Methodology for Developing Population Projections – Technical Briefing

BCC Population Health Management

May 2, 2024

This technical briefing outlines the methodology used to develop the population projections in the 2023/2024 Director of Public Health annual report "From numbers to narratives: Exploring the changes between the 2011 and 2021 Census in Birmingham and the implications for the future health of our city". At the time of writing the report, local authority projections using 2021 census data were not available. This briefing is being shared as an additional resource for transparency and to support others if developing similar models.

1 Introduction

The Office for National Statistics (ONS) produces population projections every two years, which are broken down by gender and age structure. The projection is essential for various levels of planning purposes and has been widely used by government sectors, local authorities, and researchers. Population projection also helps to inform decision making on resource prioritisation and targeted interventions for future needs in health care services.

The ONS currently employs the cohort component method to produce nation population estimate [Office for National Statistics, 2016]. The equation is written as follows:

$$Population_{t+1} = Population_t + Births_t - Deaths_t + Immigrants_t - Emigrants_t \tag{1}$$

The simplified age-structured version of the method is written as follows:

$$Population_{a+1,t+1} = Population_{a,t} * SurvivalRate_{a,t} + NetMigration_{a,t} \tag{2}$$

$$Population_{0,t+1} = \sum_a women_{a,t} * FertilityRate_{a,t} \tag{3}$$

The population aged 0 which is the infants in the next time point is the sum of overall ages a of the number of women currently aged a times the Fertility rate of the corresponding age in the current period. The key point from this equation is that it is deterministic and requires assumptions about future fertility, mortality, and migration rates. The ONS currently obtains these future rates from expert panel. The assumptions for the vital rates for population projection are made based on a combination of analyzing trends and making extrapolations, as well as seeking advice from experts in the field.

The cohort component of population produces a single value of the population projection. This does not assess uncertainty in which a single value is not helpful enough for decision making when one wants to be confident about the accuracy of the results. The ONS assesses uncertainty by producing additional variants based on alternative assumptions on the vital rates. For example, the ONS assumes a high and low fertility rate as the upper and lower bound range of value that the projection can vary. Expert opinions no wonder can provide valuable insights when data is limited or difficult to obtain. However, experts are not very good at producing forecasts from scratch. The need of a probabilistic projection can provide a better assessment of uncertainty which allows for a more accurate representation of the potential projection. This enables decision-makers to avoid risks when making decisions.

2 Method

Bayesian Probabilistic Projection is a model that was firstly developed for United Nations to produce probabilistic population projections for most countries [Raftery et al., 2014]. This method makes the total fertility rate (TFR) and life expectancy probabilistic, migration rate is kept deterministic as the packages do not support for migration rate to be probabilistic.

2.1 TFR

The TFR is modeled in 3 phases, the first phase is the beginning of the TFR transition in which the TFR is high and increasing. Phase 2 TFR is where the rate starts to decline from high levels and below the 2.1 replacement level (United Nation's assumption). Phase 3 is the post-fertility transition period.

Phase 1 of the TFR transition is not considered to be in part of the simulation process. Starting from Phase 2, a double logistic function with added stochasticity. The package then models the TFR for different countries from a ”world distribution”, the wpp2022 dataset(the United Nation's dataset has the latest TFR for United Kingdom) includes 237 countries or areas. The importance of a ”word distribution” is that when modelling the TFR for a chosen country, it can borrow strength from data for other countries to make the model hierarchical. In addition, estimating country specific's double logistic curve is unstable and the data for a single country is often sparse. By using a hierarchical model, the estimation process can be more robust and reliable.

Phase 3 is where the prediction of the future TFR begins. However, the model creators have placed the UN's assumption that the TFR will tend to increase back towards replacement level of TFR=2.1 after phase 2 in the long run. They model this by using a single first-order autoregressive model (AR1), namely

$$TFR_{c,t+1} = \mu + \rho(TFR_{c,t}) + b_{c,t} \tag{4}$$

The equation implies that the TFR in the next time point depends on the TFR in the present plus a constant/mean and white noise. The mean is equal to the approximate replacement level of 2.1 as being predetermined by the model creators. It means that the predicted TFR will tend to increase to 2.1 and fluctuate around this level in the long-run.

