A Bayesian framework for estimating disease risk due to exposure to uranium mine and mill waste on the Navajo Nation

Presenter: Lauren Hund, Ph.D.
Translator from English into Russian: Elena O'Donald, Ph.D.

The University of New Mexico
Albuquerque, NM, USA

Wednesday, November 19, 2014

- **Oversight, Review and Approval.** All research presented here is reviewed and approved by Navajo Nation Human Research Review Board (NNHRRB) and UNM Human Research Review Committee.

- **Collaborators.** This presentation contains joint work with Ed Bedrick, Curtis Miller, Gabriel Huerta, Teddy Nez, Sandy Ramone, Chris Shuey, Miranda Cajero, and Johnnye Lewis.

- **Funding Acknolwedgement.** This work was funded by research and infrastructure grants from the following: The National Institute for Environmental Health Sciences (NIEHS) RO1 ES014565; R25 ES013208; P30 ES-012072; DHHS/NIH/NCRR: # 1UL1RR031977; and Contracts and in-kind support from USEPA Region 9; and Navajo Nation Environmental Protection Agency (NNEPA).

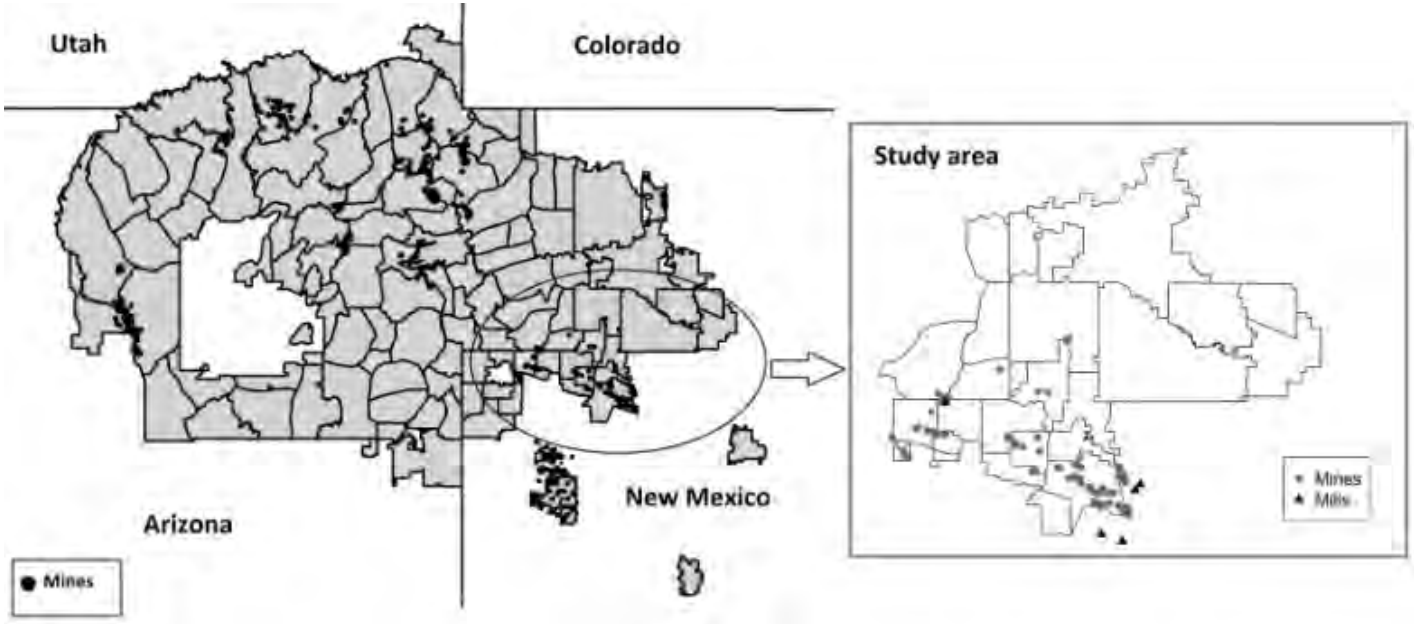
Goal

Assess whether kidney disease, diabetes, and hypertension are associated with uranium mining activities on the Navajo Nation.

Background

- Rates of chronic disease are elevated among the Navajo though the chronic health impacts of exposure to unremediated sites is unknown.
- The DiNEH team conducted a cross-sectional survey ($n = 1,304$) to learn about health, sources of uranium waste exposures, and water- and land-use practices.
- This paper presents results attempting to quantify the health impacts of uranium exposure on chronic disease among the Navajo.

