Government of **Western Australia**
Department of **Health**

Bayesian modelling methods used for Public Health Atlas indicators

Epidemiology Directorate
Public and Aboriginal Health Division
Department of Health WA

August 2024

Contents

List of Tables

List of Figures

Abbreviations

ASPR	Age-group Specific Rate
ASR	Age Standardised Rate
ASRA_ST	Age Standardised Rate by Age Spatio-Temporal model
BOD	Burden of Disease
CI	Credible Interval
DALY	Disability Adjusted Life Years
EP	Exceedance Probability
ESS	Effective Sample Size
HD	Health District
HDI	Highest Density Interval
HMDC	Hospital Morbidity Data Collection
HPC	High Performance Computer
HR	Health Region
HWSS	Health and Wellbeing Surveillance System
LGA	Local Government Area
MCMC	Markov Chain Monte Carlo
PHA	Public Health Atlas
RSE	Relative Standard Error
SRR	Standardised Rate Ratio
ST	Spatio-temporal
WA	Western Australia
WMrP_ST	Weighted Multilevel Regression and Poststratification Spatio-Temporal model
YLD	Years Lived with Disability
YLL	Years of Life Lost

1. Introduction

This document describes the Bayesian modelling methods used to estimate epidemiological measures for small areas such as Local Government Areas (LGAs) and Health Districts (HDs) for the Western Australian (WA) Public Health Atlas (PHA).

1.1 What is Bayesian modelling?

A Bayesian model is a statistical model that uses probability to represent uncertainty within the model, both the uncertainty regarding the output (e.g., estimated disease/risk factor prevalence, counts, age standardised rates, age group specific rates and standardised rate ratios) and the uncertainty regarding the input (e.g., raw data and parameters such as socioeconomic status, remoteness, and service accessibility) to the model.

Compared to conventional small area analysis methods, Bayesian methods have the following advantages:

- ✓ With the inclusion of prior distributions (i.e., existing evidence), researchers can include structured assumptions about spatial and temporal relationships to improve estimation where there are not enough cases to derive reliable estimates. Traditionally if there is insufficient data, estimates for the area would not be reported.
- ✓ Gaps in data can be filled where the issue of reliability using conventional methods to derive epidemiological measures has not been solved.
- ✓ Level of uncertainty can be reported for the measures/indicators for an area (this usually cannot be reported using conventional methods).
- ✓ Different models can be fitted based on different measure characteristics.

1.2 Levels of geography modelled using Bayesian methods

In areas with small population sizes and disease/condition counts, reliable epidemiological measures cannot be derived. This is the case for some LGAs and HDs within WA. To produce small area estimates with increased stability and certainty at the LGA and HD levels, Bayesian spatio-temporal modelling was used to obtain estimated disease/condition counts (modelled/fitted counts, not observed values) to derive associated epidemiological measures.

Due to the ready availability and high reliability of data and measures at Health Region (HR) and State geographical levels, measures presented at these levels were not derived via Bayesian modelling but computed directly using observed disease/condition counts to calculate associated measures.

2. Data and model types

Three types of data (administrative data, survey data, and burden of disease (BOD) data) were modelled using Bayesian methods and are presented in the PHA. For a full list of indicators presented in the PHA for each data type, please refer to the PHA Data Dictionary.

Please see Section 4 (*Interpreting model output*) for further information on (i) understanding the difference between a credible interval (CI) and a confidence interval, and (ii) interpreting the 'Comparison to State' measure, both of which are mentioned in the following section.

Additionally, please refer to the Appendix for mathematical notations used in this document.

2.1 Administrative data

Administrative data includes all indicators other than those mentioned in the HWSS and BOD sections of the PHA Data Dictionary. It includes indicators such as potentially preventable hospitalisations, tobacco-related hospitalisations, and hospitalisations due to injury and poisoning, among others, sourced from the WA Hospital Morbidity Data Collection (HMDC) as well as aetiological fractions derived by the WA Department of Health Epidemiology Directorate. It also includes death data from the WA Mortality Dataset and cancer incidence data identified through the WA Cancer Registry.

Key measures reported in the PHA for administrative data include modelled counts (with associated 95% CIs for modelled LGA and HD level data), Age Standardised Rates (ASRs) and associated 95% CIs for LGA and HD data and 95% confidence intervals for HR data, and Age group Specific Rates (ASPRs). Another key measure reported in the PHA is a 'Comparison to State' of disease/condition rates. For LGA

and HD level data the comparison to State is based on the exceedance probabilities (EPs) of the posterior draws (i.e., samples), to identify whether the disease/condition ASR is higher, lower, or similar between the specific LGA/HD compared to the State ASR. For HR level data, the comparison to State is based on the Standardized Rate Ratio (SRR) values which were calculated using raw unmodelled counts for the specific HR. For more information on interpreting the comparison to State variable, please refer to Section 4.2. Raw unmodelled data at the HR level was suppressed in the following cases: (i) if the count was less than 6, the count was suppressed to protect privacy and data confidentiality, (ii) if the count was less than 20, the ASRs were suppressed because the derived rates were unreliable, (iii) if the count was less than 5 the 'Comparison to State' variable was suppressed because the derived estimate was unreliable. Administrative data at the HR level also included data by Aboriginality and combined years (2011-2015, 2016-2020 and 2011-2020). Table 1 below outlines what measures are displayed in the PHA by geographical level.

Table 1. Administrative data measures presented in the PHA by geographical level

Geography Level	Measures (by area, year, and sex)
LGA	Modelled count and 95% CI
	ASR and 95% CI
	ASPR
	Comparison to State (based on posterior draw EPs)
HD	Modelled count and 95% CI
	ASR and 95% CI
	ASPR
	Comparison to State (based on posterior draw EPs)
HR	**Measures (by area, year/combined years, sex, and Aboriginality)**
	Raw count (counts <6, were suppressed)
	ASR and 95% confidence interval (ASRs where count <20, were suppressed)

	ASPR
	Comparison to State (based on SRR 95% confidence interval) (Comparison to State where count <5, were suppressed)

Bayesian modelled measures for LGA and HD level data were derived from the modelled counts obtained using a spatio-temporal (ST) model, which for identification purposes, was named *ASRA_ST* (Age Standardised Rate by Age Spatio Temporal model). This model was used to estimate counts, which were then used to calculate ASRs, ASPRs and the associated credible intervals for these measures, after adjusting for spatial and temporal variations. To reiterate, please note the ASRs and ASPRs were not estimated directly via the modelling process, but rather the counts (both age group specific and total counts) were estimated via the modelling process and subsequently ASRs/ASPRs were calculated based on these modelled/fitted counts.

