Supplementary material

Supplementary figures and tables

Geography of Australia

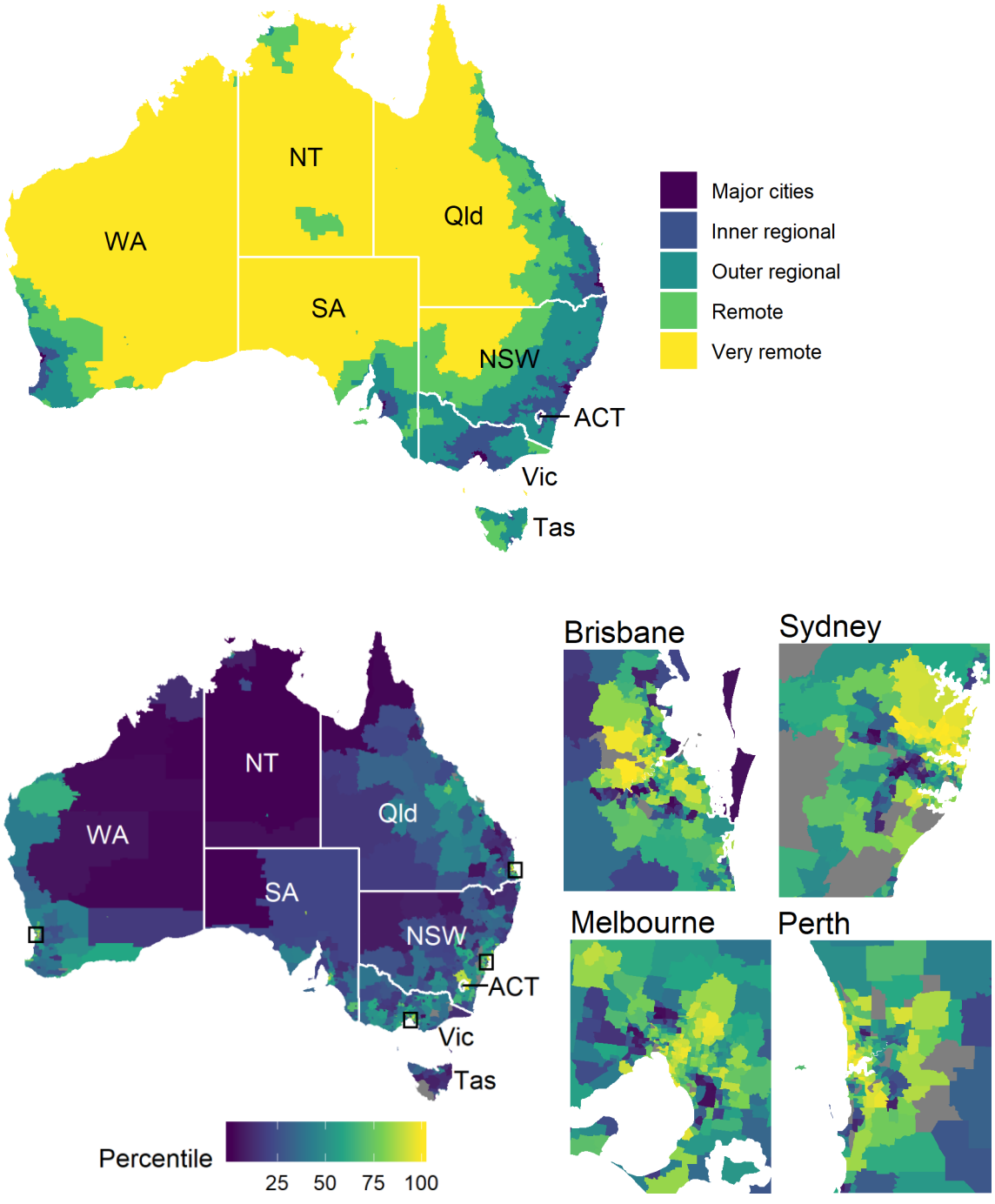

Figure S1; Maps of Australia with state and territory boundaries featuring remoteness categories (top) and each area's IRSAD percentile (bottom), with insets of the largest state capitals. Grey areas have small populations and the IRSAD for these areas cannot be published for privacy reasons. These areas are parkland, military or industrial areas.

