Title: Cancer incidence and soil arsenic exposure in an historic gold mining area in Victoria, Australia: a geospatial analysis

[Revised] Manuscript ID JESEE-11-1282

Authors:

Dora Claire Pearce[a,b], PhD

Kim Dowling[a], PhD

Malcolm Ross Sim[c], PhD

Authors' affiliations:

[a] University of Ballarat, Mt Helen, Victoria, Australia

[b] University of Melbourne, Melbourne, Victoria, Australia

[c] Monash University, Melbourne, Victoria, Australia

Corresponding author:

[b] Dr Dora C. Pearce

Melbourne School of Population Health, The University of Melbourne

Level 3, 207 Bouverie Street, Melbourne, VIC 3010, Australia

Phone: +61 3 9035 3343

Fax: +61 3 9349 5815

Email: dpearce@unimelb.edu.au

Running title: Cancer incidence and soil arsenic level

Abstract

Soil and mine waste around historical gold mining sites may have elevated arsenic concentrations. Recent evidence suggests some systemic arsenic absorption by residents in the goldfields region of Victoria, Australia. Victorian Cancer Registry and geochemical data were accessed for an ecological geographic correlation study, 1984-2003. Spatial empirical Bayes smoothing was applied when estimating standardised incidence ratios (SIRs) for cancers in 61 statistical local areas. The derived soil arsenic exposure metric ranged from 1.4 to 1 857 mg/kg. Spatial autoregressive modelling detected increases in smoothed SIRs for all-cancers of 0.05 (95% CI, 0.02-0.08) and 0.04 (0.01-0.07) per 2.7-fold increase in the natural log-transformed exposure metric for males and females, respectively, in more socioeconomically disadvantaged areas; for melanoma in males (0.05 [0.01-0.08] adjusted for disadvantage), and females (0.05 [0.02-0.09] in disadvantaged areas). Excess risks were estimated for all-cancers (relative risk 1.21 [95% CI, 1.15-1.27] and 1.08 [1.03-1.14]), and melanoma (1.52 [1.25-1.85] and 1.29 [1.08-1.55]), for males and females, respectively, in disadvantaged areas in the highest quintile of the exposure metric relative to the lowest. Our findings suggest small but significant increases in past cancer risk associated with increasing soil

arsenic in socioeconomically disadvantaged areas, and demonstrate the robustness of this geospatial approach.

Word count 200

Key words: cancer; soil; arsenic; exposure; geospatial

Introduction

Arsenic exposure through contaminated drinking water has been linked to cancers including bladder, colon, kidney, liver, lung, and skin (Chen C.J. et al, 1985; Tchounwou et al, 2003). Arsenic's mode of action as a multiorgan human carcinogen has not been fully elucidated (Hughes, 2009), although inhibition of DNA repair mechanisms may enable it to act as a cocarcinogen (Andrew et al, 2006). The soil environment has also been identified as a contributor to arsenic exposure (Hinwood et al, 2003; Wickre et al, 2004; Pearce et al, 2010). Typically associated with gold mineralisation, residual arsenic contamination from historical gold mining activity persists in the goldfields region of Victoria, Australia, where elevated arsenic concentrations have been observed in mine waste and some

residential soils, surface and ground waters (DMID, 1991; Hinwood et al, 1999; Smith E. et al, 2003; Pearce et al, 2010). Socioeconomic status may bias ecological studies, not only because of poorer health outcomes, but because lower socio-economic status increases the chance of living closer to sources of environmental pollution (Jolley et al, 1996; Carstairs, 2000). Further, exposure misclassification due to crude exposure metrics may undermine the likelihood of identifying a true relationship between an environmental hazard and disease outcomes (Sim, 2002; Elliott and Wartenberg, 2004).

In a previous aggregation of data from 22 geographical regions irregularly distributed across Victoria, each with evidence of elevated environmental arsenic concentrations or visible mine tailings, small but significantly increased risk of all-cancers, prostate and breast cancers, melanoma and chronic myeloid leukaemia were identified (Hinwood et al, 1999). However, potential limitations of this study may have arisen from a lack of adjustment for socioeconomic status, low statistical power due to sparse cancer data in some rural areas over the 10 year period 1982-1991, and a reliance on categorisation of environmental arsenic contamination to determine exposure classification in these mainly disconnected regions. Further investigation was therefore warranted to confirm associations observed with melanoma and leukaemia, not previously linked to arsenic exposure; prostate cancer, typically associated with arsenic-contaminated drinking water (Wu et al, 1989); and lack of expected associations with lung and bladder cancers.

We aimed to investigate associations between arsenic in soil and mine waste and *a priori* selected cancers in the goldfields region of Victoria at a refined spatial resolution using an improved exposure metric, taking into account socioeconomic disadvantage (SED). Further objectives were to: 1. assess the validity of this approach by comparing associations detected with those previously observed in

Victoria; and 2. evaluate the robustness of associations detected with derived exposure metrics using spatial autoregressive modelling.

Materials and methods

Study design

This small-area ecological study investigated the incidence of all-cancers, and *a priori* selected cancers, and their association with two derived soil arsenic exposure metrics. Underpinning this analysis was the availability of 20 years of Victorian Cancer Registry (VCR) data, 1984 to 2003, georeferenced to statistical local areas (SLAs), and geochemical data from the University of Ballarat and GeoScience Victoria comprising soil arsenic concentrations with geospatial coordinates of sampling sites, facilitating derivation of exposure metrics. We used the Geocentric Datum of Australia (GDA94) as the spatial coordinate system, map projection VICGRID94, and SLAs defined in 2001 by the Australian Standard Geographical Classification (Australian Bureau of Statistics [ABS], 2001a) as the areal unit of analysis.

Study site and population

Our study population comprised residents in 61 contiguous SLAs in Victoria, Australia (Figure 1): 36 core SLAs in the goldfields region, and 25 peripheral SLAs for comparative purposes. VCR data, including VCR tumour ID, gender, age in 5-year groupings, SLA of residence at time of diagnosis, and International Classification of Diseases, Tenth Revision (ICD-10) code for topography and site of tumour, was used to investigate all-cancers (C00-D47), and common cancers individually: colon cancer (C18), prostate (C61), and melanoma (C43), comprising 2% of skin cancers in Australia and included because non-melanocytic skin cancers are not reported (Thursfield and Giles, 2007;

Thursfield et al, 2008). Due to sparse populations in rural SLAs, it was impracticable to investigate rarer cancers, such as kidney, bladder, and chronic myeloid leukaemia, individually. We therefore investigated selected cancer groupings: leukaemia (C91-C95); cancers of the trachea, bronchus and lung (C33, C34); and urinary tract cancers (C64-C68).