Abandoned mines, milling sites, and waste piles remain in more than 50% of the chapters.



Exposure

Two distinct periods of exposure are prevalent on the Navajo Nation:

- historic **active** exposure from the mining period and
- current **legacy** exposure ongoing in relation to the unremediated waste sites.

Exposures during the active mining period were likely at higher concentrations and likely occurred through different exposure pathways than ongoing legacy waste exposure.

	Female		Male		Total	
	%	n	%	n	%	n
Disease outcomes						
Ever had **kidney disease**	4.9	36	5.5	31	**5.1**	67
Ever had **diabetes**	27.3	201	22.2	126	**25.1**	327
Ever had **high blood pressure**	35.5	261	36.4	207	**35.9**	468
Active Exposure						
Worked in uranium mine	2.4	18	19.4	110	**9.8**	128
Worked in uranium mill	0.4	3	3.7	21	**1.8**	24
Worked on reclamation or hauled ore	0.3	2	4.6	26	**2.1**	28
Washed or handled clothes	21.9	161	20.1	114	**21.1**	275
Lived in mining camp	3.1	23	4.4	25	**3.7**	48
Any active exposure	22.8	168	28.4	161	**25.2**	329
Legacy Exposure						
Used materials from abandoned site	14.4	106	16.7	95	**15.4**	201
Sheltered livestock in abandoned mine	1.2	9	2.6	15	**1.8**	24
Herded livestock near contaminated site	12.6	93	12.7	72	**12.7**	165
Contacted contaminated water	11.7	86	14.3	81	**12.8**	167
Played near contaminated site	10.3	76	14.1	80	**12.0**	156
Played on uranium tailings pile or dump	11.0	81	14.6	83	**12.6**	164
Any legacy exposure	25.7	189	33.1	188	**28.9**	377

Modeling framework

Goal: Examine data associations between chronic disease outcomes (kidney disease, diabetes, and hypertension) and mining exposures within an epidemiological causal inference framework.

Logistic regression is the standard model for binary outcomes.

Important modeling considerations include:

1. Bayesian versus frequentist model.
2. Multivariate (model all diseases together) versus univariate (model each disease separately) outcome model.
3. Appropriate effect measures for communicating risk.
4. Method for confounding adjustment - propensity score methods (weighting, adjustment, or matching) versus regression adjustment.
5. Sensitivity analysis for unmeasured confounding.

Bayesian modeling framework

We model the multivariate vector of disease outcomes (kidney disease, diabetes, and hypertension) using a Bayesian multivariate t-link model (O'brien and Dunson 2004) to quantify disease-exposure associations, controlling for confounding variables.

Rationale for multivariate Bayesian model choice:

- The study team hypothesizes a common mechanism underlying these diseases that is related to environmental exposures.
- Similar to a copula model, resulting in marginal logistic regression models and log-odds ratio parameter interpretations.
- With approximation, posterior sampling is simple and efficient.
- Avoid large-sample frequentist assumptions.

Multivariate t-link model

Basic idea: Model the vector of binary disease outcomes using a latent multivariate-t random variable.

- With appropriate choice of the t-parameters, regression coefficients have approximate log-odds ratio interpretations.
- Model for the linear predictor:

$$v_i^j = E_i \beta_E^j + \boldsymbol{X}_i^j \boldsymbol{\beta}^j$$

  - E_i is binary exposure (active or legacy, depending on the model).
  - $\boldsymbol{X}_i = [\boldsymbol{X}_i^K, \boldsymbol{X}_i^D, \boldsymbol{X}_i^H]$ is a vector of confounders for $j \in \{K, D, H\}$.

- Using this approximation, computation of full conditionals is based on simple Gibbs and Metropolis Hastings steps.
- Specify informative g-priors on the regression coefficients (Hanson *et. al* 2014) and informative normal priors on the correlation parameters of $\boldsymbol{R}$.

Effect estimates

1. Disease-specific **conditional odds ratio** for exposed versus unexposed, $OR^j = exp(\beta_E^j)$ for $j \in \{K, H, D\}$.

2. Disease-specific **risk difference** (average treatment effect).
   - Using a counterfactual framework, define $p_I^j(e)$ as the prevalence of disease in the sampled population if everyone had exposure level e and estimate posterior density of $p_I^j(1) - p_I^j(0)$.

3. **Multiple disease risk difference**.
   - Define $p^{M_t}(e)$ as the sample prevalence of having t or more diseases if everyone had exposure level e and estimate density of $p^{M_t}(1) - p^{M_t}(0)$.

