Mortality-analysis checklist

 An updated version of this checklist can be found at https://www.longevitas.co.uk/information-matrix-page/importance-checklists

Use this procedural checklist to ensure the data are valid and the model conclusions are robust. Using this checklist saves time overall by catching bad data early, reducing mistakes and minimising the need for rework.

1. Purpose, data nature and source

Q1.1 What is the project name?
Knowing the project name helps in ensuring there are no conflicts of interest.

Q1.2 Who is the data owner? Is there a chain of permission for use?
Knowing the data owner helps in ensuring there are no conflicts of interest.

Q1.3 What is the purpose of processing?
Examples include setting a reserving basis, pricing a bulk-annuity or reinsurance transaction, or scientific research.

Q1.4 Are there any potential conflicts of interest?
Conflicts of interest can only be identified if the client discloses enough detail in Q1.1, Q1.2 and Q1.3.

Q1.5 For commercial bases, what is the effective date of the valuation or transaction?
Any time trend will need to be forecast past the end of the usable data to the effective date.

Q1.6 Where do the data come from?
Data ideally come directly from the administration system with the minimum of filtering or pre-processing.

Q1.7 Are fields available to support deduplication[1]?
Examples include names, full postcodes or client identifiers. Note that such fields may make the data personally identifiable information (PII)[2].

Q1.8 Are there unusable parts of the experience?
Examples include past dates of archival of historic deaths.

Q1.9 Are there systematically missing deaths?
If so, consider using the `EarliestActivityDate` *or* `LatestUsableDate` *fields[3] [4].*

Q1.10 Is only part of the pension being (re)insured, or only some lives?
If so, specify the `ValuationAmount` *field[3] to value only the benefits being (re)insured.*

Q1.11 Have the benefits been indexed or increased over time?
Benefits to deaths and other exits need to be revalued to the end of the exposure period.

2. Longevitas configuration

- Set the deduplication merging options according to the data available in Q1.7.
- Set the interest rate or yield curve for valuation.
- Set the model boundaries.
  For example, set the "Reject beyond age" to 105 (say) to control for age errors[5].
- Set the gradient calculation method to derivatives.
- Set the Hermite parameters, e.g. `x0`=50, `x1`=110, `x2`=50, `x3`=105.

[1] For background details on deduplication, see https://www.longevitas.co.uk/information-matrix?tags=deduplication&year=all
[2] Note that various tools exist to depersonalise PII, including the "Transform on Download" HOWTO in the **Resources** area of your account.
[3] See Section 5.1.8 of the Longevitas User Guide in the **Resources** area of your account.
[4] See the chapter headed "Exposure calculations" in the Technical Guide in the **Resources** area of your account.
[5] For details on how age errors become problematic, see https://www.longevitas.co.uk/information-matrix-page/new-year-new-insights

3. Data processing

Q3.1 Do the data contain personally identifiable information[2] (PII) under the terms of GDPR or similar regulation?

Q3.2 What is the GDPR (or similar) basis for processing the data?
"Legitimate interest of the data controller"[6] may be used for setting reserving bases or deriving pricing bases for risk transfers like bulk annuities or longevity swaps.

Click on File Reports and click on "File Validation":

- Q3.3 What is the reject rate for uploaded records?
- Q3.4 Is there a bias in reject rate between deaths and survivors?
- Q3.5 What is the nature of the rejections? Could the records be corrected?
  Use the "Only Rejects" download under "Deduplication" to send rejected records with reason codes back to the data source.

4. Sense-checks of data

Click on File Reports and click on "File Validation":

- Q4.1 Are there suspiciously common dates of birth?
  Examples might be heaping for dates like 1st January, or lots of people "born" on 1901-01-01.
- Q4.2 Is there heaping in the date of death?
  This might indicate deaths processed without the true date of death.
- Q4.3 Is there heaping in the commencement dates?
  Heaping can validly occur for mass early-retirements as part of business restructuring. However, it might equally indicate the loss of actual commencement dates during a system migration.
- Q4.4 Is there suspicious heaping in the surnames?
  Examples include dummy surnames like "SPOUSE".
- Q4.5 Is there suspicious heaping in the postcodes?
  Examples include the changing of address on death to avoid mailing the deceased.

5. Deduplication

Click on File Reports and click on "Deduplication":

- Q5.1 Are there material numbers of duplicates?
  A "duplicate" in this sense is a valid second benefit record, e.g. a pensioner with both a main benefit and a surviving spouse's benefit.
- Q5.2 Are there significant numbers of conflicts, i.e. where a life is marked alive on one record and dead on another?
  This can indicate poor data quality or an insufficiently robust deduplication key.

[6] See https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/

6. Data audit

- Run a data audit
- Click on the  icon and note any warnings highlighted by the expert system.

Click on the  icon:

- Q6.1 Does the semi-parametric estimator[7] show the expected seasonal variation?
  In the Northern Hemisphere we expect peak winter mortality in January and a summer trough in July.

- Q6.2 Does the semi-parametric estimator exhibit the expected covid-19 shocks?