Australia is spatially diverse, with extremely low population density over most of the land area, with regions of high density at the coast, near state capitals. This is reflected in the remoteness categories shown in the Supplementary material in Figure S1, with the state and territory boundaries for Western Australia (WA), the Northern Territory (NT), South Australia (SA), Queensland (Qld), New South Wales (NSW), the Australian Capital Territory (ACT), Victoria (Vic) and Tasmania (Tas). Socioeconomic status, shown by percentile of Index for Relative Socioeconomic Advantage and Disadvantage (IRSAD) for small areas in Figure S1, is associated with remoteness, however there are considerable differences within metropolitan areas. The insets show the IRSAD percentiles for metropolitan areas around Brisbane in Queensland, Sydney in NSW, Melbourne in Victoria and Perth in WA. The Australian Capital Territory (ACT) is the only state or territory without a sea border but is classified as major city or inner regional and has very high socioeconomic status as the population has a large proportion of professionals. Hospitals and health systems are largely administered by the states and territories and pathology services tend to be state based.

Maps of SIRs and EHRs for persons, females and males

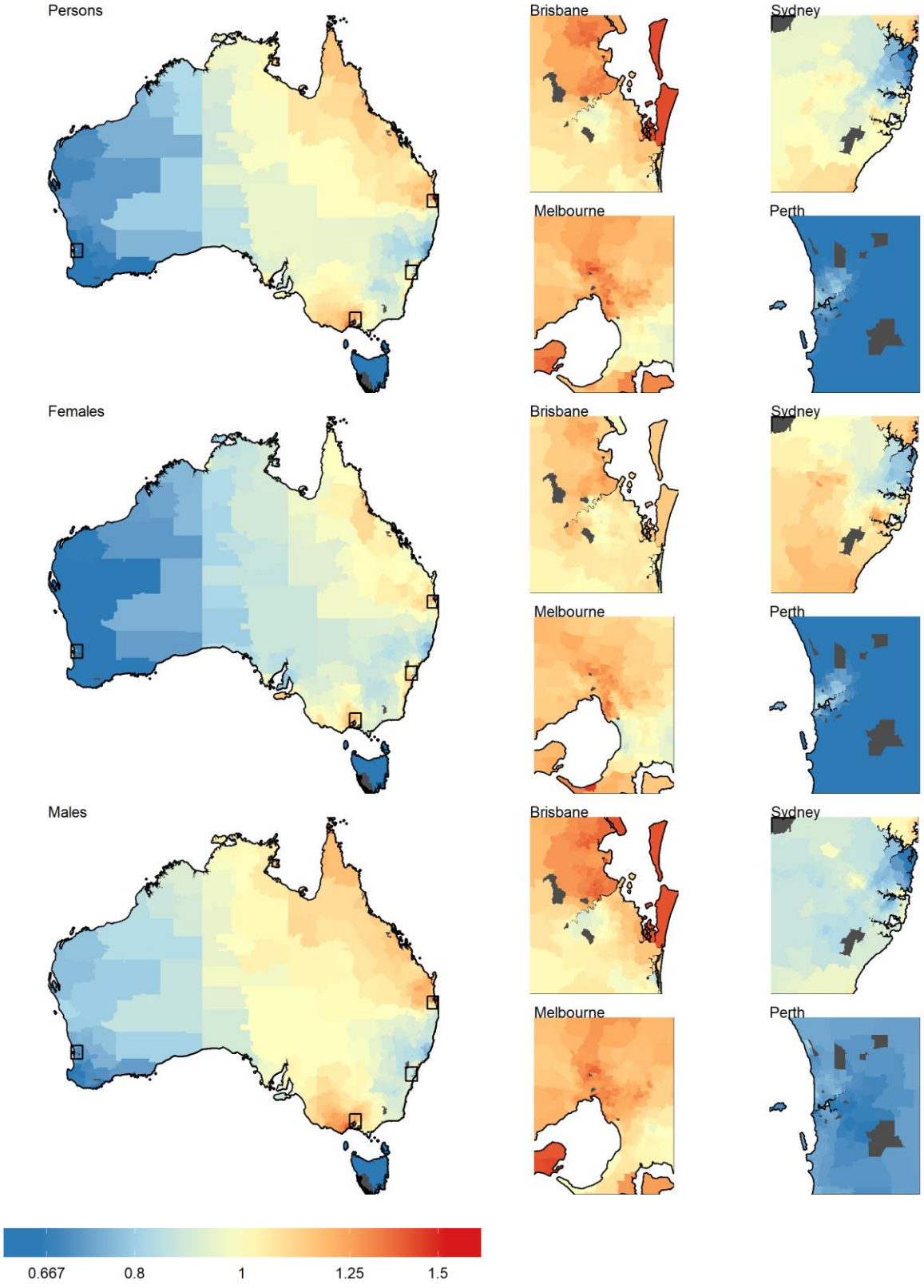

Figure S2; Maps of the standardised incidence ratios (SIRs) for classic myeloproliferative neoplasms by geographic area for persons (top), females (middle) and males (bottom) for all Australia (left) and insets of the major metropolitan areas (right). An SIR with value 1 indicates incidence is equal to the Australian average.

Figure S3; Maps of the excess hazard ratios (EHRs) for classic myeloproliferative neoplasms by geographic area for persons (top), females (middle) and males (bottom) for all Australia (left) and insets of the major metropolitan areas (right). An EHR of 1 indicates survival is equal to the Australian average.

CrIs for EHRs: no evidence of a spatial difference in hazard

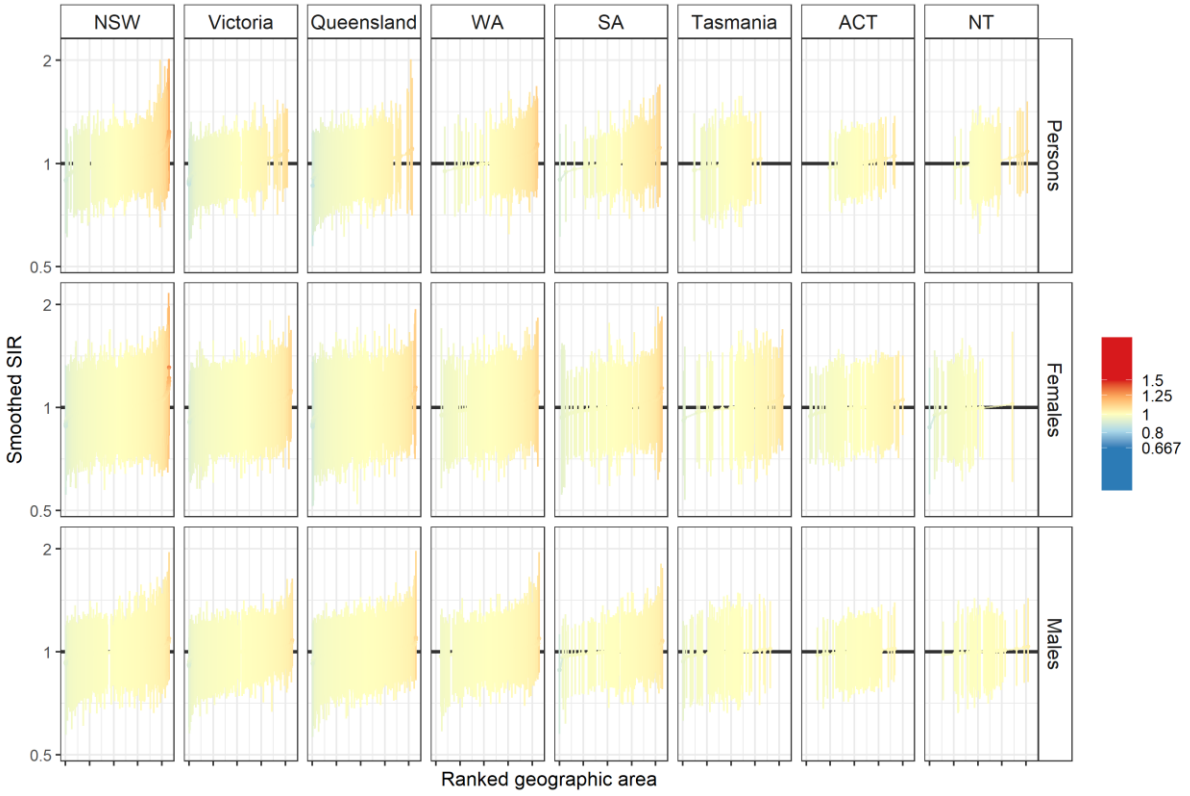

Figure S4; The 80% credible intervals for the excess hazard ratios (EHRs) for classic myeloproliferative neoplasms by sex and state, with a horizontal line indicating an EHR of 1 (representing the Australian average). The states and territories include New South Wales (NSW), Western Australia (WA), South Australia (SA), Australian Capital Territory (ACT) and the Northern Territory (NT).

