Causal Inference in Epidemiology Using Bayesian Methods: The Example of Meta-Analysis of Statins and Fracture Risk

Lawrence McCandless
lmccandl@sfu.ca

Faculty of Health Sciences, Simon Fraser University, Vancouver Canada

Summer 2013
Overview
This talk is about Bayesian Causal Inference.

Meta-analysis of observational studies is a wonderful and neglected corner of causal inference.
How many people have heard of the drug called Statins?

Are these drugs really so beneficial and safe that almost everyone over 40 should take them? James Kingsland investigates

Statins for all

SO YOU think you're healthy? You are in your 40s, feel right as rain, normal blood pressure, normal cholesterol, pretty good diet, occasional exercise. How would you react if your doctor suggested you take a powerful drug every day for the rest of your life?

The drug, known as a statin, will lower your cholesterol even further and reduce your risk of a heart attack or stroke. According to one recent estimate, most men and many women over 40 could benefit from the drugs. If you disease has little to do with lowering cholesterol and instead results from their anti-inflammatory properties, leading some to dismiss them as "expensive aspirin". So could the rush to put millions more people on statins be a costly mistake?

The association between cholesterol, its transport in the bloodstream by a protein called low-density lipoprotein and heart disease is fairly well established. Cholesterol in the form of LDL, so-called "bad cholesterol", can infiltrate the walls of coronary arteries...
What are Statins?

Statins are blockbuster cholesterol drug (sales of $1 billion/year).

They reduce the risk of strokes and cardiovascular events by 25%-30%.

There has been intensive study of the health benefits of statins that are not related to cardiovascular disease, such as reducing arthritis, cancer, and fractures (e.g. hip fractures in the elderly).

Pharmaceutical companies can make a lot of $$$ from such research.
This talk concerns a tale of statins and fracture risk

Scientific question: Do statins reduce the risk of fractures (e.g. hip fractures)???
FACTS:

1. In 2000, three observational studies in JAMA and Lancet showed that statin use was associated with huge (>50%) reductions in fracture risk in elderly patients.

2. In the next 5 years, a torrent of observational studies were published that seemed to confirm the finding.

3. Statins increase bone formation in rats, which gives some biological basis of causation.
HOWEVER:

1. Causal inference from observational studies is prone to bias.

2. A 2007 re-analysis of 4 randomized trials of cardiovascular endpoints found NO reduction in fracture risk among statin users.

3. **Controversy in the literature:** Two epidemiologic analyses of the same database came to opposite conclusions about the health benefits of statins!

   Study #1 Fracture OR=0.12 95% CI (0.04 - 0.41)
   Study #2 Fracture OR=0.59 95% CI (0.31 - 1.13).

de Vries (2006) *Int J Epidemiol*
The statins and fractures controversy is now a textbook example of a "failure of epidemiology"

SPECIAL NEWS REPORT

Epidemiology Faces Its Limits

The search for subtle links between diet, lifestyle, or environmental factors and disease is an unending source of fear—but often yields little certainty

news about health comes thick and fast days, and it seems al-constitutorily con- ritory. In January of 2012, for instance, a news study found a sign-ificant association be- tween residential radon exposure and lung cancer. nadian study did not.

on the press for its reporting of epidemiology, and even on the public "for its unrealistic expectations" of what modern medical research can do for their health. But many epidemiologists interviewed by Science say the problem also lies with the very nature of epidemiologic studies—in Rothman, editor of the journal Epidemiology, "We're pushing the edge of what can be done with epidemiology."

With epidemiology stretched too far or beyond, says Dimitrios Trichopoulos, head of the epidemiology department at Harvard School of Public Health, "We will inevitably generate false positive results and false negative results "with disturbing frequency." Most epidemiologists are aware of the problem, he adds, "and tend
Postscript: Why were all the epidemiological studies biased?

Widespread belief: residual unmeasured confounding

There are some confounders that they didn’t control for.

What confounders??

Measurements of a healthy lifestyle and health seeking behaviours, including

- regular exercise
- healthy diet
- BMI
- Smoking
- Alcohol
- Use of preventative therapies (Mammogram, prostate cancer screening ...
Unmeasured Confounding

Epidemiologists call this **Health User Bias** (Henessey 2000 JAMA).

In fact, Statin users have lower risk of being in car crashes (Dormuth 2009, Circulation).
<table>
<thead>
<tr>
<th>First Author, Year (Reference Number)</th>
<th>Inclusion and Exclusion Criteria</th>
<th>Effect Measure</th>
<th>Confounder Adjustment Technique</th>
<th>Measured Covariates Available for Control of Confounding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chan 200022</td>
<td>Inclusion: Women aged ≥60 years. Exclusion: Previous major trauma, cancer, fracture, use of HRT, bisphosphonates, calcitonin, anticonvulsants, thyroid hormone.</td>
<td>Odds ratio</td>
<td>Conditional logistic regression</td>
<td>Sex, age, health maintenance organization, chronic disease score, number of hospital admissions in previous year, use of antipsychotics, hypnotics, antidepressants, thiazide diuretics, hypoglycemics, systemic steroids, other LLDs.</td>
</tr>
<tr>
<td>Meier 200023</td>
<td>Inclusion: Age 50 to 89 years. Exclusion: Osteoporosis, osteomalacia, cancer, alcoholism, use of bisphosphonates.</td>
<td>Odds ratio</td>
<td>Conditional logistic regression</td>
<td>Sex, age, date, general practice, years in database, smoking, BMI, oral or inhaled corticosteroids, HRT, number of physician visits before the fracture date.</td>
</tr>
<tr>
<td>Wang 200024</td>
<td>Inclusion: Age ≥ 50 years. Exclusion: Previous hip fracture.</td>
<td>Odds ratio</td>
<td>Logistic regression</td>
<td>Sex, age, race, health insurance, heart disease, heart failure, hypertension, diabetes, cancer, Charlson comorbidity score, healthcare utilization (number of medications, days hospitalized, days in nursing home, number of physician visits), use of HRT, oral corticosteroids, thiazides, psychoactive medication, other LLDs.</td>
</tr>
<tr>
<td>van Staa 200125</td>
<td>Inclusion: Age ≥ 50. Exclusion: None.</td>
<td>Odds ratio</td>
<td>Conditional logistic regression</td>
<td>Sex, age, general practice, smoking, BMI, diabetes, rheumatoid arthritis, hyperthyroidism, congestive heart failure, seizures, anemia, dementia, depression, psychotic disorder, cerebrovascular accident, COPD, use of anticonvulsants, Non-steroidal anti-inflammatory drugs, methotrexate, HRT, thiazides, anxiolytics/hypnotics, antipsychotics, antidepressants, anti-Parkinson drugs, systemic and inhaled corticosteroids, bronchodilators, other LLDs.</td>
</tr>
</tbody>
</table>
Suppose that there was a single binary confounder that was unmeasured in all 17 studies.

Could confounding bias account for the discrepancy between 17 observational studies versus the 4 RCTs?

Is that possible?
Sensitivity analysis in a meta-analysis

**Objective:** In this talk I will describe a new Bayesian procedure to adjust for bias in a meta-analysis (unmeasured confounding, and selection bias).

The procedure is an example of Bayesian bias modelling for observational studies:


Findings published in:

McCandless (2013) Statins and fracture risk: Can we quantify the healthy-user effect? *Epidemiology*

McCandless (2012) Meta-analysis of observational studies with unmeasured confounders. *Int J Biostatistics*
The Model: How to Adjust for a Missing Confounder?

Data:

- Log relative risk estimates $y_1, y_2, \ldots, y_k$, with
- Standard error estimates $\sigma_1, \sigma_2, \ldots, \sigma_k$

for the association between a dichotomous exposure and dichotomous outcome in $k$ different observational studies.

Studies could be case-control, cross-sectional or cohort studies.

We assume that $y_j$ been adjusted for a set of confounders in the $j^{th}$ study, which may depend on $j$. 
The Model: How to Adjust for a Missing Confounder?

Pooling RRs, ORs and HRs:

The incidence of osteoporitic fractures is roughly 5 cases/1000 person years.

