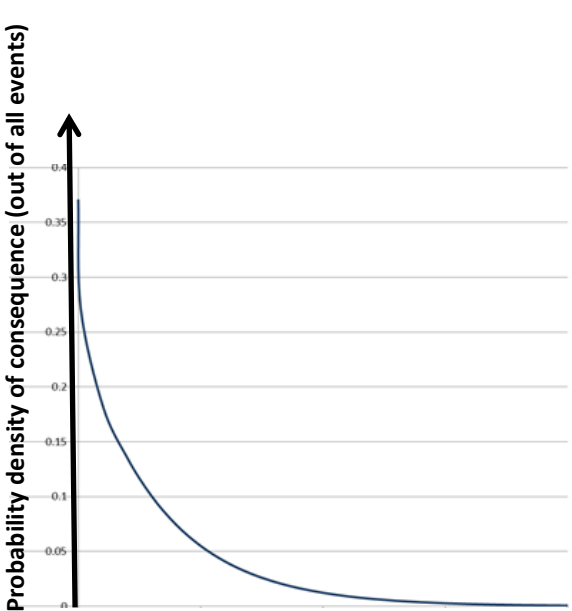
Consequence	Description
Safety	Impact of direct harm to public/personnel as a result of failure mode
Environment	Impact of failure mode taking into account the sensitivity of the geographical area

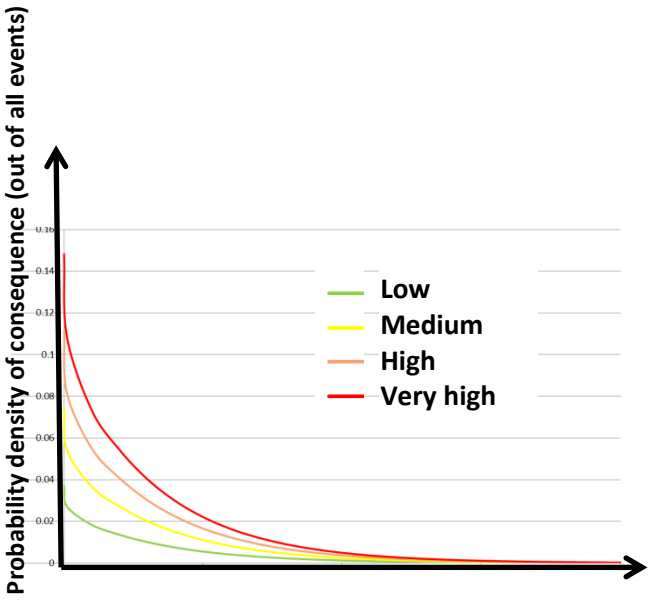
	local to the asset
Financial	Cost of the intervention needed to address and resolve the failure
System	The impact on the network of the failure and any subsequent intervention required

The probability of (material) consequence

There are currently limited details in the modified NOMs methdology[4] about the probability of the (material) consequence. It is however stated in the overall approach that the methodology will take the value of one overall cost of (material) consequence against each event type to calculate the monetised risk for each asset: $risk = \sum P(C_j) \times C_j$, in which C_j is the (monetised or material) consequence of event C_j - note that C used for both the event itself and the (material) consequence of the event without causing confusion (see Eqn. (1)). In practice however, for the same asset type and event type, the cost of the (material) consequence can still be different. This is due to possible reasons such as (1) natural variability between one event to another (e.g. the return to service time after a fault can be different for the same type of fault and asset due to other factors such as resource availability, parts availability, etc.); and (2) differentiation between the location of the assets or the role it plays in the system (such as proximity to a school/a river, or the demand of the substation, etc.). As shown in Figure 10(a), the natural variability can be represented by a probability distribution regarding the different costs of an event. It is a conceptual model that those events consist of a large number of small impact events and small number of large impact events. Figure 10 also (conceptually) covers the situation that when an event happens there is no material consequence to reflect a probabilistic (material) consequence (e.g. a TRIP event followed by delayed auto re-close). While only a single value is used for the cost of (material) consequence, it is assumed that (although not explicitly written in the document[4]) such natural variability can be sufficiently addressed by using an expected value of the probability distribution as shown in Figure 10(a). Instead of directly calculating the expected value, it might be worth to start with the probabilistic distribution of the events to give test those assumptions. This exercise has however got modelling and data implications which will be discussed in a later part of the section.

Testing case 4.1.1: Review and understand the validity of using a representative value to model the cost of (material) consequences. Special care is needed when modelling the events that have got no material consequence (in order to reflect probability of (material) consequence).





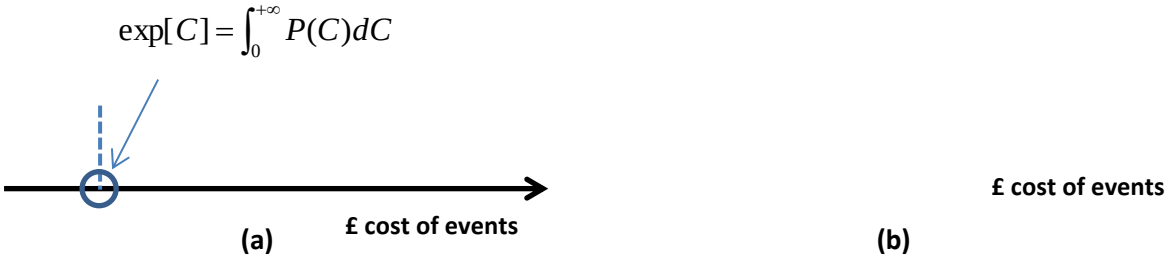


Figure 10 A conceptual model illustrating the variation of the cost of (material) consequence, the probability of the (material) consequence, as well as the relationship between criticalities and consequences.

Modelling criticalities

For the variability associated with asset locations (physical location or the location the system, point (2) in the paragraph above), its (material) consequence needs to be addressed using asset-specific information such as criticality. Note that other modifiers can be also used to give desired granularities while the concept of criticality is well accepted and is used in the current context. As shown in Figure 10(b), the cost for the same type of event against individual assets (of the same type) can be further categorised using different criticalities. It is expected that events associated with higher criticalities will be given a larger expected value to represent the cost of event (i.e. material consequence). Extra care needs to be taken when introducing criticality into the cost of (material) consequence. The existing criticality definition measures the exposure and vulnerability [3] that implicitly involve the probability of an event materialising - a higher ranking of exposure is linked with higher probability of (material) consequence. For example, the low criticality probability distribution shown in Figure 10(b) might have a large proportion of zero cost. When incorporating the criticality measures in the probability of the (material) consequence, it is important to review the process, including the definitions and assumptions to make sure the modelling is consistent and compatible. For example, it is important to review the process that the probability of (material) consequence is not double counted through both the cost of consequence (zero cost region in Figure 10) and the criticality (exposure).