VCR data were georeferenced to SLAs defined in 1991 until 2002, thereafter to 2001 SLAs. Although digitised boundaries were modified between censuses held at five year intervals, a time series profile was available for 1991, 1996, and 2001 census data georeferenced to 2001 SLAs (ABS, 2002a). Historical census data for 1986 SLAs and 1991 Census Districts (CDs) were also obtained (ABS, 1986, 1991).

Geographic data conversion

A weighting scheme to convert VCR data from 1991 to 2001 SLAs was derived by overlaying 2001 SLA digital boundaries (ABS, 2002a) with 1991 CD and SLA boundaries (ABS, 2005a, 2005b) and Vicmap Address, geocoded digital address data current in 2004, supplied by Spatial Information Infrastructure, Department of Sustainability and Environment, Victoria, as exemplified in Figure 2. Some misalignment between historical and current digital boundaries arises due to differing data sources (ABS, 2005b). Source 1991 CD populations were allocated to target 2001 SLAs based on the proportional distribution of address points within CD segments (Simpson, 2002), enabling estimation of population proportions of 1991 SLAs allocated to target 2001 SLAs. Population estimates for the 61 study SLAs were significantly ($P < 0.001$) correlated with ABS data (Spearman's rho 0.997 and 0.999), with mean (standard deviation [SD]) percentage differences of 0.25% (2.8%) and 0.25% (2.7%) for males and females, respectively.

This weighting scheme was subsequently used to convert VCR data from 1991 SLAs to 2001 SLAs. SLA boundaries from the 1986 census approximated 1991 SLAs within the study area (ABS, 2005a), enabling conversion of 1986 demographic data.

Cancer incidence rate estimation

Prior to conversion, VCR data were aggregated by gender and 1991 SLA for five-year periods centred on census years: 1984-1988, 1989-1993, 1994-1998, and 1999-2002. VCR data for 2003 were combined with converted data for 1999-2002. Observed cases were summed for 1984-2003 within each 2001 SLA. Cancer incidence rates in corresponding periods were estimated for Victoria in 5-year age-sex strata, except for two 10-year age-sex strata encompassing 30-39 and 40-49 year age bands used in the 1986 census. Using indirect standardisation (Bland, 1987), the expected number of cancer cases in each 2001 SLA was calculated by applying age-sex stratum-specific rates to individual SLA population structures and summing to obtain the total expected cases for 1984-2003. Standardised incidence ratios (SIRs) were calculated as $SIR_i = O_i/E_i$, where O_i is the observed and E_i is the expected, number of cases for SLA_i as an estimate of relative risk (RR) (Elliott and Wartenberg, 2004). Approximate 95% Poisson confidence intervals (CI) were estimated for SIRs according to Bland (1987) when observed cases exceeded 100, and Ury and Wiggins (1985; cited in Dean et al, 2008), if less than 100.

Soil Arsenic Exposure Metrics

Geochemical data provided by the University of Ballarat comprised soil sampling undertaken during systematic background surveys, and at sites of historical mining activity, interest or concern (n = 270) between 2000 and 2006. Single sites were sampled in four of 20 study SLAs investigated. Soil arsenic concentrations, determined by inductively coupled plasma mass spectroscopy (ICP-MS) after aqua

regia digest, or instrumental neutron activation analysis (INAA), ranged from 0.9 mg/kg to 16 600 mg/kg (geometric mean 50 mg/kg). GeoScience Victoria, Department of Primary Industries, supplied the Victorian Geoscientific Data Package DVD current in 2006, which included arsenic concentrations determined from the 1960's to the 1990's using various assay techniques, including X-ray fluorescence spectroscopy, hydride generation atomic absorption spectroscopy, and acid digest followed ICP-MS, from which additional geochemical data for 49 study SLAs (n = 68 497) were obtained. Sometimes collected using arsenic as a pathfinder for gold to facilitate mineral exploration (Sahoo and Pandalai, 2000), sampling sites were typically highly clustered and irregular. Arsenic concentrations ranged from 0.1 mg/kg to 18 900 mg/kg (geometric mean 12.5 mg/kg). Overall, data were available for 51/61 (84%) study SLAs.

Kriging, the optimal geostatistical technique for spatial interpolation at unsampled sites (Waller and Gotway, 2004), is constrained by excessively large and clustered data sets (Briggs, 1992). We calculated the geometric means of soil arsenic concentrations within each SLA, and assigned these values to geospatial coordinates of SLA centroids. Where two data sources were available, the highest value was assigned to reflect the worst case scenario. Soil arsenic values were highly skewed, ranging from 1.4 mg/kg to 1 857 mg/kg (median 12.9 mg/kg), the latter from a single sample of mine tailings.

Exclusion of three extreme values facilitated estimation of the empirical variogram (Snowden et al, 1994). An exponential variogram function was subsequently used in ordinary kriging to predict soil arsenic values at all SLA centroids, using the natural logarithmic transform of all available (n = 51) data points (Snowden et al, 1994; Kaluzny et al, 1998). Reliability of predictions was assessed by calculating the ratio of geometric means of predicted (2.81, n = 61) to natural log-transformed data

(2.59, n = 51) values (Leem et al, 2006), a ratio of 1.09 indicating reliability. Predicted values were corrected for bias when back-transformed (Waller and Gotway, 2004).

The quality of exposure metrics depends on the representativeness of input data and validity of the interpolation method (Elliott and Wartenberg, 2004). Because of the heterogeneous nature of soil contamination, two exposure metrics were derived to enable an assessment of exposure misclassification. Exposure metric 1 comprised soil arsenic values where available and predicted values in unsampled SLAs. Exposure metric 2 comprised back-transformed predicted values only.

Socioeconomic disadvantage

The Index of Relative Socioeconomic Disadvantage (IRSD) derived from the 2001 census, with lower scores indicating greater social and economic hardship (ABS, 2001b, 2002b), was used to adjust for SED. Since smoking prevalence is higher among socioeconomically disadvantaged groups (Siahpush, 2003; Quit Victoria, n.d.), this index acted as a proxy for smoking prevalence.