Testing rates for mutations in genes associated with classic MPNs

The following Table and Figure use data from the Australian Medicare Benefits Schedule (MBS). MBS Item 73325 provides funding for the characterization of mutations in the Janus Kinase 2 (JAK2) and myeloproliferative leukemia (MPL) for the diagnosis of patients with clinical and laboratory evidence of PV and ET.[1] Data were obtained from the Medicare Item Reports released by Services Australia.[2] State is based on the patient's residential address at time of testing.

Table S1; Number of tests per 100,000 population for mutations in the JAK2 and MPL genes by state or territory, between 2012 and 2016, inclusive.

State	Number of tests (per 100,000 persons)
NSW	179
Victoria	167
Queensland	256
WA	130
SA	162
Tasmania	144
ACT	118
NT	43
Total	182

[1] Medicare Benefits Schedule – Item 73325. Retrieved December 16, 2020. http://www9.health.gov.au/mbs/fullDisplay.cfm?type=item&q=73325

[2] Medicare Statistics, Medicare Item Reports. Services Australia, Australian Government. Retrieved December 16, 2020. http://medicarestatistics.humanservices.gov.au/statistics/mbs_item.jsp

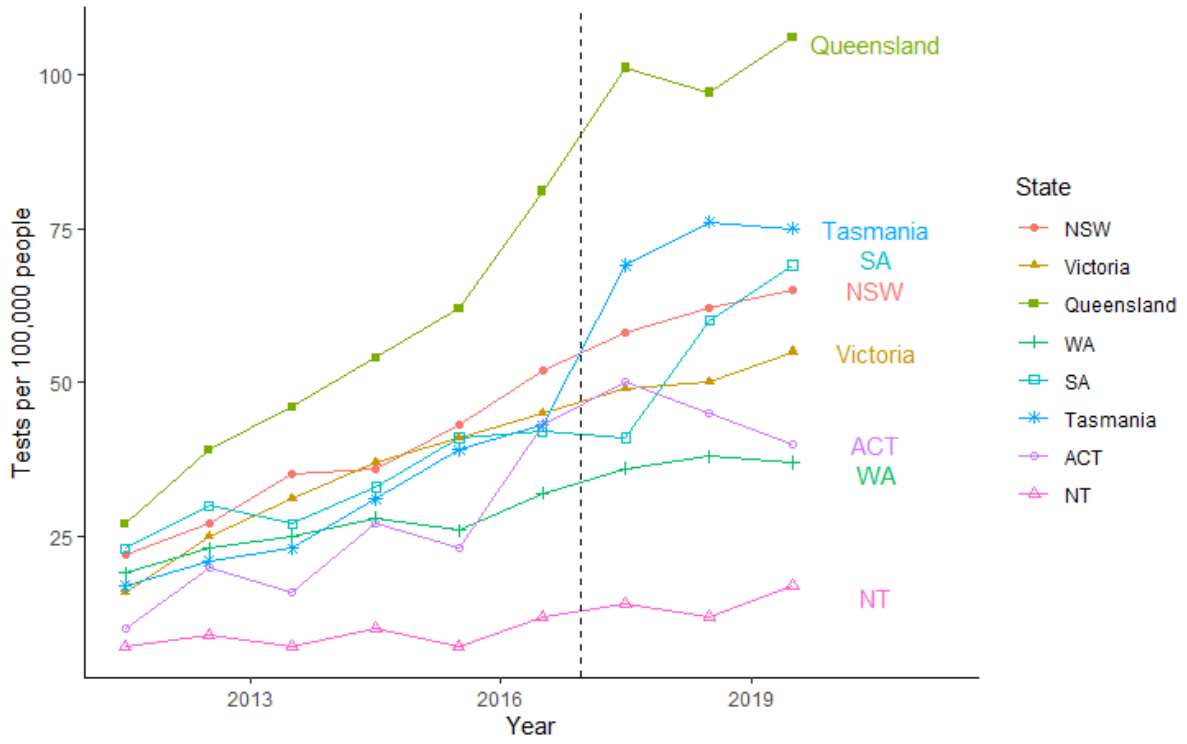

Figure S5; Rates of government-subsidised tests for mutations in the JAK2 and MPL genes by state plotted against financial year. The dashed vertical line indicates the limit of the study period, after which cancer incidence data were not available.