Testing case 4.1.2: Review the process when introducing criticality (or any other asset specific information) into the cost of (material) consequence. At the same time, make sure the current criticality definition [2] is appropriately translated into the probability of (material) consequence.

Comparison studies for the cost of (material) consequences

The current NOMs methodology [2] focuses on the cost of consequences for the “catastrophic failure” by using external referenced £cost of events. The criticality is used as a factor to determine different percentiles used in the distribution for the cost of different events. Considering the fact that this approach is relatively well accepted during the stakeholder consultation, it will be instructive to perform a validation study by comparing the cost of the disruptive failure (termed as “catastrophic failure” in the current NOMs methodology [3]) between the current and the modified NOMs methodologies. The zero cost events, i.e., probability of (material) consequence are not sufficiently considered in the current NOMs methodology. This will lead to a consistently lower cost

of events in the modified NOMs methodology, except for the financial cost that a disruptive failure will always (100% probability of material consequence) lead to the same asset management activities (cost of recovery) from the TOs. A similar comparison study can be also performed between the DNOs' cost of consequence [11] and TOs' cost of consequence, where appropriate, e.g. for the same failure modes for 132 KV network. This is particularly relevant in terms of safety and environmental consequences, for which the costs should not be sensitive to the system configurations (i.e. whether it is DNO's or TO's asset). Any cost of consequences developed independently from different TOs can be also compared. Note that some cost information (e.g. cost of replacing an asset) may be confidential to each company and certain arrangement needs to be in place for some of the comparison studies.

Validation case 4.1.3: For the events associated with disruptive (catastrophic in the current NOMs methodology) failure, compare the cost of (material) consequence values between the modified NOMs methodology and current NOMs methodology, for each asset type, and each criticality level. The system, environmental and safety consequences should be consistently lower in the modified NOMs methodology while the financial consequences should be comparable.

Validation case 4.1.3: Subject to confidentiality issues, perform comparison studies against DNOs' cost of consequence and across the cost of consequences between TOs, where a like-for-like comparison is appropriate.

<u>Financial consequences</u>

Considerable data quality and data cleanse effort will be required in order to obtain the data behind the conceptual model plotted in Figure 10. Using the financial consequence as an example, it covers the cost of carrying out the interventions and recovering service (cost of recovery). A pure data-driven approach can be adopted by combining the cost information with the defect, fault, and failure database. For each TO, the cost information is relatively well recorded in the format of schemes and work orders in the TOs central data system. However, those data are not recorded for the purpose of performing statistical analysis and calibrate the cost of consequence for the failure events. For example, a number of failure modes can be addressed when carrying out routine maintenance and can be bundled together as a single work order. The maintenance script is devised against individual assets however not against failure modes or event types. Even in some cases when event types are recorded as free text, etc., they are unlikely to have been recorded to be consistent with those developed in the modified NOMs methodology. Data cleanse and data manipulation will be required to assign and categorise the cost in the work orders and schemes to the event types, through the defect, fault, and failure database. Such data exercise is also important in terms of identifying the events with zero material consequences (e.g. TRIP event followed by delayed auto re-close). It is acknowledged that not all cost information is required, as long as a reasonable sample of the data is drawn to provide the distribution of the event costs (Figure 10). An iterative approach may also be used to reduce the data cleanse and data manipulation effort. Firstly the financial consequences are calibrated for one asset type, and as many event types as possible, using the known databases and the data procedures proposed above. Those values will then be compared with those developed in the modified NOMs methodology (details unclear as yet) to validate the process and the logic. Depending on the outcomes of the validation exercises for one asset type,

different strategies can be carried out: for instance, whether to carry on performing the (time consuming) data exercise or to accept the values proposed in the modified NOMs methodology.

Validation case 4.1.4: For the financial consequence, perform the data cleanse and data manipulation exercise by combining databases (work orders, scheme costs, defect, fault, and failure database, etc.) to categorise the cost information. A statistical analysis can be then performed to validate the logic and process in the modified NOMs methodology. Considering the effort required for the data exercise, it is advised to start with a small number of asset types and event types.

Calibration case 4.1.5: If large deviations are identified, a review process will be needed to explain and justify those differences, followed by (if needed) the above data cleanse procedures and statistical analysis to calibrate the cost of financial consequence values.

Cost of consequences from the FMEA

Another source of information is from the FMEA process. During the FMEA workshops, the "impact" of the events is recorded for each asset type, and event type. Those "impacts" are assessed in terms of score scaling from 1 to 5, which is associated with a cost range detailed in Table 3. Recalling that any information from the FEMA workshop is always assessed against a group of assets, there will be a lack of asset specific information. The term "impact" can be considered as loosely defined, as it is unclear, when the FMEA workshop is run, how the "impact" can be assessed using a "group of assets" without considering the asset specific information such as criticality. For example, the "impacts" would be different for the assets with the same type but different locations while it is unclear which "typical" location those "impact" are assessed against during the workshops. On the other hand, judging from Table 3, it is also unlikely that the "probability of consequence materialise" is considered in the FEMA worships - for example, the safety risks only materialise when there are personnel around during the event, the cost of the event, independent of its scale, can be always zero when the consequence does not materialise. Bearing the above caveats in mind, the outputs of "impact" in the FMEA workshop can provide a range estimation regarding the cost of consequences. It is logical to assume that, when the workshop is held, the collective opinions are always based on some (implicit) common ground therefore the relationship (ratios) between different event types and different criticality categories (financial, safety, environmental, system[16]) is more credible than the absolute values. Those ratios can be used to perform some envelope calculations to calibrate and validate the cost of consequences proposed in the NOMs methodology. This approach is especially valuable when some of the costs are difficult to calibrate directly from the database. For example, defects can be rectified in a bundled manner using a single work order and difficult to slip the cost of the work orders on a defect-by-defect basis. The cost of those "difficult-to-obtain" events can be inferred from those "easier-to-obtain events" (on an envelope basis) to validate and calibrate the values of the cost of (material) consequences.

[16] Cost of system consequence can be more complicated due to the system redundancy and will be discussed in a separate section.

Table 3 The (indicative) guidance used in the FMEA workshops to provide a cost ranking to the events identified in the workshops (for illustration purpose only).

