

Использование Байесовой системы моделирования
для оценки рисков возникновения заболеваний
среди индейцев Навахо в результате их работы и
проживания вблизи урановых шахт и мест
скоплений отходов предприятий
по добыче и переработке урана

Докладчик: Лорен Ханд, Ph.D.

Переводчик с английского языка: Елена О'Дональд, Ph.D.

Университет Нью-Мексико (The University of New Mexico)
Альбукерке, штат Нью-Мексико, США

Среда, 19 Ноября 2014 года

- **Одобрение на проведение исследовательских работ.** Все исследовательские работы, представленные в настоящей презентации, были рассмотрены и одобрены этическими комитетами индейцев Навахо и Университета Нью-Мексико.
- **Со-авторы.** Материалы, представленные в данной презентации, включают результаты совместной работы с коллегами: Ed Bedrick, Curtis Miller, Gabriel Huerta, Teddy Nez, Sandy Ramone, Chris Shuey, Miranda Cajero и Johnnye Lewis.
- **Финансирование проекта.** Данное исследование было финансировано грантами Национального института экомедицинских наук (The National Institute for Environmental Health Sciences (NIEHS) RO1 ES014565; R25 ES013208; P30 ES-012072; DHHS/NIH/NCRR: #1UL1RR031977; контрактами и поддержкой USEPA Region 9; а также Агентством индейцев Навахо по защите окружающей среды (Navajo Nation Environmental Protection Agency (NNEPA)).

Проверить гипотезу о том, что заболевания почек, диабет и гипертония индейцев Навахо связаны с деятельностью расположенных на их территории предприятий по добыче и обработке урана.

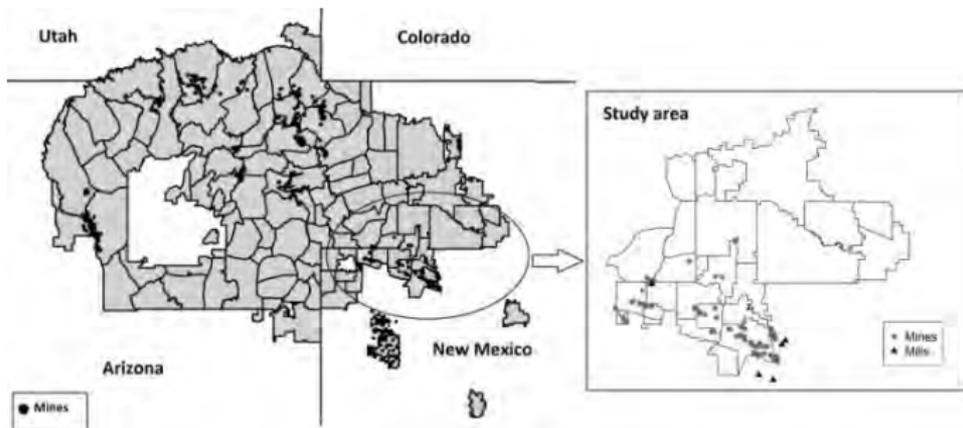
Предыстория

Наблюдается рост показателей хронических заболеваний индейцев Навахо, однако, причинно-следственная связь между токсинами загрязнённых территорий и хроническими заболеваниями среди населения пока ещё не изучена.

Группа исследователей проекта DiNEH провела кросс-секционное анкетирование населения ($n = 1,304$), затрагивающее темы здоровья населения, источников уранового загрязнения, а также использования водных и земельных ресурсов.

В настоящей презентации представлены результаты попытки количественной оценки уранового воздействия на хронические заболевания у индейцев Навахо.

Заброшенные шахты, фабрики и места скопления отходов горно-добывающей промышленности находятся в более чем половине населённых пунктов проживания индейцев Навахо.



Виды неблагоприятного воздействия окружающей среды на индейцев Навахо в основном подразделяются на:



- 1) **Активную форму** воздействия на человека во время его работы на горнодобывающих предприятиях (в прошлом) и
- 2) **Пассивную форму** - отравление отходами этих предприятий, лежащими на свалках (происходит в настоящее время).