Most valid basis for diagnosis by state or territory

Table S2; Percentage of individuals whose most valid basis for diagnosis was histology by state or territory.

State	Histology (%)
NSW	49
Victoria	34
Queensland	63
WA	46
SA	43
Tasmania	64
ACT	79
NT	45
Total	48

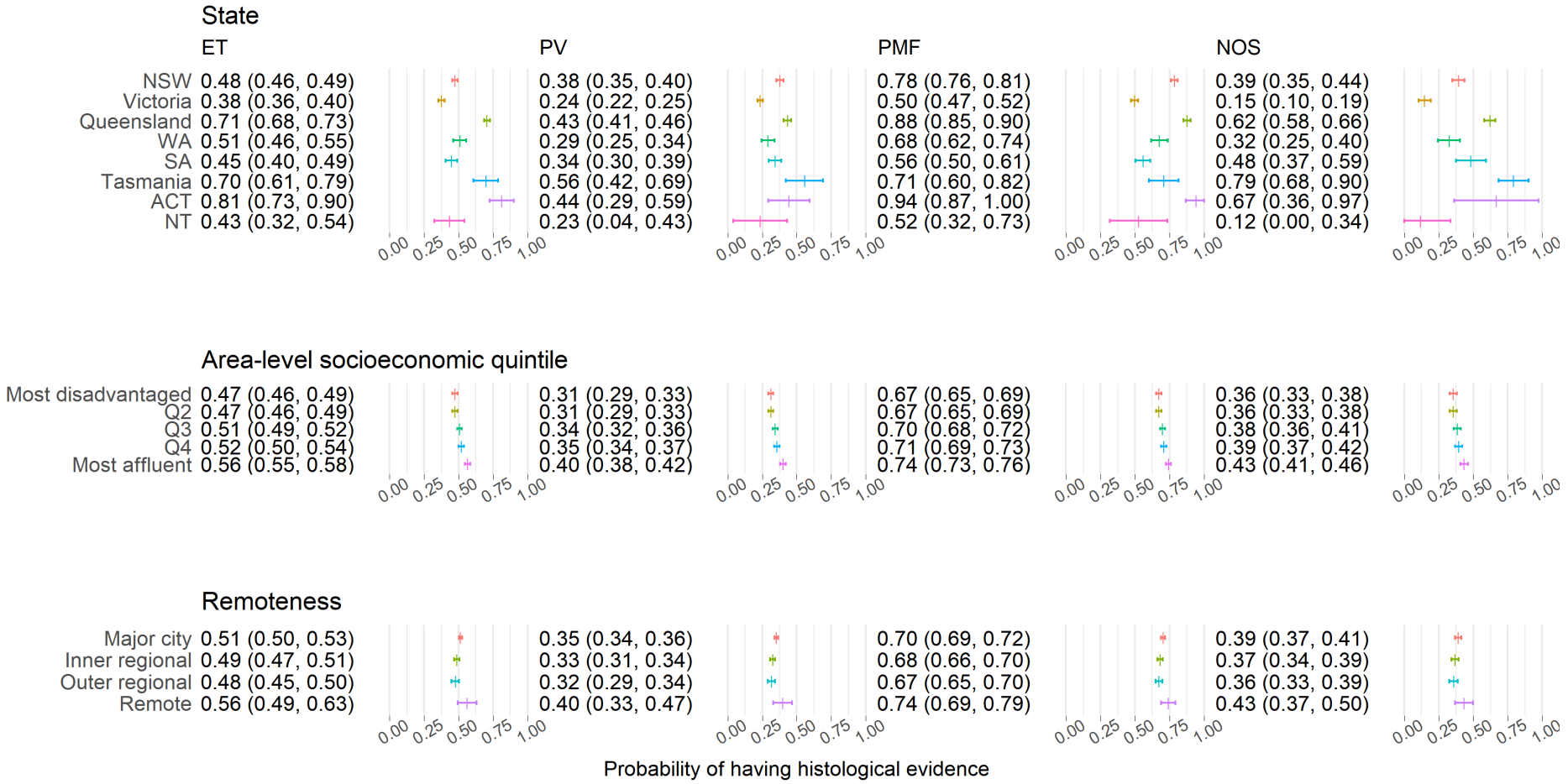

Figure S6; Modelled probabilities of having histologic evidence for diagnosis by MPN subtype and state or territory, area-level socioeconomic quintile or remoteness category. The states and territories include New South Wales (NSW), Western Australia (WA), South Australia (SA), Australian Capital Territory (ACT) and the Northern Territory (NT).

