Appendix 1 – Further mathematical details on obtaining interpretable measures of variation from mixed effects logistic and Poisson regression models

When fitting an empty random intercept model as described in this paper only two parameters are estimated. The first is the constant term, β_0, which describes the mean on the log rate scale for Poisson models and the mean on the log-odds scale for logistic models. The second parameter is the variance (or sometimes the standard deviation) of the random intercept, σ_b^2. These two parameters define the fitted normal distribution. Centiles on the log-odds or log-rate scale can be calculated from these parameters as with any normal distribution i.e.

$$\beta_0 + \Phi^{-1}(p)\sigma_b \qquad\qquad 1$$

$\beta_0 + \Phi^{-1}(p)\sigma_b$, where $\Phi^{-1}(p)$ is the appropriate centile of the inverse cumulative normal distribution. For example to estimate the 75th centile on the log-rate or log-odds scale we must first obtain the inverse cumulative normal distribution $\Phi^{-1}(0.75)$ which is approximately equal to 0.674, and apply that to Equation 1, i.e. $\beta_0 + 0.674\sigma_b$. The next step is to transform onto the rate or proportion scale. For a rate indicator this is simply a matter of exponentiating the centile i.e.

$$e^{(\beta_0 + \Phi^{-1}(p)\sigma_b)} \qquad\qquad 2$$

In the case of proportion indicators the inverse logit transform, $e^y/_{e^y + 1}$, should be applied giving

$$\frac{e^{(\beta_0 + \Phi^{-1}(p)\sigma_b)}}{e^{(\beta_0 + \Phi^{-1}(p)\sigma_b)} + 1} \qquad\qquad 3$$

The same approach can be used to calculate odds ratios or rate ratios between two centiles. As above we calculate the centiles of interest from the fitted distribution on the log-odds or log-rate scale. We then subtract one from the other to obtain a log-rate ratio or a log-odds ratio and then exponentiate to obtain a rate ratio or odds ratio. Assuming centiles equidistant from the 50th centile are used, this can be simplified as

$$e^{(2\Phi^{-1}(p)\sigma_b)} \qquad\qquad 4$$

As an example the rate ratio or log ratio for the 75th compared to the 25th centile is given by $e^{(1.348\sigma_b)}$, where 1.348 is twice 0.674 (the 75th centile of the inverse cumulative normal distribution discussed above). The related MOR is given by

$$\text{MOR} = e^{\sqrt{2}\sigma_b \Phi^{-1}(0.75)} \qquad\qquad 5$$

# Appendix 2– Graphical representation of random intercept for proportion indicators

One approach involves calculating the probability density function (PDF) on the modelled scale for a set of evenly spaced values on the proportion scale and then rescaling for plotting on the proportion scale. The process is outlined in Appendix Figure 1 (using larger spaced x-values than would be used in reality). The first step is to assign evenly spaced x-values suitable for plotting the distribution on

the native proportion scale (Appendix Figure 1a). These x-axis values are then transformed to the log-odds scale (Appendix Figure 1b). Note that the evenly spaced dots (Appendix Figure 1a) are no longer evenly spaced once the transformation is made (Appendix Figure 1b). The PDF is then calculated using the output of the mixed effects model as parameters in a normal distribution. Here we are using a mean of 0.5 (equivalent to 62% on the percentage scale) and standard deviation of 1 (Appendix Figure 1c). The PDF then needs scaling for use on the proportion scale to account for the different spacing of x-axis values. We need to maintaining a constant the area under the curve both in each segment of the curve and in total (total area under the curve should always be equal to 1). The blue and red shaded regions in Figures 6c and 6d relate to the same parts of the distribution. In Appendix Figure 1c the blue region is approximately twice as tall as the red region while being approximately half as wide, indicating that the areas of the two shaded regions are similar. In Appendix Figure 1d (following transformation and scaling) the red and blue areas have the same width by definition and we see that the heights are similar thus reproducing the similar areas seen in Appendix Figure 1c. Note that if the distribution is to be shown against a percentage (0-100) rather than a proportion (0-1) the density should be divided by 100 to maintain an area under the curve of 1.

The scaling required to transform a PDF on the log-odds scale, f_{lo}, to a PDF on the proportion scale, f_p, is given by

$$f_p = \frac{f_{lo}(e^{-2x_{lo}} + 2e^{-x_{lo}} + 1)}{e^{-x_{lo}}} \qquad\qquad 1$$

Where x_{lo} is the log-odds value of the x-axis point being transformed.
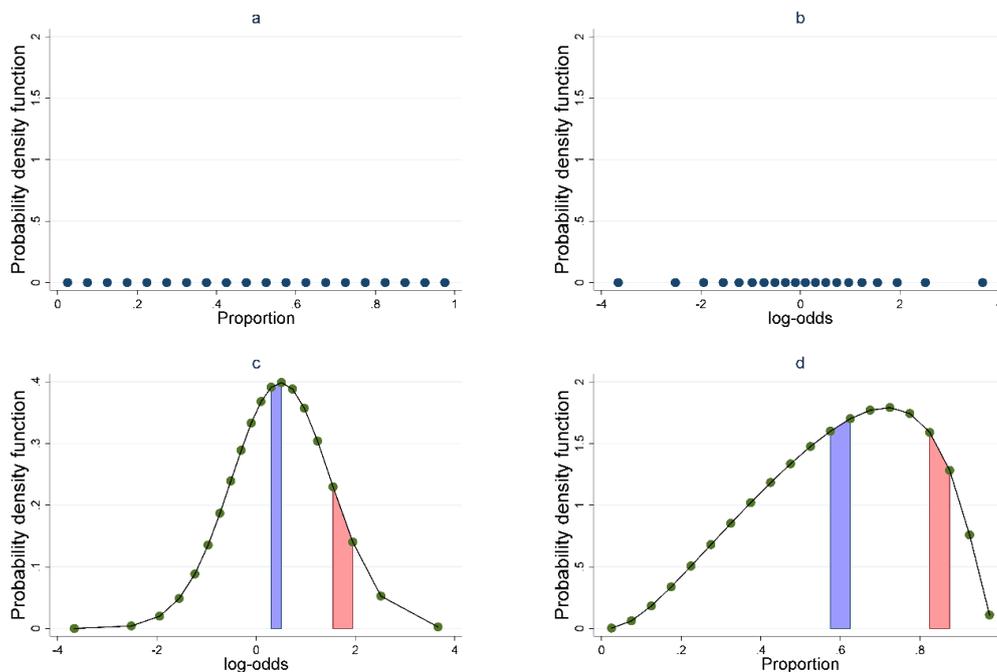
Appendix Figure 1: Illustration of how to plot the modelled distribution from a mixed effects logistic regression on the proportion scale. Full details are given in the text.