The input data required for this model included both population estimates and raw counts by area, year, and age. The output data was the modelled/fitted counts by area, year, and age, which were then used to calculate ASRs/ASPRs accordingly.

Separate models were fit for males and females for each disease/condition. The modelled counts from these two models were then combined to calculate modelled counts and derive associated measures for persons. The equation and output measures for the model are presented in Table 2, and notations explained in Appendix Table A1. Further detailed calculation processes can be found in the deliverable document by Hogg & Cramb (2022) (see link in Section 7) specifically produced for this project. Note the space-time interaction term (δ_{it}) is absent in the *ASRA_ST* model equation as it was removed due to convergence issues (as mentioned in Section 3.2).

Table 2. Administrative data model equation and output measures presented in PHA for LGA and HD level data

Model	Equation	Output measure in PHA (by area, year, and sex)
ASRA_ST		Modelled count (μ_{ita}) with 95% CI
		ASR with 95% CI

	$\log \mu_{ita} = \log(N_{ita}) + \alpha + X_{ita}\beta + \theta_i + \gamma_t$	ASPR
		Comparison to State

For administrative data where aetiological fractions were applied (e.g., alcohol, drug, tobacco related hospitalisations/deaths), a gamma distributed variable was used instead of a Poisson variable to allow for non-integer counts. There was a strong multiplicative relationship between age group and population size for some combinations of geography and age group. To allow the model sufficient flexibility to accurately model counts for these combinations (alcohol deaths at the LGA level, tobacco deaths at the HD and LGA levels and tobacco hospitalisations at the HD level) an interaction term between age group and population was added. This prevented significant over or under estimation of total counts for these conditions which occurred with the additive model.

2.2 Survey data

Survey data consists of data obtained from the WA Health and Wellbeing Surveillance System (HWSS). It includes indicators related to lifestyle behaviours such as current smoking, and alcohol consumption at levels considered high risk for long-term and short-term alcohol related harm, among others. A full list of survey data indicators included in the PHA can be viewed in the PHA Data Dictionary in the HWSS section.

Key measures reported in the PHA for survey data include prevalence by area, year and sex, and prevalence by age group with associated 95% CIs for modelled LGA and HD level data and associated 95% confidence intervals for HR level data. Please note, prevalence measures with a Relative Standard Error (RSE) greater than 50% or a prevalence of zero (for HR level data) were excluded and not presented in the PHA due to privacy policies, or to withhold an unreliable prevalence value (see Hogg & Cramb (2022) for further information on modelled RSE). Additionally, a ‘Comparison to State’ measure is reported. For LGA and HD level modelled data the comparison to State is based on the EPs of the posterior draws, to identify whether the prevalence is higher, lower, or similar between the specific LGA/HD compared to the State. For HR level data, the comparison to State is based on evaluating the 95% confidence intervals of the HR with the State prevalence. For more information on interpreting the

comparison to State variable, please refer to Section 4.2. Table 3 below outlines what measures for survey data are displayed in the PHA by geographical level.

Table 3. Survey data measures presented in the PHA by geographical level

Geography Level	Measures (by area, year, and sex)
LGA	Modelled prevalence with 95% CI
	Modelled prevalence by age group with 95% CI (by area and year only)
	Comparison to State (based on posterior draw EPs)
HD	Modelled prevalence with 95% CI
	Modelled prevalence by age group with 95% CI (by area and year only)
	Comparison to State (based on posterior draw EPs)
HR	Raw prevalence with 95% confidence interval
	Raw prevalence by age group with 95% confidence interval (by area and year only)
	Comparison to State (based on prevalence 95% confidence interval comparison of HR and State)

The *Weighted Multilevel Regression and Poststratification (WMrP_ST)* model was used for survey data. The same *WMrP_ST* model is also used for BOD data described in the next section. Note that all survey measures are binary (i.e., yes, or no) and the *WMrP_ST* model is a logistic model where the dependent variable is a binary outcome measure (e.g., current smoker or not). The advantage of this model is that it incorporates survey weights into the model. The input data required for this model included individual level survey data, post strata data (all unique combinations of the covariates used in the model and census populations), and raw survey weights available from HWSS data. Note that the model itself did not output counts, but probabilities for the sampled individuals. It was only after the model parameters were used to predict probabilities for the post strata data that counts could be derived. The counts were then converted to prevalence. The equation and output measures for the model are presented in Table 4.

Table 4. Survey data model equation and output measures presented in PHA for LGA and HD level data

Model	Equation	Output measure in PHA for all (by area, year, and sex)
WMrP_ST*	$\text{logit}(p_{jit}) = \alpha + X_{jit}\beta + \theta_i + \gamma_t$	Modelled prevalence with 95% CI
		Modelled prevalence by age group with 95% CI (by area and year only)
		Comparison to State

* In deriving the probability of having a specific outcome/risk factor (e.g., smoking or alcohol drinking), a variety of event-level (i.e., individual record for those surveyed) and area-level fixed effect factors below are included in the modelling.

- Event-level factors:
  - Age group * sex interaction term
- Area-level factors:
  - Total proportion of female population
  - Remoteness (Note: HD models do not have the remoteness covariate)
  - Socio-economic disadvantage
  - Indigenous population proportion
  - Proportion of people with low income
  - Proportion of people with tertiary education
  - Proportion of population with occupation as labourer or other manual workers
  - Total proportion of males aged 35-39 years
  - Total proportion of females aged 15-19 years

Note that the fixed term factors can be changed according to the health condition/risk factor and be determined by running ordinary least square regression models and assessing their statistical significance in relation to outcome/risk factor measures. These factors are not included in the *ASRA_ST* models used for administrative and BOD data as these data types do not have event level data like survey data, where each record represented a person surveyed. *ASRA_ST* model used data aggregated by geographical area, age group and year where each line of data represented a group of people for a geographical area, age group and year combination. Factors such as socioeconomic disadvantage and tertiary education status among others, could therefore not be determined at an individual scale for administrative and BOD data

modelled using *ASRA_ST*. Additionally, the same factors are not used for BOD data modelled using *WMrP_ST*. Fewer factors were used in the *WMrP_ST* model for BOD data as the addition of factors/interactions with the 18 age groups in BOD data created computational issues.