Confounding adjustment

Conditional on a small set of known confounders, we use Bayesian model averaging (BMA) to average over models with various functional forms.

- With a small set of confounders, modeling the outcome is often considered a safer strategy than modeling the assignment mechanism.
- We used BMA to account for uncertainty in how the confounding adjustment should occur.
- Model weights are defined as $p(M_k|D) \propto f(\boldsymbol{y}|M_k)p(M_k)$.

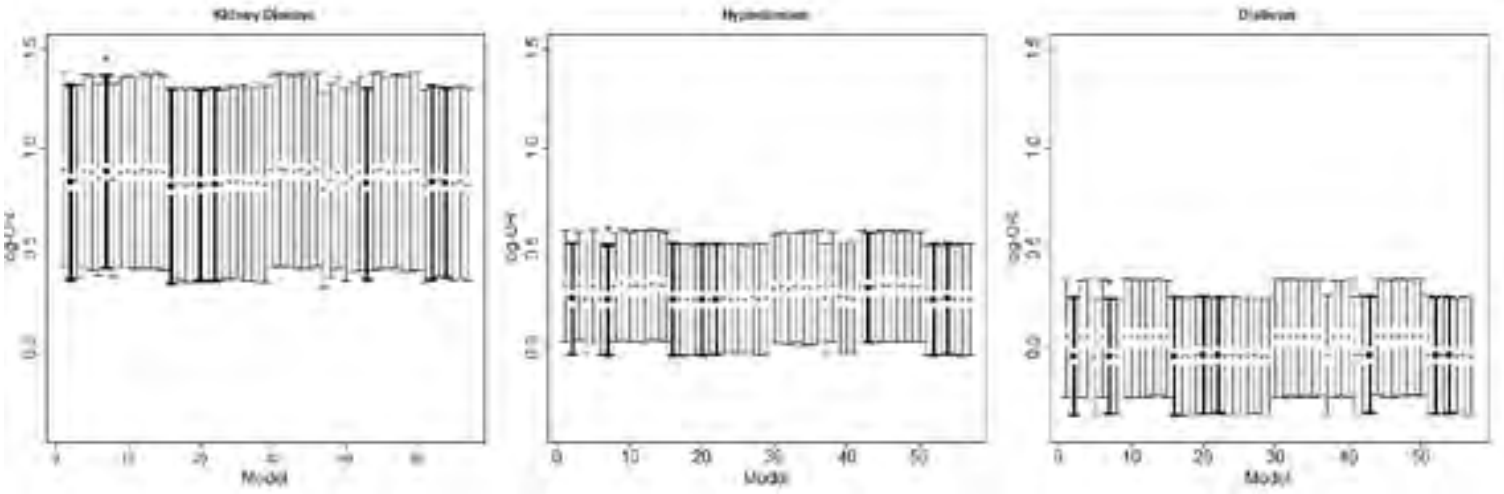
Analysis

- **Outcomes:** Kidney disease, diabetes, hypertension.
- **Exposures:** Active, legacy (both with and without controlling for active exposure).
- **Confounding variables:** Age, sex, family history of each disease, and education.

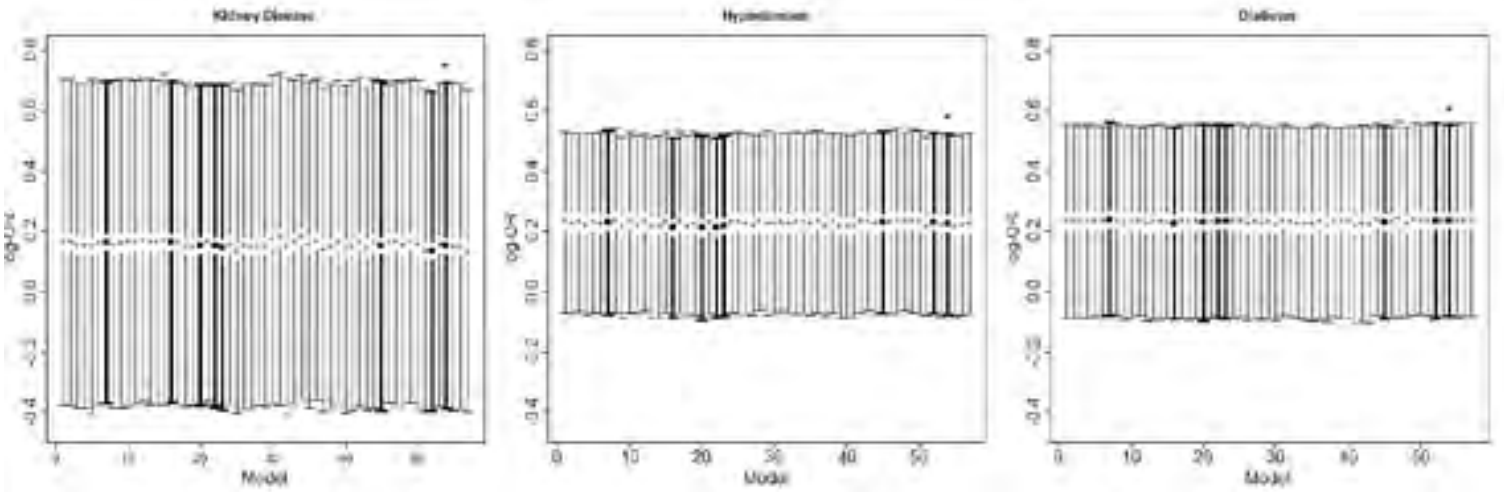
The posterior mean, standard deviation, and 95% credible intervals (CI) of the effect estimates are presented.