However, this long term assumption is not applicable for us to model regional level of population with two reasons. Firstly, we are only projecting our population 10 years forward in time, while the model creators project the population up to 2100, a long term replacement level is not suitable for a short-term projection. Secondly, UN's assumption on the replacement level at 2.1 might not be applicable to the UK as well as Birmingham. Having looked at the ONS website, the experts assume the future will be 1.53 in short-run and 1.59 in long run. As the model uses long run TFR, we have to fine tune the values in the AR1 model such that the predicted TFR values fluctuate at around 1.6 in the short term prediction. The reason of choosing 1.6 instead of 1.5 is that Birmingham has slightly higher historical TFR than the England average.

2.2 Life Expectancy

Life expectancy is modelled in the same way as TFR. A double logistic function is fitted to project expected gains in life expectancy. The model creators generate male life expectancy in condition on the projections of female life expectancy. We have to be careful about the joint projections of female and male because female always tend to have higher life expectancy. It is regulated by smoothing the gap between female and male. The gap is widen for female at around age 75, and then narrows thereafter.

These projections are then converted into age- and sex-specific mortality rates used in a cohort component method.

2.3 Bayesian Population Projection

In order to generate probabilistic forecasts, a total of 100 simulations were conducted for the total fertility rate (TFR) for every five-year interval between 2021 and 2031, as well as for the joint trajectories of male and female life expectancy, based on their posterior predictive distribution. These were transformed into age- and gender-specific mortality rates. The migration schedules were constructed from the total migration counts of the United Kingdom derived from the wpp2022. We have supplied the Birmingham sex-specific migration shares for the process. These simulations were then used to create a joint probabilistic projection using the cohort component method.

2.4 Rescaling the population estimate

As far as we know that Census is undertaken by the ONS every ten years, it implies that the population in between the Census year are estimated. However, the ONS has not updated their estimates for years prior to 2021, leading to an imbalanced data point between the most recent Census in 2021 and past estimates. This has resulted in significant fluctuations in certain age groups and could potentially affect the accuracy of population projections. To address this issue, we propose recaling the estimates for the years between Census years 2021 and 2011 and smoothing out the data points to ensure that population changes are more consistent and predictable.

We define the 2011 and 2021 census population values for some age range to be P11 and P21 respectively. We denote the i-th ONS estimate as Ei and define the change since the 2011 census to be:

$$\epsilon_i = E_i + P_{11} \tag{5}$$

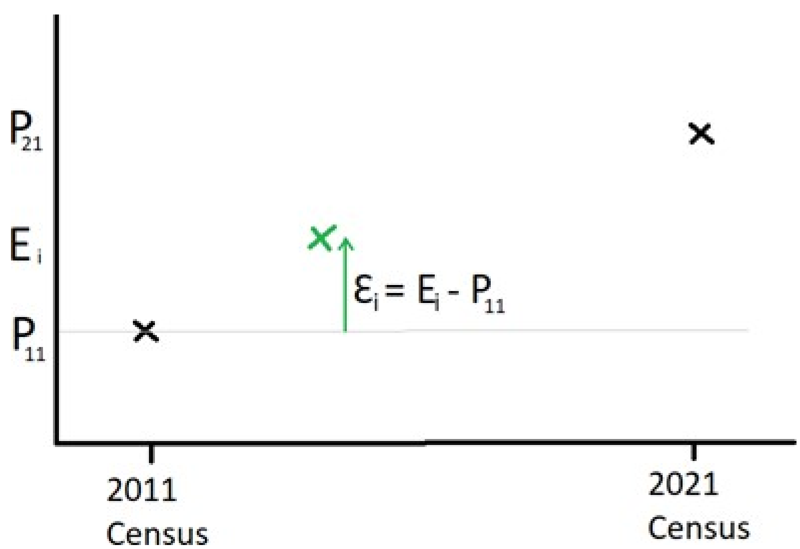

Figure 1: Method used to recale population projections.

We can then rescale all ϵi values for each age bracket by a single value α:

$$\epsilon'_i = \alpha\epsilon_i \tag{6}$$

where the prime (′) denotes that the value has been re-scaled1. We then want to choose α such that the re-scaled ONS estimate for 2021 is equal to the census value, i.e.

$$E'_21 = P_{21} \quad \Rightarrow \quad \epsilon'_{21} - P_{11} = P_{21} \tag{7}$$

Substituting in Eq. 5 and rearranging we find

$$\alpha = \frac{P_{21} - P_{11}}{\epsilon_{21}} \quad \Rightarrow \quad \alpha = \frac{P_{21} - P_{11}}{E_{21} - P_{11}} \tag{8}$$

Our re-scaled ONS estimates are then given by

$$E_i = P_{11} + \alpha(E_i - P_{11}) \tag{9}$$

Note that the value of α is likely to vary between different age groups.