2.3 Burden of disease data

BOD data is modelled using both administrative data sourced from the WA Mortality Dataset and HMDC, and survey data from the HWSS. A full list of BOD data indicators included in the PHA can be viewed in the PHA Data Dictionary in the BOD section.

Key measures reported for BOD data include YLL and YLD counts at the LGA, HD and HR geographical levels, YLL and YLD ASRs at the LGA, HD and HR geographical levels with associated 95% CIs presented at the LGA and HD levels only, and a 'Comparison to State' measure at the LGA and HD levels only. The comparison to State measure is based on the EPs of the posterior draws to identify whether the YLL or YLD ASRs are higher, lower, or similar between the specific LGA/HD compared to the State. For more information on interpreting the comparison to State measure, please refer to Section 4.2.

Additionally at the HR level, Disability Adjusted Life Years (DALY) counts, and ASRs are also reported. DALY counts/ASRs are the sum of associated YLL and YLD counts/ASRs. At the HR level, YLL, YLD and DALY counts of minor category diseases/conditions are presented in a pie chart as a percentage contribution to the total YLL, YLD or DALY count for the intermediate category of interest for the total population. For each HR and year, the leading ten conditions of YLL, YLD and DALY burden are also displayed in the PHA by sex, as both a count, percentage, and ASR (per 100,000). Additionally, the contribution (%) of total YLL, YLD and DALY burden is presented by disease groups (intermediate category) in a tree-map figure.

Table 5 below outlines what measures for BOD data are displayed in the PHA by geographical level.

Table 5. Burden of disease data measures presented in the PHA by geographical level

Geography Level	Measures (by area, year, and sex)
LGA	Modelled YLL count with 95% CI
	YLL ASR with 95% CI
	Modelled YLD count with 95% CI
	YLD ASR with 95% CI
	Comparison to State (based on posterior draw EPs)
HD	Modelled YLL count with 95% CI
	YLL ASR with 95% CI
	Modelled YLD count with 95% CI
	YLD ASR with 95% CI
	Comparison to State (based on posterior draw EPs)
HR	Raw YLL count
	YLL ASR
	Contribution (%) of minor category diseases/conditions to overall YLL count for intermediate category of interest for total population (by area and year only)
	Leading 10 conditions of YLL burden (count (%) and ASR (per 100,000)
	Contribution (%) of disease groups (intermediate categories) to overall YLL burden
	Raw YLD count
	YLD ASR
	Contribution (%) of minor category diseases/conditions to overall YLD count for intermediate category of interest for total population (by area and year only)
	Leading 10 conditions of YLD burden (count (%) and ASR (per 100,000)

	Contribution (%) of disease groups (intermediate categories) to overall YLD burden
	Raw DALY count
	DALY ASR
	Contribution (%) of minor category diseases/conditions to overall DALY count for intermediate category of interest for total population (by area and year only)
	Leading 10 conditions of DALY burden (count (%) and ASR (per 100,000)
	Contribution (%) of disease groups (intermediate categories) to overall DALY burden

For LGA and HD level data, two models were used to obtain measures: (i) *ASRA_ST* and (ii) *WMrP_ST*. Model use was determined by the type of input data. For disease/conditions using administrative data the *ASRA_ST* model was used, and for survey data, *WMrP_ST* was used. Fitted counts obtained from the *ASRA_ST* model were then used to calculate YLL or YLD and fitted probabilities from the *WMrP_ST* model used to calculate YLD from survey data.

Like administrative data, separate models were fit for males, females, and persons for each disease/condition for *ASRA_ST* models. *WMrP_ST* models however were fit to all data combined and subsequently the final estimates were calculated by sex. Note the *ASRA_ST* model used data aggregated by geographical area, age group and year where each line of data represented a group of people for a geographical area, age group and year combination; while *WMrP_ST* models used raw event level data where each record represented a person surveyed. The equations and output measures for both models are presented in Table 6.

Table 6. Burden of disease data model equations and output measures presented in PHA

Model	Equation	Output measure in PHA (by area, year, and sex)
ASRA_ST ***(YLL and YLD)***	$\log \mu_{ita} = \log(N_{ita}) + \alpha + X_{ita}\beta + \theta_i + \gamma_t$	Modelled YLL count with 95% CI and YLL ASR with 95% CI
		Modelled YLD count with 95% CI and YLD ASR with 95% CI
		YLL/YLD comparison to State
WMrP_ST* **(YLD using survey data only)**	$YLD_{it}^{(d)} = \sum_{\alpha}\sum_{h} \hat{p}_{ita}^{(d)} N_{ita} p_h e_h$	Modelled YLD count with 95% CI and YLD ASR with 95% CI
		YLD comparison to State

* In deriving YLD, a variety of event-level (i.e., individual record for those surveyed) and area-level fixed effect factors below are included in the modelling.

- Event-level factors:
  - Age (in age group)
  - Sex
- Area-level factors:
  - Remoteness (Note that HD level models do not have the remoteness factor)
  - Socio-economic disadvantage
  - Indigenous population proportion
  - Proportion of people with low income
  - Proportion of people with tertiary education

Due to the large number of age groups required for age standardised YLD estimates, YLD *WMrP_ST* models did not have an age group and sex interaction as fixed effect factors but this was included as a random effect (unlike the HWSS *WMrP_ST* models which contained the age and sex term as fixed term factors, but not as a random effect). Fewer factors were used in the *WMrP_ST* model for BOD data as the addition of factors/interactions with the 18 age groups in BOD data created computational

issues. As such, the *WMrP_ST* BOD fixed effect factors were scaled back in comparison to the *WMrP_ST* model used for survey data.

3. Convergence checks

Convergence refers to the stabilisation of the Markov Chain Monte Carlo (MCMC) algorithm which estimates the posterior distribution of the Bayesian models by drawing a very large number of posterior draws. There are a wide range of MCMC algorithms, however the one used to run the Bayesian models in this instance are called random walk or Gibbs samplers.

(For more on MCMC see: https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo).

The values of two primary model diagnostic measures from each model were used to determine if a model had/had not converged in the first instance: (i) R-hat and (ii) Effective Sample Size (ESS).

