Multivariate Bayesian meta-analysis: joint modelling of multiple cancer types using summary statistics

Farzana Jahan (f.jahan@hdr.qut.edu.au)
Queensland University of Technology https://orcid.org/0000-0002-0300-6315
Earl W Duncan
Queensland University of Technology
Susanna M Cramb
Queensland University of Technology
Peter D Baade
Cancer Council Queensland
Kerrie Lee Mengersen
Queensland University of Technology

Research

Keywords: cancer incidence, cancer atlas, online estimates

Posted Date: June 18th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-36298/v1

Version of Record: A version of this preprint was published on October 17th, 2020. See the published version at https://doi.org/10.1186/s12942-020-00234-0.

RESEARCH

Multivariate Bayesian meta-analysis: joint modelling of multiple cancer types using summary statistics

Farzana Jahan[1*], Earl W. Duncan[1], Susana M. Cramb[2], Peter D. Baade[3] and Kerrie L. Mengersen[1]

*Correspondence:
f.jahan@hdr.qut.edu.au
[1]ARC Centre of Excellence in Mathematical and Statistical Frontiers, School of Mathematical Sciences, Science and Engineering Faculty, Queensland University of Technology, QLD 4001 Brisbane, Australia
Full list of author information is available at the end of the article

Abstract

Background: Cancer atlases often provide estimates of cancer incidence, mortality or survival across small areas of a region or country. A recent example of a cancer atlas is the Australian cancer atlas (ACA), that provides interactive maps to visualise spatially smoothed estimates of cancer incidence and survival for 20 different cancers over 2148 small areas across Australia.

Methods: The present study proposes a multivariate Bayesian meta-analysis model, which can model multiple cancers jointly using summary measures without requiring access to the unit record data. This new approach is illustrated by modelling the publicly available spatially smoothed standard incidence ratios (SIR) for multiple cancers in the ACA divided into three groups: common, rare/less common and smoking-related. The multivariate Bayesian meta-analysis models are fitted to each group in order to explore any possible association between the cancers in three remoteness regions: major cities, regional and remote areas across Australia.

Results: Substantive correlation was observed among some cancer types. There was evidence that the magnitude of this correlation varied according to remoteness of a region. High risk areas for specific combinations of cancer types were identified and visualised from the proposed model.

Conclusions: Publicly available spatially smoothed disease estimates can be used to explore additional research questions by modelling multiple cancer types jointly. These proposed multivariate meta-analysis models could be useful when unit record data are unavailable because of privacy and confidentiality requirements.

Keywords: cancer incidence; cancer atlas; online estimates

Background

Cancer atlases are the geographical representation of cancer incidence, mortality or survival to describe the cancer burden scenario across/between areas of a country, sub-region or group of countries with accompanying descriptive and analytical statistics [1]. The atlases are useful tools for showing geographic patterns of cancers [2] and have made significant contributions in cancer research [3]. A cancer atlas can be one of the methods to identify cancer patterns or risk factors [1]. Examples of early cancer atlases include, the Atlas of cancer mortality for U.S. counties, 1950–1969 [4], Atlas of U.S. cancer mortality among whites, 1950–1980 [5], Atlas of U.S. cancer mortality among non-whites, 1950–1980 [6] and U.S. cancer mortality rates and trends, 1950–1979 [7]. Cancer atlases started to be published online in

recent times, such as: Atlas of Cancer in India [8], NCI Cancer Atlas [9], Cancer Atlas of the United Kingdom and Ireland [10], the U.S. Atlas of Cancer Mortality [11], Atlas of Cancer in Queensland [12] and the Australian Cancer Atlas [13]. These atlases not only provide important information about the geographical variation in cancer burden but can also motivate different etiological questions about cancers. Most of the available cancer atlases modelled each cancer separately (univariate modelling) to obtain age standardised rates or indirect standardised ratios for incidence and hazard ratios or similar for survival for each cancer across the small areas.

One recent cancer atlas is the Australian cancer atlas (ACA). The ACA provides point and interval estimates of cancer diagnosis (incidence) and relative survival for 20 cancers over 2148 small areas (Statistical area level 2, SA2 [14]) across Australia along with interactive maps to visualise geographic patterns in cancer incidence and survival. The estimates used to produce the maps are based on an underlying Bayesian spatial model of the observed population data aggregated to the SA2 level; for details of the underlying methodology, please see [15]. All the smoothed estimates of cancer incidence and survival available in the ACA were obtained by univariate modelling of each cancer type separately.

There has been growing interest in joint modelling of two or more cancer types in order to explore the shared and divergent trends among the cancers in terms of geographic patterns and risk factors [16]. The most popular joint model for identifying the common risk factors of multiple disease is the shared component model [17], where instead of a multivariate model for jointly modelling two diseases, the underlying risk surface is separated into a disease specific risk component and a shared component. For example, [18] applied multivariate disease mapping of seven prevalent cancer types in Iran using a shared component model. A joint-analysis of the spatio-temporal variation of the six age-gender (three ages groups (0–14, 15–64, and 65 and over) and gender (male, female)) mortality risks was performed by [19] using a shared component spatio-temporal model. Bayesian shared component spatio-temporal models for male and female lung cancer was applied to analyse the spatio-temporal variation of lung cancer diagnosis [20, 21].

Other multivariate approaches for modelling multiple cancers are also available. Use of mixture factor models in modelling multivariate cancer outcomes was introduced by [22]. [23] also developed a latent mixture model for modelling four types of carcinoma and explored the spatial correlation structures among the cancer types between 300 geographic units in England, Scotland and Wales. A spatio-temporal mixture model was proposed to analyse the space-time variation in respiratory cancers in the state of South Carolina [24]. [25] proposed a hierarchical Bayesian factor model for spatially correlated data to explain across and within county correlations of cancer incidence rates by assuming that all different cancer types (12 for females and 10 for males) share one or more spatially correlated common factors. The model was to age-standardised cancer incidence rates by sex in 56 counties of Scotland. Most of these modelling approaches used unit level data from population based cancer registries, but this data can be difficult to access due to confidentiality and privacy requirements of data custodians.

More recent work has proposed ways to use summary measures, instead of raw unit record files, when modelling, such as by applying an extended Gamma-Poisson

model [26]. The authors showed an algorithm to extract data from several sources and analyse the summary statistics. However, the algorithm and model is applicable for univariate response variable. Additionally, [27] proposed new statistical models for analysis of summary estimates for symbolic data analysis. These models considered any symbols, such as random lists, histogram or intervals, derived from aggregating individual level data and performed statistical inferences for the symbols. One of the limitations of the symbolic data analysis approach is the problem of evaluating high dimensional integral over data space. There is further scope for improvement to existing methods and development of new methods in order to model the estimated summary information without accessing the unit record data.

In an earlier study, Bayesian hierarchical meta-analysis models for each of the 20 cancers were fitted separately and the pattern of incidence according to remoteness categories (major cities, regional and remote areas) was explored. The univariate meta-analysis model, if extended to accommodate multiple selected cancers in the same model, can be employed to identify possible association between selected cancers and could also help in detecting small areas where multiple cancer types have higher incidence rates jointly.

There has been only one study, to the best of our knowledge, which has studied the relationship between two cancers, namely colorectal and breast cancer, using summary measures from a cancer atlas to explore the factors responsible for the observed association [28]. This was a simplistic graphical comparison of ranked age-standardised cancer death rates, supplemented with a literature review to provide some etiologic hypotheses and suggest new opportunities of research in order to explore the association between the two cancers.

In the present study, instead of considering only two specific cancer types, multiple cancers from the ACA were chosen and the relationship is evaluated using posterior correlation matrices obtained by fitting a multivariate Bayesian hierarchical meta-analysis model. In addition to investigating the relationship among multiple cancers, the areas with higher risk of multiple cancers are also identified. The meta-analysis uses the spatially smoothed estimates from ACA, since these are publicly available. The proposed multivariate models in this study are expected to provide a more comprehensive understanding of relationships between the incidence of different cancer types. Using the hierarchical structure, we examine differences or similarities in observed relationships among groups of cancers across broad remoteness regions in Australia.

Methods

The proposed multivariate Bayesian meta-analysis model is described in the context of the ACA. The ACA is a freely accessible and interactive online platform, showing the spatial variation in standardised incidence and survival for 20 cancers across Australia (for a complete list, please see Appendix A.2). The ACA provides the point estimates for the standardised incidence ratios (SIR) and excess hazard ratios (EHR) and their 95% credible intervals for each of the 20 cancers in each of 2148 geographical areas (SA2) covering Australia.

Model Formulation

Let y_{ijk} and s^2_{ijk} denote the estimated mean and variance of the log standardised incidence ratio (SIR) respectively for the ith cancer, jth small area and kth category, where $i = 1, 2, 3, ..., n$ and n is the number of cancers included in the multivariate model, $j = 1, 2, 3, ..., J$, J is the total number of areas and $k = 1, 2, ..., K$ and K is the number of categories of interest. In our analysis of the ACA, $J = 2148$ and $K = 3$ ($k = 1$ if the jth SA2 is a major city, $k = 2$ if the jth SA2 is a regional area and $k = 3$ if the jth SA2 is a remote area) and n takes on different values according to the analysis; see below.

The remoteness information is obtained from the remoteness structure provided by Australian Bureau of Statistics in each Statistical Area level 1 (SA1, which aggregate to form SA2s) as a five-category index (major cities, inner regional, outer regional, remote and very remote) [29]. We assigned one remoteness area to each SA2 based on SA1 population sizes before combining the inner and outer regional areas, as well as remote and very remote areas, into regional and remote, respectively. Among the 2148 SA2s considered in the ACA, 1242 are major cities, 810 are regional and 96 are classified as remote areas.

In the ACA, the values of y_{ijk} and s^2_{ijk} are the outputs of a Bayesian spatial model. Hence, we model y_{ijk} as follows:

$$y_{ijk} \sim N(\mu_{ijk}, \sigma^2_{ijk}) \tag{1}$$

where, μ_{ijk} is the true value of the $log(SIR)$ for the i^{th} cancer, j^{th} SA2 and k^{th} region and with associated variance σ^2_{ijk}. Here we are not modelling the raw data but the estimated statistics for each small area which are provided by the ACA.

Now, μ_{ijk} can be further modelled as a multivariate normal distribution:

$$\mu_{ij(k)} \sim MVN(\mu_{i(k)}, \Sigma_{(k)}) \tag{2}$$

where, $\mu_{i(k)}$ is the region-specific means for k^{th} region and i^{th} cancer and $\Sigma_{(k)}$ denotes the covariance matrix accounting for the covariance among the means in the same region and different cancers. This hierarchy is added in the model to address the research question involving identifying patterns of cancer incidence in different regions.

The region-specific means for the i^{th} cancer, $\mu_{i(k)}$ can be further modelled hierarchically (see Appendix A.3), but for the sake of this study, we will consider modelling up to this level and will focus on the posterior means and the posterior covariance matrices associated with different cancers in each region. The aim is to explain how the relationship between the cancers varies with respect to major cities, regional and rural areas.