Score	Financial (examples)	Safety (examples)	Environment (examples)	Security of Supply (examples)
1 Very low	£0- £x	Minor Injury / Near Miss / Neg.	Negligible environmental impact	Negligible disruption
2 Low	£x-£y	Lost Time injury / HSE letter of concern	Minor environmental impact eg Localised spillage	contract customers in DNs disrupted I&C or <250 domestics / Minor disruption to Op Systems. Tariff customers in DNs disrupted
3 Moderate	£x-£z	Major Injury eg RIDDOR reportable	Major incident eg contamination of water courses / EA Letter of Concern	Short term failure / Firm contract customers disrupted I&C or >250 domestics
4 Significant	£z-£u	Fatality / HSE Enforcement Notice	EA Enforcement Notice / Improvement Notice issued	DNs disrupted / major outage for significant period of time
5	Above £u	Multiple Fatality / HSE Enforcement Notice	EA Prohibition Notice	NTS disrupted / total system outage for lengthy period

Validation case 4.1.6: Where possible, use the "impact" of events recorded from the FMEA workshops together with Table 3 to provide an "envelope estimation" to the cost of (material) consequences (proposed in the modified NOMs methodology). Bear in mind that the FMEA "impact" values do not have asset specific information and they are unlikely to have considered the probability of consequence materialised (i.e. those zero consequence events).

Calibration case 4.1.7: Where possible, take advantages of the cost ratios from the FMEA workshop (between different event types and consequence types) to calibrate those "hard-to-obtain" cost values from those "easy-to-obtain" cost values.

4.2 Safety and environmental consequences

There is very limited information about the cost of safety and environmental consequences in the current version of the modified NOMs methodology[4]. They may be tested, validated, and calibrated using the general approaches that are detailed in the above section. It is however expected to have more data implications in this area, not due to data quality but to the data size. Those safety, and environmental related asset failure events are extremely rare in the transmission system, making it difficult to analyse using conventional statistical analysis. Specifically, the safety related events are well recorded in databases such as LTI (lost time injury) and public safety track record; while the environmental related events can be also found in the report such as SF6 (RRP) reporting and historical environmental agency fines. The cost of those rare events can be treated as a complete data set and the costs of those events are normally clearly documented in each TOs event report and the "total safety/environmental cost to the TOs[17]" in "the past 10 years" (specified by ref. [2]) can be obtained.

A direct statistical analysis to link the costs with event type and asset type suffers from small data size hence lack of statistical significance. A modelling approach can be established to validate the safety and environmental costs proposed in the modified NOMs methodology. The total risk cost in terms of safety and environmental consequences over the past 10 years can be calculated from the assembly of safety and environmental risk cost across each individual asset, by knowing the probability of events and the cost of (material) consequence of the events. This validation approach requires a monetised risk model and assumes the probability of events are tested and validated using the approaches proposed in section 3. It is worth noticing that, when 10 years' data is used, the probability of each event type for each asset, as a function of the health score, is not a constant but follows the natural deterioration. The criticality/consequence of an asset may also change due to various reasons (e.g. newly build housing next to substations, etc.). Year on year information will be needed in order to ensure the validity of the modelling results to provide a refined validation case (although some simplifications may be assumed for example the criticality change is relatively small and may be ignored).

Validation case 4.2.1: Construct a monetised risk model to calculate the overall risk cost by assembling asset level risk cost and compare it with the historical cost data, to validate the safety and environmental consequences derived in the NOMs methodology (which is against each asset and each event type).

Calibration case 4.2.2: If significant deviations are found, re-calibration of those cost of (material) consequences will be required to satisfy validation case 4.2.1. This can be achieved using techniques proposed in calibration cases 4.1.7, i.e. the ratio between events generated from the FMEA workshops. Such calibration will give range estimations of the values and reducing the range to a single value may require further assumptions (e.g. the range distributed following a normal distribution, etc.).

[17] It is possible that those costs are incurred outside the TOs, such as cost to the insurance providers, etc. Those cost should be consider as part of the cost of consequences. It is also possible the cost is not available.

4.3 System consequences

To calculate the system consequences (and the material cost of those consequences) on an asset by asset basic is not a straightforward task. The asset topology in the transmission system with build-in redundancy and the post-event activities carried out by asset owners and system operators make it a complex task to pin-down each individual asset and evaluate the impact on the system of the asset failure. Direct calibration from historical data for the cost of system consequences can be difficult due to similar reasons. The system consequence data, namely, loss of supply, disconnection of generation, constraint cost (of boundary transfer or generation), energy reserve and auxiliary cost (e.g. reactive powers), are not recorded on an asset by asset basis (e.g. circuit by circuit, substation by substation, region by region, or national). The modified NOMs methdology[4] proposed a way of assessing system consequence using a bottom-up approach on a probabilistic manner. It takes advantage of the latest understandings of the assets regarding the probability of failure and probability of events, as well as the return to service time (RTS) for each asset type and event type. The current section of testing, validation, and calibration is not aiming to dive into those details of the system consequence methodology but to provide validation and assurance using alternative methods including data and modelling. The philosophy of performing the validation study is not to go through every single asset in the system but sufficient cases to provide assurances from various aspects. Data driven approach can be performed in the similar way to the validation case 4.2.1, that the overall risk can be calculated from the NOMs methodology using a monetised risk model and compare the results with the historical data. On the other hand, there are a number of system reliability studies and models built in the past, providing benchmarks to the system consequence proposed in the modified NOMs methodology. Each element involved in the system consequence methodology (in ref [4]), i.e. RTS time, customer disconnection, generation compensation, constraint costs, and reserve cost, auxiliary service cost, are discussed separately.

<u>Return to service time</u>

The RTS time, as an output from the FMEA workshop while an important input to calculate the system consequence, need to be validated. The RTS time is understood as an average value for the asset type and event type. It is deemed as a representative value that the variability between outages of the same type can be ignored when calculating the system consequence. This assumption, together with the values of the RTS time, can be tested and validated using the outage data that is recorded in each TO's database. The probability of events can be multiplied with the RTS time, on an asset by asset basis, to give the expected asset unavailability (this is an important reliability measure and more details can be found in ref [9]). Statistical analysis on outage data associated with maintenance, repairs, trips, and replacement schemes can also be used to calculate the circuit unavailability. A comparison can be carried out between the circuit unavailability and asset unavailability for validation purposes. It is also understood that a circuit outage can be used to address multiple assets and may address multiple activities (such as maintenance or repair jobs) hence the overall asset unavailability may be large than the overall circuit unavailability (failures may happen concurrently, however the outage will bundle the failures that either happen concurrently or in sequence but within a given time window). Some moderate data cleanse and data manipulation effort will be needed to assign outages to asset types and event types. Other information can also be taken into account such as asset health. This is already demonstrated as achievable, that the TOs have been reporting the repair related circuit outages (average circuit unreliability, or ACU) since at

least 2010, broken down by asset types and asset health index. Some understanding of the transmission system is also required to perform those data cleanse and data manipulation. For example, to enable a like-for-like comparison, it is important to assign a suitable duration to each asset if one outage addresses multiple assets. The RTS time of the circuit must be limited by the longest RTS time of an asset in the circuit while the RTS time for other assets may be considerable smaller. The free text recorded with each outage may be used to provide valuable information. The outage durations, when broken down by asset types, and event types, will be compared with those calculated from the RTS time and probability of events from the NOMs methodology. Due to the objectivity and good data size, calibration can be also carried out on such a data driven fashion.