3.1 R-hat ($\hat{R}$)

- It is recommended to run multiple, independent MCMC algorithms for the same model, called chains. All models in this instance, were run with 4 independent chains. Each chain starts with a different set of initial parameter values. It can be ascertained if convergence is acceptable by comparing the behaviour of the posterior draws from the different chains.
- $\hat{R}$ compares the behaviour of the posterior draws from the different chains. Well behaved chains should converge to the same area of the parameter space regardless of the initial parameter values used.
- Separate chains that converge to the same density are described as “mixing well”.
- $\hat{R}$ is calculated by taking the average of within-chain variances and comparing this to the variances of all the chains mixed together by taking the square root of the mixture variance divided by the average within-chain variance (Gelman & Shirley, 2011).
- $\hat{R}$ is always greater or equal to 1.

- At convergence, the chains will have mixed, so that the distribution of the simulations between and within chains will be identical and the ratio $\hat{R}$ should equal 1.
- If $\hat{R}$ is greater than 1, this implies that the chains have not fully mixed and that further simulation with an increased number of iterations might increase the precision.
- An $\hat{R}$ less than or equal to 1.01 was used as a cutoff to denote convergence as recommended by Vehtari et al. (2021).

3.2 Effective Sample Size (ESS)

- ESS considers the dependence in the posteriors and estimates the number of independent posterior draws. Ideally, we would like the posterior draws to be independent however given the nature of the Gibbs sampling method, not all posterior draws are independent.
- A highly correlated or inefficient MCMC algorithm would give very low values of ESS, which can be artificially increased/improved by taking more posterior draws.
- A crude rule of thumb, as recommended by rstan (Stan Development Team 2022) and Hogg & Cramb (2022), is that all model parameters should have an ESS larger than the number of chains multiplied by 100. In this instance as each model was run with 4 independent chains, the recommended cut-off is to have an ESS larger than 400 for all model parameters.

If $\hat{R}$ and ESS values fell within the assigned cut-off values, further checks were performed on model outputs to determine if the modelled results were appropriate to include in the PHA.

This included:

1. Assessing trace plots of the posterior draws of model parameters
2. Comparing observed counts/rates to the modelled/fitted counts/rates to see if the modelled results were plausible
3. Assessing residual plots to examine the relationship between standardised residuals of the posterior and the fitted values, to ensure there are no systematic patterns in the residuals

4. Sensitivity analysis to compare priors/hyperpriors[1] to ensure the posterior distribution is similar regardless of prior/hyperprior choices
5. Posterior predictive checks to determine model fit

If convergence was not achieved, the following steps were taken to improve convergence:

1. Increase number of iterations (i.e., run the algorithm for longer)
   - Models were initially run at 100,000 iterations, if the model did not converge, iterations were increased to 200,000, and subsequently 300,000 etc. until convergence was reached
2. Increase the number of iterations to discard at the start of chains (i.e., burn-in[2])
3. Increase the level of thinning
   - Thinning was set to achieve at least ~10,000 usable draws i.e., if running a model with 100,000 iterations, thinning was set to 10 (i.e., every 10th draw

[1] In Bayesian statistics a hyperprior is a prior distribution placed on the hyperparameters of a hierarchical model. Hierarchical models are a type of Bayesian model where parameters of the model are themselves random variables with their own distributions. Hyperparameters are parameters of the prior distributions e.g., if a model parameter is assumed to follow a normal distribution, the mean and variance of the normal distribution are its hyperparameters.

Hyperpriors are used to express uncertainty about the hyperparameters themselves. Instead of fixing hyperparameters to specific values, which might introduce bias or overconfidence in the model, hyperpriors allow these hyperparameters to vary according to another probability distribution. This approach is particularly useful when there is little prior knowledge about the hyperparameters or when one wishes to remain as noncommittal as possible regarding their values. Hyperpriors are therefore a powerful tool in Bayesian modelling, offering a way to incorporate uncertainty at multiple levels of the model. They enable analysts to build more flexible and robust models that can better capture the complexities of real-world data.

[2] In a Bayesian model, we are often mainly interested in the posterior distribution, as it describes our knowledge about the parameters of interest given our priors and after having seen the data.

Often this posterior distribution is not tractable analytically, but we can still *sample* from it. Based on, e.g., the average of the samples, we can then approximate a posterior mean of the parameters. This is often done with so called Markov chain Monte Carlo (MCMC) methods used for this project. The idea is to devise a strategy of sampling such that, when producing draws via this chain, these draws will be "almost" draws from the posterior distribution from which we want to sample (if we can directly sample from the posterior distribution, we will do so, but often, that is not possible) provided the chain has run long enough.

So, to make sure that it has run "long enough" we discard the initial draws - *the burn ins* - that may still be affected by where we initialised the chain and hence not yet be "trustworthy" draws from "almost" the posterior distribution.

kept, resulting in 10,000 usable draws. If running model with 200,000 iterations, thinning was set to 20 etc.)

4. Increase the frequency of the MCMC algorithms automatic adaption phase (i.e., the step size for sampling)
5. Reducing the complexity of models by removing the space-time interaction term (this interaction term was not included for all *ASRA_ST* models)

If all attempts to improve convergence as outlined above failed, and a model for an individual disease/condition did not converge, the disease/condition was excluded from the PHA and therefore not shown in the drop-down disease/condition list. Its counts were however rolled up into the disease major category and still included at the major category level. For further information and examples of non-convergence vs convergence, please refer to the deliverable document by Hogg & Cramb (2022).

4. Interpreting model output

4.1 Credible intervals

In Bayesian statistics, credible intervals (CIs) are a way to quantify the uncertainty or variability in the estimates of unknown parameters. Unlike frequentist confidence intervals, which are based on the long-run behaviour of repeated sampling, Bayesian CIs are derived from the posterior distribution of the parameter, which represents our updated knowledge about the parameter after considering the observed data and any prior information.

To understand CIs, suppose we have a Bayesian model where we are interested in estimating the mean (μ) of a population. We have some prior belief about the possible values of μ, which is expressed as a prior distribution. After collecting data, we update our knowledge about μ and obtain the posterior distribution, which incorporates both the prior distribution and the observed data (Figure 1).