По сравнению с пассивной формой, активное воздействие, скорее всего, характеризовалось более высокими концентрациями токсинов и иными траекториями воздействия на человека.

	Женщины		Мужчины		всего	
	%	n	%	n	%	n
<u>Болезни в настоящем или прошлом</u>						
Заболевания почек	4.9	36	5.5	31	5.1	67
Диабет	27.3	201	22.2	126	25.1	327
Гипертония	35.5	261	36.4	207	35.9	468
<u>Активное воздействие (AB)</u>						
Работа в урановой шахте	2.4	18	19.4	110	9.8	128
Работа на фабрике переработки урана	0.4	3	3.7	21	1.8	24
Утилизация и перевозка руды	0.3	2	4.6	26	2.1	28
Стирка и контакт с одеждой шахтёров	21.9	161	20.1	114	21.1	275
Проживание в шахтёрском городке	3.1	23	4.4	25	3.7	48
Имеется хотя бы одно AB	22.8	168	28.4	161	25.2	329
<u>Пассивное воздействие (ПВ)</u>						
Использование материалов с шахт	14.4	106	16.7	95	15.4	201
Держат скот в заброшенных шахтах	1.2	9	2.6	15	1.8	24
Выпас скота в зоне загрязнения	12.6	93	12.7	72	12.7	165
Контакт с загрязнённой водой	11.7	86	14.3	81	12.8	167
Играли в зоне загрязнения	10.3	76	14.1	80	12.0	156
Играли на кучах урановых отбросов	11.0	81	14.6	83	12.6	164
Имеется хотя бы одно ПВ	25.7	189	33.1	188	28.9	377

Цель: В рамках эпидемиологической причинно-следственной системы исследовать связь между хроническими заболеваниями (болезнями почек, диабетом и гипертонией) и воздействием результатов деятельности горно-добывающих предприятий. Для оценки бинарных результатов (двоичного типа) обычно используется логическая регрессия. При выборе стратегии моделирования важно учесть следующее:

- 1. Байесова модель (Bayesian model) против фреквентистской модели (frequentist model).**
 - ❖ Фреквентистский подход к вероятностной модели основан на изучении доли успешных экспериментов (в нашем случае - заболеваний) при стремлении к бесконечности общего числа экспериментов.
 - ❖ Байесовский подход к вероятностному моделированию предполагает самообучение (обновление сложившихся представлений в свете полученного опыта) и рассматривает вероятность с позиции степени доверия, используя так называемое правдоподобие (likelihood) - вероятность данных при условии зафиксированных параметров модели.
- 2. Многомерная (multivariate) модель против одномерной (univariate) модели.** В многомерной модели все три заболевания (зависимые переменные) рассматриваются вместе, во взаимосвязи друг с другом. Одномерные модели строятся отдельно для каждого заболевания.
- 3. Адекватная оценка рисков заболеваний и форма их представления.** Нужно получить общепринятые в литературе количественные результаты оценки риска здоровью под воздействием вредных факторов окружающей среды.
- 4. Выбор метода корректировки ошибок смешения (confounding adjustment):** метод подбора контрольной группы по индексу соответствия (A: propensity score methods) против корректировки регрессии (B: regression adjustment). Метод (A) моделирует экспозицию как функцию ошибок смешения, а метод (B) моделирует результат (outcome) как функцию ошибок смешения.
- 5. Анализ чувствительности к неизмеренным факторам смешения,** оказывающим влияние на зависимую переменную (заболевание).

Байесова система моделирования

Для того, чтобы определить количественные характеристики связей между неблагоприятным воздействием окружающей среды и заболеванием населения, а также учесть переменные смешения, оказывающие своё влияние на эти заболевания, мы воспользовались Байесовой многомерной моделью с t -распределением (**Bayesian multivariate t-link model**: O'Brien and Dunson 2004). С её помощью мы моделируем многомерный вектор выходных параметров системы (зависимых переменных) – заболеваний почек, диабета и гипертонии.