Calculating marginal survival

The model is:

$$d_i \sim Poisson(\mu)$$

$$\log(\mu - d^*) = \log(y) + x\beta$$

$$h(t, x, \beta) + h^*(t) = \frac{\mu - d^*}{y} = e^{x\beta}$$

Where h is the excess mortality and h^* is the age- and sex-matched population mortality.

Stata outputs $e^{x\beta}$ using the following syntax:

```
margins state##sex##end irsad##sex##end, expression(exp(predict(xb
nooffset)))
```

This syntax provides the marginal $h_t = e^{x\beta} - h_t^*$ by state and area-level socioeconomic quintile (IRSAD) for each sex and risk interval (t) and the corresponding standard error, $SE(h_t)$.

The interval-specific relative survival ratio (r_t) can be calculated as follows:

$$r_t = \exp(-\exp(x\beta) + h_t^*) = e^{-h_t}$$

We can then calculate the 5-year relative survival ratio:

$$r_{5yr} = \prod_{i=1}^{5} r_i = \prod_{i=1}^{5} e^{-h_i} = \exp\left(-\sum_{i=1}^{5} h_i\right)$$

We can calculate the variance of the cumulative excess mortality as follows:

$$\mathrm{Var}\left(\sum_{i=1}^{5} h_i\right) = \sum_{i=1}^{5} Var(h_i) = \sum_{i=1}^{5} \left(SE(h_i)\right)^2$$

Hence, the 95% confidence interval for the 5-year cumulative excess mortality is:

$$\sum_{i=1}^{5} h_i \pm 1.96 \sqrt{\sum_{i=1}^{5} \left(SE(h_i)\right)^2}$$

And the 95% confidence interval of the 5-year relative survival is:

$$\exp\left(-\sum_{i=1}^{5} h_i \pm 1.96 \sqrt{\sum_{i=1}^{5} \left(SE(h_i)\right)^2}\right)$$

Spatial modelling

Statistical models

Bayesian Leroux spatial models were fitted to the incidence and survival data for males, females and persons as per the Australian Cancer Atlas.[26] Briefly, the number of observed cases, y_i, for each area, i, was modelled as a Poisson process, as shown in the equation below, where E_i was the expected number of cases, θ_i was the log standardised incidence rate (SIR), β_0 was the global log SIR and S_i was the spatial random effect.

$$y_i \sim Pois\left(E_i e^{\theta_i}\right)$$

$$\theta_i \sim \beta_0 + S_i$$

The expected number of cases was age-standardised as below, where y_k was the total number across Australia of cases in age group k, N_k is the population size of age group k in Australia and N_{ik} is the size of the population in area i.

$$E_i = \sum_k \frac{y_k}{N_k} N_{ik}$$

The spatial random effect, S_i, was given a Leroux prior, as shown below, where ρ is the proportion of the spatial random effect that was autocorrelation between areas, w_{ij} is 1 if areas i and j are neighbouring, but otherwise 0 and σ_S^2 is a variance.

$$S_i | S_{\setminus i} \sim N\left(\frac{\rho \sum_j w_{ij} S_j}{\rho \sum_j w_{ij} + 1 - \rho}, \frac{\sigma_S^2}{\rho \sum_j w_{ij} + 1 - \rho}\right)$$

The following prior distributions were applied:

$$\beta_0 \sim N(0, 100\,000)$$

$$\sigma_S^2 \sim IG(1, 0.01)$$

$$\rho \sim Uniform(0, 1)$$

The survival model was a Bayesian Leroux spatial piecewise relative survival model. Since survival is poorer for PMF than ET or PV, the spatial survival model was adjusted for MPN subtype, with an indicator variable indicating whether the cancer was PMF or not. The number of deaths for age group k and area i and follow-up interval j, d_{kji}, was Poisson with mean μ_{kji}:

$$d_{kji} \sim Poisson(\mu_{kji})$$

The expected number of deaths, d^*_{kji}, was calculated using population mortality; y_{kji} was the person-time at risk; β_{0j} was the intercept for each follow-up year; the β_k were effects for each age group, β_s was the effect for sex, when modelling survival of persons and β_h was the effect for cancer subtype, where h had two levels: the first representing cases of PMF and the second ET or PV. The spatial random effect, S_i, was given a Leroux prior as above. Then μ_{kji} was modelled as follows.

$$\log(\mu_{kji} - d^*_{kji}) = \log(y_{kji}) + \beta_{0j} + \beta_k + \beta_h + \beta_s + S_i$$

$$\beta_{0j} \sim N(0, 0.01)$$

$$\beta_k \sim N(0, 0.01)$$

$$\beta_s \sim N(0, 0.01)$$

$$\frac{1}{\sigma_S^2} = \tau_S^2 \sim N(0, 0.2)$$

$$\rho \sim Uniform(0, 1)$$

The excess hazard ratio (EHR) for each area was calculated as $EHR_i = \exp(S_i)$.

Model convergence was checked using Geweke diagnostics[3] for the parameters and spatial effects and inspecting the trace and density plots of the posterior distributions for each parameter. Moran's I was used to test for autocorrelation in the residuals.[4]

[3] Geweke, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bayesian Statistics 4 (ed JM Bernado, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK.

[4] Li, Hongfei; Calder, Catherine A.; Cressie, Noel (2007). Beyond Moran's *I*: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. *Geographical Analysis*. **39** (4): 357–375.

Code for the spatial survival model

The data set consists of a row for each small area, MPN subtype, risk year, age group and sex.

The WinBUGS code is saved to a file location (this location is specified in the R code below). The R code calls the WinBUGS code that runs the model.

R code

```
library(R2WinBUGS)

# Fixed values
bugs.dat <- list(
            N = N,            # Number of areas (SA2s)
            T = T,            # Number of risk years
            N.d = N.d,        # Number of data rows (N * T * # of covariates)
            d = d,            # Number of deaths
            d.star = d.star,  # Expected number of deaths due to causes other than cancer of interest
            y = y,            # Person-time at risk offset
            RiskYear = RiskYear,
            Area = Area,
            adj = adj$adj,
            num = adj$num,
            cum = c(cumsum(adj$num) - adj$num, sum(adj$num)),
            sumnum = sum(adj$num),
            agegp2 = x1,
            agegp3 = x2,
            agegp4 = x3,
            subtype = x4,
            sex = x5,
            num.covariates = 5)

# Initial values
inits <- function() {list(
                        alpha = rep(-3, 5),
                        u = rep(0, N),
                        sigma.u2 = 0.2,
                        rho = 0.5,
                        beta = rep(0, num.covariates))}

# Parameters to monitor in WinBUGS
parameters <- c("alpha", "u", "sigma.u2", "beta", "rho")

# Run WinBUGS
bugs(
            data = bugs.dat,
            inits = inits,
            parameters.to.save = parameters,
            model.file = "Leroux.bug",
            n.chains = 1,
            n.iter = 150000,
            n.burnin = 50000,
            n.thin = 10,
            debug = FALSE,
            bugs.directory = "C:/WinBUGS14/",
            program = "WinBUGS",
            DIC = FALSE)
```

WinBUGS Code

```
model{
 for(i in 1:N.d){
                    d[i] ~ dpois(mu[i])
                    mu[i] <- d.star[i] + d.excess[i]
                    log(d.excess[i]) <- log(y[i]) + alpha[RiskYear[i]] + beta[1] * agegp2[i] + beta[2] * agegp3[i]
+ beta[3] *  agegp4[i] + beta[4] * subtype[i] + beta[5] * sex[i] + u[Area[i]]
                    }

 # Leroux prior for spatial random effects
 for(j in 1:N){
                  u[j] ~ dnorm(mean.u[j], prec.u[j])
                  A[j] <- (rho * num[j] + 1 - rho)
                  prec.u[j] <- A[j] / sigma.u2
                  mean.u[j] <- rho * sum(W.u[cum[j] + 1:cum[j+1]]) / A[j]
                  }

 for(h in 1:sumnum){
                          W.u[h] <- u[adj[h]]
                          }

 # Other priors
 sigma.u2 ~ dnorm(0, 0.2)I(0,)
 rho ~ dunif(0, 1)
 for(t in 1:T){
                  alpha[t] ~ dnorm(0, 0.01)
                 }

 for(k in 1:5){
                  beta[k] ~ dnorm(0, 0.01)
                  }
}
```