Validation case 4.3.1: Use asset outage databases to calculate the asset unavailabilities and compare with those calculated using the probability of events and the RTS of the events. The comparison study can be also be broken down by asset types, event types, and asset health index. The broken- down exercise requires further data cleanse and data manipulation effort similar to those when report ACUs during regulatory reporting.

Testing case 4.3.2: Use outage database combined with asset database (asset health index, etc.) to test the assumptions around RTS that is generated by the FMEA process, e.g. the RTS can use a single value to represent all the events/assets of this type; the RTS time is insensitive to asset condition, etc. Such testing cases will rely on the assumption log from the modified NOMs methodology.

Calibration case 4.3.3: If large discrepancies are found, the outage database will need to be used to calibrate the RTS time using the data cleanse and data manipulate approach discussed above.

Direct customer disconnection - approach 1

From the transmission system point of view, the customer disconnection (i.e. loss of supply, loss of generation) is associated with events that occur on more than one circuit. When an asset fails, how customers are disconnected and how the demand/supply can be recovered is determined by a number of factors, including local system topology, the ability for the failure to be isolated and recovered, emergency response, and the demand/supply level at the time of the event. As an example, the mechanism of loss of supply is illustrated in Figure 11, showing the relationship between the asset, system, asset management activities, and the costumers. The methodology adopted in the NOMs methodology provides a bottom-up approach in calculating the cost for customer disconnection due to asset failures, by taking into consideration the system redundancy and the probability of material consequence. Similar to the safety and environmental events, the customer disconnection is also a rare event which can't be calibrated for each asset by directly using the historical database. Similar to case 4.2.1, a monetised risk model can be used to aggregate the probability of event and cost of event, from the asset level to the system level, and then compared against (the aggregated cost) with the historically recorded cost information. Using loss of supply as an example, the historical events for MWhrs lost due to failures of lead assets, (the cost can be obtained when multiplied with the unit cost of MWhr failed to transmit[18]), are available for the past

[18] The unit cost per MWHr is normally an economical figure that is agreed between parties therefore can be treated as explicit. For example, the transmission license [1] sets the loss of supply incentive is £16,000 per MWhr in 2009/10 prices.

20 years. The dataset should be complete regarding the historical events. But assigning the asset type to those events will require subject matter knowledge. For example, some post event analysis may be needed if an event is triggered by failure of a lead asset and a non-lead asset (failed concurrently). It might be unclear how the cost of consequence is split between the two assets and only the lead assets should be used for a like-for-like comparison. If such a split is proved to be difficult, the historical data analysis will give at least an upper and lower limit of expected aggregation from the (asset level) monetised risk model and can be used to validate the cost of consequences derived from the modified NOMs methodology.

Validation case 4.3.4: Use TOs' historical customer disconnection database to validate the cost of system consequences from the NOMs methodology. Some data manipulations will be required to differentiate the cause of the events. This validation case can be used for disconnecting either side of supply and demand.

Direct customer disconnection - approach 2

During the past decades, considerable efforts have been devoted to model the reliability of the electricity transmission system. This type of system reliability model is achieved by 1) setting up a model incorporating the real system topology on an asset by asset level and/or the running arrangement of the substation; 2) details of scenarios and that will lead to full or partial loss of transmission service, as well as how the asset owners' and system operators' respond to restore the service; and, 3) probabilities of the scenarios in (2) will occur in the asset level that is configured in (1). Some examples of the system reliability modelling can be found in ref. [13, 14]. One of the key inputs is the (asset level) probability of the failure events and the return to service time. Depending on the scale and the assumptions of the models, extra information may be required such as common mode failure or collateral damage (see the section 3.2 for the discussion around interdependence), weather related failures, failures of non-lead assets such as protection system, maintenance schedule and emergency response, etc. The system reliability models have got some merit by giving indications how reliable the local system is and the relative importance of each asset regarding the total system unreliability. The outcomes of the modelling are often used for decisions such as system reinforcement, outage management, post-event risk assessment, etc. However, when considering an increased number of assets, the models can get complicated very quickly therefore only a small part of the transmission system is often modelled. The results of the modelling are also heavily reliant on the input data which is regarded as the limiting factor for the predictive power of the whole model.

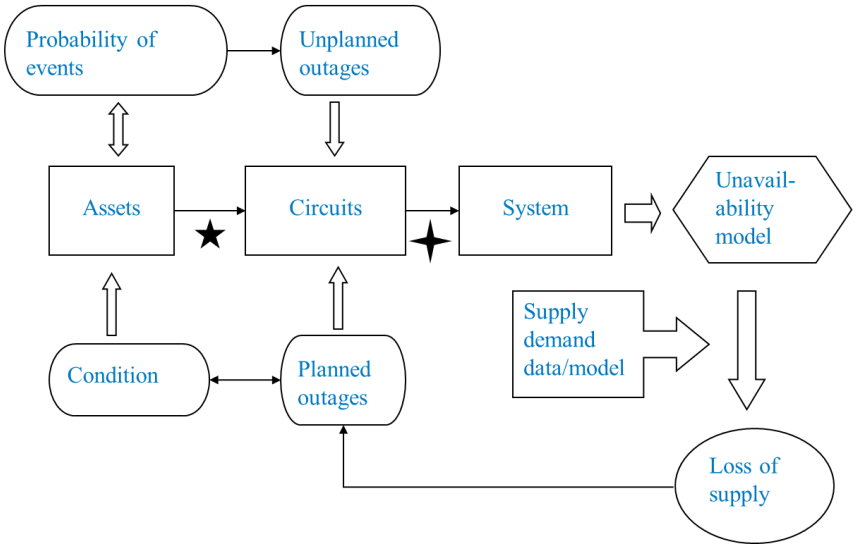


Figure 11 An schematic model used in the system reliability studies regarding the loss of supply events.