BMA summary: Posterior mean and 95% CIs for log-OR across models.

Kidney Disease **Hypertension** **Diabetes**



Active exposure



Legacy exposure, adjusting for active exposure

Results

	$\hat{OR}^j$ (95% CI)	$\hat{ATE}^j$ (sd)	95% CI	P
Kidney Disease				
Active	2.33 (1.43, 3.79)	.051 (.017)	.019, .086	.000
Legacy	1.62 (1.00, 2.64)	.028 (.015)	-.000, .058	.027
Legacy, adjusted	1.16 (0.68, 1.98)	.008 (.015)	-.020, .039	.291
Hypertension				
Active	1.28 (0.97, 1.69)	.046 (.027)	-.006, .100	.041
Legacy	1.33 (1.02, 1.74)	.054 (.026)	.003, .105	.019
Legacy, adjusted	1.25 (0.92, 1.70)	.043 (.029)	-.015, .100	.072
Diabetes				
Active	0.96 (0.72, 1.28)	-.006 (.024)	-.053, .041	.610
Legacy	1.19 (0.89, 1.59)	.028 (.025)	-.019, .077	.124
Legacy, adjusted	1.26 (0.92, 1.74)	.039 (.027)	-.015, .092	.077

Exposure health effects summary. Results include the posterior mean and sd of the disease-specific risk difference $\hat{ATE}^j$ (sd); a 95% CI; and the posterior probability that $\hat{ATE}^j < 0$ (P).

	$\hat{ATE}^{M_t}(sd)$	95% CI	P
Active			
$t = 1+$	.049 (.032)	-.013, .112	.066
$t = 2+$	.020 (.017)	-.013, .054	.128
$t = 3$	.003 (.001)	.001, .006	.000
Legacy			
$t = 1+$	.076 (.031)	.016, .134	.006
$t = 2+$	.039 (.018)	.006, .076	.010
$t = 3$	.002 (.001)	.000, .005	.014
Legacy, adjusted			
$t = 1+$	.068 (.035)	-.003, .134	.032
$t = 2+$	.038 (.029)	-.001, .076	.027
$t = 3$	.001 (.001)	-.001, .003	.153

Multiple disease risk differences for $t = 1+, 2+,$ or 3 diseases. Results include the posterior mean and sd $\hat{ATE}^{M_t}(sd)$, a 95% CI, and posterior probability that the $ATE^{M_t} < 0$ (P).

Conclusion

Main results:

- Known risk factors for chronic disease are also important risk factors in this population.
- Evidence of associations between mining exposures and chronic disease after controlling for known risk factors.

Limitations:

- Self-reported exposure and outcome.
- Cross-sectional convenience sample.
- Aggregated definition of exposure.

Conclusion

We use a Bayesian causal inference framework to assess the relationship between exposure to uranium mine waste and chronic disease.

- Multivariate t model is computationally efficient and results in useful multivariate disease summaries.
- Model selection procedures often occur ‘behind the scenes’ and BMA facilitates transparency in methods.
- BMA approach assumes the set of confounders is known and computational intensity increases with number of confounders.
- Counterfactual “causal inference” framework helps establish conditions for a causality and informs sensitivity analyses.

Future directions

- Assess health effects associated with other exposure sources, such as contaminated drinking water or wind-blown dusts.
- Incorporate data from clinical assessments of disease on sub-sample of 267 participants.

Contact Information

Presenter:
Lauren Hund
E-mail: lbhund@gmail.com

Translator from English into Russian:
Elena O'Donald
E-mail: eodonald@unm.edu