The priors for the model parameters can be specified as follows:

σ^2_{ijk} can take a prior that utilises the uncertainty information from the estimates available in the atlas as,

$$\sigma^2_{ijk} \sim \frac{\nu s^2_{ijk}}{\chi^2_{(\nu)}} \tag{3}$$

where the degrees of freedom of the χ^2 distribution are chosen to reflect the prior degree of certainty in these estimates [30]. Following the rationale of [30], a common choice of ν is 2, which will be used in this study.

The prior for the variance covariance matrix Σ_k is described by an inverse Wishart distribution as

$$\Sigma_{(k)} \sim IW(V, n) \tag{4}$$

which can be written equivalently as:

$$\tau_{(k)} \sim W(\Gamma, n) \tag{5}$$

where $\tau_{(k)} = \Sigma_k^{-1}$, is the precision matrix for kth region, which is a Wishart prior with degrees of freedom n set equal to the number of cancers considered in the model and the scale matrix Γ is specified as an identity matrix so that the priors are minimally informative [31].

Selection of cancer types for Multivariate Model

The proposed multivariate models are fitted for each of the groups mentioned below.

Group 1: Most Common Cancer types

Among the cancer types reported in ACA, the most common are, prostate, breast, colorectal (bowel), melanoma and lung cancer. These five cancer types account for around 60% of all cancers diagnosed in Australia [32]. To fit the proposed multivariate model, we grouped these common cancer types into subgroups as follows:

- Model 1: Lung, melanoma and bowel cancers : 1(a): for males, 1(b): for females and 1(c): for persons
- Model 2: Lung, melanoma, bowel and prostate cancers for males
- Model 3: Lung, melanoma, bowel and breast cancers for females

Group 2: Less Common and rare cancers

According to Cancer Australia, most cancer types, except breast, prostate, bowel, lung and melanoma, can be classified as rare or less common [33].A rare cancer is defined as a type of cancer that has less than 6 cases per year per 100,000 population, whereas a less common cancer is defined as one that has between 6 and 12 cases per year per 100,000 population [33].

According to the most recent age standardised incidence rates per 100,000 population for Australia [34], the rare cancer types, among the selected cancer types in ACA, include liver cancer for females (4.7) and oesophageal cancer for females (3.6). The less common cancers include brain cancer (males: 9.1, females: 6.0 and persons: 7.5) , cervical cancer for females (7.1), head and neck cancer for females (8.6), kidney cancer for females (9.4), liver cancer for all persons (8.7), oesophageal cancer for males (8.7) and persons (6.2), stomach cancer for females (6.4) and persons (9.3) and thyroid cancer for males (6.5). We created the following subgroups for the less common/rare cancers to fit the proposed model to each group:

- Model 4: Liver and oesophageal cancer for females

- Model 5: Brain, oesophageal, thyroid cancers for males
- Model 6: Brain, Cervical, head and neck, kidney and stomach cancers for females
- Model 7: Brain, liver, oesophageal and stomach cancer for persons

Group 3: Cancers associated with smoking
One of the most studied cancer risk factors is smoking, which has been shown to cause several types of cancer. The following cancers are found to be related to smoking [35–40], which form the last group for fitting the proposed model:

- Model 8: Lung, liver, pancreatic, stomach, kidney, oesophageal and head and neck cancers : 8(a): for males, 8(b): for females and 8(c): for persons

Model Implementation

A total of 12 multivariate Bayesian meta-analysis models were run for the different combinations of cancer types in `R version 3.6.0` [41] using the package `R2jags version 0.5-7` [42]. The MCMC model output was summarised in R using the `coda` package [43]. The JAGS code for the model is given in the appendix.

Three parallel MCMC chains, each with 100,000 iterations with a burn in period of 10000 iterations were run to fit the proposed models. Convergence was examined using visual diagnostics for the parameters of interest $\mu_{ij(k)}, \mu_{i(k)}$ and Σ_k.

Model Inferences

From the posterior distributions of the parameters of interest from each of the fitted Bayesian meta-analysis models, the following inferences were drawn in this study.

Comparing the posterior mean $log(SIR)$ $(\mu_{ij(k)})$ for the group of cancers for each SA2, we identified those SA2s for which all cancers in a group had higher incidence compared to the Australian average.

From the matrix of posterior means, we were able to evaluate the behaviour of a group of cancers in different regions. The posterior covariance matrix for each of the regions was used to obtain the correlation between all possible pairs of cancers in a group within and across different regions (major cities, regional and remote areas). An asymptotic χ^2 test was used to test the equality of multiple correlation matrices [44].

$$C^2 = \sum_{i=1}^{k} \left(\frac{1}{2} tr(Z_i^2) - dg^{'}(Z_i) S^{-1} dg(Z_i) \right) \sim \chi^2_{(k-1)p(p-1)/2} \tag{6}$$

where $Z_i = \sqrt{n_i}\bar{R}^{-1}(R_i - \bar{R})$, $\bar{R} = (n_1 R_a + ... + n_k R_k)/n = \bar{r}_{ij}$, $S = (\delta_{ij} + \bar{r}_{ij}\bar{r}^{ij})$, $\bar{r}^{ij} = \bar{r}_{ij}^{-1}$, $R_1, R_2, ..., R_k$ are sample correlation matrices based on k independent samples of sizes $n_1, n_2, ..., n_k$ from $p-$variate normal populations, δ_{ij} is the Kronecker delta and $dg(Z_i)$ denotes the diagonal of a square matrix Z_i of correlation coefficients.

Using Jennrich's test, we identified which cancers had substantially different correlation matrices in urban, regional and remote Australia.

Using the model inferences, high risk areas for each cancer and the groups of cancers are identified. High risk areas are defined as the SA2s having an SIR substantially larger than the Australian average. Several options are possible to identify the areas, but here posterior probabilities (PP) are used. The posterior probability that an estimated SIR of a particular cancer is greater than the national average can be calculated for each SA2. It is defined as the ratio of the number of MCMC iterations in which the modelled SIR is above 1, divided by the the total number of iterations [15]. SA2s with PP ≥ 0.80 can be considered as a high risk area for a cancer [45].

Results

The posterior means with 95% credible interval for $\mu_{i(k)}$ of each group of cancer types in each of the 3 remoteness categories (namely major cities, regional and rural areas) under each of the 12 models are shown in Figures 1 to 3 (for the actual values of posterior means, see Supplementary Information.).

Figures 1-3 demonstrate how different cancers have different incidence patterns over different regions of Australia. For example, Figure 1, the highest melanoma incidence has occured in regional areas, whereas lung cancer has higher incidence in remote areas (for males, females and persons). Figure 2 (Model 6), remote areas had the highest incidence of cervical and head and neck cancers among persons on average.

[Figures 1,2 3 should be inserted here]

The mean posterior correlation matrices for each model in the three different regions are shown in Tables 1 - 3. When two cancers have positive correlation, it means that incidence patterns for both cancer types are similar in that particular region. If the 95% credible interval of the correlation coefficient includes zero, it is assumed that no substantive correlation is present between the incidence patterns of the pair of cancer types under consideration.

[Tables 1,2 3 should be inserted here]

The posterior correlation matrices for most common cancer types (Models 1a, 1b, 1c, 2 & 3) are presented in Table 1. In Table 1, we can see that the correlation coefficients of melanoma and lung cancer are negative in major cities (for males: model 1(a) and persons: model 1(c)) and these are not substantially correlated in regional and remote areas. Some more examples of correlation in different regions: negative correlation between prostate and lung cancers in cities and no correlation in regional and remote areas (Table 1: Model 2) and significant positive correlation between breast cancer and melanoma in cities and no substantive correlation in regional and remote areas (Table 1: Model 3).

In Table 2, the posterior correlation matrices for rare and less common cancers are reported. We can observe no substantive correlation between liver and oesophageal cancer for females in all three regions (Model 4). Thyroid and brain cancer have negative correlation in major cities but no correlation in regional and remote areas (Model 5). Head and neck cancer and cervical cancer have a significant positive correlation in major cities and regional areas but none in remote areas (Model 6).

Table 3 shows the 7×7 correlation matrices for smoking related cancers for males, females and persons (Model 8a, 8b and 8c). As can be seen, correlation can

substantially differ between the same pairs of cancers across major cities, regional and remote areas. For example, stomach and lung cancers have significant positive correlations for males and persons (model 8a and 8c) in major cities but there is no substantive correlation between these cancers in regional and remote areas. Similarly, lung and kidney cancers for males, females and persons (Model 8a, 8b and 8c) have significant positive correlation in major cities and weak or no correlation in regional and remote areas. There are also similar correlations across different regions among pairs of cancers. For instance, lung, head and neck cancers are positively correlated in major cities, regional and remote areas for persons (Model 8c).

Clearly, different models have some similarities and dissimilarities according to pairwise correlation. From Jennrich's test (Table 4), substantive differences among the correlation matrices were found for the majority of models including most common cancers (Models 1a, 1b, 1c, 2 & 3), rare and less common cancers (Models 6 & 7) and smoking related cancers (Models 8a, 8b & 8c) (Table 4). For the other less common/rare cancers (Models 4and 5) care should be taken due to small sample sizes.

[Table 4 should be inserted here]

Using the model inferences, we have identified small areas (SA2s) with higher incidence of a cancer or a group of cancer types; namely high risk areas. Figures 4-6 visualise the high risk areas around Australia for each group of cancer types. For example, for most common cancers (group 1, model 1), the spatial map in figure 4 shows high risk areas for each cancer type of the group individually as well as jointly, (lung, melanoma and bowel individually; lung and melanoma; lung and bowel; lung, melanoma and bowel jointly, for persons, Model 1c). The cluster of areas having high risk for group of cancers are also identified under each model (see Figures 4-6 and Tables 5 & 6). To enable a clearer view, four insets of the full map are shown alongside.

[Figures 4-6 should be inserted here]

[Tables 5 and 6 should be inserted here]

From Table 5, we can see 76 SA2s out of 2148 are identified as high risk areas for all three cancers considered in Model 1 (for persons). There are 22 SA2s around Australia which have higher incidence for lung and melanoma, 123 SA2s for lung and bowel, 116 SA2s for melanoma and bowel jointly. For more information on the number of SA2s having substantially higher SIRs for individual and joint cancers under selected models, refer to Tables 5 and 6. Only three models out of the twelve are illustrated to show the high risk areas in maps in this section. In figure 3, only groups of cancer types with 20 or more SA2s are showed in the map. For more details, refer to the supplementary material.

Discussion

A multivariate Bayesian meta-analysis model was proposed in the present study to model multiple cancers jointly to identify any existent relationships among the cancers. The advantages of this model include that it incorporates the uncertainty of the input modelled summary estimates, it allows for easy identification and visualisation of areas with high risk for different combinations of cancers, and it is readily extendable.