This type of system reliability modelling can however be ideal to be used to provide testing, validation and calibration cases for the cost of system consequence in the NOMs methodology. An illustration of the modelling approach is shown in Figure 12, in which only the circuits (defined as from bus-bar to bus-bar) are modelled in the asset level granularity. It is assumed that the majority of the reliability related events occur at the circuits (while bus bars are relatively reliable). The circuit unavailability can be calculated by knowing the configuration of the circuit, the probabilities of events and the return to service time of each asset. A Monte Carlo analysis will then be used to translate the asset unavailability to the circuit unavailability while some analytic solution is also available[19]. Note that such circuit unavailability is similar to that detailed in case 4.3.1, with the exception that some failures may not necessarily and/or immediately lead to an outage. The system configuration will then be incorporated into the model to translate the circuit unavailability to the system unavailability. Finally, the model calculates the loss of supply by knowing the demand profile, the planned outages such as maintenance schedule, and other implications such as weather impact. A simplified modelling scheme is illustrated in Figure 13.

[19] Some assumptions may be required for analytical solutions to calculate the system consequences. For example, the system consequence for a circuit failure should not be the sum of the system consequence for individual assets in the circuit. Once an asset is out of service the whole circuit will be out of service - those two events should have same system consequences. At the same time, when two assets in a same circuit are out of service concurrently, only one system consequence should be considered. This may lead to some double counting during those concurrent events however due to the extremely small probabilities of those concurrent events the material impact is regarded as negligible. This assumption is also implicitly used in the DNO methodology [11] - the probability of failure for linear assets (cables and overhead lines) is modelled as per kilometre.

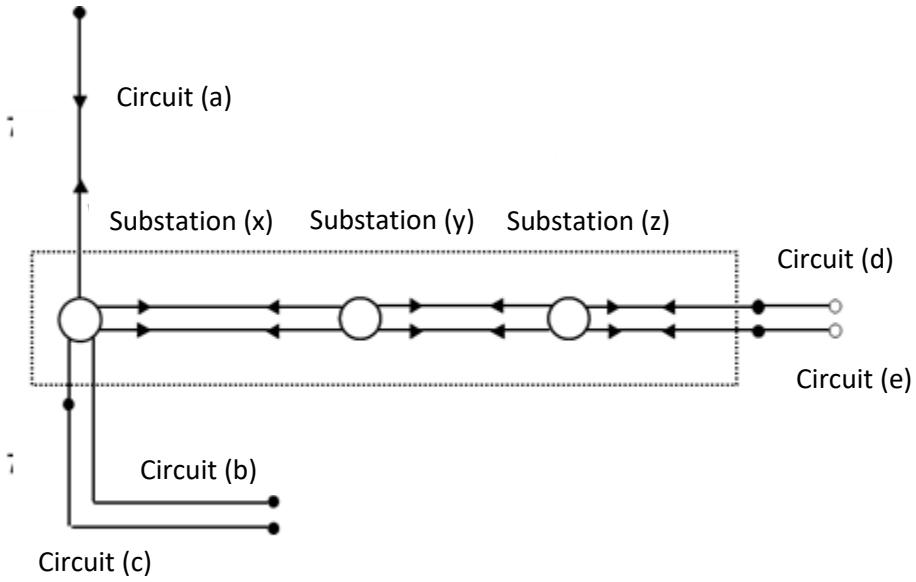


Figure 12 An illustration of the topologic of the transmission system that is used in the type of reliability modelling.

A validation case is designed here by using the existing system reliability model and inputs from the (previous) case studies, to calculate the loss of supply for some substations or local regions. The results can be compared with the loss of supply calculated using the modified NOMs methodology for the assets in same substations and regions. A testing case can be also performed by using the existing system reliability model but inputs from the NOMs methodology, to give assurance to the process of the NOMs methodology calculating system consequence of customer disconnection.

Validation case 4.3.5: Use established system reliability modelling to validate the cost of material consequence for substations and local regions on the system. It is proposed to carry out at least two different regions to perform this validation exercise to avoid the situation of ad hoc modelling inputs.

Testing case 4.3.6: Use established system reliability modelling but the input data from the modified NOMs methodology, namely, the probability of events and the return to service time, to test the process of calculating the system consequence regarding the customer disconnection.

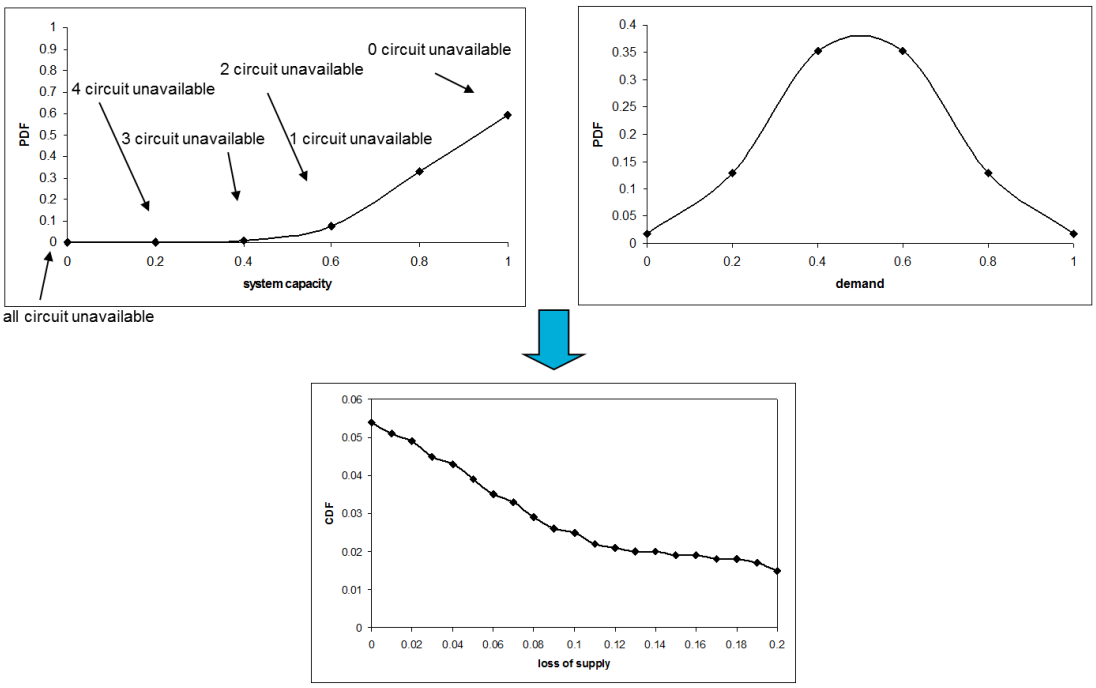


Figure 13 An demonstration of how the loss of supply can be modelled using probabilistic method by combining circuit unavailability and demand profile.