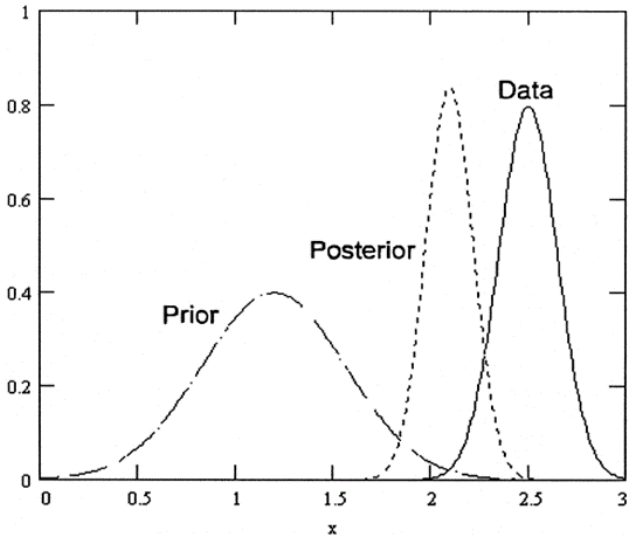

Figure 1. The relationship between Bayesian prior distribution, the data, and the posterior distribution (Matthews, 2001).

A CI is a range of values from the posterior distribution that is associated with a certain degree of credibility or probability. It represents a range of plausible values for the parameter, given the data and the prior information. For example, a 95% CI is constructed such that it contains the true parameter value with a probability of 0.95. (See a comparison of traditional 95% confidence interval vs. Bayesian 95% credible interval in https://www.statsdirect.com/help/basics/confidence_interval.htm).

The construction of CIs involves determining the limits of the interval based on the probability density of the posterior distribution ascertained from the MCMC algorithms. The Bayesian models used by the Department of Health uses the Highest Density Interval (HDI), which constructs the CI to contain the highest density. This is a pragmatic method for skewed distributions (Figure 2).

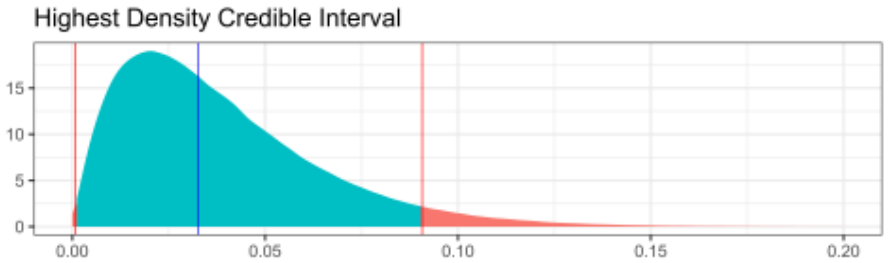

Figure 2. Credible Interval (red vertical lines) constructed using Highest Density Interval. The median is indicated by the blue line (Hogg & Cramb, 2022).

It's important to note that credible intervals (CIs) in Bayesian analysis, unlike frequentist statistics, measure uncertainty using available data and prior information, and are not based on repeated sampling. CIs offer a probabilistic assessment of a parameter's plausible values, influenced by the choice of prior distribution, and facilitate decision-making based on the entire posterior distribution rather than just point estimates.

Figure 3 shows the estimated ASRs and credible intervals for age standardised potentially preventable hospitalisation rates by health districts in WA. The length of the bar represents the ASR and the two whiskers, the lower and upper credible intervals. The wider the credible interval range, the more uncertain the ASR estimate, meaning the ASR for the area is less reliable compared to other areas. (Note that the darker the bar, the higher the ASR). Usually, areas with smaller populations have more uncertainty in their rates, often resulting in wider credible interval ranges. For example, as shown in Figure 3, the CI width from a small population area like Southern Wheatbelt is wider than that for an area with large population like Inner South, especially if their rate estimates are similar.

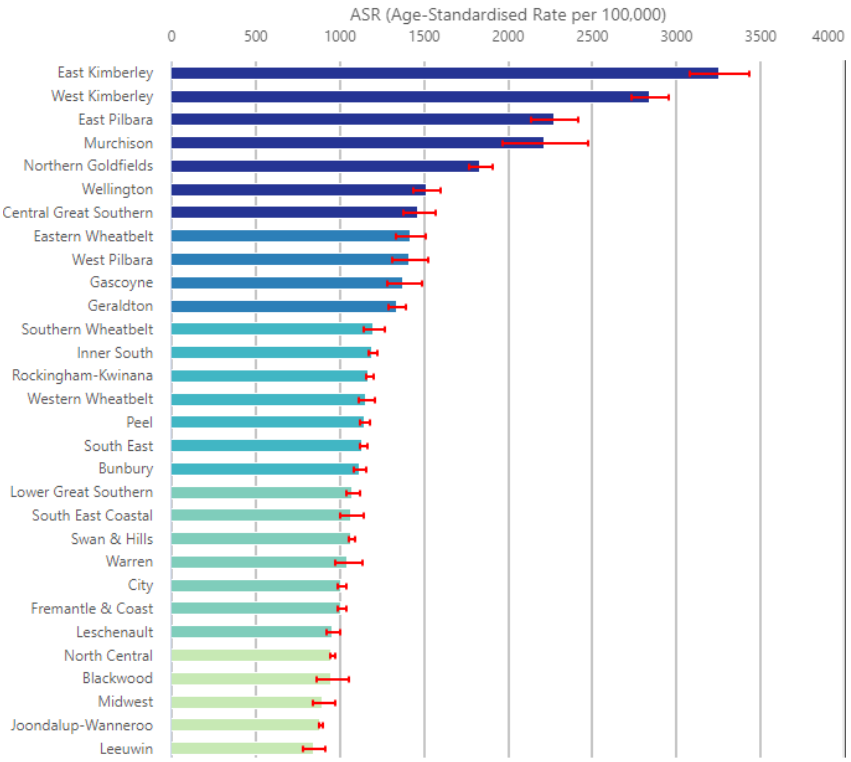

Figure 3. Comparison of credible Intervals for ASR by health district in WA in PHA.

