**Estimating a suitable value of**

For the mechanistic selection methodologies (smooth selection and truncation selection), the standard deviation of the Normal distribution of the Polygenic Resistance Score (PRS) is manually input into the model as a parameter. In the standard methodology, this is accounted for in the exposure scaling factor as part of the calibration process.

There is therefore a need to determine what are suitable values of the population standard deviation of the population mean PRS. In order to calculate this, mean bioassay mortality results (with a descriptor of the spread of the data – either standard deviation, standard error or 95% confidence interval) needs to be obtained for WHO cylinder bioassays conducted on wild mosquitoes.

WHO cylinder bioassay values were obtained from:

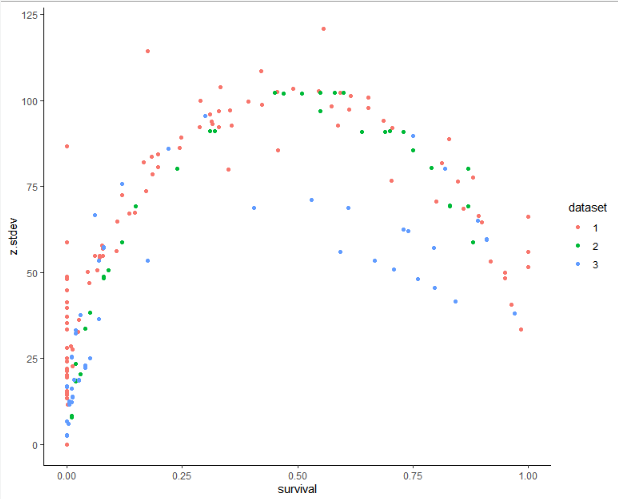
1. (Thomsen et al., 2014)
2. (Wat’Senga et al., 2018)
3. (Alemayehu et al., 2017)

These papers provided mean mortality estimates with 95% confidence intervals for WHO cylinder bioassays as percentages. Where the number of bioassays was not given, but the number of mosquitoes were, this was estimated by dividing the total number of mosquitoes by 25 and rounding to give an integer, as 25 mosquitoes is the recommended number of mosquitoes per bioassay according the WHO insecticide susceptibility testing guidelines [REF]. The data is from both the *An. gambiae* and *An. funestus* complexes. Data was extracted for all insecticides the WHO bioassays were conducted for.

The 95% confidence interval and sample size (number of bioassays) was then used to calculate the standard deviation (<https://handbook-5-1.cochrane.org/chapter_7/7_7_3_2_obtaining_standard_deviations_from_standard_errors_and.htm>).

Where would be the number of bioassays conducted.

This then gives the standard deviation of the bioassay results as a percentage. This was then divided by 100 to give the proportion. This value was then put into the Michaelis-Menten (Hill variant) equation (Equation 1(a)) to convert this standard deviation to a Polygenic Resistance Score.



This looks like the standard deviation behaves very strange at the 0 survival mark. There is an overall inverse U shape to the graph, such that SD appears to increase until 50% survival, before then decreasing again between 50% and 100% survival. There are a few explanations for this behaviour:

1. At low survivals the expected Normal distribution values of z include individuals with a z less than zero, these individuals would still have a survival of 0%.
2. Limited by the 0 to 100% survival probabilities.

To account for this one solution is to use the standard deviation estimates for mean survivals between 25-75% which allow for sufficient room either side of the mean to not be cut off by the maximum and minimum survival in a bioassay.

It should be noted that the model is preferentially designed to run for novel insecticides, withdrawing insecticides at 10% bioassay survival. Therefore, restricting the analysis to only include replicates where the bioassay survival was 10% or less (90% or greater bioassay mortality) is likely to be more useful in terms of assessing a standard deviation that would be representative of the PRS the model is most likely to encounter. This is also important giving the curved relationship between bioassay survival and the Polygenic Resistance Score; such that a 1 unit increase in the PRS does not equate to exactly the same unit increase survival across all values of the PRS. Therefore restricting the analysis to values where the mean PRS would be 100 or less reduces the impact of this curved relationship.

Chart, box and whisker chart

Description automatically generated

Figure 1 Histogram of the standard deviation of the Polygenic Resistance Score from field sampled populations. Green indicates all insecticides combined.

Mean = 28.5 (IQR: 16.1 – 40.2)

For smooth selection:

Use standard deviation values in the 15-40 range and exposure scaling factor 1 or 10. See which gives, when using 1 insecticide in sequence, a calibration that gives the most frequent time of 100 generations.