Constraint cost

Theoretically, the constraint cost can be also calculated using the system reliability modelling technique detailed in the above section. Similar to those shown in Figure 13, after an asset failure, the total transmission capacity across all boundary circuits (of region that the asset is belonged to) may be compromised. The reduced boundary transfer capacity can be compared with the demand to be transmitted across the regions to calculate the MWhr of energy constraint, which is then translated to £cost through the charging schemes. This is however a huge task as every single asset within the region needs to be considered by the system reliability model. Even if only the assets in the boundary circuits are considered, to calculate the capacity loss when losing a boundary circuit also needs to take those non-boundary circuits into consideration. This is because the system operator is able to re-configure the topology within the region to maximise the boundary transfer capacity when losing one or more boundary circuits. While establishing/seeking this type of model may be difficult, an alternative and simplified validation case is proposed in this section. For a number of boundary circuits with known configuration and known transmission capacity, the loss of boundary transfer capacity after losing one circuit must be larger or equal to the total capacity minus the capacity of the circuit. Such logic will provide an envelope calculation for the constraint cost regarding boundary transfer: the constrained cost calculated from the modified NOMs method must be smaller or equal to assuming the above case. Firstly, the probability of failure and return to service time (assuming the values of which are already validated) for each asset are used to calculate the circuit unavailability for all the boundary circuits. Then the boundary capacity can be calculated in a probabilistic manner using the envelope calculation detailed above. Finally the energy constrained can be obtained by comparing the reduced boundary transfer capacity with the demand

profile. The energy constrained calculated using the approach proposed in this section will be higher than the real energy constrained which is detailed in the NOMs methodology hence provides a validation case from the upper bound.

Validation case 4.3.7: Use a simplified linear model proposed in the current section to provide an envelope validation (upper bound) for the constraint cost upon asset failure. Ideally two sets of boundary circuits are considered to avoid the situation of using ad hoc inputs.

Reserve cost and auxiliary service cost

It is extremely difficult to perform a system reliability modelling study to evaluate the reserve cost and auxiliary service cost, as those costs can be incurred from across the whole system. Inspired by the validation case 4.3.4, it is proposed that the reserve cost and auxiliary service cost proposed in the NOMs methodology be validated against the recorded database. On the one hand, the monetised risk model can be established using the probability of failure and the cost of consequence, on an asset by asset basis. On the other hand, the recorded cost data including reserve cost, and auxiliary service cost in the past 10 years [2] are used to compare with the outcomes from the risk model. Some data manipulation effort is require to make sure those events are associated with failures of lead assets and the comparisons are like-for-like. The post event analysis reports will be a good data source and manual work is expected to gather useful data from those reports. The validation case proposed here also tests the validity of the simplifications and assumptions (which are regarded as essential given the complexity and scale of the problem) proposed in the NOMs methodology associated with those consequences. This approach is also regarded as applicable to validate constraint cost on top of the validation case 4.3.7.

Validation case 4.3.8: Use the historical events associated with constraint costs, reserve cost and auxiliary service cost to validate those aspects of system consequences from the modified NOMs methodology. Some data manipulation effort might be required to differentiate the cause of the events.

Loss of supply to vital infrastructures

The modified NOMs methodology (implicitly) assumes that the cost to consumers from a normal loss of supply event is included in the ENS (energy not supplied) incentive [1] which is passed from TOs' revenue back to the society and consumers. The ENS [1] approach ignores the nonlinear effect of the loss of supply impact and the different cost profile to different types of consumers (e.g. industrial vs. domestic). The flat rate[19] in ENS serves the purpose as a representative value that highlights the performance of the TOs, i.e. volume of energy (failed to) transmitted. On the other hand, the modified NOMs methodology[4] proposes an additional cost item if an asset supports a vital infrastructure, which includes transport hubs, economic key points, and particularly sensitive COMAH sites. Specifically, the modified NOMs methodology uses three different unit costs for different vital infrastructure types. Instead of using cost of energy per unit MWhrs, the unit cost per minute is used to calculate loss of supply to those vital infrastructures. A quick research concluded that there is currently no international standard regarding the treatment and the cost of those consequences. It is therefore important to 1) test the assumption that the "unit cost per minute" model leads to an appropriate cost of overall consequence regarding vital infrastructures, and 2) validate the values of "unit cost per minute" against different types of vital infrastructures. Assuming

that the duration of loss of supply on an asset-by-asset basis is already validated using the methodology proposed above (cases 4.3.4-4.3.6), it is proposed to use historical loss of supply events to test the assumption and to validate the unit costing values. For example, the current NOMs methodology [3] detailed a number of loss of supply events associated with different types of vital infrastructures. The event costs and duration of those events can be plotted together, with other known econometrical case studies (for example, ref. [15]) to test and validate the modified NOMs methodology. It is also worth to understand and justify whether cost of losing supply to vital infrastructure is double counted with the cost of losing supply to normal customers covered by the ENS incentive[19].

Testing case 4.3.9: Use external referenced event reports to extract the cost information and duration information. This data can be used to test the assumption whether the cost of events is suitable to be represented by duration, i.e. unit cost (per minute), for different types of vital infrastructures.

Validation case 4.3.10: If the above case can be justified, the same dataset can be further used to validate the values of unit cost (per minute) proposed in the modified NOMs methodology.

5. The monetised network risk

5.1 Introduction

The final element in the NOMs common methodology is the overall network risk. It is an aggregation from the monetised risk calculated for individual assets, which is calculated from the framework that monetised risk is equal to probability events multiplied by the (material) consequence of failure (Figure 4). This is a well-accepted framework that is used in the DNOs' NOMs framework [11] (also accepted by the Authority), in which the overall network risk is calculated directly via summation across the asset register. Such direct summation implicitly assumes the (condition related) asset failures are independent while some common mode failures or collateral damage are regarded as none material. The calculation of the overall network risk is therefore, based on this assumption, a pure deductive process. By definition if all the elements in the preceding parts of this document are tested, validated, and calibrated, the aggregated monetised network risk will be automatically tested, validated, and calibrated. A number of further validation cases are however proposed in this section to provide extra assurance to the methodology, as well as to demonstrate some overarching asset management principles embedded in the monetised risk methodology.

5.2 Life cycle costing models

Managing assets throughout their whole life is one of the key concepts in the modern asset management framework. The life cycle cost of an asset or asset system presents a single economic evaluation of the total costs, risks, and other business impacts associated with the ownership over its life cycle [16]. The life cycle costing model takes into consideration the cost of acquiring an asset,

operational and maintenance costs, risk costs of asset failures and consequences, de-commission cost, and other costs such as lost opportunities, residual liability after de-commission, etc. The TOs frequently perform studies assessing the cost and risk of asset ownership in the pre-construction phase such as pioneering. Either quantitative models such as life cycle costing ([e.g. ref. [17]) or qualitative evaluations such as whole life value (e.g. ref. [18]) are used to strike a balance between risk, cost, and performance.