Наш выбор многомерной Байесовой модели основан на следующих соображениях:

- ❖ Согласно нашей гипотезе, в основе этих заболеваний лежит общий механизм, связанный с неблагоприятным экологическим воздействием. Наша модель позволяет учесть взаимные связи между всеми тремя заболеваниями.
- ❖ Как мы говорили ранее, для оценки бинарных зависимых переменных обычно используется логистическая регрессия (logistic regression). Выбранная нами модель основана на одномерных логистических регрессиях. В результате мы получаем общепринятые в эпидемиологии отношения шансов и в то же время учитываем корреляционные связи между тремя заболеваниями.
- ❖ Байесовы модели порой бывают достаточно сложны. Однако модель с t -распределением допускает определённые приближения, тем самым позволяя упростить процедуру выборки и делая её более эффективной. В результате нам становятся не нужны большие объёмы выборок, принятые при использовании фреквентистского (frequentist) подхода моделирования.

Многомерная модель с t-распределением

Основная идея: Моделировать вектор бинарных выходных параметров (в нашем случае – трёх заболеваний), используя латентную многомерную случайную переменную t .

- ❖ Если правильно выбрать t -параметры, то коэффициенты регрессии могут быть приблизительно интерпретированы как логит (логарифмированное отношение шансов: log-odds ratio).
- ❖ Модель для линейной независимой переменной (прогнозирующего параметра) выглядит так:

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

... E_i - бинарная экспозиция (активная или пассивная, в зависимости от модели).

... $\mathbf{X}_i = [\mathbf{X}_i^K, \mathbf{X}_i^D, \mathbf{X}_i^H]$ – вектор ошибок смешения для $j \in \{K, D, H\}$.

Где:

K=kidney disease – заболевание почек,

H=hypertension – гипертония и

D=diabetes - диабет.

Оценка эффекта неблагоприятного воздействия 10

- 1 Индивидуальное для каждого заболевания **условное отношение шансов (conditional odds ratio)** для подвергнувшихся неблагоприятному воздействию против неподвергнувшихся:

$$OR^j = \exp(\beta^j_E) \text{ for } j \in \{K, H, D\}.$$

Где: K=kidney disease – заболевание почек, H=hypertension – гипертония и D=diabetes - диабет.

- 2 Индивидуальная для каждого заболевания **разница рисков (risk difference)** – это средняя величина эффекта воздействия (average treatment effect).

Используя контрафактическую среду (counterfactual framework), определяем $p^j_i(e)$ как распространенность (prevalence) заболевания в выборке, как если бы уровень неблагоприятного воздействия на каждого человека был равен e , а оценочная апостериорная плотность (estimate posterior density) равна $p(1) - p(0)$.

В контрафактической среде данные анализируются в сослагательной форме: “если бы... то бы..”

- 3 **Разница риска нескольких заболеваний (multiple disease risk difference).**

... Определяем $p^{Mt}(e)$ как распространённость в выборке (sample prevalence) наличия t или более заболеваний, как если бы уровень неблагоприятного воздействия на каждого человека был равен e , а оценочная плотность (estimate density) равна $p^{Mt}(1) - p^{Mt}(0)$.

Корректировка ошибок смешения (confounding adjustment)

Полагаясь на небольшое число известных нам факторов смешения (влияющих на рассматриваемые нами заболевания), мы используем **Байесову модель усреднения (Bayesian model averaging, BMA)** для усреднения моделей с различной функциональной формой.

- ❖ При наличии небольшого числа факторов смешения зачастую лучше смоделировать выходные параметры (в нашем случае - заболевания), чем составлять модель механизма присвоения (assignment mechanism).
- ❖ Мы использовали модель BMA для того, чтобы принять в расчёт то, что мы точно не знаем, как именно нам следует провести корректировку ошибок смешения.
- ❖ Величины веса модели определяются как:

$$p(M_k | D) \propto f(\mathbf{y} | M_k) p(M_k).$$

Выходные параметры системы (зависимые переменные, *outcomes*): болезни почек, диабет и гипертония.