The proposed model was illustrated by joint modelling of multiple cancers in different groups formed from the 20 cancers included in the ACA. The most common cancers (Models 1,2,3) and the smoking related cancers (Model 8) were found to have significantly different correlation matrices across major cities, regional and remote areas. These findings imply that additional factors influencing cancer incidence in the three different regions may be present. Among the less common and rare cancers group, models 6 and 7 have a significantly different correlation matrix in each of the three regions. The correlation coefficients in each of the correlation matrices represent the correlation between incidences of pairs of cancers within each cancer group and region.

Mostly in the published literature, multivariate meta-analyses of cancer have focused on exploring the relationship between risk factors/prognostic factors and specific cancers [37, 46–50]. The present study is the first of its kind identifying correlation between the incidence of pairs of cancer types in selected groups. While some of the obtained results support the already known facts, some of the results are new and could create opportunities for further investigation into the reasons for the observed patterns. For instance, the smoking related cancers are modelled jointly (Model 8) for male, females and persons. These cancers are expected to have positive correlation, yet significant negative correlation was observed between oesophageal and liver cancer incidence (for males & persons in major cities) as well as kidney, head and neck cancers (for persons in remote areas). It may be that these cancers are predominantly driven by risk factors that could not be included in our analysis, such as obesity (for oesophageal cancer) or chronic hepatitis viral infection (for liver cancer). These models can be used to identify unexpected negative correlations for further investigation.

The proposed multivariate Bayesian hierarchical meta-analysis model is applied to model the publicly available smoothed estimates of multiple cancers jointly. Such an approach is useful when the raw data are unavailable and can be used to answer additional research questions of interest. However, when applying these models to groups of cancers, one consideration is that the choice of cancers can have a noticeable impact on estimates obtained (see Supplementary material). For instance, in Figure 1, Melanoma for females has slightly different estimates in model 1(b) and model 3. This results from the multivariate nature of the proposed model and the covariance structure within the group. Since correlation is a standardised form of covariance, the precision of estimates and the correlation between cancers are related. Since the choice of cancer types included may influence the results, we recommend comparing the multivariate results with the univariate results (using an approach such as [51]). Also, because this model was developed for summarised modelled estimates the proposed model cannot be applied to raw incidence rates without modifying some form of spatial smoothing and changing the distributional assumptions at different levels.

Conclusions

This study presents a novel use of Bayesian meta-analysis for multivariate modelling of reported cancer incidence estimates. The modelling technique can be generalised for other disease maps or atlases. The hierarchy introduced in the model using the

remoteness structure of each of the small areas could also be replaced by any other factor of interest. For example, in the present model, if we wanted to check how the correlation among the multiple cancers differs in different states across Australia, we could use states as the hierarchy instead of the remoteness regions. Hence the proposed modelling approach is flexible for joint modelling of multiple estimated disease outcomes with different research questions of interest. The hierarchical stage could also be extended in a straightforward manner to include more than one factor. For example, we could use both states and regions in the model by including one more hierarchy in the existing model. We could also extend the model by using socio-economic status of each area as another factor of interest. The scope for this model is vast, and we anticipate it being a useful addition for analysing summary estimates in more detail.

Ethics approval and consent to participate
No ethics approval or consent to participate was required for this work.

Consent for publication
All the authors have provided consent to publish this manuscript.

Availability of data and materials
The cancer incidence data used in this research can be downloaded from the Australian Cancer Atlas website (https://atlas.cancer.org.au/). The other data set used in this research is the remoteness indexes which can be downloaded from ABS ASGS 2011 website (https://data.gov.au/dataset/ds-dga-4b208cc1-f5de-405d-af96-0777645dfc87/details?q=). Some transformations are made to get the remoteness indexes for each of the SA2s.

Competing interests
The authors declare that they have no competing interests.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the Australian Research Council Grant no. FL150100150, "Bayesian Learning for Decision Making in the Big Data Era".

Author's contributions
FJ and KM conceived the modelling approach. FJ conducted the analyses and led the write up. ED, SC, PB and KM supervised the analyses and interpretations and contributed to the write up. ED helped in coding for the spatial maps.

Acknowledgements
The authors acknowledge the work of Australian Cancer Atlas Project Team and funding bodies in developing the atlas. The authors also acknowledge the support of the Australian Research Council (ARC) Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS).

Author details
[1]ARC Centre of Excellence in Mathematical and Statistical Frontiers, School of Mathematical Sciences, Science and Engineering Faculty, Queensland University of Technology, QLD 4001 Brisbane, Australia. [2]Institute of Health and Biomedical Innovation, Queensland University of Technology, QLD 4001 Brisbane, Australia. [3]Cancer Council Queensland, 553 Gregory Terrace, Fortitude Valley QLD 4006 Brisbane, Australia.

References
1. D'Onofrio A, Mazzetta C, Robertson C, Smans M, Boyle P, Boniol M. Maps and atlases of cancer mortality: a review of a useful tool to trigger new questions. e Cancer Medical Science. 2016;10.
2. Gundersen L. Mapping it out: using atlases to detect patterns in health care, disease, and mortality. Annals of Internal Medicine. 2000;133(2):161–162.
3. Tatalovich Z, Stinchcomb DG. In: Berrigan D, Berger NA, editors. Creating Maps and Mapping Systems for Cancer Control and Prevention. Cham: Springer International Publishing; 2019. p. 59–79.
4. Mason TJ, McKay FW. Atlas of Cancer Mortality for U.S. Counties, 1950-1969. DHEW publication no. (NIH) 75-780. U.S. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health; 1975.
5. Pickle LW. Atlas of US cancer mortality among whites, 1950-1980. 87. US Dept. of Health and Human Services, Public Health Service, National US Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health; 1987.
6. Pickle LW. Atlas of US cancer mortality among nonwhites, 1950-1980. 90. US Department of Health and Human Services, Public Health Service, National Institutes of Health ; 1990.

7. Riggan WB. US cancer mortality rates and trends, 1950-1979. vol. 1. NCI/EPA Interagency Agreement of Environmental Carcinogenesis; 1983.
8. NRCP. Development of an atlas of cancer in India; 2010. Available from: http://www.ncdirindia.org/ncrp/ca/index.aspx.
9. NCI. Geographic Information Systems and Science for Cancer Control; 2017. Available from: https://gis.cancer.gov/canceratlas/.
10. GovUK. Cancer Atlas of the United Kingdom and Ireland. Office for National Statistics. 2014;https://data.gov.uk/dataset/91e37ff6-162e-47ca-8610-50b5c910d94f/cancer-atlas-of-the-united-kingdom-and-ireland. Accessed on 8/4/2019.
11. NIH. U.S. Atlas of Cancer Mortality. Division of Cancer Epidemiology and Genetics - National Cancer Institute. 2016;https://dceg.cancer.gov/research/how-we-study/descriptive-epidemiology/cancer-mortality-atlas.Accessed on 8/4/2019.
12. Cramb SM BP Mengersen KL. Queensland Cancer Atlas. Cancer Council Queensland. 2011;https://cancerqld.org.au/research/queensland-cancer-statistics/queensland-cancer-atlas/. Accessed on 10/2/2019.
13. ACA. Australian Cancer Atlas. Cancer Council Qld, Qld University of Technology, Cooperative Research Centre for Spatial Information. 2018;http://atlas.cancer.org.au/. Accessed on 5/10/2018.
14. ABS. Australian Statistical Geography Standard (ASGS): volume 1—main structure and greater capital city statistical areas. Canberra: Australian Bureau of Statistics. 2011;Available from: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS).
15. Duncan EW, Cramb SM, Aitken JF, Mengersen KL, Baade PD. Development of the Australian Cancer Atlas: spatial modelling, visualisation, and reporting of estimates. International Journal of Health Geographics. 2019;18(1):21.
16. Dabney AR, Wakefield JC. Issues in the mapping of two diseases. Statistical Methods in Medical Research;.
17. Knorr-Held L, Best NG. A shared component model for detecting joint and selective clustering of two diseases. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2001;164(1):73–85.
18. Mahaki B, Mehrabi Y, Kavousi A, Akbari ME, Waldhoer T, Schmid VJ, et al. Multivariate disease mapping of seven prevalent cancers in Iran using a shared component model. Asian Pacific Journal of Cancer Prevention . 2011;12(9):2353–8.
19. Manda SO, Abdelatif N. Smoothed temporal atlases of age-gender all-cause mortality in South Africa. International Journal of Environmental Research and Public Health. 2017;14(9):1072.
20. Cramb SM, Baade PD, White NM, Ryan LM, Mengersen KL. Inferring lung cancer risk factor patterns through joint Bayesian spatio-temporal analysis. Cancer Epidemiology. 2015;39(3):430–439.
21. Richardson S, Abellan JJ, Best N. Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK). Statistical Methods in Medical Research. 2006;15(4):385–407.
22. Bailey T, Hewson P. MIXTURES OF FACTOR MODELS FOR MULTI-VARIATE DISEASE RATES. Revstat Statistics Journal . 2011;9(1):99–114. Available from: https://pdfs.semanticscholar.org/072c/b8311019ca31cc715f6e017593dc8e49c40e.pdf.
23. Hewson P, Bailey TC. Modelling multivariate disease rates with a latent structure mixture model. Statistical Modelling. 2010;10(3):241–264.
24. Carroll R, Lawson AB, Kirby RS, Faes C, Aregay M, Watjou K. Space-time variation of respiratory cancers in South Carolina: a flexible multivariate mixture modeling approach to risk estimation. Annals of Epidemiology. 2017;27(1):42–51. Available from: https://doi.org/10.1016/j.annepidem.2016.08.014.
25. Mezzetti M. Bayesian factor analysis for spatially correlated data: application to cancer incidence data in Scotland. Statistical Methods & Applications. 2012;21(1):49–74.
26. Lee JY, Brown JJ, Ryan LM. Sufficiency revisited: Rethinking statistical algorithms in the big data era. The American Statistician. 2017;71(3):202–208.
27. Beranger B, Lin H, Sisson SA. New models for symbolic data analysis. arXiv preprint arXiv:180903659. 2018;.
28. Howell MA. The association between colorectal cancer and breast cancer. Journal of Chronic Diseases. 1976;29(4):243–261.
29. ABS. Australian Statistical Geography Standard (ASGS): Volume 5 - Remoteness Structure,, July 2011; 2013. Available from: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/remotenessstructure.
30. DuMouchel W. Hierarchical Bayes linear models for meta-analysis; 1994. Available from: https://people.eecs.berkeley.edu/~russell/classes/cs294/f05/papers/dumouchel-1994.pdf.
31. Gelman A, Hill J. Data analysis using regression and multilevel hierarchical models. vol. 1. Cambridge University Press New York, NY, USA; 2007.
32. AIHW. Cancer in Australia: In brief 2019. Cancer Series no 122 Cat no 126CanberraAIHW. 2019;.
33. Cancer Australia. Rare and less common cancers; 2014. Available from: https://canceraustralia.gov.au/about-us/news/rare-and-less-common-cancers.
34. AIHW. ACIM books & ACD pivot table. Australian Institute of Health and Welfare & Cancer Australia (AIHW). 2019;Available from: https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/acim-books.
35. Simon S. Study: Smoking Causes Almost Half of Deaths from 12 Cancer Types; 2015. Available from: https://www.cancer.org/latest-news/study-smoking-causes-almost-half-of-deaths-from-12-cancer-types.html.
36. Freedman ND, Silverman DT, Hollenbeck AR, Schatzkin A, Abnet CC. Association between smoking and risk of bladder cancer among men and women. JAMA: The Journal of the American Medical Association. 2011;306(7):737–745.
37. Bagnardi V, Blangiardo M, La Vecchia C, Corrao G. A meta-analysis of alcohol drinking and cancer risk. British Journal of Cancer. 2001;85(11):1700.
38. Sasco A, Secretan M, Straif K. Tobacco smoking and cancer: a brief review of recent epidemiological evidence. Lung Cancer. 2004;45:S3–S9.