The risk costing is an important aspect in the life cycle costing model. Due to the modern asset management ideology, it is not only a significant costing item, rather, to bear, mitigate, or transfer risks can also create business opportunities to achieve an organisation's strategic objectives. The first validation case is to use the optioneering process that is aiming to minimise annual cost throughout the life of the asset ownership. While the cost data is relatively easy to obtain from each TOs' scheme papers, work orders and regulatory reporting information, the risk data from the modified NOMs methodology can be used to calculate the risk costing to complete the life cycle costing modelling. Different options of asset management: e.g. different designs (AIS vs. GIS), different build (overhead line vs. underground), different interventions (replace vs. refurbish) can be examined to carry out case studies. The outcomes from those case studies, i.e. those preferred options, will be compared with the previous decisions made in order to validate the risk costing information in the modified NOMs methodology (some other factors and constraints may be considered for a like-for-like comparison). Another validation study can be carried out by directly comparing existing life cycle costing reports with those using the risk costing information from the NOMs methodology. For example, the IET [19] have published the whole life cost information across a number of options for overhead line and underground cables. Those published whole life costs provide a benchmark for the TOs - a valid monetised risk framework should lead to similar costing values (at least ballpark values) to the known case studies.

Validation case 5.2.1: Perform life cycle costing modelling using the inputs obtained from the NOMs methodology. Either equivalent annual cost or whole life cost can be used to compare with decisions made during the optioneering process or with existing case studies.

5.3 Validation against known risk assessments

Validation case 2.4.2 introduces a validation case based on the DNOs' NOMs methodology. It is largely a qualitative approach in which the overall network risks are broken down in different ways such as asset type, health index and criticality. Conclusions can be made by judging the outcomes of those exercises at the same time understanding some root cause analysis to verify that the model fulfils its designed functionalities. The current section follows a similar principle while a more structured validation framework is proposed for the overall monetised network risk.

The first validation case is to take advantage of the transition period between existing network risk measure, i.e. using replacement priorities, and the new risk measure, i.e. monetised values. For each transmission (lead) asset type, the monetised risk values for each asset can be aggregated and then categorised by replacement priorities. The tables used in the current regulatory reporting can be adopted to demonstrate the relationship between known risks (i.e. replacement priority) and

monetised risks. As shown in Table 4, for each asset type and each voltage level, the cells in the table show the number (volume) of assets, the monetised risk, and the monetised risk per unit calculated from the two former cells. While acknowledging that the two risk measures (replacement priority and monetised risk) won't be 100% compatible, they must follow the same trend at a portfolio level. By reviewing the values in each cell in Table 4, a (qualitative) validation case is proposed by comparing the network risks using different risk measures at a portfolio level. A similar validation case can also be carried out by breaking down the asset portfolio using the asset health index and criticality (also used in regulatory reporting), to calculate the monetised risk per unit volume of asset. These exercises can be further carried out for different points in time, including historical, current, and forecast network risks (then visualised in the way similar to Figure 3).

Validation case 5.3.1: Compare the existing risk measure and the monetised risk measure in a portfolio sense, using the £risk per unit asset measure and visualisation techniques such as illustrated in Table 4 and Figure 3. The results need to be summarised and reviewed to provide a qualitative validation. It is proposed that such exercise be carried out for the asset health index, criticalities, and replacement priorities, as well as different points in time (e.g. in the same format as the current regulatory reporting tables).

Table 4 A proposed table to visualise the relationship between monetised risk and existing risk assessments, from a portfolio point of view. The numbers in the table is for illustration purposes only.

Asset Categories	**Units**	**Reporting Year End Replacement Priority**											
		RP1			**RP2**			**RP3**			**RP4**		
400kV Network		**volume**	**Total £ risk**	**£ risk per unit**	**volume**	**Total £ risk**	**£ risk per unit**	**volume**	**Total £ risk**	**£ risk per unit**	**volume**	**Total £ risk**	**£ risk per unit**
Circuit Breaker	No.	1	£200K	£200K	10	£1000K	£100K	100	£5000K	£50K	1000	£1 M	£10K
Transformer	No.												
Reactor	No.												
Underground Cable	km												
OHL Conductor	km												
OHL Fittings	km												

The final validation case proposed is to compare the monetised risk with the intervention in each TO's Capex or Opex plan. Use the non-load related business plan for example, the monetised risk framework naturally assigns priorities against each asset across all asset types. The assets that are due to be replaced in the non-load related schemes, this includes those schemes already carried out

in the past and planned in the next (for example) 10 years, should be sitting in the high ranks when the monetised risk based prioritisation is performed.

Recalling the fact that the non-load related schemes are mainly based on the risk assessment method detailed in the current NOMs methodology [3], a logic is implicitly assumed here: the current risk assessments, although largely remaining qualitative and may be inconsistent from one to another, should have captured the majority of risks, e.g. the need cases to replace a poor condition asset. A report can be drafted to summarise the findings from the risk prioritisation study, and the comparison study with the existing intervention plans. This type of validation cases, although remaining qualitative and indicative, is crucial for the TOs asset management planning and regulatory performance: transmission assets take a long time to plan and construct and re-prioritisation of their replacement will cost a significant amount of time and resource, including cost already spent to prepare the interventions (planning, pre-construction survey, feasibility study, supplier development, etc.). Another aspect of the risk prioritisation study is to introduce the cost element. The balance between risk and cost can be optimised by looking into how much risk can be reduced by investing into the network. This is in line with the spirit from the Authority that "application of the [risk monetisation] model will lead to investment decisions that maximize benefit to consumers" [2]. It is however worth noticing that such an optimisation exercise is heavily subjected to constraints. For example, the transmission assets are inter-connected with limited system assess. Other risks also need to be taking consideration such as resource, suppliers, and deliverability. To fully follow the prioritised assets using the monetised risks framework may not be optimal in terms of the overall picture. Those constraints are (expected) also to be addressed in the modified NOMs methodology (details unclear as yet) regarding justification for under- or over-delivery against the regulatory performance target.

Validation case 5.3.2: Use the monetised risk to prioritise asset intervention and compare the results with the known list of asset intervention. The findings will be detailed in a case study report in order to provide validation to the monetised network risk in modified NOMs methodology. A further case study can be included in the report detailing an optimisation study between risk and cost.