3 Results

The population projections up to 2031 are given as line plot and table for convenience. The projection is calculated as a five year average and therefore no annual projection can be provided from this R package.

In figure 2, we can see that the adjusted TFR for Birmingham will fluctuate at around 1.6 and eventually drop to the level where it just sits above 1.5 in 2050. Having the TFR rate adjusted instead of using the UN's assumption can better predict the future TFR rate that is in line with the ONS expert panel's assumption.

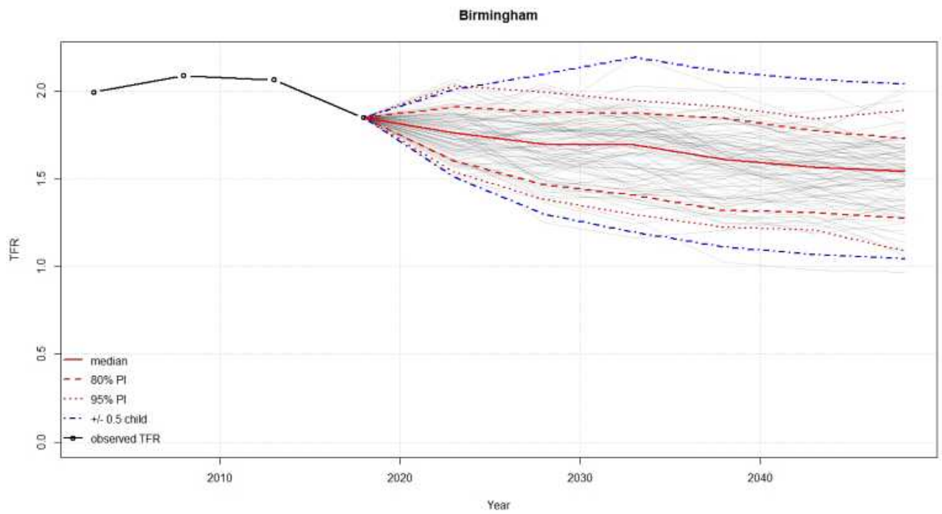

Figure 2: TFR projection for Birmingham

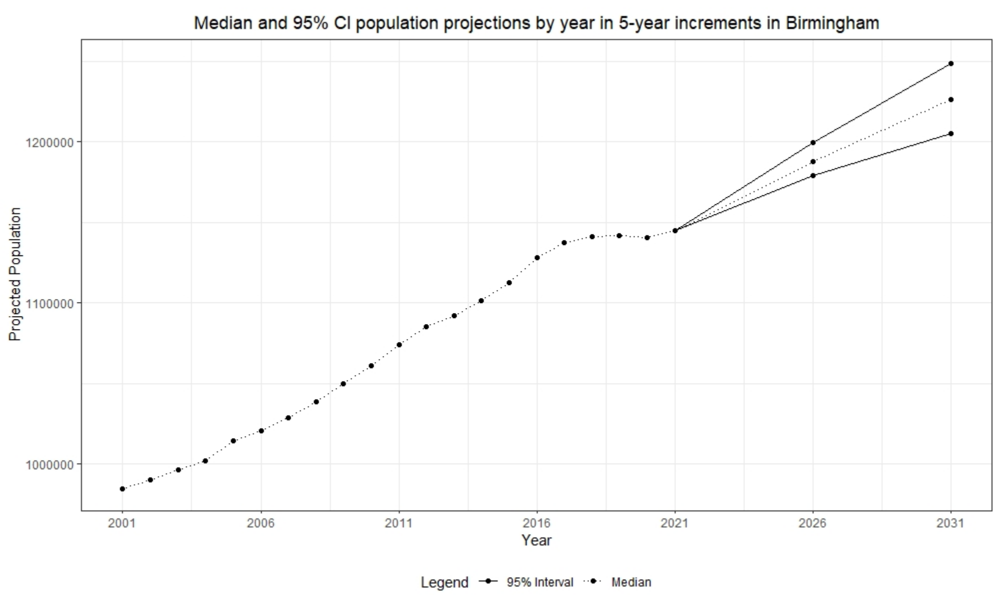

Figure 3: Probabilistic population projection by all age to 2031 in Birmingham

The projection of the Birmingham population in 2031 will be at 1.22 million and the value of the true population lies between 1.2 million and 1.24 million at the 95% CI (see table 1). The projection is further broken down into age-structured projection (see figure 2), the population in the younger cohort tends to decrease and the older population tends to increase in the next 10 years. It is because the