39. Lash TL, Aschengrau A. A null association between active or passive cigarette smoking and breast cancer risk. Breast Cancer Research and Treatment. 2002;75(2):181–184.
40. Morabia A. Smoking (active and passive) and breast cancer: epidemiologic evidence up to June 2001. Environmental and Molecular Mutagenesis. 2002;39(2-3):89–95.
41. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2018. Available from: https://www.R-project.org/.
42. Su YS, Yajima M. R2jags: Using R to Run 'JAGS'; 2015. R package version 0.5-7. Available from: https://CRAN.R-project.org/package=R2jags.
43. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News. 2006;6(1):7–11. Available from: https://journal.r-project.org/archive/.
44. Jennrich RI. An asymptotic $\chi 2$ test for the equality of two correlation matrices. Journal of the American Statistical Association. 1970;65(330):904–912.
45. Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. Environmental Health Perspectives. 2004;112(9):1016–1025.
46. Malats N, Bustos A, Nascimento CM, Fernandez F, Rivas M, Puente D, et al. P53 as a prognostic marker for bladder cancer: a meta-analysis and review. The Lancet Oncology. 2005;6(9):678–686.
47. Guan X, Wang Y, Xie R, Chen L, Bai J, Lu J, et al. p27Kip1 as a prognostic factor in breast cancer: a systematic review and meta-analysis. Journal of Cellular and Molecular Medicine. 2010;14(4):944–953.
48. Nakamura H, Ando K, Shinmyo T, Morita K, Mochizuki A, Kurimoto N, et al. Female gender is an independent prognostic factor in non-small-cell lung cancer: a meta-analysis. Annals of Thoracic and Cardiovascular Surgery. 2011;17(5):469–480.
49. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. JAMA: The Journal of the American Medical Association. 2008;300(23):2765–2778.
50. Pavia M, Pileggi C, Nobile CG, Angelillo IF. Association between fruit and vegetable consumption and oral cancer: a meta-analysis of observational studies. The American Journal of Clinical Nutrition. 2006;83(5):1126–1134.
51. Jahan F, Duncan E, Cramb S, Baade P, Mengersen K. Augenting Disease Maps: a Bayesian Meta-analysis approach. Royal Society Open Science. under review;.

Figures and Tables

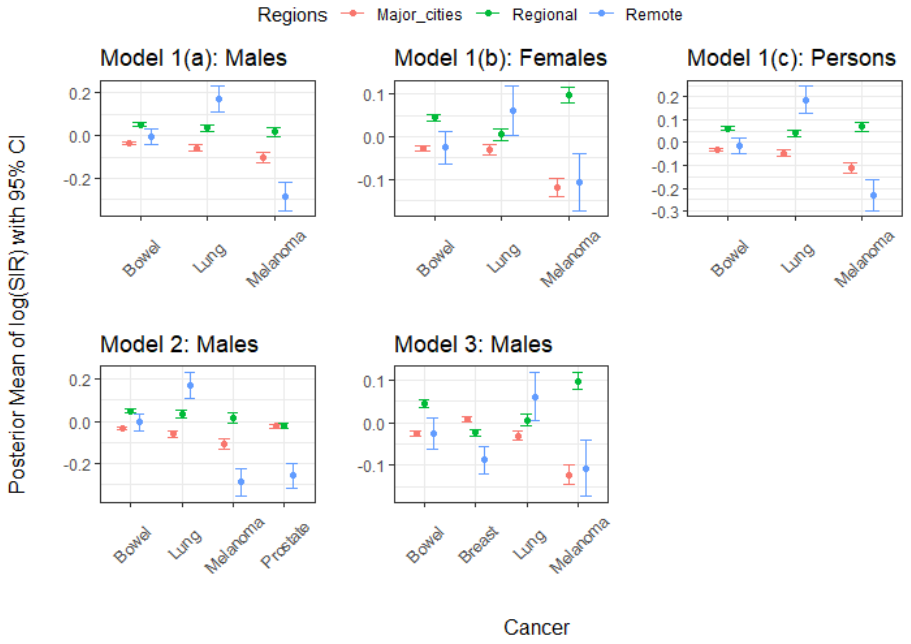

Figure 1 Posterior means and 95% credible intervals of log(SIR) by multivariate model for the most common cancers (Group 1) over remoteness regions, Australia

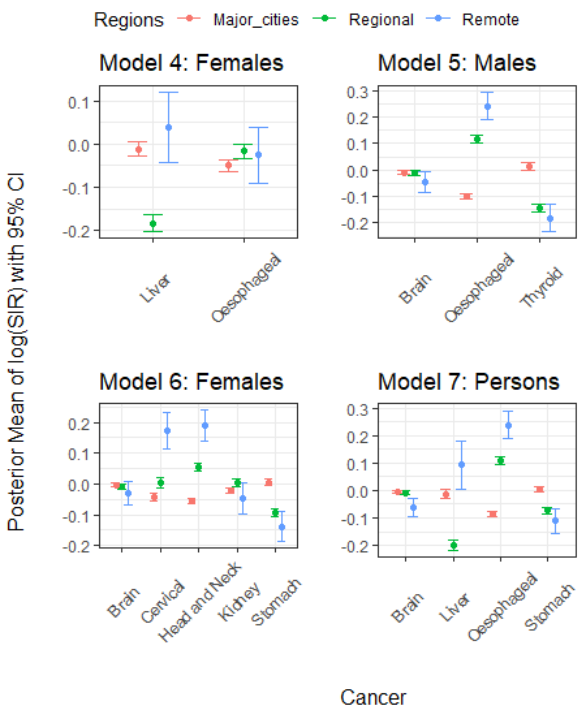

Figure 2 Posterior means and 95% credible intervals of log(SIR) by multivariate model for the less common cancers/rare cancers (Group 2) over remoteness regions, Australia

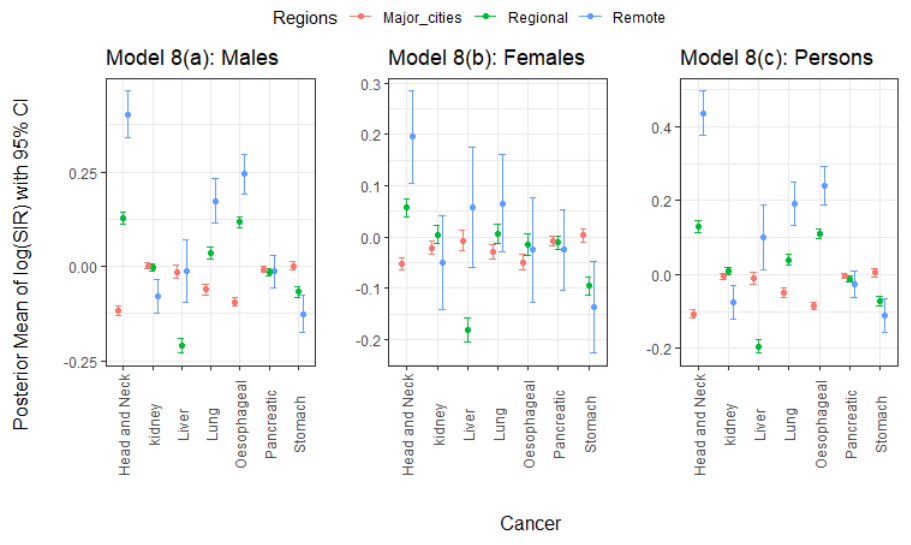

Figure 3 Posterior means and 95% credible intervals of log(SIR) by multivariate model for the smoking related cancers (Group 3) over remoteness regions, Australia

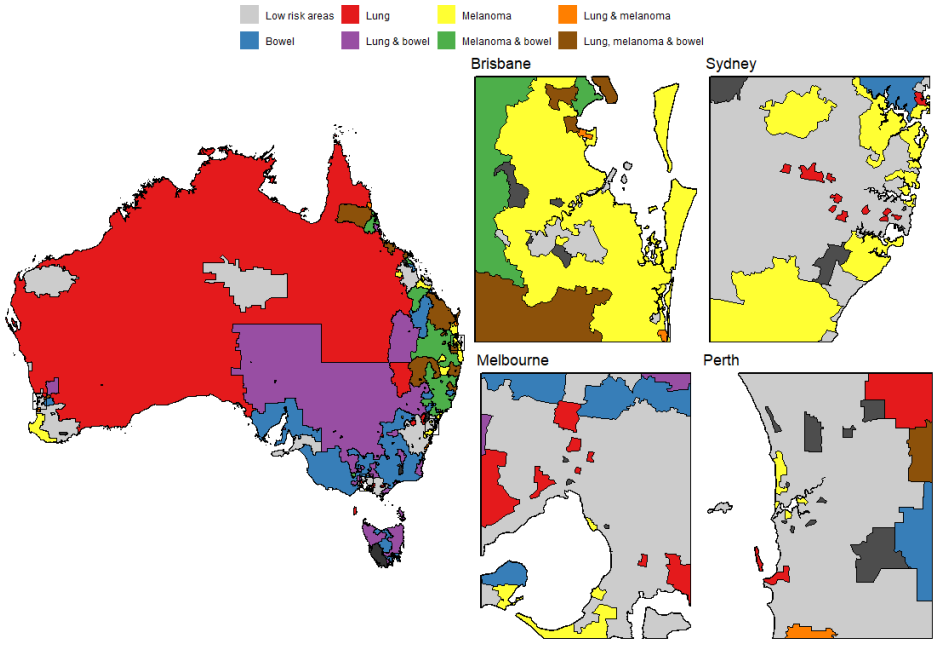

Figure 4 High risk areas for multiple cancers (most common cancers for persons: Model 1(c))