Statistical methods

The areal unit of analysis was SLA. To reduce variance in raw SIRs, we applied spatial empirical Bayes (SEB) smoothing, in which local estimators for priors for the mean and variance of SIRs are determined for each 'local' neighbourhood set for *i*, and including *i* (Anselin et al, 2006). Neighbourhood structure was defined by adjacency of SLA digital boundaries, resulting in a symmetric spatial connectivity matrix (Anselin, 2005). We estimated associations between raw or SEB smoothed SIRs of all-cancers for males and females and natural logarithmic transformations of exposure metrics 1 or 2, with spatial dependence between neighbouring SLAs modelled by fitting a

conditional autoregressive (CAR) model (Haining, 1990; Richardson and Monfort, 2000, Waller and Gotway, 2004). Simultaneous autoregressive (SAR) models were also used to estimate the associations between raw SIRs of all-cancers and exposure metric 1. Associations between exposure metric 1 and SEB smoothed SIRs of individual cancers were investigated using CAR models.

CAR models incorporated a symmetric spatial connectivity matrix defined by adjacency, in which neighbouring SLAs had a weighting of 1, zero otherwise, and SAR models incorporated an asymmetric matrix defined by 4-nearest neighbours, with neighbour weights the inverse of the distance between centroids, standardised to be between 0 and 1 (Cressie, 1993; cited in Kaluzny et al, 1998).

The IRSD score was dichotomised at the median value (1013.39) for Victorian SLAs (ABS, 2001b). More disadvantaged SLAs, with IRSD scores less than 1013.39, were coded SED = 1, and less disadvantaged, SED = 0. SED was represented as a factor with treatment contrasts. Effect modification of exposure metrics 1 and 2 by SED level was investigated by including their product terms (Greenland, 2009). We assessed the significance of the spatial parameter, ρ, and comparative fit of consecutive spatial autoregressive models, using likelihood ratio tests; with individual model fit evaluated by normality of residuals and plots of residuals and observed data against fitted (spatial and linear components) values (Haining, 1990; cited in Kaluzny et al, 1998).

We investigated large scale spatial trend in SIRs using linear and quadratic trend surface models with polynomials of X and Y location coordinates (expressed in kilometres) as explanatory variables (Kaluzny et al, 1998; Anselin, 2005). When trends were detected, polynomials of coordinates were included in regression models. If spatial autoregressive models were inestimable, model selection

was based on the Multiple R-Squared values obtained from ordinary least squares (OLS) trend surface models. Positive spatial autocorrelation, indicative of clustering of areas with similar characteristics, was assessed by Moran's I statistic (Cliff and Ord, 1981).

For comparative purposes, we stratified SLAs by quintiles of exposure metric 1 (Supplementary Table 1) and estimated aggregate SIRs by gender for more socioeconomically disadvantaged SLAs only, the two less disadvantaged SLAs in the uppermost quintile considered to be an inappropriately small and potentially biased comparator group, particularly as one was unsampled. According to the method of Breslow and Day (1987), we sought evidence of a dose-response across quintiles assuming an approximate Poisson distribution, calculating relative risk as the ratio of the SIR in the uppermost quintile relative to the lowest quintile (Supplementary Tables 2a and 2b for males and females, respectively).

The GIS used for spatial data manipulation, linkage and display in choropleth maps was MapInfo Professional for Windows Version 8.0 (MapInfo Corporation, 2005). GeoDa 0.9.5-i Beta (Anselin et al, 2004) was used to estimate SEB smoothed SIRs. Standard statistical analyses were conducted using SPSS 15.0 for Windows (SPSS Inc., 2006) and S-PLUS 6.2 for Windows Professional Edition (Insightful Corporation, 2003), and spatial statistical analyses were conducted using the module S+SPATIALSTATS Version 1.5.6 for Windows (Insightful Corporation, 2003).

Results

Study SLAs ranged in area from 12.6 to 5 451 square kilometres (mean 1 479 square kilometres). Smaller SLAs were typically more densely populated. During the study period, 26 851 male and 22

822 female cancers were diagnosed. Leukaemia had the lowest overall incidence, with 836 male and 628 female cases, with between 0 and 64 male, and 0 and 60 female, cases observed in individual SLAs.

IRSD scores ranged from 904-1083, with 26/36 (72%) of core SLAs compared with 10/25 (40%) peripheral SLAs (Fisher's exact *P* = 0.017) more socioeconomically disadvantaged. The majority of SLAs in the uppermost quintile of exposure metric 1 (10/12 [83%]) were more disadvantaged, of which 7 (70%) were core SLAs. Overall, exposure metric 1 was weakly and negatively correlated with IRSD score (Spearman's rho = -0.11 [*P* = 0.401]), although positively correlated in less disadvantaged areas (0.16 [0.434]).

Conversion of 1986 demographic data to 2001 SLAs resulted in larger population increases for 1986-1991 than other between-census periods. Of 13 SLAs with an estimated increase of >20%, 7 (54%) were core SLAs, 5 (39%) were more disadvantaged, and all had soil arsenic values below 30 mg/kg. SIRs for all-cancers did not differ significantly between these and remaining SLAs (median 94.8 versus 97.1, *P* = 0.492 for males; 93.5 versus 93.9, *P* = 0.291 for females). Of individual cancers investigated, no significant differences were detected for males. For females, the median smoothed SIR for colon cancer was lower in these SLAs (0.95 versus 1.04, *P* = 0.043). Hence, it is unlikely that underestimation of the 1986 population biased our results.

Soil arsenic values for exposure metric 1 ranged from 1.4 to1 857 mg/kg (median 15.0 mg/kg) and for exposure metric 2, from 12.0 to 299 mg/kg (median 32 mg/kg). Interpolation contributed to spatial trend in exposure metric 2 (Figure 3), and spatial autocorrelation increased over that detected in exposure metric 1 using the adjacency-based connectivity matrix (Moran's I statistic 0.47

[permutation *P*-value < 0.001] versus 0.16 [0.019], respectively. Despite this, moderate agreement between SLAs classified by quintiles of the two exposure metrics (kappa = 0.43, *P* < 0.001) was achieved. Of the 10 unsampled SLAs, nine were peripheral, and elevation of predicted values may be due to edge effects in the interpolation procedure (Waller and Gotway 2004).

For individual SLAs, SIRs for all-cancers ranged from 0.33 [95% CI, 0.26-0.42] to 1.39 (95% CI, 1.32-1.47) for males and 0.31 (95% CI, 0.18-0.48) to 1.26 (95% CI, 1.19-1.33) for females. While median values for raw and SEB smoothed SIRs for males and females were uniformly 0.97, variance was greater in raw SIRs (SD 0.20 versus 0.17 for males; 0.19 versus 0.15 for females). No large scale spatial trends or autocorrelation were detected in raw or SEB smoothed SIRs for all-cancers.

Spatial autoregressive models: all-cancers

Consistently significant product terms between SED category and exposure metrics were obtained in spatial autoregressive models for all-cancer in males. Although less consistent for females, likelihood ratio tests generally indicated that inclusion of the statistical interaction term improved model fit. Hence results are reported for soil arsenic exposure metrics nested within levels of SED (Table 1).The magnitude of regression coefficients obtained for exposure metric 1 was comparable between models with varying specifications, while larger for males than females, and for exposure metric 2. The spatial parameter, ρ, failed to achieve statistical significance, except in Model 4 for females.

Due to the consistency of these results, and since SEB smoothing reduces the inherent variance instability in rates estimated from sparse data while allowing for possible case assignment errors,

and exposure metric 1 provided a more conservative effect estimate, the following analyses for individual cancers are reported for models with SEB smoothed SIRs as dependent variables, and exposure metric 1 and SED as explanatory variables in CAR models (Table 2).

Spatial regression analysis: individual cancers

Positive spatial autocorrelation was detected in SEB smoothed SIRs of all individual cancers investigated (Moran's I statistics > 0.15, permutation *P* values < 0.035), as expected due to the smoothing procedure (Waller and Gotway, 2004). Evidence of spatial trend was detected in male colon and prostate cancers, female leukaemia and lung cancer, and male and female melanoma, with more complex trends in female colon cancer, male leukaemia, and female cancers of the urinary tract. Although polynomials of X and Y coordinates were included to model these trends (Table 2), a component of unexplained spatial variation remained in CAR models, as indicated by the significance of the spatial parameter, ρ.

Soil arsenic level was associated with male colon cancer, adjusted for SED and an increasing southerly trend and leukaemia in socioeconomically disadvantaged areas adjusted for an easterly trend. Melanoma in males was associated with soil arsenic level when adjusted for SED and an easterly spatial trend, whereas female melanoma increased in disadvantaged SLAs when adjusted for an increasing northerly trend, as did male prostate cancer. Weak evidence of an association with urinary tract cancer was detected in females when adjusted for spatial trend and SED.

Non-spatial analyses

Excess risks for all-cancers in males (RR 1.21 [95% CI, 1.15-1.27]) and females (1.08 [1.03-1.14]) were observed when comparing SIRs in the uppermost soil arsenic quintile to the lowest in more disadvantaged areas, where risks of male leukaemia (1.55 [1.15-2.14]), melanoma (1.52 [1.25-1.85]), colon (1.18 [1.01-1.38]) and prostate (1.23 [1.11-1.37]) cancers, and female melanoma (1.29 [1.08-1.55]) and colon (1.21 [1.02-1.44]) cancers, were increased (Figure 4; Supplementary Tables 2a and 2b). Increasing trends across quintiles of soil arsenic were also detected.

Discussion

At an ecological level, this study found a small but significant increase in cancer risk associated with soil arsenic level in more socioeconomically disadvantaged areas in the goldfields region of Victoria, Australia. The robustness of the associations detected between soil arsenic level and all-cancers in males and females is established by the consistency of findings using diverse methodologies: spatial autoregressive modelling with various model specifications at individual SLA level, and estimation of RRs for SLAs aggregated across quintiles of soil arsenic level in socioeconomically disadvantaged areas. The mix of cancers may have contributed to the slightly weaker associations detected for all female cancers, as the estimated spatial autoregressive models were not highly sensitive to neighbourhood structure definition or covariance model specification. For males, consistent results were also obtained for colon and prostate cancers, and leukaemia and melanoma, while for females, melanoma was most consistently associated with soil arsenic. Heterogeneity in individual susceptibilities and exposures within regions (Elliott and Wartenberg, 2004), possibly due in part to gender-specific lifestyle factors, or exposure misclassification, may have contributed to the weaker and less consistent findings for females.

Our finding of an association between soil arsenic and melanoma confirms that observed by Hinwood et al (1999). Although the most common arsenic-induced skin cancers are squamous and basal cell carcinomas (Centeno et al, 2002), the aetiology of these non-melanocytic skin cancers and cutaneous melanoma share an involvement with sun exposure (Elwood and Jopson, 1997), and the incidence of melanoma is much higher in fair-skinned ethnic groups (Sneyd and Cox, 2009). Evidence that arsenic may act as a cocarcinogen with ultraviolet radiation in causing squamous cell carcinomas (Rossman et al, 2002), and a suggestion that a reduced capacity for repair of UV light-induced DNA damage may increase the risk of cutaneous malignant melanoma, the most serious skin cancer (Wei et al, 2003), also lend support to our finding of an association between soil arsenic and melanoma. Further, Yorifuji et al (2010) found that arsenic exposure via contaminated milk powder in infancy was associated with increased risk of mortality from skin (including melanoma), liver, and pancreatic cancers, and leukaemia.

The association detected with prostate cancer also confirms previous research in Victoria (Hinwood et al, 1999). Arsenic has been shown to increase prostate cancer progression by inducing androgen-independence in human prostate cells (Benbrahim-Tallaa et al, 2005), and ingestion of arsenic contaminated water increased risk of mortality from prostate cancer (Lewis et al, 1999). The link between male leukaemia and soil arsenic found in this study is consistent with the association previously observed between environmental arsenic and chronic myeloid leukaemia (Hinwood et al, 1999), warranting further research at an individual level into possible biological mechanisms. Surprisingly, no evidence of a strong association between soil arsenic and lung cancer was obtained in this study or in earlier research in this locality (Hinwood et al, 1999), since lung cancer is typically associated with lower socio-economic status (Williams et al, 1991; Smith D. et al, 1996), and a synergistic effect between smoking and arsenic exposure has been observed (Chen C.L. et al, 2004). The weak association detected with female urinary tract cancers may be spurious due to small case

numbers, although Hinwood et al (1999) found an elevated non-significant risk of kidney cancer in arsenic-contaminated parts of Victoria.

Arsenic-contaminated drinking water is unlikely to have been a major contributor to arsenic exposures in this study. Most Victorian households currently depend on a reticulated water supply, with self-extraction from rainwater tanks, or from surface or ground waters, reportedly less than 4% in 2004-05 (ABS, 2006). Past consumption of contaminated water was possibly greater, but a previous individual level study revealed that residents with soil arsenic concentrations of up to 9 900 mg/kg consumed water with arsenic concentrations ranging from below the detection limit to 1.3 µg/L (Hinwood et al, 2004).

Evidence of unexplained spatially-varying factors in incidence of all-cancers was detected only in the CAR model of SEB smoothed SIRs for females in association with exposure metric 1, possibly reflecting confounding by social and environmental factors not included in the analysis (Elliott and Wakefield, 2000), or an artefact of the smoothing procedure. We applied SEB smoothing to individual cancers to reduce the inherent variability in SIRs (Waller and Gotway, 2004), having greatest effect on the least-stable estimates in sparsely populated areas (Elliott and Wartenberg, 2004), but possibly contributing to unexplained small-scale neighbourhood effects observed in most spatial autoregressive models.

Statistical interactions cannot be assumed to imply causal biological mechanisms or synergistic effects (VanderWeele, 2009). However, socioeconomic factors influence spatial variation in disease risk (Elliott and Wakefield, 2000; Richardson and Monfort, 2000; Bentley et al, 2008), and are typically associated with poor nutrition (James et al, 1997). Susceptibility to arsenic-induced toxicity

is increased by poor nutrition (Vahter, 2007), hence our findings may be suggestive of a causal mechanism. Suggestive of a synergistic effect (Greenland, 2009), cancer incidence was shown to be exacerbated by exposure to arsenic in soil in more socioeconomically disadvantaged SLAs. The weak negative association between soil arsenic and cancer incidence in less disadvantaged SLAs is most likely spurious, possibly due to the slightly increasing trend in socioeconomic status with arsenic level in this relatively small number of SLAs. Inflation of exposure metric values in some peripheral SLAs due to edge effects in the interpolation procedure may have exaggerated the strength of positive associations with cancer incidence in more socioeconomically disadvantaged areas and negative associations in less disadvantaged areas. However, analysis of core SLAs alone yielded reasonably consistent results (data not shown), in addition detecting significant associations with female leukaemia and colon cancer and male urinary tract cancer, although the significance of associations with female melanoma and male leukaemia was lost. It is possible that confounding by socioeconomic status biased our findings, since the majority of SLAs with higher soil arsenic values are more disadvantaged, but reduced incidence of cancers of the colon, prostate, leukaemia and melanoma may be expected in these areas (Williams et al, 1991; Smith D. et al, 1996).

Inferences from group-level cannot be assumed, explicitly or implicitly, to apply at an individual-level (Richardson and Monfort, 2000). While the spatial resolution of environmental data must correspond to health data, potentially obscuring exposure variation between areal units (Nuckols et al, 2004), exposures vary less within smaller areas, thus making grouped results more applicable to individuals (English, 1996). The geographical units selected must provide sufficient spatial resolution and rate stability while protecting confidentiality (Pickle, 2000). Data aggregation may therefore introduce ecological bias, but where distribution patterns of individuals within areal units reflect exposure levels, environmental effects occurring at an individual level may be detectable at an aggregate level (Portnov et al, 2007). Indicative of sufficient sensitivity and specificity to correctly

classify relative exposures at an aggregate level (Nuckols et al, 2004), exposure metric 1 enabled detection of relationships with geospatial variation in cancer incidence, whereas the induced trend and spatial autocorrelation in exposure metric 2 potentially increased exposure misclassification.

Ascertainment of arsenic uptake from soil and mine waste to determine individual-level exposures and cancer risk is needed to corroborate our findings, but was beyond the scope of this study. However, housing and agricultural developments in many rural communities in the Victorian goldfields region are in close proximity to sites of historical gold mining activity. Elevated soil arsenic concentrations in the geochemical databases used to derive exposure metrics likely reflect 'hot spots', indicative of the heterogeneity of soil arsenic concentration in this region, and typically associated with gold mineralisation or pesticide use (Smith E. et al, 2003). While the lowest quintile of exposure metric 1 corresponds to background soil arsenic concentrations in Australia (mean 3.9 mg/kg, range 1-8 mg/kg; Tiller, 1992), some residential soil arsenic concentrations are known to exceed the 100 mg/kg guideline (enHealth Council, 2001) in old gold mining areas (Hinwood et al, 2004; Pearce et al, 2010).

Children may be particularly vulnerable to, and at increased risk of, environmental exposures (Graeter and Mortensen, 1996), evidenced by an increased risk of kidney cancer mortality when exposed to arsenic in utero and during early childhood above that of adult exposures (Yuan et al, 2010). Ongoing arsenic uptake from soil by children resident in the study region was recently confirmed using synchrotron-based X-ray fluorescence (XRF) and X-ray absorption near-edge structure (XANES) spectroscopy applied to thin sections of children's toenail clippings (Pearce et al, 2010). XRF mapping suggested periodic arsenic exposures, and XANES spectroscopy detected two distinct arsenic species: a trivalent arsenic species, possibly methylated and bound to sulphur, and a

pentavalent arsenic species, consistent with some excretion of arsenic metabolites into the nail matrix as well as some direct exogenous absorption. A plausible geophysical route (Nuckols et al, 2004) clearly exists for the uptake of arsenic from soil in childhood.

The main limitations of our study were the potential for exposure misclassification, and our inability to adjust for individual-level confounding factors due to the inherent characteristics of ecological studies. A further limitation was our inability to account for migration in and out of individual SLAs. However, the study area comprised around half of regional Victoria, where growth exceeded that of the major urban area during much of the 1970s and 1980s, in part due to higher fertility rates and net gain from migration (Deacon, 2000). Given the long latency period for arsenic-induced cancers (Marshall et al, 2007), this likely corresponded to exposures linked to the development of cancers diagnosed during the study period. Any outward migration from socioeconomically disadvantaged areas with elevated soil arsenic levels would likely dilute cancer incidence rates. Error in SIRs, possibly due to mismeasurement of observed and expected cancer cases in SLAs, may have increased inferential uncertainty but is unlikely to have systematically altered the estimated relationships between cancer incidence and soil arsenic level, whereas mismeasurement in soil arsenic level may have attenuated the effect estimate (Gustafson, 2004).

Our findings indicate that increasing soil arsenic level was associated with a small but significant increase in past cancer risk in more socioeconomically disadvantaged areas. Although we were unable to take into account individual-level risk factors, and associations detected at an aggregate level may not apply at individual level (Elliott and Wartenberg, 2004), our ecological study has identified areas warranting further investigation. A case–control study to precisely ascertain exposures and cancer risk at individual-level in these areas is needed to corroborate our findings.

However, the consistency of the associations identified in this study using diverse methodologies, taken together with established links between various cancers and arsenic exposure, validate the geospatial approach taken to derive an exposure metric from geochemical data, aggregate cancer cases across time and changed areal boundaries, and estimate the associations between them. Recent evidence of ongoing arsenic uptake from soil during childhood in the study area suggests that these findings may represent an important public health risk, warranting close monitoring of exposures to arsenic present in soil and mining waste in the goldfields region of Victoria, Australia, and globally, to ensure that exposures are below currently acceptable levels.

Word count: 4987

Ethics approval

Ethics approval for this research was granted by the Human Research Ethics Committees of the Cancer Council Victoria and the University of Ballarat.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This research was supported by a University of Ballarat Postgraduate Award. We thank the following organizations for provision of data used in this analysis: Cancer Council Victoria for Victorian Cancer Registry data; Australian Bureau of Statistics for current and historic census data and digital

boundaries; Australian Social Science Data Archive, Australian National University, for historic census data; Geoscience Victoria, Department of Primary Industries, Victoria, for the Victorian Geoscientific Data Package; University of Ballarat for geochemical data; Australian Institute of Nuclear Science and Engineering for INAA of soils under awards AINGRA04216 and AINGRA06051; and Spatial Information Infrastructure, Department of Sustainability and Environment, Victoria, for Vicmap Digital Address Data.

Supplementary Information accompanies the paper on the Journal of Exposure Science and Environmental Epidemiology website (http://www.nature.com/jes)

References

Australian Bureau of Statistics (ABS). *1986 Census of Population and Housing: Statistical Local Area Summary File (Usual Residence) Summary File: Victoria, Table U02 Age group × sex [data file]*. Australian Bureau of Statistics: Canberra, ACT, 1986.

Australian Bureau of Statistics (ABS). *1991 Census of Population and Housing: Basic Community Profile, Census District, Victoria, Total Persons × sex [data file]*. Australian Bureau of Statistics: Canberra, ACT, 1991.

Australian Bureau of Statistics (ABS).. *Statistical Geography Volume 1 Australian Standard Geographical Classification (ASGC). ABS Cat. No. 1216.0*. Australian Bureau of Statistics: Canberra, ACT, 2001a. Available: http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/AA73DF0A91A3F71BCA256AD500017147/$File/12160_jul2001.pdf [accessed March 2005].

Australian Bureau of Statistics (ABS). *Information Paper: Census of Population and Housing - Socio-Economic Indexes for Areas (SEIFA2001), Australia, 2001. ABS Cat. No. 2039.0*. Australian Bureau of Statistics: Canberra, ACT, 2001b. Available: http://www.abs.gov.au/Ausstats/abs@.nsf/0b2680ff5fd3bf7aca256a0600054302/dcb7c23a858f99ccca2568890021c847!OpenDocument [accessed March 2005].

Australian Bureau of Statistics (ABS).. *CDATA2001: Time Series Profile, CD-ROM [data file*]. Australian Bureau of Statistics: Canberra, ACT, 2002a.

Australian Bureau of Statistics (ABS).. *Census of population and housing - Socio-Economic Indexes for Areas, 2001 (SEIFA2001), v2.0, CD-ROM [data file].* Australian Bureau of Statistics: Canberra, ACT, 2002b.

Australian Bureau of Statistics (ABS). *Historic Australian Standard Geographical Classification (ASGC) Digital Boundaries, Australia (1981-1995), CD-ROM [data file], Updated June 2005, Withdrawn Publication, Product No. 1216.0.30.001*. Australian Bureau of Statistics: Canberra, ACT, 2005a.

Australian Bureau of Statistics (ABS). *Historic Australian Standard Geographical Classification (ASGC) Digital Boundaries Files, Australia: Product Brief, ABS Cat. No. 1216.0.30.001*. Australian Bureau of Statistics: Canberra, ACT, 2005b.

Australian Bureau of Statistics (ABS). *Water Account Australia 2004-05. ABS Cat. No. 4610.0*. Australian Bureau of Statistics: Canberra, ACT, 2006. Available: http://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/6F380840F971B08DCA2577E700158A5E [accessed November 2011].

Andrew A.S., Burgess J.L., Meza M.M., Demidenko E., Waugh M.G., Hamilton J.W. *et al.* Arsenic exposure is associated with decreased DNA repair in vitro and in individuals exposed to drinking water arsenic. *Environ Health Perspect* 2006: **114**: 1193–1198.

Anselin L. and the Regents of the University of Illinois. *GeoDa 0.9.5-i Beta* [Software]. Spatial Analysis Laboratory, Urbana, IL, 2004. Available: https://www.geoda.uiuc.edu/downloadin.php [accessed 11 May 2006].

Anselin L. *Exploring Spatial Data with GeoDa: A Workbook.* Spatial Analysis Laboratory: Urbana, IL, 2005. Available: http://geodacenter.asu.edu/system/files/geodaworkbook.pdf [accessed November 2011].

Anselin L., Lozano N., and Koschinsky J. *Rate transformations and smoothing.* Spatial Analysis Laboratory: Urbana, IL, 2006. Available: http://geodacenter.asu.edu/pdf/smoothing_06.pdf [accessed November 2011].

Benbrahim-Tallaa L., Webber M.M., and Waalkes M.P. Acquisition of androgen independence by human prostate epithelial cells during arsenic-induced malignant transformation. *Environ. Health Perspect* 2005: **113**: 1134-1139.

Bentley R., Kavanagh A.M., Subramanian S.V., and Turrell G. Area disadvantage, individual socio-economic position, and premature cancer mortality in Australia 1998 to 2000: a multilevel analysis. *Cancer Causes Control* 2008: **19**: 183-193.

Bland M. *An Introduction to Medical Statistics.* Oxford University Press: Oxford, 1987.

Breslow N.E., and Day N.E. *Statistical methods in cancer research: Volume II – the design and analysis of cohort studies.* International Agency for Research on Cancer (IARC) Scientific Publications No. 82. IARC: Lyon, 1987.

Briggs D.J. Mapping environmental exposure. In: Elliott P., Cuzick J., English D., and Stern R. (eds). *Geographical & Environmental Epidemiology: Methods for Small-Area Studies*. Oxford University Press: New York, 1992, pp 158-176.

Carstairs V. Socio-economic factors at areal level and their relationship with health. In: Elliott P., Wakefield J.C., Best N.G., and Briggs D.J. (eds). *Spatial Epidemiology: Methods and Application.* Oxford University Press: Oxford, 2000, pp 51-67.

Centeno J.A., Mullick F.G., Martinez L., Page N.P., Gibb H., Longfellow D. *et al.* Pathology related to chronic arsenic exposure. *Environ. Health Perspect* 2002: **110** (Suppl 5): 883-886.

Chen C.J., Chuang Y.C., Lin T.M., and Wu H.Y. Malignant neoplasms among residents of a Blackfoot disease-endemic area in Taiwan: high-arsenic artesian well water and cancers. *Cancer Res* 1985: **45**: 5895-5899.

Chen C.L., Hsu L.I., Chiou H.Y., Hsueh Y.M., Chen S.Y., Wu M.M., and Chen C.J.: Blackfoot Disease Study Group. Ingested arsenic, cigarette smoking, and lung cancer risk: a follow-up study in arseniasis-endemic areas in Taiwan. *JAMA* 2004: **292**: 2984-2990.

Cliff A.D., and Ord J.K. *Spatial processes: Models & Applications*. Pion Ltd: London, 1981.

Deacon M. The future of population growth in Victoria. *People Place* 2000: **8**: 1-12. Available: http://arrow.monash.edu.au/hdl/1959.1/480622 [accessed November 2011].

Dean A.G., Sullivan K.M., and Soe M.M. *OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version 2.2.1,* 2008. Available: www.OpenEpi.com [accessed Jan 2009].

DMID. *Arsenic in the Environment: Stage 1 Report.* Department of Manufacturing and Industry Development: Victoria, Australia, 1991.

Elliott P., and Wakefield J.C. Bias and confounding in spatial epidemiology. In: Elliott P., Wakefield J.C., Best N.G., and Briggs D.J. (eds). *Spatial Epidemiology: Methods and Applications.* Oxford University Press: Oxford, 2000, pp 68-84.

Elliott P., and Wartenberg D. Spatial Epidemiology: Current approaches and future challenges. *Environ Health Perspect* 2004: **112**: 998-1006.

Elwood J.M. and Jopson J. Melanoma and sun exposure: an overview of published studies. *Int J Cancer* 1997: **73**(2): 198-203.

English D. Geographical epidemiology and ecological studies. In: Elliott P., Cuzick J., English D., and Stern R. (eds). *Geographical and Environmental Epidemiology: Methods for Small-Area Studies.* Oxford University Press: Oxford, 1996, pp 3-21.

enHealth Council. *Health-based soil investigation levels*. Commonwealth of Australia: Canberra, ACT, 2001. Available: http://www.health.gov.au/internet/main/publishing.nsf/content/66E7D805C1C1AD69CA2573CC0013EA68/$File/env_soil.pdf [accessed November 2011].

Graeter L.J., and Mortensen M.E. Kids are different: developmental variability in toxicology. *Toxicology* 1996: **111**: 15-20.

Greenland S. Interactions in Epidemiology: relevance, identification, and estimation. *Epidemiology* 2009: **20**: 14-17.

Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments.* Chapman & Hall/CRC: Boca Raton, Fl, 2004.

Haining R. *Spatial data analysis in the social and environmental sciences*. Cambridge University Press: Cambridge, 1990

Hinwood A.L., Jolley D.J., and Sim M.R. Cancer incidence and high environmental arsenic concentrations in rural populations: results of an ecological study. *Int J Environ Health Res* 1999: **9**: 131-141.

Hinwood A.L., Sim M.R., Jolley D., de Klerk N., Bastone E.B., Gerostamoulos J. *et al.* Hair and toenail arsenic concentrations of residents living in areas with high environmental arsenic concentrations. *Environ Health Perspect* 2003: **111**: 187-193.

Hinwood A.L., Sim M.R., Jolley D., de Klerk N., Bastone E.B., Gerostamoulos J. *et al* Exposure to inorganic arsenic in soil increases urinary inorganic arsenic concentrations of residents living in old mining areas. *Environ Geochem Health* 2004: **26**: 27-36.

Hughes M.F. Arsenic methylation, oxidative stress and cancer – is there a link? *JNCI* 2009: **101**: 1660-1661.

Insightful Corporation. *S-PLUS Version 6.2 for Windows Professional Edition* [Software]. Insightful Corporation: Seattle, WA, 2003.

Insightful Corporation. *S+SPATIALSTATS Version 1.5.6 for Windows* [Software]. Insightful Corporation: Seattle, WA, 2003.

James W.P.T., Nelson M., Ralph A., and Leather S. Socioeconomic determinants of health: the contribution of nutrition to inequalities in health. *BMJ* 1997: **314**: 1545-1549.

Jolley D.J., Jarman B., and Elliott P. Socio-economic confounding. In: Elliott P., Cuzick J., English D., and Stern R. (eds). *Geographical & Environmental Epidemiology: Methods for Small-Area Studies.* Oxford University Press: Oxford, 1996, pp 115-124.

Kaluzny S.P., Vega S.C., Cardoso T.P., and Shelly A.A. *S+SPATIALSTATS User's Manual for Windows and Unix.* Springer-Verlag: New York, 1998.

Leem J.-H., Kaplan B.M., Shim Y.K., Pohl H.R., Gotway C.A., Bullard S.M. *et al.* Exposures to air pollutants during pregnancy and preterm delivery. *Environ Health Perspect* 2006: **114**: 905-910.

Lewis D.R., Southwick J.W., Ouellet-Hellstrom R., Rench J., and Calderon R.L. Drinking water arsenic in Utah: a cohort mortality study. *Environ Health Perspect* 1999: **107**: 359-365.

MapInfo Corporation. *MapInfo Professional for Windows Version 8.0* [Software]. MapInfo Corporation: New York, NY, 2005.

Marshall G., Ferreccio C., Yuan Y., Bates M.N., Steinmaus C., Selvin S. *et al*. Fifty-year study of lung and bladder cancer mortality in Chile related to arsenic in drinking water. *JNCI* 2007: **99**: 920-927.

Nuckols J.R., Ward M.H., and Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environ Health Perspect* 2004: **112**: 1007-1015.

Pearce D.C., Dowling K., Gerson A.R., Sim M.R., Sutton S.R., Newville M. *et al*. Arsenic microdistribution and speciation in toenail clippings of children living in a historic gold mining area. *Sci Total Environ* 2010: **408**: 2590-2599.

Pickle L.W. Mapping mortality data in the United States. In: Elliott P., Wakefield J.C., Best N.G., and Briggs D.J. (eds). *Spatial Epidemiology: Methods and Applications*. Oxford University Press: Oxford, 2000, pp 240-252.

Portnov B.A., Dubnov J., and Barchana M. On ecological fallacy, assessment errors stemming from misguided variable selection, and the effect of aggregation on the outcome of epidemiological study. *J Expo Sci Environ Epidemiol* 2007: **17**: 106-121.

Quit Victoria [Website]. Smoking rates. n.d. Available: http://www.quit.org.au/ [accessed August 2007].

Richardson S., and Monfort C. Ecological correlation studies. In: Elliott P., Wakefield J.C., Best N.G., and Briggs D.J. (eds). *Spatial Epidemiology: Methods and Applications*. Oxford University Press: Oxford, 2000, pp 205-220.

Rossman T.G., Uddin A.N., Burns F.J., and Bosland M.C. Arsenite cocarcinogenesis: an animal model derived from genetic toxicology studies. *Environ Health Perspect* 2002: **110** (Suppl 5): 749-752.

Sahoo N.R., and Pandalai H.S. Secondary geochemical dispersion in the Precambrian auriferous Hutti-Maski schist belt, Raichur district, Karnataka, India. Part I: anomalies of As, Sb, Hg and Bi in soil and groundwater. *J Geochem Explor* 2000: **71**: 269-289.

Siahpush M. Socioeconomic status and tobacco expenditure among Australian households: results from the 1998-99 Household Expenditure Survey. *J Epidemiol Community Health* 2003: **57**: 798-801.

Sim M.R. Case studies in the use of toxicological measures in epidemiological studies. *Toxicology* 2002: **181-182**: 405-409.

Simpson L. Geography conversion tables: a framework for conversion of data between geographical units. *Int J Popul Geogr* 2002: **8**: 69-82.

Smith D., Taylor R., and Coates M. Socioeconomic differentials in cancer incidence and mortality in urban New South Wales, 1987-1991. *Aust N Z J Public Health* 1996: **20**: 129-137.

Smith E., Smith J., Smith L., Biswas T., Correll R., and Naidu R. Arsenic in Australian environment: an overview. *J Environ Sci Health A Tox Hazard Subst Environ Eng* 2003: **A38**: 223-239.

Sneyd M.J., and Cox B. 2009. Melanoma in Maori, Asian, and Pacific Peoples in New Zealand. *Cancer Epidemiol Biomarkers Prev* 2009: **18**: 1706-1713.

Snowden V., Moore C., and Kelly D. Using geostatistics to assist in optimizing grade control estimation at KCGM's Fimiston open pit. In: *Proceedings of the 4th Large Open Pit Mining Conference*, 1994 Sept 5-9, Perth, Australia. The Australasian Institute of Mining and Metallurgy. 1994, pp 117-131.

SPSS Inc. *SPSS 15.0 for* Window [Software]s. SPSS Inc: Chicago, IL, 2006.

Tchounwou P.B., Patlolla A.K., and Centeno J.A. Carcinogenic and systemic health effects associated with arsenic exposure: A critical review. *Toxicol Pathol* 2003: **31**: 575-588.

Thursfield V., and Giles G. *Skin Cancer*. *Canstat 44*. Cancer Council Victoria: Melbourne, Victoria, 2007. Available: http://www.cancervic.org.au/downloads/cec/CanStats/44-Skin-cancer.pdf [accessed May 2011].

Thursfield V., Farrugia H., Giles G. *Cancer in Victoria 2005*. *Canstat 45*. Cancer Council Victoria: Melbourne, Victoria, 2008. Available: http://www.cancervic.org.au/downloads/cec/CanStats/45-Cancer-in-Victoria-2005.pdf [accessed November 2011].

Tiller K.G. Urban soil contamination in Australia. *Aust J Soil Res* 1992: **30**: 937-957.

Vahter M.E. Interactions between arsenic-induced toxicity and nutrition in early life. *J Nutr* 2007: **137**: 2798-2804.

VanderWeele T.J. Sufficient cause interactions and statistical interactions. *Epidemiology* 2009: **20**: 6-13.

Waller L.A., and Gotway C.A. *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons: Hoboken, 2004.

Wei Q., Lee J.E., Gershenwald J.E., Ross M.I., Mansfield P.F., Strom S.S. *et al.* Repair of UV light-induced DNA damage and risk of cutaneous malignant melanoma. *J Natl Cancer Inst* 2003: **95**(4): 308-15.

Wickre J.B., Folt C.L., Sturup S., and Karagas M.R. Environmental exposure and fingernail analysis of arsenic and mercury in children and adults in a Nicaraguan gold mining community. *Arch Environ Health* 2004: **59**: 400-409.

Williams J., Clifford C., Hopper J., and Giles G. Socioeconomic status and cancer mortality and incidence in Melbourne. *Eur J Cancer* 1991: **27**: 917-921.

Wu M.-M., Kuo T.-L., Hwang Y.-H. and Chen C.-J. Dose-response relation between arsenic concentrations in well water and mortality from cancers and vascular diseases. *Am J Epidemiol* 1989: **130**: 1123-1132.

Yorifuji T., Tsuda T., and Grandjean P. Unusual cancer excess after neonatal arsenic exposure from contaminated milk powder. *JNCI* 2010: **102**: 360-361.

Yuan Y., Marshall G., Ferreccio C., Steinmaus C., Liaw J., Bates M. et al. Kidney cancer mortality: fifty-year latency patterns related to arsenic exposure. *Epidemiology* 2010: **21**: 103-108.

Figure Legends

Figure 1. Location map of the study area within Victoria, Australia, showing 36 core statistical local areas (SLAs), and 25 peripheral SLAs (shaded). Sources: CDATA2001 (ABS, 2002a).

Figure 2. Map showing overlays of digitised boundaries for 2001 statistical local areas (SLAs), 1991 SLAs and census districts (CDs), and Vicmap digital address points in part of the study area. Sources: CDATA2001 (ABS, 2002a); ABS, 2005a; Department of Sustainability and Environment, Victoria.

Figure 3. Choropleth maps showing quintiles of soil arsenic values for a. exposure metric 1, and b. exposure metric 2. Sources: CDATA2001 (ABS, 2002a); modified data from University of Ballarat and Geoscience Victoria.

Figure 4. Graphs show for a. males and b. females, relative risks (RR, 95% CI) for all-cancers, colon, leukaemia, lung, melanoma, and urinary tract (UTC) cancers, and prostate cancer in males, in more socioeconomically disadvantaged statistical local areas (SLAs). After stratifying SLAs by quintiles of exposure metric 1 and estimating aggregate standardised incidence ratios (SIRs) by gender, RRs were calculated as ratios of SIRs within the uppermost quintile relative to the lowest quintile.