Table 1: Median and 95% Confidence Interval of the projected population in Birmingham

Year	Median	Upper 95% CI	Lower 95% CI
2021	1144923	NA	NA
2026	1188181	1198460	1179242
2031	1224868	1246317	1203901

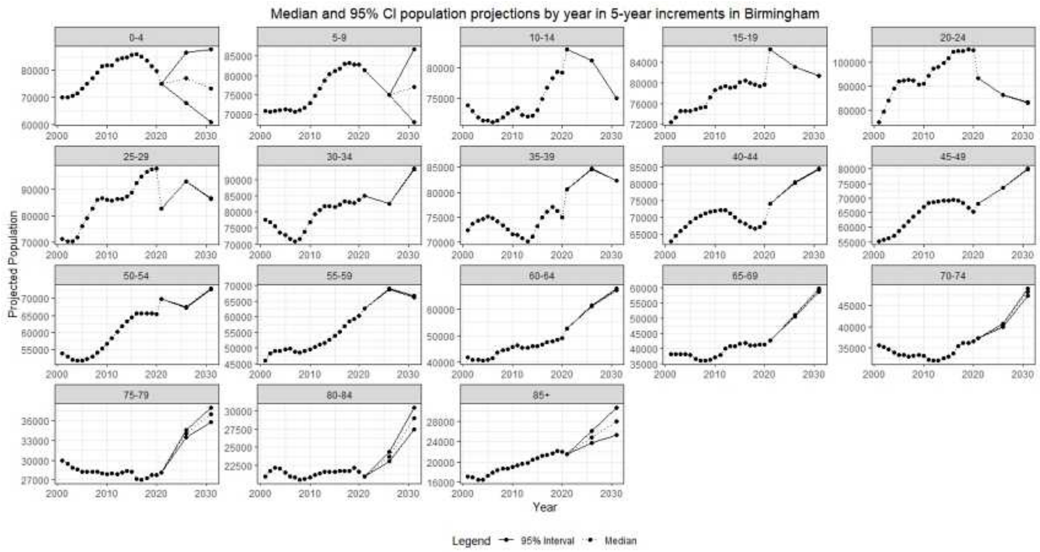

Figure 4: Probabilistic population projection by age structure to 2031 in Birmingham

future TFR is assumed to fluctuate at a lower level.

However, we can see that for some age groups, namely, 15-19, 20-24 etc, have dramatic changes in the population. As mentioned previously in the method section, it is because the ONS has not yet update the population estimate in between 2011 and 2021. Thereby, the change in population is dramatic which influences the predictive power of the model.