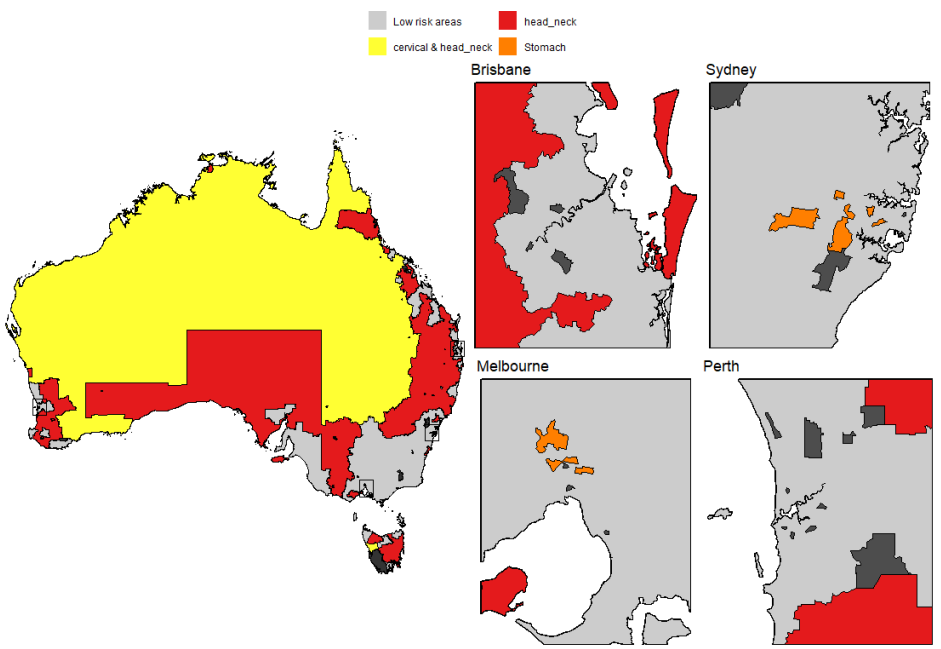

Figure 5 High risk areas for multiple cancers (less common/rare cancers for females: Model 6)

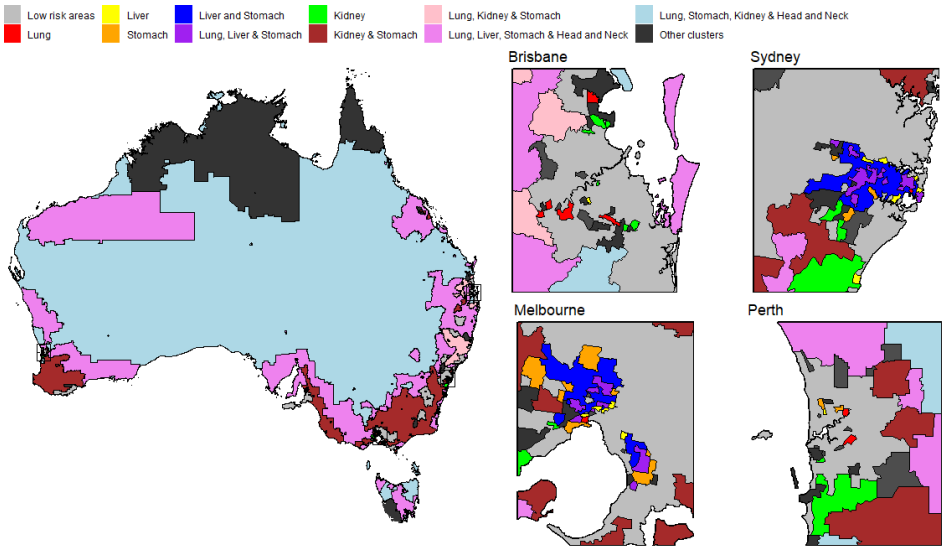

Figure 6 High risk areas for multiple cancers (smoking related cancers for males: Model 8(a))

Table 1 Posterior Correlation matrices with 95% credible intervals for most common cancers (Model 1) by region

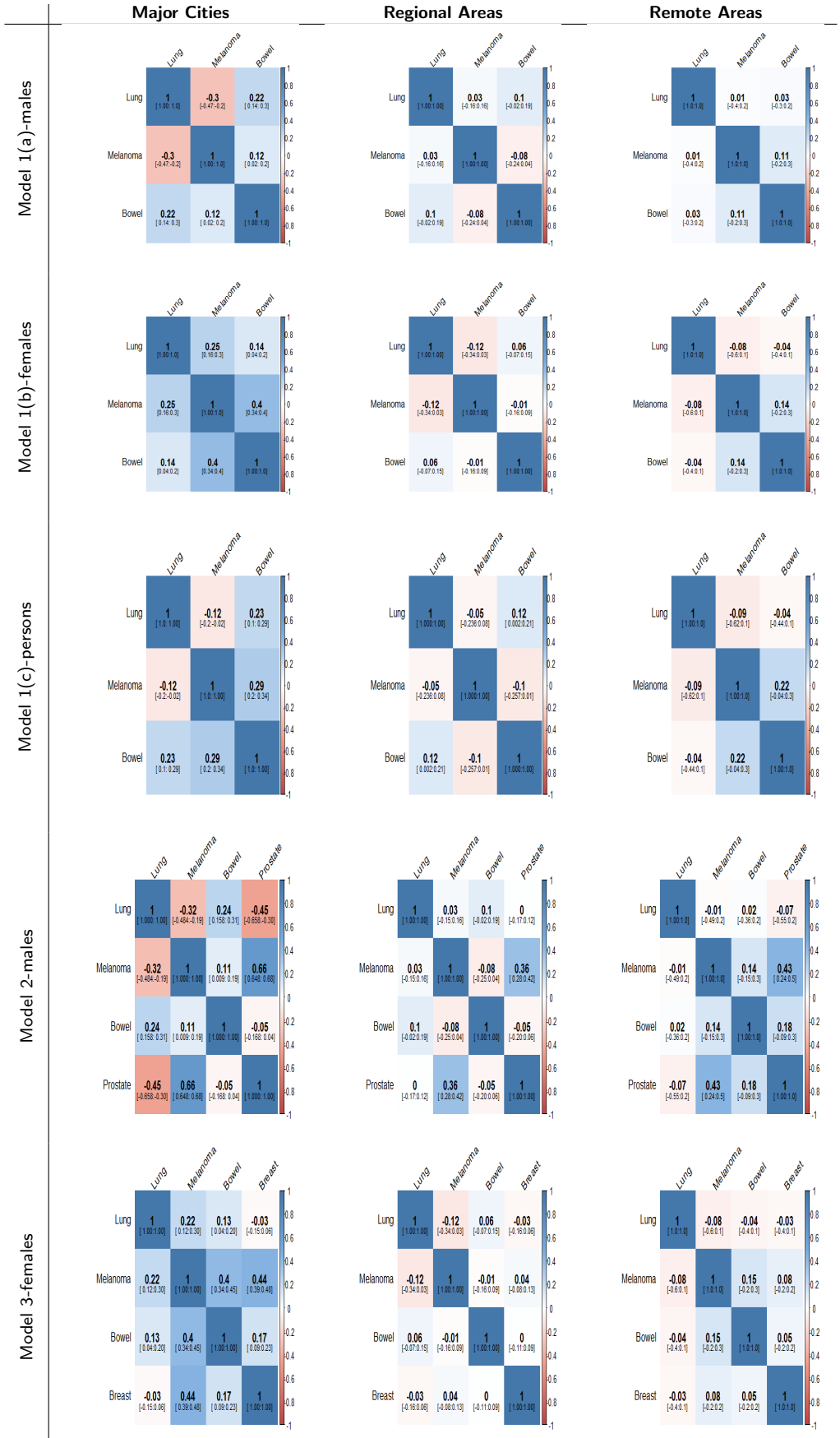

Table 2 Posterior Correlation matrices with 95% credible intervals for less common/rare cancers (Model 4,5,6, & 7) by region

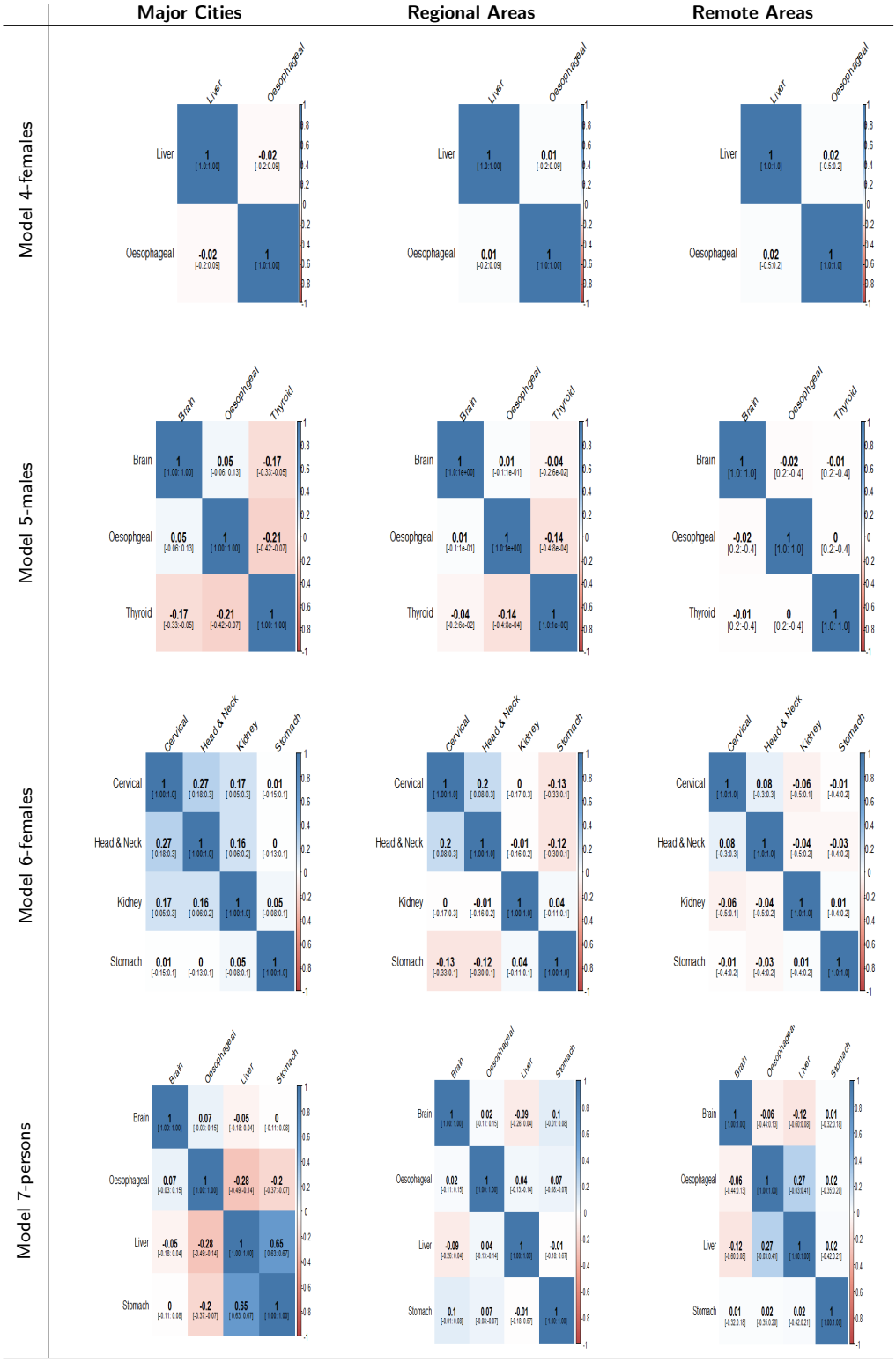

Table 3 Posterior Correlation matrices with 95% credible intervals for smoking related cancers (Model 8) by region