5.4 Testing the risk model and scenarios

The validation cases above will ideally be facilitated with a risk calculation model which takes into consideration all the elements discussed in sections 2-4, e.g. probability of failure, failure modes, cost of consequences, etc. The model calculates the monetised risk for each asset and then aggregates into the overall (monetised) network risk. The model will have two additional functionalities: 1) modelling the forecast of network risk as discussed in cases 2.4.2 and 3.5.4, and 2) modelling of the interventions discussed in cases 3.2.5-3.2.7. Considering the importance and the long-term utilisation of the model, it would be worth to firstly have a testing case to verify that the model is delivered/programmed in the way that it appropriately reflects those articulated in the modified NOMs methodology (that is approved by the Authority), as well as that it appropriately fulfil the design specifications detailed in e.g. ref. [2]. It is proposed to seek for independent review of the model, which will be examined in the comprehensive manner against all known

documentations including the outcomes from the testing, validation and calibration exercises. Either the TOs or the Authority is to commission an independent review report to test model, subjecting to time, cost implications and agreements between the Authority and the TOs.

Testing case 5.4.1: An independent review report to verify the logic and the functionalities using the methodology documentations, the design specifications, and the outcomes of the testing, validation and calibration exercise.

After the model is verified, it will be used to test against different asset scenarios and intervention scenarios. Such cases require both the forecast modelling and intervention modelling functionalities, which are tested and validated separately in an earlier context. In order to comply with the direction [2], the TOs will carry out further testing to, for example, verify that "the model works across a suitable range of credible scenarios". From the network replacement output perspective, it is important to compare the risk profile under different intervention scenarios. It is proposed to carry out a case study detailing the in-year network risk, and forecast network risk at end of RIIO-T1, for different years (e.g. 2010, 2013-2016), together with the capital plan that is submitted to the Authority each year. Both monetised risk approach and the existing approach (using asset health index and replacement priorities [3]) are used to understand the impact on the network risk due to: the movement of intervention plan; trade-off between capital plans or between risks and interventions; the cost of trading off; the risk evolvement such as asset deterioration. Other scenarios, such as the treatment load related schemes; early-life asset failure (as detailed in ref. [3]); alternative interventions such as circuit breaker refurbishment (which might lead to non-zero monetised risk after intervention, depending on the details of the NOMs methodology); trade-off between Capex and Opex, etc., can be also tested using the model. By pulling together all the above scenarios, it is possible to use the model to monetise the network risk under different scenarios (as well as combinations of different scenarios) and to examine whether the TOs have over- or under-delivered against the network replacement output target. While each scenarios can be examined individually as justified or unjustified (that will be further developed in the NOMs methodology), the ultimate functionality of the NOMs methodology, i.e. assessment of the TOs' performance, will be achieved.

Testing case 5.4.2: Develop scenarios under the guidance of the NOMs methodology: justified and unjustified over and under delivery. Evaluate the impact of those scenarios on the overall network risk and review the validities. This testing case will be based on a case study using known asset condition profiles and capital plans that were submitted to the Authority in the past, e.g. the RIIO-T1 business plan and regulatory reporting.

References

[1] Ofgem. National Grid Electricity Transmission Plc Electricity transmission licence Special Conditions.
https://epr.ofgem.gov.uk//Content/Documents/National%20Grid%20Electricity%20Transmission%20Plc%20-%20Special%20Conditions%20-%20Current%20Version.pdf

[2] Ofgem, 2016. Direction under paragraph 2L.13 of Special Condition 2L (Methodology for Network Output Measures) of the electricity transmission licence https://www.ofgem.gov.uk/system/files/docs/2016/04/160429_et_noms_direction_subsid_3.pdf

[3] National Grid, SP Transmission PLC, Scottish Hydro Electric Transmission Limited, 2016. Electricity Transmission Network Output Measures Methodology. https://www.ofgem.gov.uk/system/files/docs/2016/04/noms_methodology_clean_subsid.pdf

[4] Ofgem, 2016. Decision to direct electricity transmission licensees under Special Condition 2L of their electricity licences to modify the Network Output Measures (NOMs) Methodology. https://www.ofgem.gov.uk/system/files/docs/2016/04/160429_et_noms_direction_decision_main_doc_0.pdf

[5] UK electricity distribution networks, 2016. Validation of the DNO Common Network Asset Indices Methodology. https://www.ofgem.gov.uk/system/files/docs/2016/01/appendix_2_validation_exercise.pdf

[6] National Grid, 2016. Network Innovation Allowance Project Completion Report. Project Name: Overhead Line Conductors Condition Assessment Project ID: (10359 – NIA_NGET0140) - to appear in ENA portal.

[7] Saltelli A., 2002. Sensitivity Analysis for Importance Assessment. Risk Analysis. Vol. 22 (3): pp 1-12.

[8] Danielsson J., James K., Valenzuela M., Zer I., 2014. Model Risk of Risk Models SRC (Systemic Risk Centre) Discussion Paper No 11. ISSN 2054-538X.

[9] Rausand M., Høyland A., 2004, System reliability theory: models, statistical methods, and applications. 2nd editon, Wiley-Interscience.

[10] IEEE Standard 352-1987 - IEEE Guide for General Principles of Reliability Analysis of Nuclear Power Generating Station Safety Systems. IEEE Power and Energy Society, Reaffirmed 2010.

[11] UK electricity distribution networks, 2015. DNO Common Network Asset Indices Methodology: Health & Criticality - Draft Version 4. https://www.ofgem.gov.uk/system/files/docs/2016/01/dno_common_network_asset_indices_methodology_draft_v4_0.pdf

[12] Yi S. et al. (Ofgem). “A regulatory approach to the assessment of asset replacement in electricity networks”, Asset Management Conference 2011, IET and IAM, London.

[13] Ball, R. F., 2012. The Probability of Faults and Trips on the Grid System. National Grid.

[14] Awadallah S. K. E., 2014. Probabilistic Methodology for Prioritising Replacement of Ageing Power Transformers Based on Reliability Assessment of Transmission System. PhD thesis, University of Manchester.

[15] Ralph D., Kelly S., and Copic J. 2015. Erebos Blackout Cyber Risk Scenario for US Electrical Distribution Network. Cambridge Centre for Risk Studies.

[16] The institute of asset management, 2016. Lifecycle Manager Whole Life Costing. Subject Specific Guidelines. The IAM.

[17] Woodhouse J., 2015. Asset management decision-making: The SALVO Process. ASIN: B00WZ1H6DG

[18] National Grid, 2013. Whole life value framework. Policy Statement - internal and contract specific. PS(T) 101 - issue 2

[19] Electricity Transmission Costing Study - An Independent Report Endorsed by the Institution of Engineering & Technology. Parsons Brinckerhoff, 2012.
http://www.theiet.org/factfiles/transmission-report.cfm