4.2 Comparison to State

For all data types (administrative, survey and BOD data) the comparison to State value (higher/lower/similar) for LGAs and HDs is determined from the exceedance probabilities (EPs). This is the probability of the posterior draws being above a certain value. This was derived from the posterior draws using:

$$EP = \frac{1}{D}\sum_{d} I\left(\theta^{(d)} > c\right)$$

Where $I\left(\theta^{(d)} > c\right)$ is equal to 1 if $\theta^{(d)}$ is larger than the baseline value c and zero if $\theta^{(d)}$ is smaller than c (Hogg & Cramb, 2022). The EP was used to indicate whether the ASR, prevalence, age standardised YLL or age standardised YLD in a particular area is significantly higher than the state measure. EP values above 0.8 (i.e., 80% of the posterior) were considered likely to be above the state value and therefore 'higher'. Values below 0.2 (i.e., 20% of the posterior) on the other hand were considered likely to be below the state value and therefore 'lower'. Values between 0.2 and 0.8 were then considered to be 'similar' to the state rate.

For HR level data however, where Bayesian modelling was not used, the comparison to State value was determined differently for each of the data types as detailed below.

Administrative Data:

- Comparison to State value was determined by analysing the 95% confidence interval of the SRR for the HR of interest with the State.
- SRR is the ratio of observed disease/condition counts to expected disease/condition counts.

  $$SRR_{HR} = \frac{\sum y}{\sum E}$$

  Where y is the observed disease/condition counts in a particular HR, and E (expected counts) is the state rate multiplied by the HR population.
- An SRR greater than 1 therefore indicates that the HR rate is higher; an SRR lower than 1 indicates the HR rate is lower.
- To determine the appropriate value for the comparison to State, the 95% confidence intervals were used to ensure the difference was significant.

- If the SRR lower confidence interval value was greater than 1, then the HR rate was significantly higher than the State.
- If the SRR upper confidence interval value was less than 1, then the HR rate was significantly lower than the State.
- If the SRR confidence interval included 1 in its range, the HR rate was similar to the State.

Please note, the term SRR is used in this document for the sake of simplicity in describing this type of measure; and the estimation of similar measures such as standardised incidence ratio (e.g., for cancer incidence) or standardised mortality ratio (e.g., death data) will follow the identical process as for SRR.

Survey data:

- Comparison to State value was determined by analysing the 95% confidence interval of the prevalence for the HR of interest with the State.
  - If the prevalence lower confidence interval value was greater than 1, then the HR prevalence was significantly higher than the State.
  - If the prevalence upper confidence interval value was less than 1, then the HR prevalence was significantly lower than the State.
  - If the prevalence confidence interval included 1 in its range, the HR prevalence was similar to the State.

Burden of disease data:

- A comparison to State measure is not presented in the PHA for BOD data at the HR level.

5. Advantages of modelled data over raw unmodelled data

Modelled estimates for LGAs and HDs include estimated count, age group specific rate, age standardised rate, and prevalence. **Important note**: Modelled data are estimates only and are not actual counts, rates, or prevalence. Caution should therefore be exercised when using the modelled data.

There are several benefits of using modelled data. Firstly, modelled measures can be more stable compared to raw measures, particularly in areas with small counts and/or populations. Figure 4 shows a comparison of three prevalence measures (raw, modelled, and state prevalence) of adults who drink alcohol at levels that increase the risk of long-term harm for a particular LGA from 2011-2020. The modelled prevalence (red) is a 'smoothed' version of the raw prevalence (green). Overall heterogeneity in modelled estimates were reduced, depicting a similar trend to the State prevalence (blue). Additionally, the modelled prevalence had increased stability and certainty compared to the raw data as indicated by the narrower CIs for the modelled data compared to the wider confidence intervals observed in the raw data. This increased stability, mitigates the need for data suppression due to small counts or unreliable estimates. It also enables data users to observe clearer trends over time thereby contributing to a more comprehensive and insightful set of results.

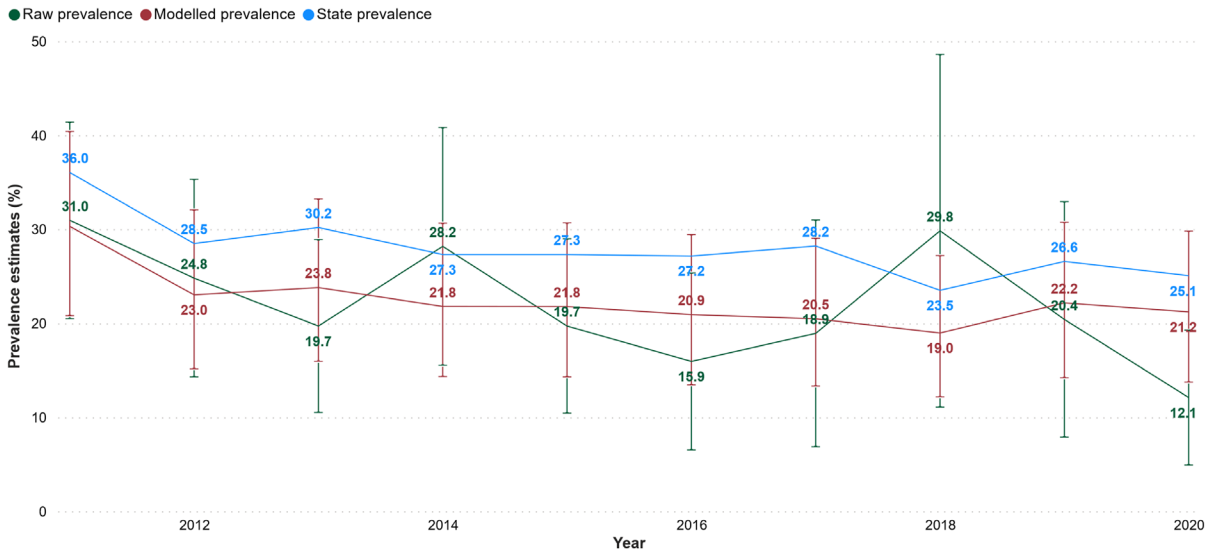

Figure 4. Raw, modelled, and state prevalence for long-term alcohol related harm in a WA LGA from 2011-2020.

To further highlight the capabilities of Bayesian methods, Figure 5 illustrates a comparison in data coverage in ASRs for hospitalisations due to injury and poisoning across WA LGAs before and after applying Bayesian methods. When mapping the raw data, Figure 5(a) shows there are several gaps in data coverage (i.e., not reportable ASRs) primarily due to small event counts and/or small population sizes. Conversely, mapping the modelled data using Bayesian methods, Figure 5(b) shows complete coverage, leaving no data gaps across WA.