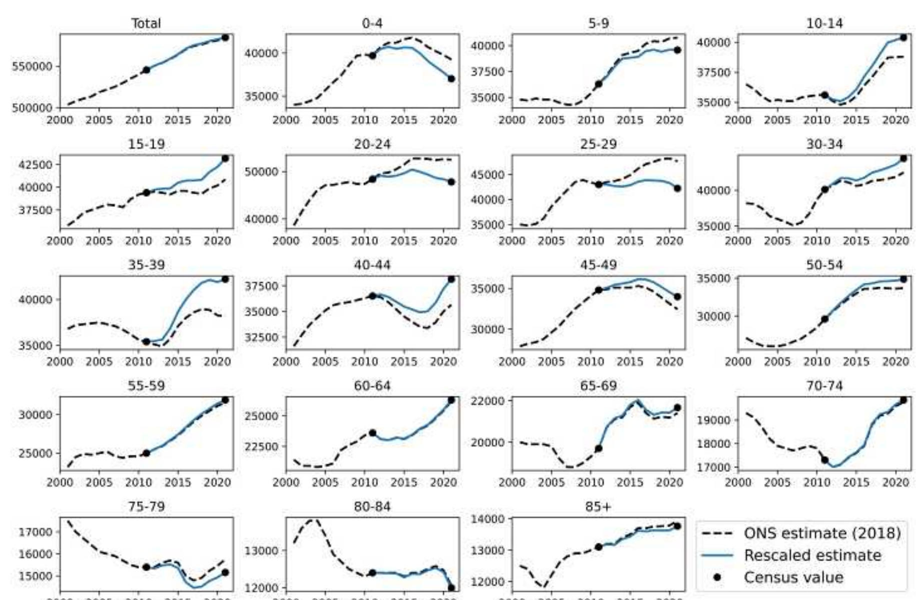

Figure 5: ONS estimates and the rescaled value of the female population by age groups

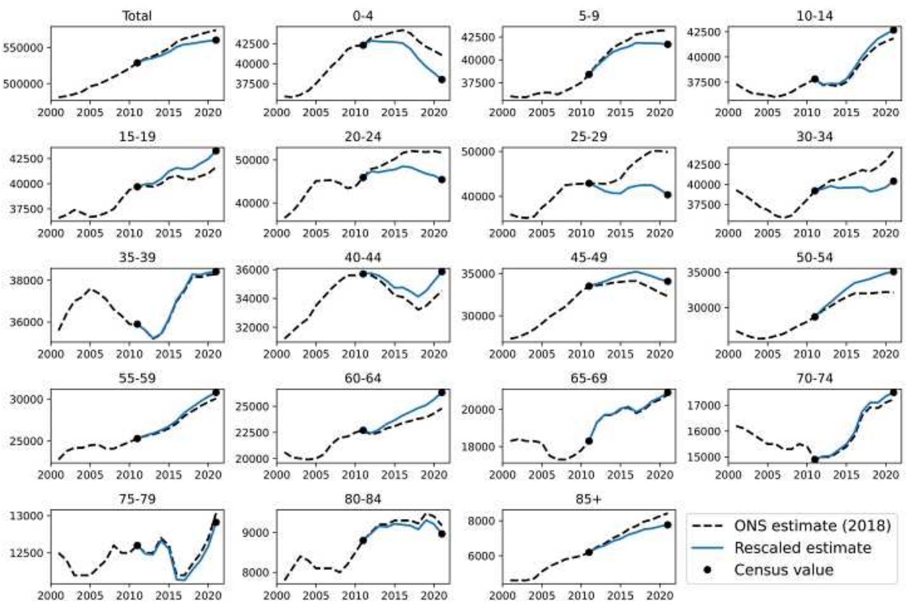

Figure 6: ONS estimates and the rescaled value of the male population by age groups

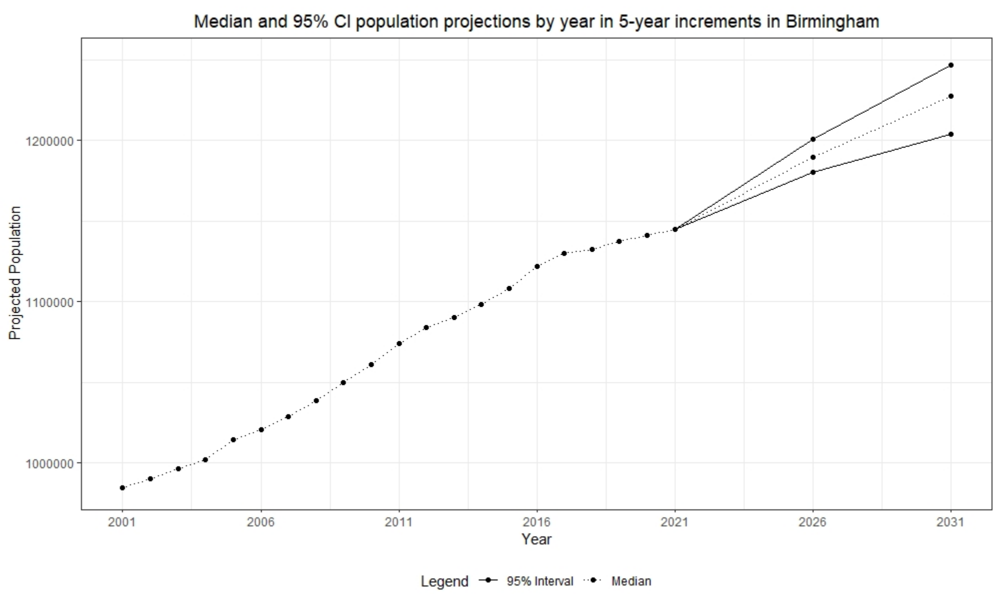

Figure 7: Probabilistic population projection by all age to 2031 in Birmingham (rescaled)