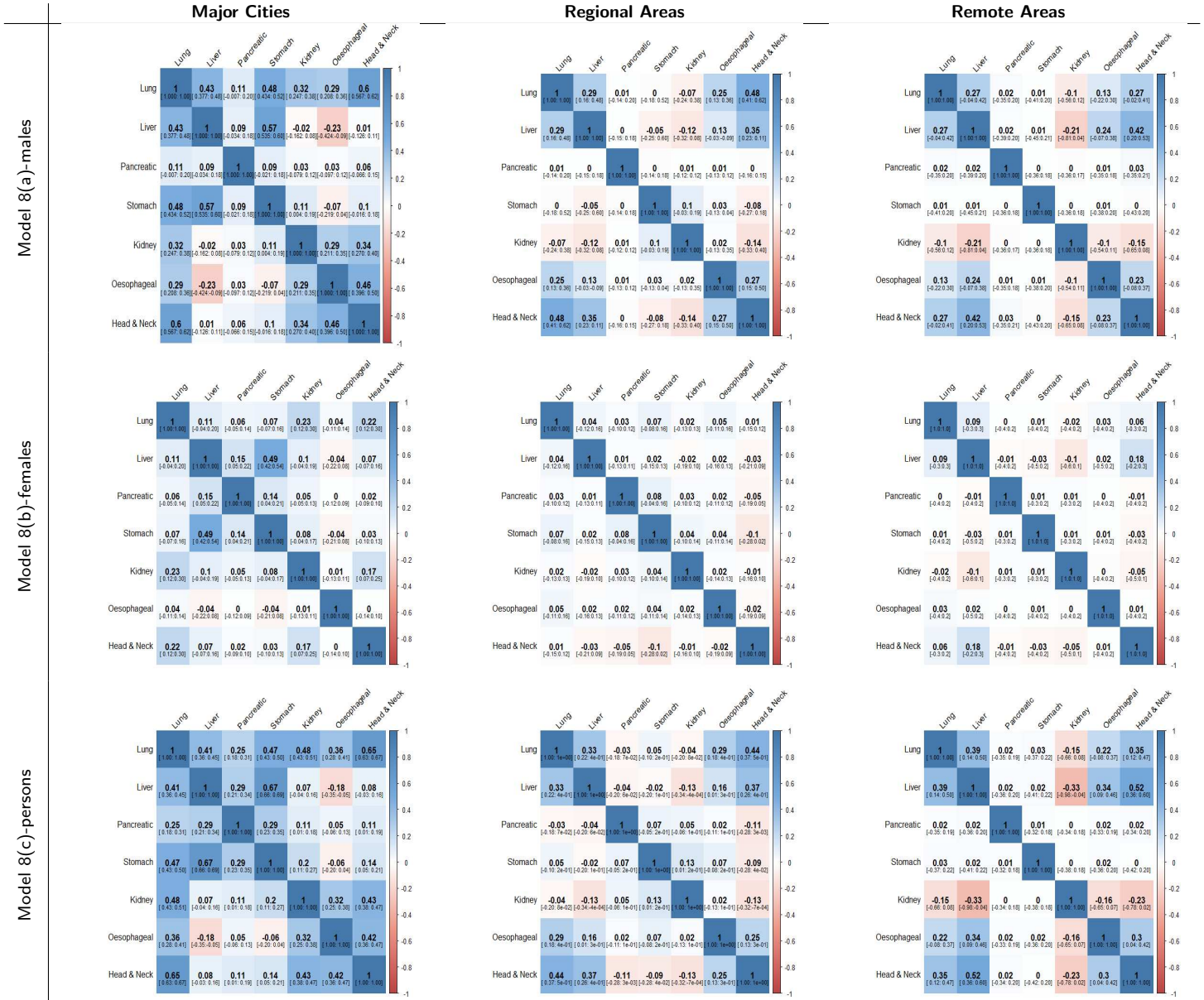

Table 4 Results of Jennrich's Test of differences in Correlation matrices applied to each group of cancers in different remoteness regions

Group	Model	Test Statistic [a]	P value
Most common cancers	1(a)	114.74	<0.0001
	1(b)	155.91	<0.0001
	1(c)	111.42	<0.0001
	2	282.87	<0.0001
	3	250.55	<0.0001
Less common and rare cancers	4	0.65	0.7225
	5	2.68	0.8481
	6	44.31	<0.0001
	7	384.79	<0.0001
Smoking related cancers	8(a)	767.74	<0.0001
	8(b)	226.35	<0.0001
	8(c)	1005.75	<0.0001

[a] Null Hypothesis: Equality of correlation matrices in major cities, regional and remote areas for each group of cancers are tested

Table 5 Number of SA2s with higher incidence for groups of cancers jointly and individually

Group	Cancer	No. of SA2s
Most Common Cancers (for persons): Model 1(c)	Lung only	143
	Melanoma only	442
	Lung & melanoma	22
	Bowel only	238
	Lung & bowel	123
	Melanoma & bowel	116
	Lung, melanoma & bowel	76
Less Common/ Rare Cancers (for females):Model 6	Head & neck only	378
	Cervical & Head and neck	79
	Stomach only	13

Table 6 Number of SA2s with higher incidence of Smoking related cancers (for males), Model 8(a) jointly and individually

Cancer	No. of SA2s
Low risk areas	990
Lung	21
Liver	24
Lung & Liver	1
Stomach	20
Lung & Stomach	11
Liver and Stomach	113
Lung, Liver & Stomach	51
Kidney	26
Lung & Kidney	17
Kidney & Stomach	2
Lung, Kidney & Stomach	5
Lung, Liver, Stomach & Kidney	10
Oesophageal	200
Lung and Oesophageal	1
Head & Neck	36
Lung & Head and Neck	7
Lung, Liver, Stomach & Head and Neck	1
Lung, Kidney & Head and Neck	13
Lung, Stomach, Kidney & Head and Neck	2
Lung, Liver, Stomach, Kidney & Head and Neck	1
Oesophageal & Head and Neck	294
Lung,Oesophageal & Head and Neck	284
Lung, Liver, Oesophageal & Head and Neck	18

Appendix

A.1 Model code in R

```
library(R2jags)
model<-"model{
for(n in 1: Nobs){
for(j in 1:Ncancer){
d[n,j]~dnorm(mu[n,j],tau.d[n,j])
tau.d[n,j]<- 1/sigma.d[n,j]
sigma.d[n,j]<-(2*se[n,j]^2)/a[n,j]
a[n,j] ~ dchisqr(2)}
mu[n,1:Ncancer]~dmnorm(theta.i[region[n],],tau.i[region[n],,])}
for(i in 1:Nregion){
theta.i[i,1:Ncancer]~dmnorm(mu.0[],tau.0[,])
tau.i[i,1:Ncancer,1:Ncancer] ~ dwish(Gamma.i[,],Ncancer)
}
}"
```

A.2 Cancers included in ACA

Table 7 Cancers included in Australian cancer atlas by sex

Cancer	ICD-10 codes	Sex
All Cancers	C00–C97, D45, D46, D47.1, D47.3–D47.5	Males, Females, Persons
Bowel Cancer	C18-C20	Males, Females, Persons
Brain Cancer	C71	Males, Females, Persons
Breast Cancer	C50	Females
Cervical Cancer	C53	Females
Head and Neck Cancer	C00–C14, C30–C32	Males, Females, Persons
Kidney Cancer	C64	Males, Females, Persons
Leukaemia	C91-C95	Males, Females, Persons
Liver Cancer	C22	Males, Females, Persons
Lung Cancer	C33-C34	Males, Females, Persons
Melanoma	C43	Males, Females, Persons
Myeloma	C90	Males, Females, Persons
Non-Hodgkin Lymphoma	C82-C86	Males, Females, Persons
Oesophageal Cancer	C15	Males, Females, Persons
Ovarian Cancer	C56	Females
Pancreatic Cancer	C25	Males, Females, Persons
Prostate Cancer	C61	Males
Stomach Cancer	C16	Males, Females, Persons
Thyroid Cancer	C73	Males, Females, Persons
Uterine Cancer	C54-C55	Females

A.3 More hierarchy in the multivariate meta-analysis model

The possible hierarchical modelling after equation (2) is shown below using the usual Bayesian hierarchical framework below. Although the parameters therein are not of interest for this particular case study, it can be appropriate for other types of data and/or research questions.

$$\mu_{i(k)} \sim MVN(\mu_i, \omega_i) \tag{7}$$

where μ_i is the overall mean of i^{th} cancer and ω_i be the variance covariance term accounting for variation between the means of different regions and same cancers. We are not interested in these parameters as we already performed univariate analysis to see the means and variation due to remoteness for each cancer separately [51].
The overall mean of i^{th} cancer, μ_i can then be modelled as:

$$\mu_i \sim N(\mu_0, \sigma_0^2) \tag{8}$$

where, μ_0 is the overall mean of all cancers in Australia, and σ_0^2 is the variance among the means of different cancers.

Additional Files

Additional file 1 — Supplementary material of Multivariate Bayesian meta-analysis: joint modelling of multiple cancers using summary measures

More results from the proposed model in form of tables and figures are provided in the supplementary material.