Входные параметры системы (независимые переменные, *exposures*) - неблагоприятное воздействие окружающей среды: активное и пассивное (оба вида экспозиции с учётом или без учёта поправки на активную форму неблагоприятного воздействия).

- ❖ **Активная форма** воздействия на человека осуществлялась в прошлом во время его работы на горно-добывающих предприятиях.
- ❖ **Пассивная форма** воздействия – происходящее в настоящее время отравление населения отходами этих предприятий.

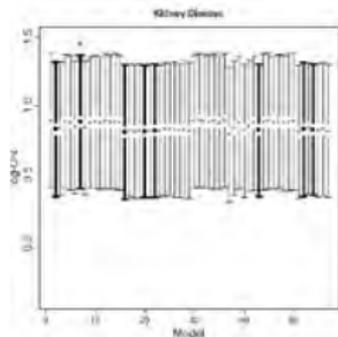
Параметры возмущения системы (параметры ошибок смешения, *confounding variables*): возраст, пол и образование участника исследования, а также наличие данного заболевания у его близких родственников.

Для оценки эффекта неблагоприятного воздействия нами вычислены: апостериорное среднее (posterior mean), стандартное отклонение (standard deviation) и 95% доверительный интервал (credible intervals, CI).

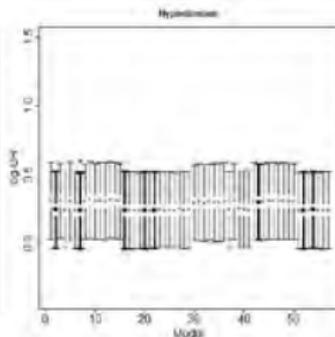
Апостериорное среднее – это среднее арифметическое (mean) распределения (posterior distribution), полученного в результате моделирования (т.е. после того, как мы учли данные).

Результаты ВМА моделей: апостериорное среднее (posterior mean) и 95% CI для логарифмированного отношения шансов (log-OR).

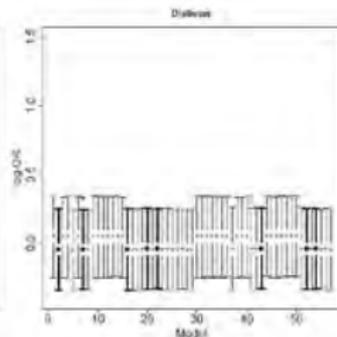
Болезни почек



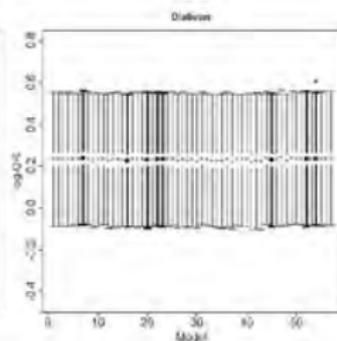
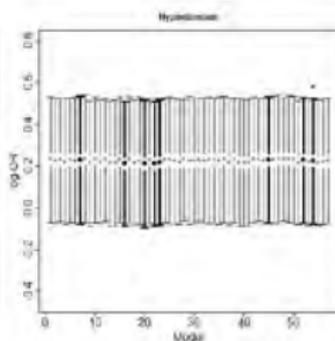
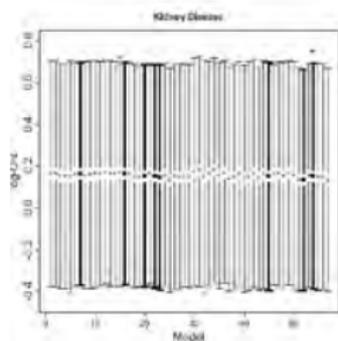
Гипертония



Диабет



Активное неблагоприятное воздействие



Пассивное воздействие, с поправкой на активное неблагоприятное воздействие

Результаты моделирования

Воздействие:
 АВ=активное,
 ПВ=пассивное

14

	\hat{OR}^j (95% CI)	\hat{ATE}^j (sd)	95% CI	<i>P</i>
<u>Болезни почек</u>				
АВ	2.33 (1.43, 3.79)	.051 (.017)	.019, .086	.000
ПВ	1.62 (1.00, 2.64)	.028 (.015)	-.000, .058	.027
ПВ с учётом АВ	1.16 (0.68, 1.98)	.008 (.015)	-.020, .039	.291
<u>Гипертония</u>				
АВ	1.28 (0.97, 1.69)	.046 (.027)	-.006, .100	.041
ПВ	1.33 (1.02, 1.74)	.054 (.026)	.003, .105	.019
ПВ с учётом АВ	1.25 (0.92, 1.70)	.043 (.029)	-.015, .100	.072
<u>Диабет</u>				
АВ	0.96 (0.72, 1.28)	-.006 (.024)	-.053, .041	.610
ПВ	1.19 (0.89, 1.59)	.028 (.025)	-.019, .077	.124
ПВ с учётом АВ	1.26 (0.92, 1.74)	.039 (.027)	-.015, .092	.077

Влияние экспозиции на здоровье. Результаты включают апостериорное среднее (posterior mean) и стандартное отклонения (sd) индивидуальной для каждого заболевания разницы риска (risk difference) \hat{ATE}^j (sd); 95% CI (95% доверительный интервал), а также апостериорную вероятность того, что $\hat{ATE}^j < 0$ (*P*).

Воздействие:

AB=активное,
 ПВ=пассивное

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

Разницы риска нескольких заболеваний (multiple disease risk differences) для *t* = 1+, 2+, или 3 заболеваний. Результаты включают апостериорное среднее (posterior mean) и стандартное отклонение (sd) $\widehat{ATE}^{M_t} (sd)$, а также 95% CI (95% доверительный интервал) и апостериорную вероятность (posterior probability) того, что $ATE^{M_t} < 0$ (*P*).

Основные результаты:

- ❖ Известные факторы риска хронических заболеваний [человека] также присутствуют и являются важными факторами риска среди данного населения (индейцев племени Навахо).
- ❖ Установлена связь между воздействием результата деятельности предприятий по добыче и обработке [урана] и хроническими заболеваниями человека (с учётом поправки на известные факторы риска).

Ограничения исследования (limitations):

- ❖ Данные о факторах благоприятного воздействия и наличия хронических заболеваний были собраны в результате анкетирования населения, самостоятельно заполнившего анкеты.
- ❖ В исследовании использована кросс-секционная нерепрезентативная выборка (взятая из соображения удобства исследования) - cross-sectional convenience sample.
- ❖ При анализе данных использовано составное определение неблагоприятного воздействия (aggregated definition of exposure).

Для установления связи между неблагоприятным воздействием результатов деятельности предприятий по добыче урана и хроническими заболеваниями населения этих регионов мы воспользовались Байесовой системой причинно-следственного анализа.

- ❖ Многомерная модель с t -распределением эффективна с точки зрения вычислительного процесса и позволяет сделать практические выводы о взаимодействии нескольких выходных параметров (в нашем случае – трёх заболеваний).
- ❖ Процесс выбора подходящей модели зачастую скрыт от пользователя, но Байесова модель усреднения (Bayesian model averaging, BMA) способствует открытости этого процесса.
- ❖ Используемая моделью BMA подход подразумевает, что нам известны факторы смещения (влияющие на рассматриваемые нами заболевания). С увеличением числа этих факторов также растёт интенсивность вычислений.
- ❖ Выбранная нами система причинно-следственного анализа является контрафактической (то есть анализирующей данные в сослагательной форме: “если бы... то бы..”). Такой подход позволяет создать условия для рассмотрения причинно-следственных связей и предоставить данные для анализа чувствительности, выявляющего критические показатели проекта.

- ❖ Изучить последствия воздействия на здоровье человека других источников загрязнения, таких как питьевая вода и переносимая ветром пыль.
- ❖ Включить в анализ объективные, клинические данные по заболеваемости 267 участников исследования, включённых в подгруппу нашей выборки.

Докладчик:

Lauren Hund

E-mail: lbhund@gmail.com

Переводчик с английского языка на русский язык:

Elena O'Donald

E-mail: eodonald@unm.edu