Figure 5 and Figure 6 show the comparisons between the ONS population estimate and the rescaling values for male and female population by age groups. The blue lines indicate the recaled value and the black broken lines indicate the ONS estimates. If there is no recaling process, the black broken line will directly connect to the latest 2021 census which causes dramatic changes in population. The blue line shows a smoothing process in between the pass values the the latest census 2021 value as we can see the shift the population structure is a lot more consistent.

Figure 7 and 8 used the rescaled value from the blue lines in figure 5 and 6. The population projection with rescaled value shows a smoother transition from period to period for each age group. Especially the 15-19 age groups show a smoother population projection in the future. Table 2 shows the exact number of the predicted value in the future population. The rescaled projection of the Birmingham population in 2031 will be at 1.227 million and the value of the true population lies between 1.2 million and 1.24 million at the 95% CI (see table 1). The rescaled version of projection is similar to

Table 2: Median and 95% Confidence Interval of the projected population in Birmingham (rescaled)

Year	Median	Upper 95% CI	Lower 95% CI
2021	1144923	NA	NA
2026	1189342	1200419	1180218
2031	1227335	1246395	1204052

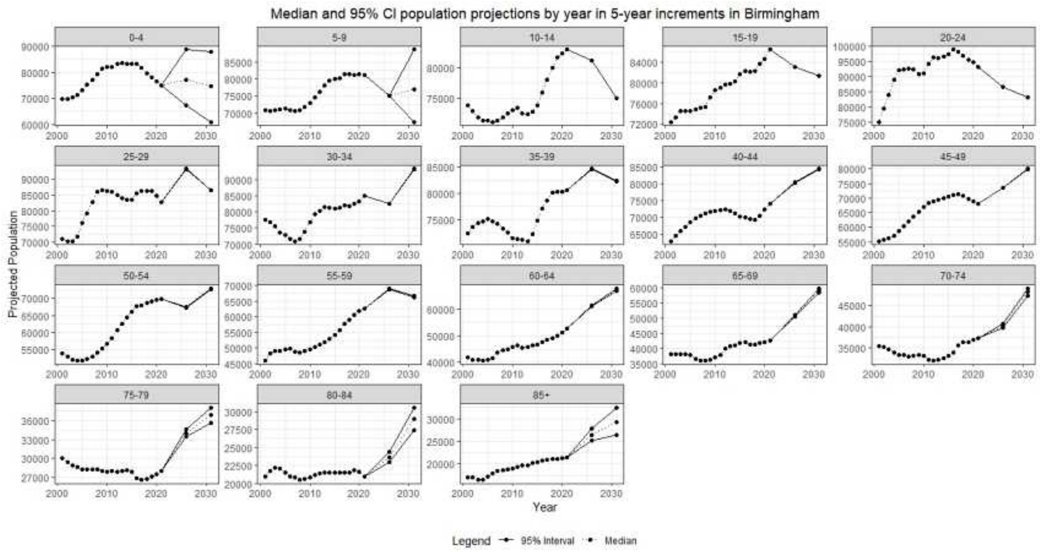

Figure 8: Probabilistic population projection by age structure to 2031 in Birmingham (rescaled)

the earlier projection (figure 3 and 4) with less than 0.2% discrepancy in the 2031 median projection value.

4 Conclusion

The method used is based on the ONS method, with additional uncertainty incorporated, as such it considers all the vital rates and includes an assessment of the uncertainty of future population projections. By developing this approach and projections, based on the latest Census 2021 data, decision makers can make more informed decisions when planning for the future.

References

[Office for National Statistics, 2016] Office for National Statistics (2016). Methodology used to produce the national population projections. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/methodologies/methodologyusedtoproducethenationalpopulationprojections. Accessed: 2024-05-02.

[Raftery et al., 2014] Raftery, A., Alkema, L., and Gerland, P. (2014). Bayesian population projections for the united nations. *Stat Sci*, 29(1):58–68.