Figures

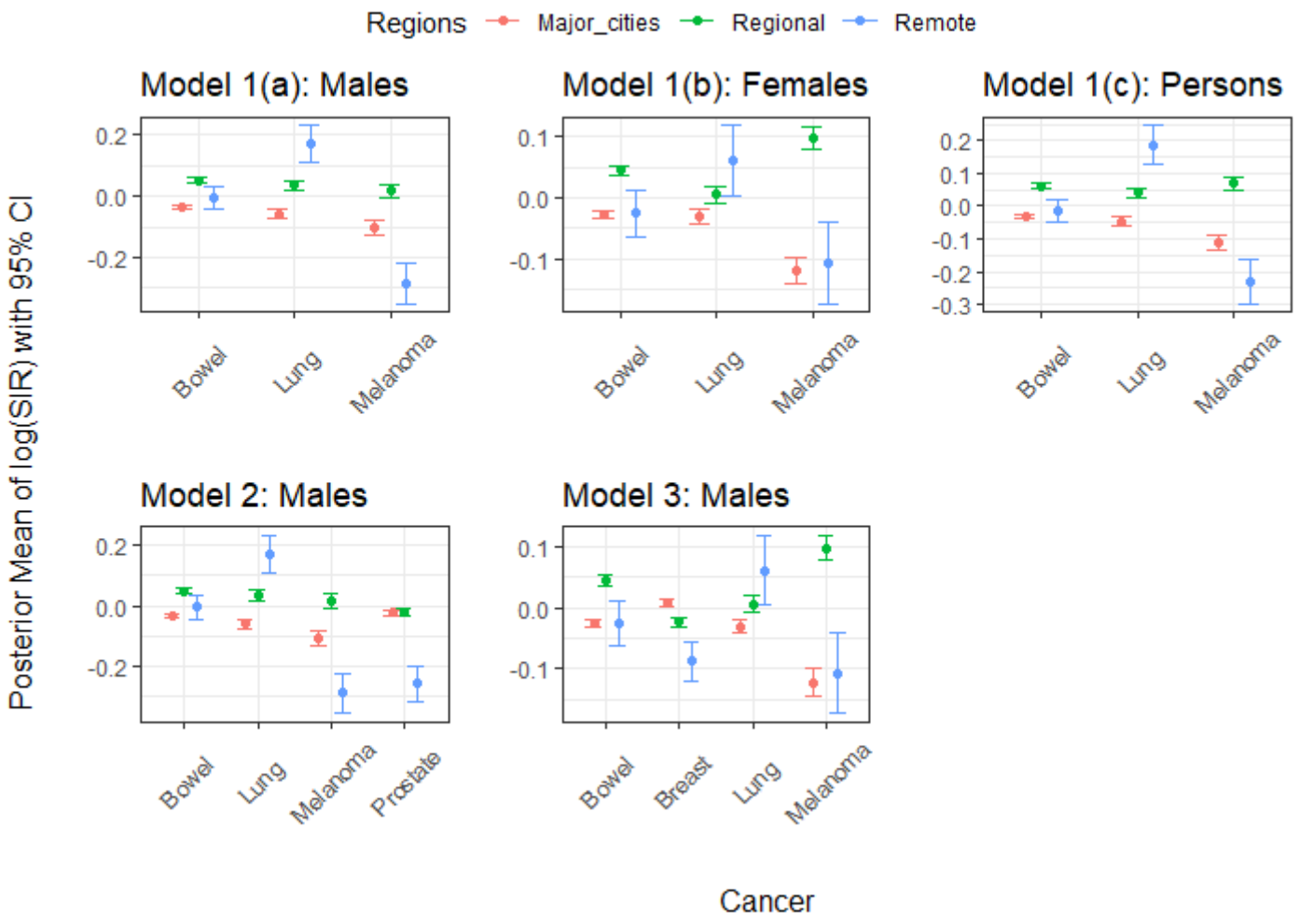

Figure 1

Posterior means and 95% credible intervals of log(SIR) by multivariate model for the most common cancers (Group 1) over remoteness regions, Australia

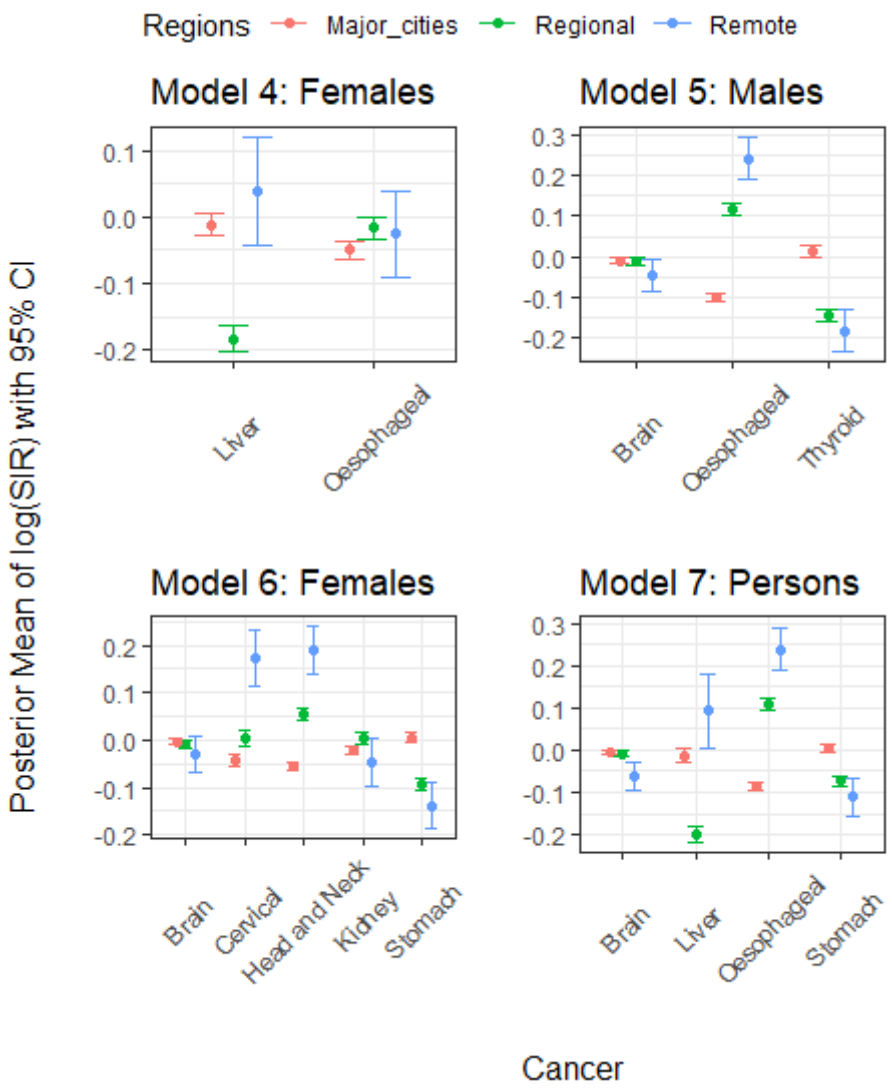

Figure 2

Posterior means and 95% credible intervals of log(SIR) by multivariate model for the less common cancers/rare cancers (Group 2) over remoteness regions, Australia

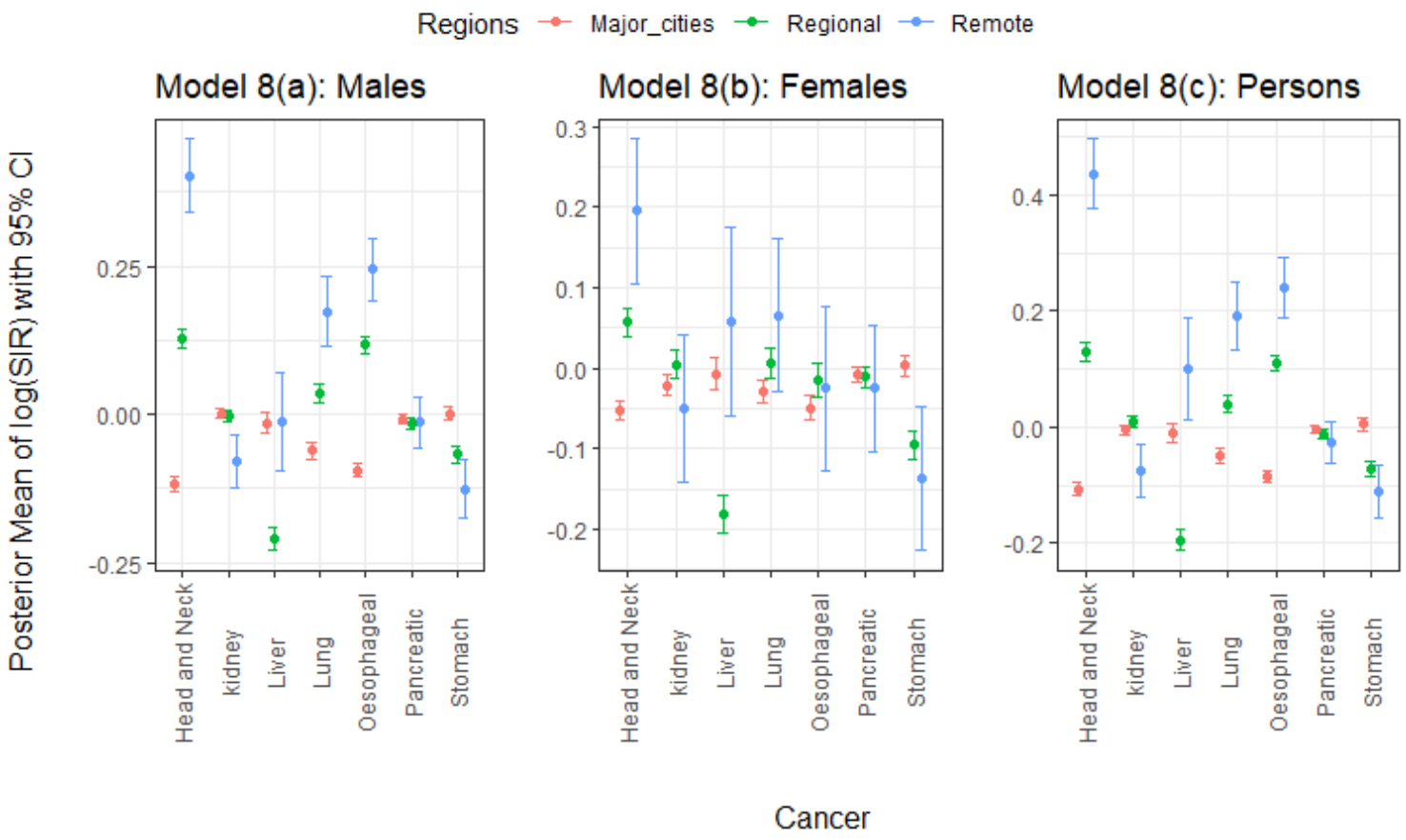

Figure 3

Posterior means and 95% credible intervals of log(SIR) by multivariate model for the smoking related cancers (Group 3) over remoteness regions, Australia