7. Identifying usable data ranges

- Fit a Perks model with the unrestricted age range and full date range.
- Use only Age as a risk factor initially.

Using the  tab:

- Q7.1 Does the plot of log(crude mortality rate) exhibit the expected "hockey-stick" shape above age 50?
  Plot Actual deaths/Exposure against Age on a log scale.

- Q7.2 Are there suspicious patterns above age 95?
  An example would be falling mortality rates.

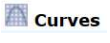
- Q7.3 Are there any discontinuities suggestive of missing deaths?
  Plot Actual deaths against Time.
  Plot Actual deaths/Exposure against Time.
  Plot Deviance residuals against Duration.

Q7.4 What is the usable age range?
Start modelling from the age where appreciable exposures and deaths occur.
Stop modelling before ages demonstrate suspicious patterns in Q7.2.

Q7.5 What is the usable date range?
See answers to Q6.1, Q6.2 and Q7.3 and remember to have equal numbers of each season[8].

8. Identifying subtle data-quality issues

- Fit a Perks model with the restricted date range and from age 65 upwards.
- Enable Kaplan-Meier under  .
- Use only Age and Gender as risk factors.

Q8.1 Is there clear separation between male and female Kaplan-Meier curves at all ages?
Plot the Kaplan-Meier survival curves on the  *tab to check[9]*

Q8.2 Is there crossover between male and female Kaplan-Meier curves at high ages? If so, this can indicate “immune” cases that could distort estimates of age interactions.

Q8.3 How are late-reported deaths handled?
Examples include ignoring the last six months of experience, fitting a delay function[10] or using a flexible spline basis.

[7] See https://www.longevitas.co.uk/information-matrix-page/visualising-data-quality-time
[8] For details of seasonal variation and its importance, see https://www.longevitas.co.uk/information-matrix?tags=season&year=all
[9] For examples of good and bad data, see https://www.longevitas.co.uk/information-matrix-page/spotting-hidden-data-quality-issues
[10] For details of delay functions, see https://www.longevitas.co.uk/information-matrix-page/allowing-reporting-delays

9. Testing risk factors

- Perform initial investigation using the Makeham-Perks model for speed.
- Under Advanced Options:
  - Ensure Kaplan-Meier, bootstrapping and mis-estimation options are switched off to shorten run times.
  - Set the rate-table effective date to be *six months after*[11] the answer in Q4.
  - Set a suitable revaluation rate for deaths and other exits using the answer to Q10.
- Start with simple Age model, then Age+Gender, then Age+Gender+Time.
- Use Reuse to add risk factors one at a time, rejecting ones that increase the AIC and keeping ones that lower it.
  Beware of factors that seem suspiciously powerful, as data corruption may have made them proxies for the death status.
- Ignore age interactions at this stage.
- Use the Model Optimiser to define pension-size bands.
  Use "Optimised grouping" to perform an ordinal optimisation on the pre-calculated SizeBand factor.
- Consider re-optimising any groupings to check that mappings haven't changed due to adding risk factors.

10. Finalising the model

- Reuse the last Makeham-Perks model with the desired risk factors.
- Switch to a Hermite I or II model[12] to make mortality differentials automatically narrow with age.
- Under Advanced Options:
  - Set the rate-table effective date to be *six months after*[8] the answer in Q4.
  - Set a suitable revaluation rate for deaths and other exits using the answer to Q10.
  - Enable both bootstrapping and mis-estimation.
- Include the Gender:Oldest interaction so that males & females have different limiting mortality rates.
- Include the AgeTimeTrend parameter to estimate the portfolio's own improvements.
- From the finished model open the Valuation section and click on Mis-estimate to find the equivalent percentage of a standard table for the best estimate and a confidence interval for it.

If there is no mis-estimation icon in the Valuation, you forgot to enable mis-estimation under Advanced Options.

Q10.1 Is the bootstrap A/E percentage close to 100% for amounts as well as lives?
Lives-weighted bootstrap results are always close to 100% for models fitted by maximum likelihood, but a model needs an amounts-weighted bootstrap result close to 100% to be suitable for financial purposes.

11. Final checks

Q11.1. Is the best-estimate basis consistent with the traditional amounts-weighted A/E analysis?
Use the A/E link to generate a traditional actuarial A/E analysis by lives and amounts.

Q11.2 Is the best-estimate basis consistent with the Kaplan-Meier curves for a simple age-gender model?
Use the Kaplan-Meier survival-curve plot in the Curves tab for the model in Q8.1 with a standard-table overlay to verify the lives-based A/E percentages.

12. Further reading

See the User Guide or Technical Guide, both of which are available in the Resources section of your Longevitas account.

See our published papers on actuarial mortality analysis at https://www.longevitas.co.uk/published-papers

[11] For a q_x rate to apply over a one-year period [0,1), any mortality improvement in the model needs to project to time t=0.5.
[12] For details of Hermite-spline models, see https://www.longevitas.co.uk/information-matrix-page/hermite-model-mortality