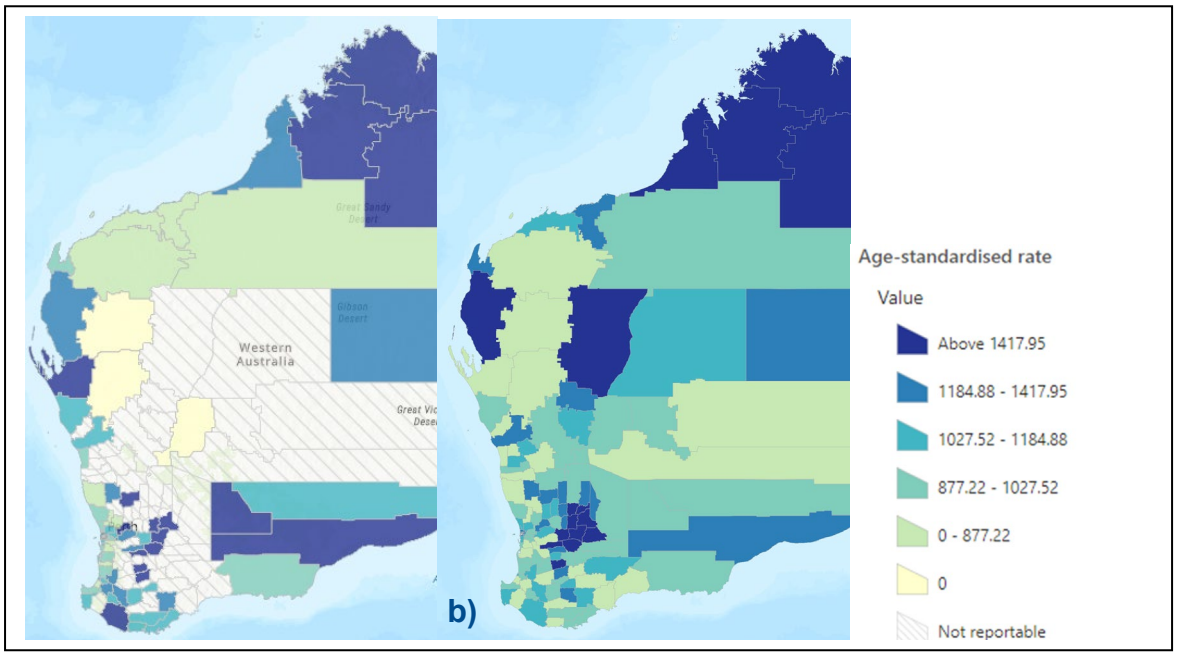

Figure 5. Data coverage across WA LGAs before (a) and after (b) using Bayesian methods to estimate ASRs for hospitalisations due to accidental falls in 2019.

6. Disadvantages and limitations

All statistical analysis methods come with disadvantages and limitations. Bayesian modelling methods and processes are no exception.

A key disadvantage of Bayesian modelling using MCMC, is that it requires a large amount of computational power and human resources. Some models such as *WMrP_ST* may take several days to run on a laptop computer. This can be particularly time consuming if there are many diseases/conditions where modelling is required as separate models will need to be run for each disease/condition and sub-category of interest i.e., separate models will need to be run for a particular major category and similarly if the disease/condition is further broken down by intermediate and minor

categories. Additionally separate models will need to be run for each of the sex categories (male, female, total). In some instances, models may need to be rerun several times after adjusting model parameters through trial and error (as described in Section 3) to reach convergence which can also be a time-consuming process. To improve efficiency, it is beneficial to run multiple models simultaneously, which requires the use of a high-performance computer (HPC) however access to such resources may be limited due to accessibility and/or cost. In addition to running several models, analysts tasked with producing Bayesian estimates, must also have some level of understanding of Bayesian modelling methods and processes, R software and HPC use/language. This requires a significant investment in human resources in terms of time and funding to train analysts.

The modelling itself also comes with some limitations in that the priors used in the models for the spatio-temporal terms in this project, potentially smooth over all adjacent areas and time points, with the extent of smoothing determined by the data itself. This may lead to rate estimates for individual years being all above or all below the actual state rate estimate, particularly where the state rate fluctuates distinctly between years, as shown in Figure 6, however the smoothed stable estimate is a desirable outcome as it allows estimation of the underlying "real" rate.