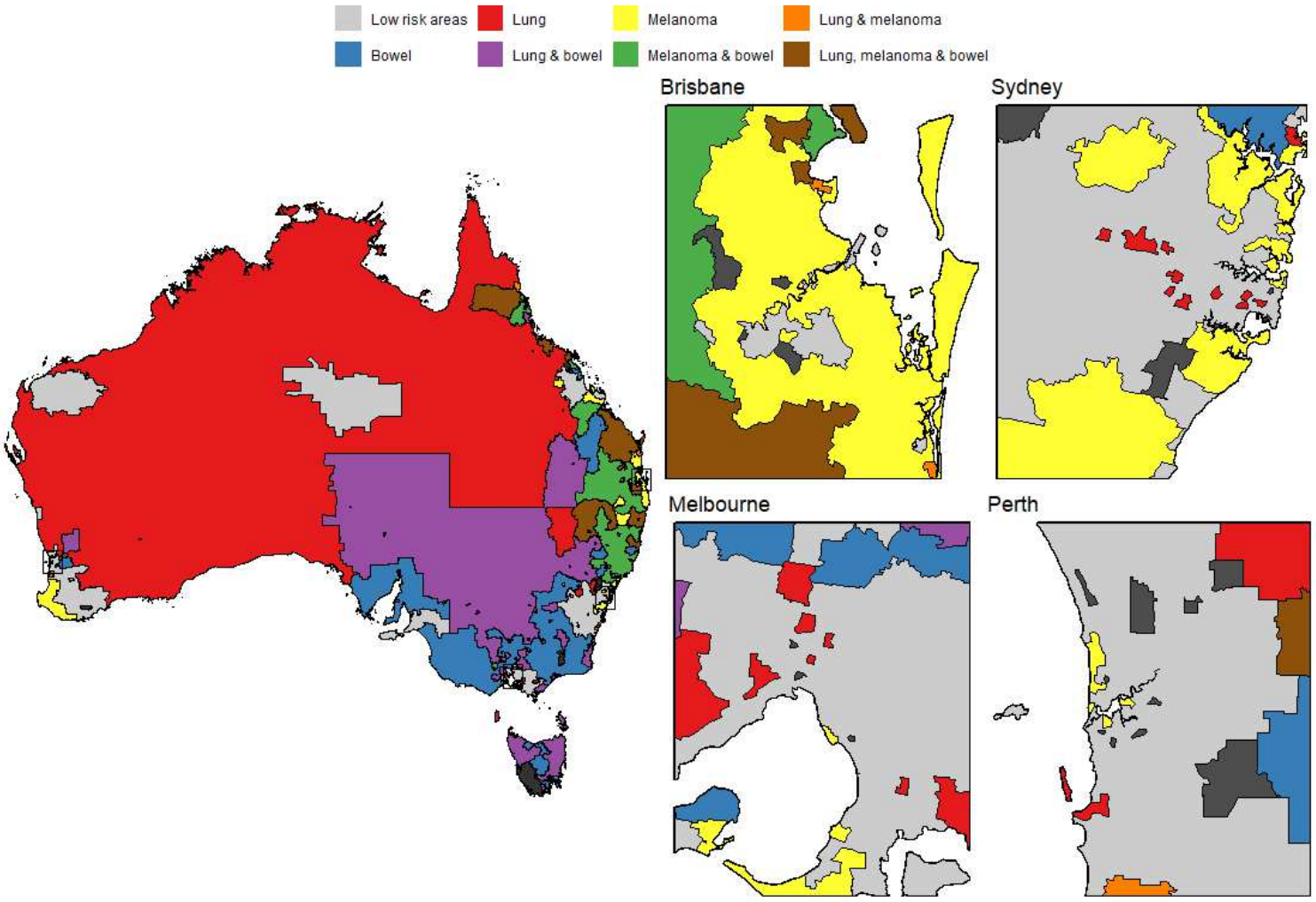

Figure 4

High risk areas for multiple cancers (most common cancers for persons: Model 1(c))

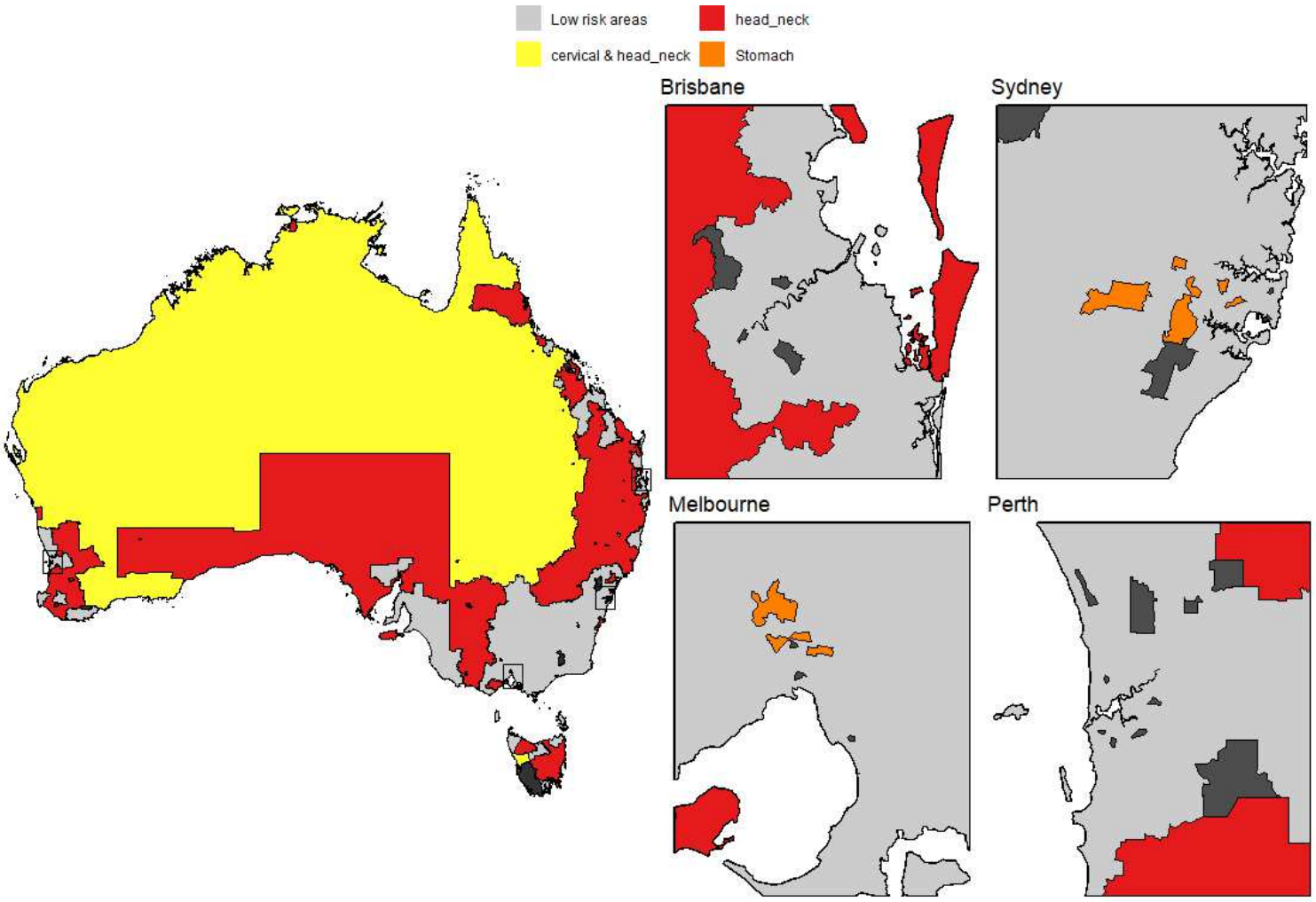

Figure 5

High risk areas for multiple cancers (less common/rare cancers for females: Model 6)

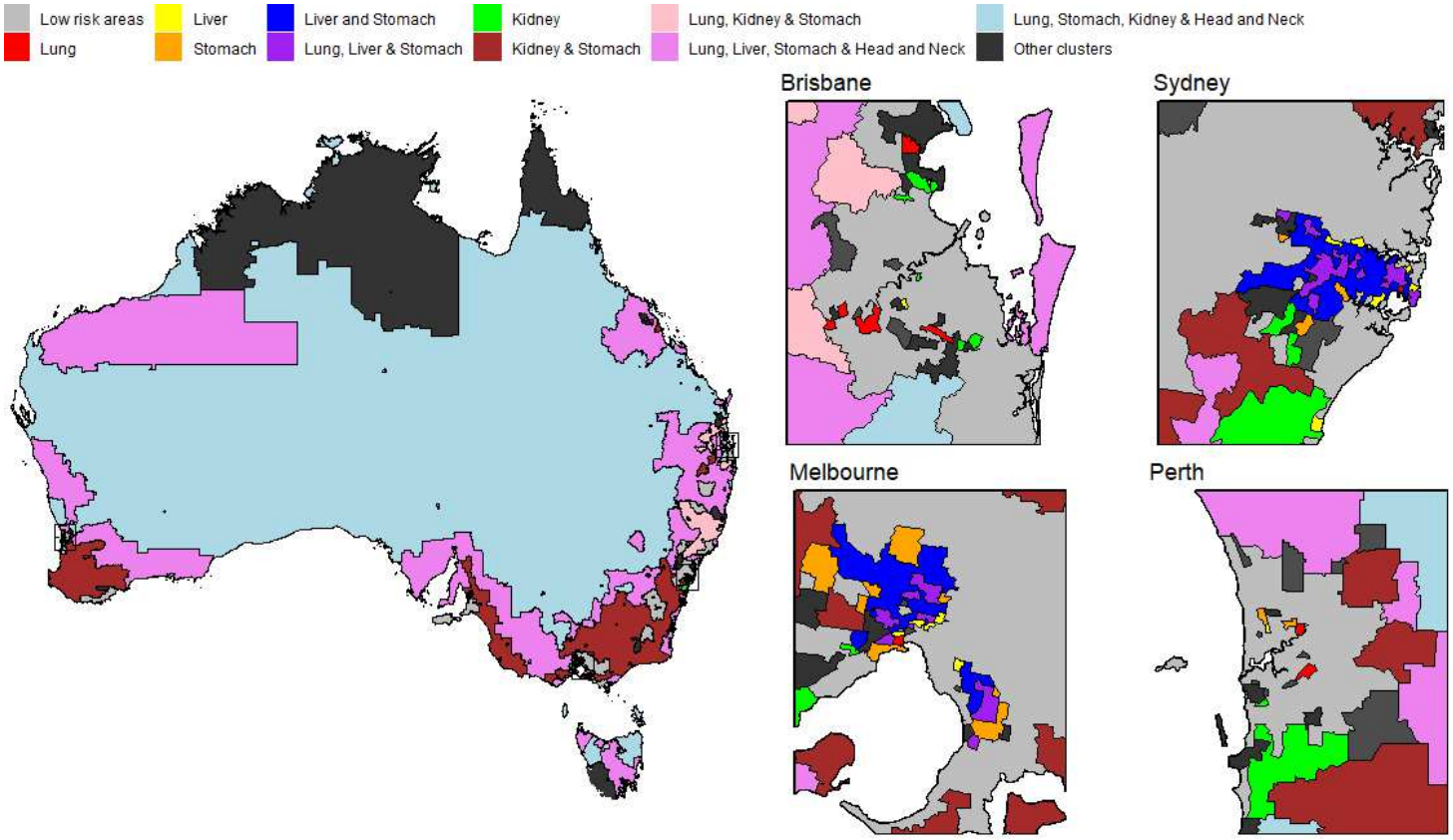

Figure 6

High risk areas for multiple cancers (smoking related cancers for males: Model 8(a))

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterialMultivariateMetaAnalysis.pdf