Published by Oxford University Press on behalf of the International Epidemiological Association
© The Author 2013; all rights reserved.
International Journal of Epidemiology 2013;**42**:1862–1872
doi:10.1093/ije/dyt182

Arsenic metabolism efficiency has a causal role in arsenic toxicity: Mendelian randomization and gene-environment interaction

Brandon L Pierce,[1,2]* Lin Tong,[1] Maria Argos,[1] Jianjun Gao,[1] Farzana Jasmine,[1] Shantanu Roy,[1] Rachelle Paul-Brutus,[1] Ronald Rahaman,[1] Muhammad Rakibuz-Zaman,[3] Faruque Parvez,[4] Alauddin Ahmed,[3] Iftekhar Quasem,[3] Samar K Hore,[5] Shafiul Alam,[3] Tariqul Islam,[3] Judith Harjes,[6] Golam Sarwar,[3] Vesna Slavkovich,[4] Mary V Gamble,[4] Yu Chen,[7] Mohammad Yunus,[5] Mahfuzar Rahman,[3] John A Baron,[8] Joseph H Graziano[4] and Habibul Ahsan[1,2,9]

[1]Department of Health Studies and [2]Comprehensive Cancer Center, University of Chicago, Chicago, USA, [3]UChicago Research Bangladesh, Dhaka, Bangladesh, [4]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, USA, [5]International Center for Diarrheal Disease Research, Bangladesh, Dhaka, Bangladesh, [6]Bioinformatics Service Center, Dartmouth Medical School, USA, [7]Division of Epidemiology, Department of Population Health, New York University, New York, USA. [8]University of North Carolina, Lineberger Comprehensive Cancer Center, Chapel Hill, USA and [9]Departments of Medicine and Human Genetics, University of Chicago, Chicago, USA

*Corresponding author. Center for Cancer Epidemiology and Prevention, University of Chicago, 5841 South Maryland Avenue, Suite N101, MC2007, Chicago, IL 60637, USA. E-mail: brandonpierce@uchicago.edu

Accepted 5 August 2013

Background Arsenic exposure through drinking water is a serious global health issue. Observational studies suggest that individuals who metabolize arsenic efficiently are at lower risk for toxicities such as arsenical skin lesions. Using two single nucleotide polymorphisms (SNPs) in the 10q24.32 region (near *AS3MT*) that show independent associations with metabolism efficiency, Mendelian randomization can be used to assess whether the association between metabolism efficiency and skin lesions is likely to be causal.

Methods Using data on 2060 arsenic-exposed Bangladeshi individuals, we estimated associations for two 10q24.32 SNPs with relative concentrations of three urinary arsenic species (representing metabolism efficiency): inorganic arsenic (iAs), monomethylarsonic acid (MMA) and dimethylarsinic acid (DMA). SNP-based predictions of iAs%, MMA% and DMA% were tested for association with skin lesion status among 2483 cases and 2857 controls.

Results Causal odds ratios for skin lesions were 0.90 (95% confidence interval [CI]: 0.87, 0.95), 1.19 (CI: 1.10, 1.28) and 1.23 (CI: 1.12, 1.36) for a one standard deviation increase in DMA%, MMA% and iAs%, respectively. We demonstrated genotype-arsenic interaction, with metabolism-related variants showing stronger associations with skin lesion risk among individuals with high arsenic exposure (synergy index: 1.37; CI: 1.11, 1.62).

Conclusions We provide strong evidence for a causal relationship between arsenic metabolism efficiency and skin lesion risk. Mendelian randomization can be used to assess the causal role of arsenic

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022

exposure and metabolism in a wide array of health conditions. Developing interventions that increase arsenic metabolism efficiency are likely to reduce the impact of arsenic exposure on health.

Keywords Arsenic, arsenic metabolism, Mendelian randomization, gene-environment interaction, *AS3MT*

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022

Background

Arsenic exposure through drinking water has been a serious public health issue in many countries around the world, including Bangladesh, India, Taiwan, Mexico, Argentina, Chile and the USA, affecting approximately 140 million individuals worldwide.[1] In addition to being a known carcinogen with established effects on bladder, liver, skin and kidney cancer risk,[2–7] arsenic has been reported to increase risk for cardiovascular conditions,[8] pregnancy complications,[9] neurological conditions[10,11] and overall mortality.[12,13] In general, these associations are quite clear at high levels of exposure (>300 μg/l in water) but less consistent at lower levels of exposure. Skin lesions are the classical sign of arsenic toxicity, and risk for such lesions shows a dose-response relationship with arsenic exposure.[14] Skin lesions are believed to reflect susceptibility to arsenic-related disease, including cancer.[5]

Arsenic (As) consumed in drinking water enters the blood as inorganic arsenic (iAs): As^{V} and As^{III}. iAs then undergoes a series of oxidation and methylation steps to produce monomethylarsonic acid (MMA) and then dimethylarsinic acid (DMA). These reactions involve S-adenosyl methionine (SAM) as the methyl donor and are catalyzed by methyltransferase enzymes, most notably, arsenite methyltransferase (*AS3MT*). These arsenic species (iAs, MMA and DMA) are excreted in urine and can be measured and expressed as percentages of total urinary arsenic (iAs%, MMA% and DMA%). These percentages are measures of an individual's capacity to methylate arsenic, which shows considerable inter-individual variation.[15] Urinary iAs% and MMA% are positively correlated whereas DMA% shows strong negative correlations with iAs% and MMA%.[16–18] Thus, high DMA% (and low iAs% and MMA%) represent enhanced methylation capacity whereas high iAs% and MMA% (and low DMA%) represent low methylation capacity. Increased methylation capacity has been hypothesized to decrease susceptibility to arsenic-related toxicity, presumably due to enhanced excretion and lower toxicity of DMA relative to iAs and MMA.[15] Several epidemiological studies have reported that increased methylation capacity is inversely associated with risk for arsenical skin lesions.[17–21]

In previous work, our group[16] and others[22,23] have shown that genetic variation in the 10q24.32 region (containing *AS3MT* and other genes) influences arsenic methylation capacity in Bangladeshi individuals, after adjusting for arsenic exposure. We demonstrated that multiple single nucleotide polymorphisms (SNPs) in this region are independently associated with urinary DMA% and MMA%, with one SNP showing evidence of association with skin lesion risk through interaction with arsenic. In this work, using a larger sample size compared with our prior study, we are now able to rigorously address the hypothesis that arsenic methylation capacity (measured as arsenic species percentages) has a causal role in arsenic toxicity (measured as skin lesions) using a Mendelian randomization (MR) approach (Figure 1). MR is a method for assessing the causal effect of a risk factor on a disease outcome using genetic determinants of the risk factor as instrumental variables (IVs).[24,25] If the IVs are valid, the MR estimate represents the causal component of the observed association, not the components due to unmeasured confounding variables.

Methods

Study participants

Subjects genotyped for this work were participants in one of two ongoing studies of arsenic exposure in Bangladesh: the Health Effects of Arsenic Longitudinal Study (HEALS)[26] and the Bangladesh Vitamin E

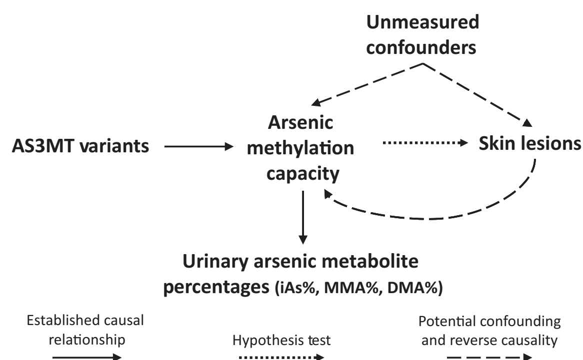

Figure 1 A causal diagram representing the relationships among 10q24.32 genetic variants, arsenic methylation capacity, urinary arsenic metabolites and arsenical skin lesion risk in an arsenic-exposed population. For simplicity, arsenic exposure is not shown, but the hypothesized effect of arsenic methylation capacity on skin lesions would occur through an interaction with arsenic exposure

and Selenium Trial (BEST).[27] HEALS is a prospective investigation of health outcomes associated with arsenic exposure through drinking water in a population-based cohort of 11 746 adults (age 18–75 years) in Araihazar, Bangladesh, a rural area east of the capital city, Dhaka (described in detail in reference 26). A total of 10 970 wells in the study area were tested for arsenic, and individuals were assigned an arsenic exposure level based on their reported primary drinking well. BEST is a 2×2 factorial randomized chemoprevention trial evaluating the long-term effects of vitamin E and selenium supplementation on non-melanoma skin cancer risk among 7000 individuals with arsenic-related skin lesions living in Araihazar (the same geographical area as HEALS), Matlab and surrounding areas. BEST uses many of the same study protocols as does HEALS, in particular those for exposure assessment and biospecimen collection. Additional details on these studies can be found in the Supplementary data (available at *IJE* online).

Eligibility

Participants included in this work are a subset of HEALS and BEST participants from Araihazar that were genotyped at >250 000 SNPs using Illumina's HumanCytoSNP-12 v2.1 chips ($n = 5340$). Analyses of urinary arsenic species were conducted among 2060 genotyped HEALS participants with available data on DMA%, MMA% and iAs% (measured at baseline). This sample includes all 1313 individuals included in our prior study.[16] Analyses of skin lesion status was conducted among 5340 genotyped individuals (2483 skin lesion cases and 2857 controls) selected from both studies, including the 2060 HEALS individuals with metabolite data. This sample includes all 1085 cases and all 1794 controls included in our prior study. Skin lesion cases included individuals with keratosis, melanosis and leukomelanosis. A summary of the participants included in this study is provided in Figure 2.

Laboratory methods

DNA extraction and quality control (QC) for HEALS and BEST blood samples have been described previously.[16] Genotyping was conducted in two batches of approximately equal size, containing approximately equal numbers of skin lesion cases and controls. Samples were genotyped on Illumina HumanCytoSNP-12 v2.1 chips (~300 000 SNPs). QC was performed using PLINK,[28] the details of which are contained in the Supplementary data at *IJE* online. After QC, high-quality genotype data were available for 5354 individuals and 257 747 SNPs.

In both HEALS and BEST, urinary arsenic was measured using a graphite furnace atomic absorption spectrometry method with a very low limit of detection (~1 µg/l) in a single laboratory at Columbia University.[29] For a subset of HEALS participants, urinary arsenic metabolites (arsenobetaine, arsenocholine, arsenite, arsenate, MMA and DMA) were distinguished as described by Ahsan *et al.*,[18] using a high-performance liquid chromatography method for separation of arsenic metabolites, followed by detection using inductively coupled plasma-mass spectrometry with dynamic reaction cell. The total concentrations of iAs (arsenate and arsenite), MMA and DMA in urine were expressed as a percentage of total arsenic, after subtracting asenobetaine and arsenocholine (i.e. nontoxic organic arsenic from dietary sources). In HEALS, arsenic in drinking water was measured at baseline using graphite furnace atomic absorption spectrometry (and inductively coupled plasma-mass spectrometry for samples with concentrations below the limit of detection).

Skin lesion assessment

In both HEALS and BEST, a structured protocol was used to ascertain skin lesions, implemented by study physicians who had undergone training for the detection and diagnosis of skin lesions. The study physician recorded the presence or absence of melanosis (hyperpigmentation), leucomelanosis (hypopigmentation) or keratosis (a hyperkeratotic thickening of the skin typically on the palms and soles) and the location of the lesion.[30] For the present case-control analysis, skin lesion status was defined as any type of skin lesion detected at either baseline (in HEALS or BEST) or during follow-up (HEALS).

Statistical methods

Associations between SNPs in the 10q24.32 region (coded as 0, 1 or 2 minor alleles) and urinary arsenic species percentages (measured at baseline) were estimated using mixed linear models[31] to account for cryptic relatedness in this population (described in the Supplementary data available at *IJE* online). Logistic regression was used to estimate observational associations between arsenic species percentages and skin lesion status using data on 632 cases and 4036 controls with available urinary arsenic species data (all from HEALS). As previously reported, there is very little evidence of population structure in our sample (based on principal components analysis, after removal of one individual from each related pair),[16] thus no adjustments for principle components were required. Imputation of unmeasured genotypes did not produce imputed SNPs that were stronger predictors of arsenic methylation capacity than our genotyped SNPs, so results for imputed SNPs are not presented.

We used an MR approach to assess the causal effect of methylation capacity (measured as urinary arsenic species percentages) on skin lesion risk using sequential regressions. In the first-stage mixed linear model ($n = 2060$), the arsenic methylation capacity variable was regressed on two genotyped 10q24.32 SNPs (i.e. the IVs) adjusting for sex, age, smoking, body mass index and genotyping stage. Each arsenic

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022

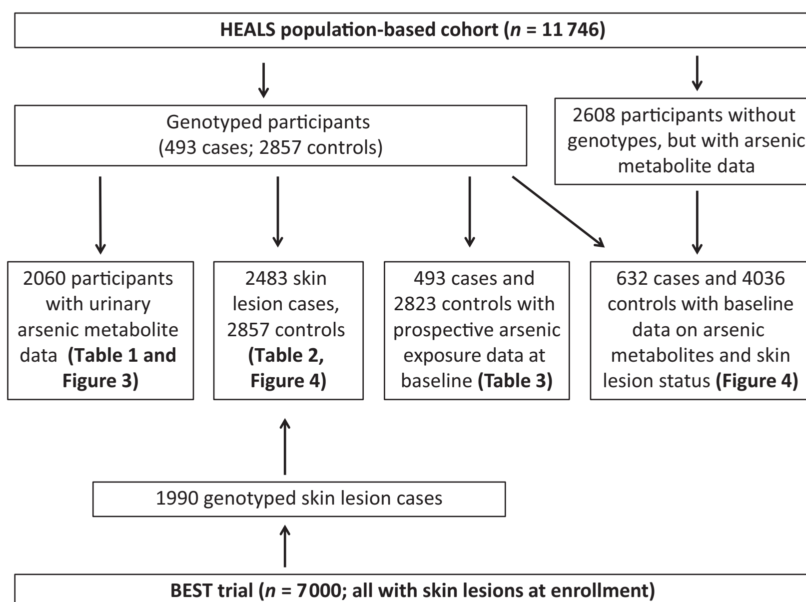

Figure 2 A description of the participants included in this study

species measure (DMA%, MMA% and iAs%) was used to represent arsenic methylation capacity in three separate MR models. rs9527 and rs11191527 were selected as IVs because these SNPs best represent the two independent association signals in this region for each metabolite. Using two IVs in the same model (each coded as 0,1 or 2 risk alleles), rather than a single IV (such as a weighted SNP allele score), allowed us to make no assumptions regarding relative effect sizes of these SNPs, avoiding the loss of power associated with mis-specifying allele weights.[32] In the second-stage logistic regression (2483 cases and 2857 controls), skin lesion status was regressed on the predicted values of arsenic metabolites based on the beta coefficients from the first-stage regression, again adjusting for the same covariates as the first stage. Our recent work has shown that this two-stage procedure with linear and logistic regression maintains appropriate type I errors rates but may be very slightly biased towards the null;[33] thus, our estimates may be somewhat conservative.

Observational (cross-sectional) association estimates between each arsenic species measure and skin lesion risk were estimated using a subset of HEALS participants with available data on urinary arsenic species and skin lesion status, many of whom did not have SNP data (632 skin lesion cases and 4036 controls). Differences between observational and MR estimates were tested in Stata using the *suest* command (seemingly unrelated regression).

Gene-environment interaction analysis was conducted among a set of 493 cases and 2823 controls (all from HEALS) with prospective data on arsenic exposure from drinking water (measured prior to arsenic mitigation efforts[34] using logistic regression (Figure 2). In logistic regression models, arsenic exposure was modelled as tertiles and genotypes were modelled as binary variables indicating minor allele carriers and non-carriers. Individuals were assigned to 'low', 'intermediate' and 'high' methylation capacity categories based on their genotypes at rs9527 and rs11191527. Additive interaction was estimated in two ways: the synergy index and relative excess risk due to interaction (RERI).[35,36] Additional details of SNP-arsenic interaction methods are provided in the Supplementary data at *IJE* online.

Results

Associations between SNPs and arsenic metabolism

Associations between SNPs in the 10q24.32 region and urinary arsenic species percentages (measured at baseline) were estimated using data on 2 060 HEALS participants, described in Table 1. Consistent with prior findings,[16,22,23] 10q24.32 variants showed association signals for both MMA% and DMA% ($P < 10^{-10}$) and a weaker signal for iAs% ($P = 5 \times 10^{-5}$). For DMA%, MMA% and iAs%, the SNP showing the smallest *P*-value was rs9527 ($P = 1.1 \times 10^{-11}$, $P = 2.5 \times 10^{-11}$ and $P = 5.8 \times 10^{-5}$, respectively). After conditioning on rs9527, there were strong secondary signals for all three arsenic species, with rs11191527 best representing this secondary signal for DMA% ($P = 8.7 \times 10^{-10}$), MMA% ($P = 2.9 \times 10^{-6}$) and iAs% ($P = 2.9 \times 10^{-6}$) (Supplementary Figure 1, available as Supplementary data at *IJE* online). After adjusting for these two independent association signals, no additional evidence of association was observed for any of the three arsenic species (i.e. all $P > 10^{-4}$ for each species across the entire 2-Mb region surrounding the association

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022

Table 1 Mean urinary arsenic species percentages (reflecting arsenic methylation capacity) for 2060 HEALS participants with arsenic metabolite data

Characteristic	%	iAs%	MMA%	DMA%
Sex				
Male	50.4	14.9	**15.3**	**69.8**
Female	49.6	15.2	**11.6**	**73.2**
Age (yr)				
⩽30	26.2	**16.1**	**11.9**	71.9
31–39	23.9	**15.8**	**13.1**	71.0
40–48	26.4	**14.9**	**13.8**	71.2
⩾49	23.5	**13.3**	**15.1**	71.7
BMI (kg/m^2)				
<17.38	25.0	15.4	**14.3**	**70.3**
17.38–19.03	25.0	15.3	**13.9**	**70.8**
19.04–21.37	25.0	15.1	**13.3**	**71.7**
>21.37	25.0	14.6	**12.3**	**72.9**
Cigarette smoking				
Current	34.8	**15.0**	**14.9**	**70.1**
Former	8.5	**15.3**	**12.2**	**72.4**
Never	56.7	**13.9**	**15.4**	**70.7**
Urinary arsenic (μg/g)				
<101	25.1	**14.0**	**13.4**	**72.7**
101–185	24.9	**14.6**	**13.0**	**72.4**
186–320	25.0	**15.5**	**13.4**	**71.1**
⩾321	25.1	**16.3**	**14.0**	**69.6**
Water arsenic(μg/l)				
<9.0	25.2	**13.9**	**12.9**	**73.2**
9.0–49.9	24.7	**15.0**	**13.1**	**71.9**
50.0–127.9	25.1	**15.4**	**13.6**	**70.9**
⩾128	25.0	**16.0**	**14.1**	**69.9**

The participants with metabolite data described in this table are all HEALS cohort members and are a subset of the larger case-control sample described in Table 2. Bold numbers represent P-values <0.05 for global test of association for each characteristic with each arsenic metabolite using t-test or ANOVA.

signals). For these two SNPs (rs9527 and rs11191527), which are in mild linkage disequilibrium (D′= 0.26; r^2 = 0.04), the alleles associated with increases in DMA% are associated with decreases in MMA% and iAs%, as expected based on correlations among the metabolite measures (Figure 3). Results remained the same after exclusion of 68 prevalent skin lesion cases from the analysis. These results suggest that the relative concentrations of iAs, MMA and DMA are influenced by a common set of underlying causal variants, and that these variants represent an individual's capacity to produce DMA, the end product of arsenic metabolism. We acknowledge the two lead SNPs reported here are not the top SNPs reported in our prior

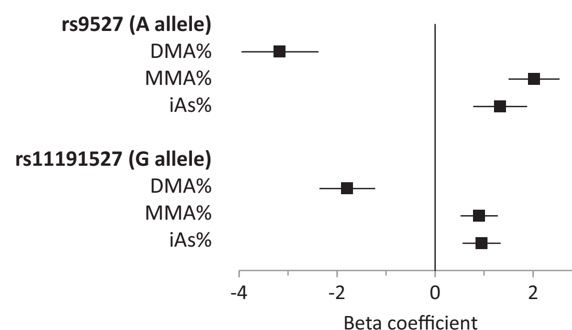

Figure 3 Associations between 10q24.32 SNPs and arsenic species percentages. Estimates for SNPs were generated in multivariate linear regression models that included sex, age categories and both SNPs (n = 2060). The observed allele frequencies for the low-efficiency alleles of rs9527 (A) and rs11191527 (G) are 0.08 and 0.84, respectively

smaller study. Those SNPs remain strongly associated with methylation capacity, but based on the larger sample size used in this work, the previously reported SNPs no longer have the lowest P-values for the two signals in this region. The previously reported SNPs remain adequate tagSNPs for the two association signals we observe.

Mendelian randomization

The direction of the observational association between arsenic methylation capacity and skin lesion risk (Figure 4) is consistent with associations reported in prior work,[17,19–21] including our prior report,[18] with increased methylation capacity (high DMA% and low MMA% and iAs%) being associated with decreased skin lesion risk. Our IV SNPs show little evidence of association with the potential confounding factors shown in Tables 1 and 2, supporting the validity of these SNPs as IVs that are useful for MR (Supplementary Table 1, available as Supplementary data at *IJE* online). MR analysis indicated that a causal relationship underlies this association, with each of the three metabolites producing a causal OR with P <0.0001 and a direction consistent with the observational associations (Figure 4). The first-stage F-statistics from all models were >15, ensuring that our results are free from appreciable weak-IV biases.[32,37] Results excluding covariates adjustments were similar to those from adjusted analyses. 10q24.32 is a well-established arsenic metabolism locus;[23] however, in order to ensure that our MR estimates were not biased due to the fact that we selected our IV SNPs using a subset of the data we used for MR, a cross-validation analysis (described in the Supplementary data at *IJE* online) was used to confirm that this phenomenon did not account for the MR results. For DMA% and MMA%, this cross-validation approach produced results very similar to those obtained in the primary MR analysis (Supplementary Table 2, available as Supplementary data at *IJE*

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022

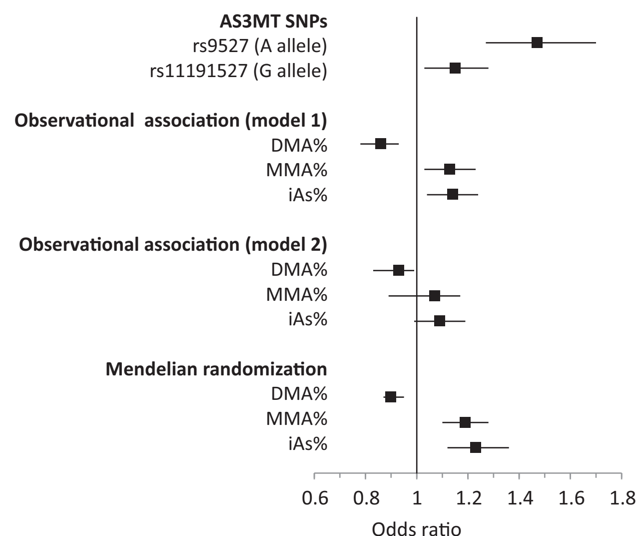

Figure 4 Associations for 10q24.32 SNPs and arsenic species percentages in relation to skin lesion risk. Estimates for SNPs were generated in multivariate logistic regression models that included sex, age categories and both SNPs coded additively (2483 cases and 2857 controls). Observational estimates for arsenic species percentages were generated using logistic regression and cross-sectional data on HEALS participants with arsenic metabolite and skin lesion status data, measured at baseline (632 cases and 4036 controls). Model 1 included adjustments for age and sex, and model 2 included additional adjustments for smoking and body mass index. Observational ORs correspond to a one standard deviation increase in the metabolite. For the Mendelian randomization estimates, first-stage regressions were estimated using 2060 HEALS participants with arsenic metabolite data and SNP data (rs9527 and rs11191527), adjusting for age, sex, smoking, body mass index and genotyping stage. F-statistics were 45, 38 and 17 for DMA%, MMA% and iAs%, respectively. The second-stage logistic regression models were fit using data on 2483 cases and 2857 controls with SNP data, adjusting for the same covariates included in the first-stage regression. MRORs correspond to a one standard deviation change in the arsenic species percentage

online). The effects of our two SNPs on iAs% were not strong enough for cross-validation analyses to be conducted (see Methods and Discussion). In addition, we conducted MR analysis using only one SNP at a time and obtained similar results to those obtained when using both SNPs (Supplementary Table 3, available as Supplementary data at *IJE* online), further suggesting our SNPs are valid IVs (see Discussion). For MMA% and iAs%, there was evidence that fully-adjusted observational estimates (model 2) and MR estimates were different ($P = 0.09$ and 0.02, respectively), but not for DMA% ($P = 0.97$).

SNP-Arsenic interaction

In SNP-arsenic interaction analyses, there was suggestive evidence that the individual effects of rs9527 and rs11191527 on skin lesion risk were stronger among individuals with higher exposure, measured as arsenic in drinking water (Table 3). When information on these SNPs was combined into 'high', 'intermediate' and 'low' genetically-defined categories of arsenic methylation capacity, there was strong evidence that the association between SNP-based methylation capacity and skin lesions risk was stronger among individuals with higher exposure, as measured by interaction on the additive scale (synergy index = 1.37, CI: 1.11, 1.62; RERI = 0.39, CI: 0.00, 0.79), consistent with the hypothesis that individuals with higher arsenic metabolism capacity have lower

Table 2 Participant characteristics for all skin lesion cases and controls with genetic data

Characteristics	*n* (%)	Skin lesion cases[a] (%)	Controls[b] (%)
n	5340	2483	2857
Sex			
Male	2580 (48.3)	1450 (58.4)	1130 (39.6)
Female	2760 (51.7)	1033 (41.6)	1727 (60.4)
Age (yr)			
⩽30	1348 (25.2)	345 (13.9)	1003 (35.1)
31–39	1273 (23.8)	514 (20.6)	759 (26.6)
40–48	1413 (26.5)	756 (30.5)	657 (23.0)
⩾49	1306 (24.5)	868 (35.0)	438 (15.3)
BMI (kg/m^2)[c]			
<17.56	1324 (24.9)	621 (25.1)	703 (24.8)
17.56–19.29	1333 (25.1)	608 (24.6)	725 (25.6)
19.30–21.69	1322 (24.9)	603 (24.3)	719 (25.4)
>21.69	1331 (25.1)	645 (26.0)	686 (24.2)
Cigarette smoking			
Current	1575 (29.5)	799 (32.2)	776 (27.2)
Former	453 (8.5)	285 (11.5)	168 (5.8)
Never	3312 (62.0)	1399 (56.4)	1913 (67.0)
Urinary arsenic[c] [creatinine adjusted] (μg/g)			
<90	1324 (25.0)	686 (27.6)	638 (22.5)
91–177	1327 (25.0)	522 (21.0)	805 (28.4)
178–344	1321 (25.0)	521 (21.0)	800 (28.3)
⩾345	1322 (25.0)	733 (29.4)	589 (20.8)
Data available for a subsample of participants[d]			
Water arsenic[d] (μg/l)			
n	3316	493	2823
⩽7.00	841 (25.4)	71 (14.4)	770 (27.3)
7.01–45.25	818 (24.7)	97 (19.6)	721 (25.5)
45.26–119	830 (25.0)	138 (28.0)	692 (24.5)
⩾120	827 (24.9)	187 (38.0)	640 (22.7)

[a]Cases were selected from HEALS and BEST.
[b]Controls were selected from HEALS.
[c]BMI and urinary arsenic measures have <1% missing data.
[d]Water arsenic was measured in HEALS only.

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022

arsenic toxicity risk. Consistent with this interaction, we also demonstrate that the observational association between arsenic methylation capacity and skin lesion risk is stronger among individuals with higher exposure levels (Supplementary Table 4, available as Supplementary data at *IJE* online). Interestingly, the effect of rs11191527 on methylation capacity appears to depend on arsenic exposure status, whereas the effect of rs9527 does not appear to vary substantially across exposure strata (Supplementary Table 5, available as Supplementary data at *IJE* online). Thus, even in the low exposure group ($<16\,\mu g/l$), we observe an association between rs9527 and DMA%. However, this is not entirely unexpected, as the effects of 10q24.32 variants on arsenic methylation capacity have been observed in prior studies of individuals with low exposure.[38] Our low exposure group includes very few truly unexposed participants.

Discussion

In this work, we have used a MR approach to provide strong evidence that arsenic methylation capacity (as measured by either DMA%, MMA% or iAs%) is causally related to risk for arsenic toxicity (as measured by arsenical skin lesions). We estimate that one standard deviation increase in urinary DMA% (~7–8 percentage points) results in an approximately 10% decrease in the odds of skin lesions. In contrast, one standard deviation increase in MMA% and iAs% would result in 20% and 25% increases in the odds of skin lesions, respectively. We also show that arsenic has a weaker association with skin lesion risk among individuals with high methylation capacity genotypes, consistent with our MR results. The effect of rs11191527 on arsenic methylation capacity may vary by arsenic exposure, so these estimates represent the average treatment effect across all levels of arsenic exposure.

Several observational studies have reported that increased methylation capacity is inversely associated with arsenic toxicity,[17–21] but none have used an MR approach. The key advantage of MR in this context is that MR estimates are not susceptible to biases stemming from unmeasured confounding. It is possible that a wide array of unmeasured (or poorly measured) lifestyle and host factors (e.g. nutrition, adiposity) influence both one's capacity to methylate arsenic and the effectiveness of host defence against arsenic toxicity and skin lesion development, inducing a non-causal association between methylation capacity and skin lesion status. Under the assumption that the 10q42.32SNPs used in this work are valid IVs, we can interpret association estimates derived from MR as causal effects. Consequently, interventions that increase arsenic methylation capacity are very likely to reduce arsenic toxicity. Several micronutrients related to one-carbon metabolism have been proposed as potential modulators of arsenic metabolism,[39] including folate. Randomized trials have provided strong evidence that folate supplementation increases arsenic methylation capacity among individuals with low plasma folate.[40,41]

This work builds on our prior work[16] on this topic in two ways. First, we have increased the sample size of our prior SNP association study of arsenic metabolites, and we can now show that only two association signals for arsenic methylation capacity are observed in the 10q24.32 region, and these two signals are observed regardless of which metabolite measure is used to represent arsenic methylation capacity. Thus, we can use the same IVs for each arsenic methylation phenotype. Second, by expanding the sample size of individuals with both SNP and skin lesion data, we now have the statistical power to conduct MR analyses that convincingly show that methylation capacity is causally related to skin lesions risk, with both SNPs providing consistent evidence.

Whereas the identity of the causal variant in the 10q24.32 region remain unknown, it is likely that they affect the function of AS3MT, a gene with a clear role in arsenic metabolism. However, other genes in this region have also been suggested as having potential roles arsenic methylation capacity (such as *C10orf32*, *CNNM2*, *NT5C2* and *USMG5*).[42] Our group[16] and others[42] have shown that metabolism-associated variants may influence local gene expression, a potential mechanism by which these variants could influence arsenic methylation capacity.

The validity of our MR approach relies on the assumption that our two SNPs are not related to skin lesion risk independently of the methylation capacity phenotype.[24,25] Although this assumption cannot be proven, its validity is supported by several observations. First, the magnitude and direction of the SNPs' associations with arsenic methylation capacity and skin lesion risk are consistent, with rs11191527 having a weaker associaitons than rs9527 for both phenotypes. Thus, our two SNPs produce similar effect estimates when analysed in separate two-stage models, taking into account the size of the confidence intervals and the different effect sizes for our two SNPs (Supplementary Table 3, available as Supplementary data at *IJE* online). If one of the IVs were invalid, these estimates would be expected to be different.[43] Second, there is clear biological plausibility for SNPs near *AS3MT* to have effects that are specific to arsenic methylation capacity, supporting the validity of the IV assumptions.

In addition to the advantages of MR for isolating causal relationships, 10q24.32 variants also capture the effect of arsenic methylation capacity over a large portion of the life course. The association between 10q24.32 variants and arsenic methylation capacity is present for all age groups in our data, i.e. from 18 to 70 years of age; thus, its association with skin lesion risk may reflect the cumulative effects of a lifetime of low (or high) methylation capacity. Urinary metabolite percentages, on the other hand, may vary

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022

Table 3 Odds ratios and 95% confidence intervals for the association between *AS3MT* SNPs and skin lesion status (493 cases and 2823 controls) stratified by arsenic exposure (measured in drinking water)

Water arsenic tertiles (μg/l)	*AS3MT* genotype			Multiplicative Interaction P	RERI (CI)[a]	Synergy index (CI)[a]
	rs9527					
	GG	GA or AA				
0.1–16	1.00 (Ref)	0.85 (0.44, 1.63)		0.06	0.45 (0.09, 0.82)	1.93 (0.60, 3.26)
17–87	1.00 (Ref)	1.08 (0.64, 1.82)				
>88	1.00 (Ref)	1.69 (1.12, 2.56)				
	rs11191527					
	GG	GA or AA				
0.1–16	1.00 (Ref)	1.14 (0.67, 1.92)		0.56	−0.19 (−0.49, 0.11)	0.69 (0.32, 1.07)
17–87	1.00 (Ref)	0.65 (0.43, 1.00)				
>88	1.00 (Ref)	0.85 (0.60, 1.22)				
	SNP-based arsenic methylation capacity					
	High	Intermediate	Low			
0.1–16	1.00 (Ref)	1.01 (0.56, 1.83)	0.63 (0.23, 1.74)	0.13	0.39 (0.00, 0.79)	1.37 (1.11, 1.62)
17–87	1.00 (Ref)	1.45 (0.91, 2.30)	1.73 (0.84, 3.56)			
>88	1.00 (Ref)	1.31 (0.86, 1.99)	1.85 (1.00, 3.42)			

OR, odds ratio; RERI, relative excess risk for interaction; SNP, single nucleotide polymorphism.
[a]Interaction analyses are adjusted for both SNPs, age and sex
In the absence of additive interaction, the RERI and the synergy index have values of 0 and 1, respectively.

over time due to changes in factors such as arsenic exposure or nutritional status, and may not consistently reflect an individual's long-term average capacity to methylate arsenic.

Our MR analysis is limited by the fact that we only have metabolite data for a subset of 2060 genotyped HEALS cohort participants, which includes very few cases ($n = 67$). However, because this is a randomly selected sub-cohort (most of whom were free of skin lesions at the time of measurement), the association estimates obtained (Figure 3) represent the associations present in the population and are appropriate for predicting arsenic metabolite percentages in larger samples of cases and controls. Simulated data from our group show that these subsample estimators are unbiased when IVs are strong ($F > 10$). We are also limited by the fact that urinary arsenic metabolites percentages are a proxy for the true underlying construct of interest, arsenic methylation capacity. Whereas urinary metabolites percentages are a well-established measure of methylation capacity, these percentages may be influenced by arsenic exposure[18] and perhaps additional unmeasured factors that contribute to measurement error. However, one advantage of MR is that classical measurement error in the risk factor does not result in bias or substantial reductions in power.[33] Power for detecting arsenic-SNP interaction was limited by the fact that we had prospective arsenic exposure data for only 493 genotyped cases and 2823 controls (from HEALS). This lack of prospective exposure data for cases from BEST prohibited us from performing IV analyses that were stratified by arsenic exposure.

Ideally, the genetic factors used as IVs should be identified in an independent dataset, as the association estimates SNPs identified in a genome-wide association study may be biased away from the null (i.e. winner's curse),[44] potentially biasing the MR estimate. However, because 10q24.32 variants are well-established as determinants of arsenic metabolism across several populations,[23] our SNP selection can be viewed as a regional association scan. Thus, this potential bias is likely much less than is expected for a genome-wide search. However, to protect against this potential bias, we used a cross-validation procedure (described in the Supplementary Methods, available as Supplementary data at *IJE* online) to show that we arrive at the same conclusion when using independent datasets for IV discovery and MR analysis (Supplementary Table 2, available as Supplementary data at *IJE* online). This cross-validation procedure was possible for DMA% and MMA%, but not for iAs%, as the associations between 10q24.32 SNPs and iAs% were too weak to be

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022

consistently detected when using only half of the first-stage sample.

The variants used for MR in this work may not be the best predictors of methylation capacity in other populations. Variation in this region has been linked to methylation capacity in several populations,[23] but the causal genetic variation underlying the 10q24.32 association signals remains unknown. Due to differences in linkage disequilibrium patterns and allele frequencies across populations, and potential differences in the underlying causal variation, future studies should consider the possibility that other 10q24.32 SNPs may be stronger predictors of arsenic methylation capacity than the SNPs described here. In addition, the effects of 10q24.32 variants on arsenic methylation capacity may vary by exposure status (Supplementary Table 5, available as Supplementary data at *IJE* online). Our results suggest that the association between rs11191527 and DMA% is weak or absent among individuals with low exposure ($<16\,\mu g/l$), indicating that this variant may not be a strong IV for methylation capacity in populations with low exposure. Additional research on this potential SNP-arsenic interaction is needed, and researchers considering using 10q24.32 variants for MR should consider the implications of such an interaction for power and interpretation of results.

In summary, we have used MR to provide strong evidence that arsenic metabolism efficiency has a causal role in arsenic toxicity. We further demonstrate this causal relationship by showing that the effects of metabolism-related variants are stronger among individuals with high arsenic exposure. The MR methods demonstrated here represent a powerful approach for making causal inferences about the effects of arsenic and arsenic metabolism on health. This approach may be valuable for studying health conditions where the role of arsenic is less clear, such as diabetes,[45] cardiovascular conditions,[8] non-malignant respiratory illnesses[46] and neurological diseases,[10,11] even when prospective data on arsenic exposure are not available. Developing interventions that increase arsenic methylation capacity are very likely to reduce the impact of arsenic exposure on health.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

This work was supported by National Institutes of Health (NIH) grants R01ES020506, P42ES010349, R01CA102484, R01CA107431 and P30CA014599.

Conflict of interest: None declared.

KEY MESSAGES

- Epidemiological studies suggest that arsenic metabolism efficiency is inversely associated with risk for arsenical skin lesions, presumably due to enhanced excretion of arsenic metabolites.
- Genetic variants in the 10q24.32 region are known to influence arsenic metabolism efficiency, and these SNPs can be used to assess the causal relationship between metabolism efficiency and arsenic toxicities using a Mendelian randomization approach.
- Using data on two 10q24.32 SNPs in an arsenic-exposed Bangladeshi cohort, we demonstrate that the association between arsenic metabolism efficiency and arsenical skin lesions is very likely to be causal.
- We demonstrate evidence of additive interaction between 10q24.32 genotype and arsenic exposure in relation to skin lesion risk, providing additional support for this causal relationship.
- Developing interventions that increase arsenic metabolism efficiency are very likely to reduce the impact of arsenic exposure on health.

References

1 World Health Organization. *United Nations Systhesis Report on Arsenic in Drinking Water*. Geneva: WHO, 2001.
2 Celik I, Gallicchio L, Boyd K *et al*. Arsenic in drinking water and lung cancer: a systematic review. *Environ Res* 2008;**108:**48–55.
3 Mink PJ, Alexander DD, Barraj LM, Kelsh MA, Tsuji JS. Low-level arsenic exposure in drinking water and bladder cancer: a review and meta-analysis. *Regul Toxicol Pharmacol* 2008;**52:**299–310.
4 Liu J, Waalkes MP. Liver is a target of arsenic carcinogenesis. *Toxicol Sci* 2008;**105:**24–32.
5 Yu HS, Liao WT, Chai CY. Arsenic carcinogenesis in the skin. *J Biomed Sci* 2006;**13:**657–66.
6 Chen CJ, Chen CW, Wu MM, Kuo TL. Cancer potential in liver, lung, bladder and kidney due to ingested inorganic arsenic in drinking water. *Br J Cancer* 1992;**66:**888–92.

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022

[7] Yuan Y, Marshall G, Ferreccio C *et al*. Kidney cancer mortality: fifty-year latency patterns related to arsenic exposure. *Epidemiology* 2010;**21:**103–08.
[8] States JC, Srivastava S, Chen Y, Barchowsky A. Arsenic and cardiovascular disease. *Toxicol Sci* 2009;**107:**312–23.
[9] Vahter M. Effects of arsenic on maternal and fetal health. *Annu Rev Nutr* 2009;**29:**381–99.
[10] Vahidnia A, van der Voet GB, de Wolff FA. Arsenic neurotoxicity – a review. *Hum Exp Toxicol* 2007;**26:**823–32.
[11] Brinkel J, Khan MH, Kraemer A. A systematic review of arsenic exposure and its social and mental health effects with special reference to Bangladesh. *Int J Environ Res Public Health* 2009;**6:**1609–19.
[12] Argos M, Kalra T, Rathouz PJ *et al*. Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in Bangladesh (HEALS): a prospective cohort study. *Lancet* 2010;**376:**252–58.
[13] Sohel N, Persson LA, Rahman M *et al*. Arsenic in drinking water and adult mortality: a population-based cohort study in rural Bangladesh. *Epidemiology* 2009;**20:** 824–30.
[14] Argos M, Kalra T, Pierce BL *et al*. A prospective study of arsenic exposure from drinking water and incidence of skin lesions in Bangladesh. *Am J Epidemiol* 2011;**174:** 185–94.
[15] Tseng CH. A review on environmental factors regulating arsenic methylation in humans. *Toxicol Appl Pharmacol* 2009;**235:**338–50.
[16] Pierce BL, Kibriya MG, Tong L *et al*. Genome-wide association study identifies chromosome 10q24.32 variants associated with arsenic metabolism and toxicity phenotypes in Bangladesh. *PLoS Genet* 2012;**8:**e1002522.
[17] Kile ML, Hoffman E, Rodrigues EG *et al*. A pathway-based analysis of urinary arsenic metabolites and skin lesions. *Am J Epidemiol* 2011;**173:**778–86.
[18] Ahsan H, Chen Y, Kibriya MG *et al*. Arsenic metabolism, genetic susceptibility, and risk of premalignant skin lesions in Bangladesh. *Cancer Epidemiol Biomarkers Prev* 2007;**16:**1270–78.
[19] Lindberg AL, Rahman M, Persson LA, Vahter M. The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. *Toxicol Appl Pharmacol* 2008;**230:**9–16.
[20] Valenzuela OL, Borja-Aburto VH, Garcia-Vargas GG *et al*. Urinary trivalent methylated arsenic species in a population chronically exposed to inorganic arsenic. *Environ Health Perspect* 2005;**113:**250–54.
[21] Gao J, Yu J, Yang L. Urinary arsenic metabolites of subjects exposed to elevated arsenic present in coal in Shaanxi Province, China. *Int J Environ Res Public Health* 2011;**8:**1991–2008.
[22] Engstrom K, Vahter M, Mlakar SJ *et al*. Polymorphisms in arsenic (+III oxidation state) methyltransferase (AS3MT) predict gene expression of AS3MT as well as arsenic metabolism. *Environ Health Perspect* 2011;**119:**182–88.
[23] Agusa T, Fujihara J, Takeshita H, Iwata H. Individual variations in inorganic arsenic metabolism associated with AS3MT genetic polymorphisms. *Int J Mol Sci* 2011; **12:**2351–82.
[24] Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32:**1–22.
[25] Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;**27:**1133–63.
[26] Ahsan H, Chen Y, Parvez F *et al*. Health Effects of Arsenic Longitudinal Study (HEALS): description of a multidisciplinary epidemiologic investigation. *J Expo Sci Environ Epidemiol* 2006;**16:**191–205.
[27] Verret WJ, Chen Y, Ahmed A *et al*. A randomized, double-blind placebo-controlled trial evaluating the effects of vitamin E and selenium on arsenic-induced skin lesions in Bangladesh. *J Occup Environ Med* 2005;**47:**1026–35.
[28] Purcell S, Neale B, Todd-Brown K *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81:**559–75.
[29] Nixon DE, Mussmann GV, Eckdahl SJ, Moyer TP. Total arsenic in urine: palladium-persulfate vs nickel as a matrix modifier for graphite furnace atomic absorption spectrophotometry. *Clin Chem* 1991;**37:**1575–79.
[30] Ahsan H, Chen Y, Parvez F *et al*. Arsenic exposure from drinking water and risk of premalignant skin lesions in Bangladesh: baseline results from the Health Effects of Arsenic Longitudinal Study. *Am J Epidemiol* 2006;**163:** 1138–48.
[31] Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 2012; **44:**821–24.
[32] Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2010;**39:**1037–45.
[33] Pierce BL, Vanderweele TJ. The effect of non-differential measurement error on bias, precision and power in Mendelian randomization studies. *Int J Epidemiol* 2012; **41:**1383–93.
[34] Chen Y, van Geen A, Graziano JH *et al*. Reduction in urinary arsenic levels in response to arsenic mitigation efforts in Araihazar, Bangladesh. *Environ Health Perspect* 2007;**115:**917–23.
[35] Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology* 1992;**3:**452–56.
[36] Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. Confidence intervals for measures of interaction. *Epidemiology* 1996;**7:**286–90.
[37] Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med* 2011;**30:**1312–23.
[38] Lindberg AL, Kumar R, Goessler W *et al*. Metabolism of low-dose inorganic arsenic in a central European population: influence of sex and genetic polymorphisms. *Environ Health Perspect* 2007;**115:**1081–86.
[39] Hall MN, Gamble MV. Nutritional manipulation of one-carbon metabolism: effects on arsenic methylation and toxicity. *J Toxicol* 2012;**2012**; article ID 595307.
[40] Gamble MV, Liu X, Ahsan H *et al*. Folate and arsenic metabolism: a double-blind, placebo-controlled folic acid-supplementation trial in Bangladesh. *Am J Clin Nutr* 2006;**84:**1093–101.
[41] Gamble MV, Liu X, Slavkovich V *et al*. Folic acid supplementation lowers blood arsenic. *Am J Clin Nutr* 2007;**86:** 1202–09.
[42] Engstrom KS, Hossain MB, Lauss M *et al*. Efficient arsenic metabolism – the AS3MT haplotype is associated with DNA methylation and expression of multiple genes around AS3MT. *PLoS One* 2013;**8:**e53732.
[43] Glymour MM, Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022

evaluating the instrumental variable assumptions. *Am J Epidemiol* 2012;**175:**332–39.

44 Zhong H, Prentice RL. Correcting 'winner's curse' in odds ratios from genomewide association findings for major complex human diseases. *Genet Epidemiol* 2010;**34:** 78–91.

45 Navas-Acien A, Silbergeld EK, Streeter RA, Clark JM, Burke TA, Guallar E. Arsenic exposure and type 2 diabetes: a systematic review of the experimental and epidemiological evidence. *Environ Health Perspect* 2006;**114:** 641–48.

46 Smith AH, Marshall G, Yuan Y, Liaw J, Ferreccio C, Steinmaus C. Evidence from Chile that arsenic in drinking water may increase mortality from pulmonary tuberculosis. *Am J Epidemiol* 2011;**173:**414–20.

47 Thornton T, McPeek MS. ROADTRIPS: case-control association testing with partially or completely unknown population and pedigree structure. *Am J Hum Genet* 2010;**86:**172–84.

48 Skrondal A. Interaction as departure from additivity in case-control studies: a cautionary note. *Am J Epidemiol* 2003;**158:**251–58.

Downloaded from https://academic.oup.com/ije/article/42/6/1862/739417 by U.S. Department of Justice user on 16 August 2022