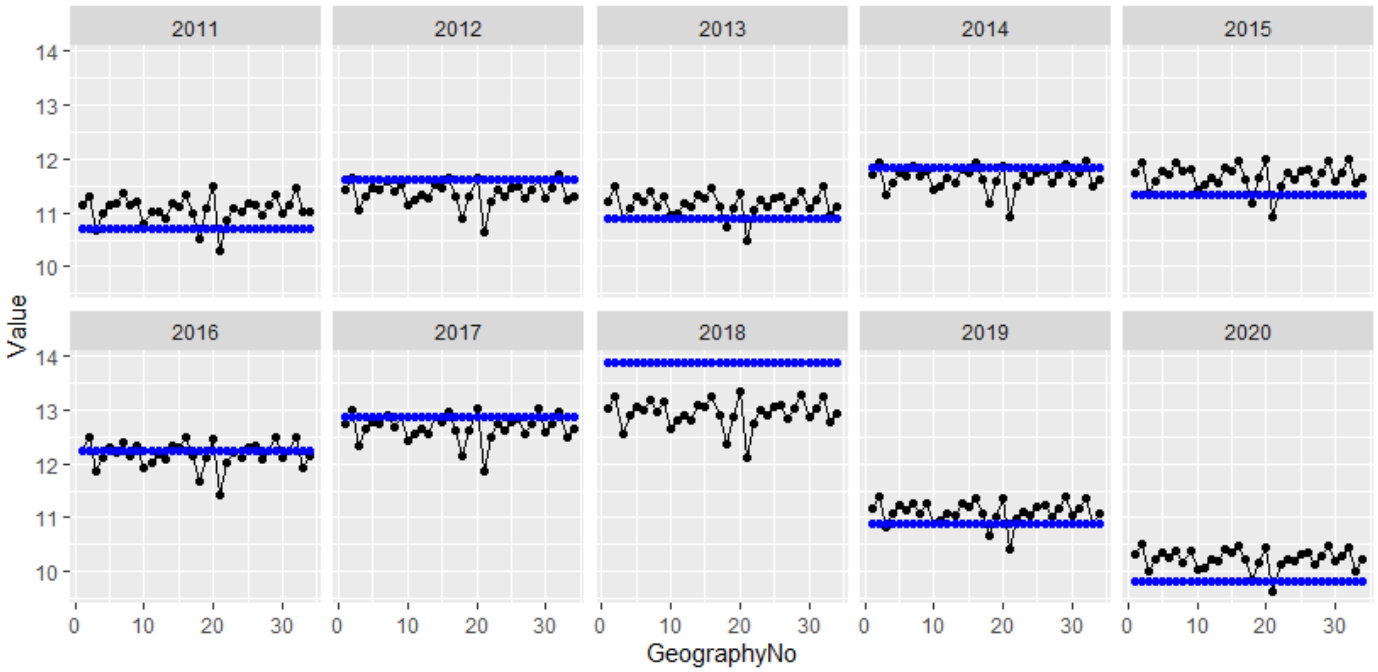

Figure 6. Modelled rate estimates by health district (black dots) and raw state rate estimates (blue). The raw rate fluctuates clearly from 2017 to 2019, with model smoothing giving 2018 health district estimates lower than the state raw rate.

The spatial priors used in this process have previously been shown to work well in the Australian context (Cramb et al., 2020) however, if discontinuities in rates between adjacent areas are expected, alternative smoothing priors that allow for large differences in rates between neighbours are needed. While many priors that allow discontinuities are available, these have resulted in convergence difficulties for certain areas when used with sparse Australian health data (Cramb et al., 2020), so were not considered for this project.

While it is important to acknowledge these disadvantages and limitations, the benefits of producing Bayesian modelled data, outweigh its potential downfalls. The modelled estimates allow for a more complete picture of population health outcomes and trends across WA to be observed which in turn, provides essential epidemiological measures to inform public health planning, policy, and decision making that would otherwise not be available.

7. Further information

Bayesian modelling project Deliverable 2: Modelling recommendations

https://www.health.wa.gov.au/~/media/Corp/Documents/Health-for/Population-health/Bayesian-modelling-project-Deliverable-2---Modelling-recommendations.pdf

Publication: 'Improving the spatial and temporal resolution of burden of disease measures with Bayesian models'

Hogg, J., Staples, K., Davis, A., Cramb, S., Patterson, C., Kirkland, L., Gourley, M., Xiao, J., & Sun, W. (2024). Improving the spatial and temporal resolution of burden of disease measures with Bayesian models. *Spatial and Spatio-temporal Epidemiology, 49, 100663*. https://doi.org/10.1016/j.sste.2024.100663.

Australian Cancer Atlas

The Australian Cancer Atlas (https://atlas.cancer.org.au/) has detailed information on how the Bayesian modelling is conducted and how output from the models is interpreted. The methodological document (Duncan et al, 2020) is a useful reference.

References

Cramb, S., Duncan, E., Baade, P., & Mengersen, K. (2020). A comparison of Bayesian spatial models for cancer incidence at a small area level: Theory and performance. *Case Studies in Applied Bayesian Data Science: CIRM Jean-Morlet Chair, Fall 2018*. *Springer, Cham, pp. 245-274*. https://eprints.qut.edu.au/203868/

Duncan, E., Cramb, S., Baade, P., Mengersen, K., Saunders, T., & Aitken, J. (2020). Developing a Cancer Atlas using Bayesian Methods: A Practical Guide for Application and Interpretation. https://atlas.cancer.org.au/developing-a-cancer-atlas/

Gelman, A., & Shirley, K. (2011). Handbook of Markov Chain Monte Carlo: http://www.mcmchandbook.net/HandbookChapter6.pdf

Hogg, J., & Cramb S. (2022). Bayesian modelling project, Deliverable 2: Modelling recommendations. (unpublished)

Hogg, J., Staples, K., Davis, A., Cramb, S., Patterson, C., Kirkland, L., Gourley, M., Xiao, J., & Sun, W. (2024). Improving the spatial and temporal resolution of burden of disease measures with Bayesian models. *Spatial and Spatio-temporal Epidemiology, 49, 100663*. https://doi.org/10.1016/j.sste.2024.100663.

Matthews, R. (2001). Methods for Assessing the Credibility of Clinical Trial Outcomes. *Drug Information Journal*, *35(4):1469-1478*. https://doi.org/10.1177/009286150103500442

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Burkner, P-C. 2021. Rank-Normalization, Folding, and Localization: An Improved Rˆ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, *16(2):667-718*. https://doi.org/10.1214/20-BA1221

Appendix

Table A1. Legend for notations used in equations

Notation	Description
μ_{it}	Fitted value for area ($_i$) and year ($_t$)
μ_{ita}	Fitted value for area ($_i$), year ($_t$), and age ($_a$)
N_{ita}	Population for area ($_i$), year ($_t$), and age ($_a$)
X_{ita}	The design matrix of indicators for area ($_i$), year ($_t$), and age ($_a$)
X_{jit}	Fixed effect design matrix for survey weights
$YLD_{it}^{(d)}$	Years Lived with Disability for area ($_i$) and year ($_t$) for dth posterior draw
E_{it}	Expected counts for area ($_i$) and year ($_t$)
$\sum_{a}$	Sum across age groups
$\sum_{h}$	Sum across health states
$\sum y$	Sum of observed disease/condition counts
$\sum E$	Sum of expected disease/condition counts
$\hat{p}_{ita}^{(d)}$	Proportion of people in area ($_i$), year ($_t$), and age group ($_a$) for dth posterior draw
p_h	Proportion of all persons with the condition that are in health state h
e_h	Health state specific disability weight
β	Coefficients for fixed effects
θ_i	Combined spatial random effects
γ_t	Temporal random effects
δ_{it}	Space-time random effects
p_{jit}	Probability for survey weights
α	Intercept
SRR_{HR}	Standardised Rate Ratio at the Health Region geographical level
D	Total number of posterior draws
$\sum_{d}$	Sum of posterior draws
I	Identity function (I is equal to 1 if $\left(\theta^{(d)} > c\right)$ is true, or equal to 0 if $\left(\theta^{(d)} > c\right)$ is false)
$\theta^{(d)}$	d^{th} posterior draw of theta
c	Baseline value

This document can be made available in alternative formats on request for a person with disability.