When the disease incidence is low, then the relative risk is approximated by the odds ratio and by the hazard ratio (Jewell 2000).

Thus we allow for the possibility that some of the $y_1, \ldots, y_k$ are adjusted log ORs (from case-control/cross sectional studies via unconditional logistic) or are adjusted HRs (from cohort studies via Cox proportional hazards models).
The Model: How to Adjust for a Missing Confounder?

Bayesian random effects meta-analysis (Carlin 1992, Gelman et al. 2003):

Assume

\[ y_j \sim N(\theta_j^*, \sigma_j^2). \quad \text{for } j \in 1 : k \]

where \( \sigma_j^2 \) is assumed to be known.

**The quantity \( \theta_j^* \):**

The log relative risk for the association between statin use and fractures in study \( j \), conditional the set of covariates that are measured in study \( j \).
The Model: How to Adjust for a Missing Confounder?

Suppose that we assume that relative risks are estimated from a loglinear model. Following confounder model of Lin et al. (1997) *Biometrics*, we conceptualize

Reduced model

\[
\log P(Y = 1|X, C_j) = \alpha_j^* + \theta_j^* X + \xi_j^* T C_j
\]

Full model

\[
\log P(Y = 1|X, C_j) = \alpha_j + \theta_j X + \xi_j^T C_j + \lambda U
\]

We can alternatively use logistic or Cox PH regression.
The Model: How to Adjust for a Missing Confounder?

When incidence of disease is low, then we have

$$\theta_j^* \approx \theta_j + \Omega(\text{RR}_j, p_{1j}, p_{0j})$$

where

$$\Omega(\text{RR}_j, p_{1j}, p_{0j}) = \log \frac{\text{RR}_j \times p_{1j} + (1 - p_{1j})}{\text{RR}_j \times p_{0j} + (1 - p_{0j})}.$$

with bias parameters: $$(\text{RR}_j, p_{1j}, p_{0j})$$

Formula applies when $\theta_j^*$ and $\theta_j^*$ are obtained from logistic or Cox regression models provided the disease incidence is low.
The Model: How to Adjust for a Missing Confounder?

This gives us the model:

\[ y_j \sim N\left(\theta_j + \Omega(\text{RR}_j, \rho_{1j}, \rho_{0j}), \sigma_j^2\right) \quad \text{for } j \in 1 : k \]

with parameters:

\[ \theta = (\theta_1, \ldots, \theta_k) \]
\[ \text{RR} = (\text{RR}_1, \ldots, \text{RR}_k) \]
\[ \rho_1 = (\rho_{1,1}, \ldots, \rho_{1,k}) \]
\[ \rho_0 = (\rho_{0,1}, \ldots, \rho_{0,k}) \]

Not identifiable because in study \( j \) we have 1 estimate \( y_j \) and four parameters \((\theta_j, \text{RR}_j, \rho_{1,j}, \rho_{0,j})\)
Priors for $\theta$ and $(RR, p_1, p_0)$

The exposure effects $\theta$:

$$\theta_j \overset{\text{IID}}{\sim} N(\mu, \tau^2) \quad \text{for } j \in 1 : k$$

$$\mu \sim N(0, 10^3)$$

$$\tau \sim \text{Uniform}(0, 10^3)$$

The quantities $\mu$ and $\tau$ are the usual parameters from a random effects meta-analysis.
Priors for $\theta$ and $(RR, p_1, p_0)$

The bias parameters $(RR, p_1, p_0)$:

\[
\begin{align*}
\log RR_j & \overset{\text{IID}}{\sim} N(\mu_{RR}, \tau_{RR}^2) \\
\logit(p_{j0}) & \overset{\text{IID}}{\sim} N(\mu_{p0}, \tau_{p0}^2) \\
\logit(p_{j1}) & \overset{\text{IID}}{\sim} N(\mu_{p1}, \tau_{p1}^2).
\end{align*}
\]

We require hyperparameters $(\mu_{RR}, \tau_{RR}^2), (\mu_{p0}, \tau_{p0}^2)$ and $(\mu_{p1}, \tau_{p1}^2)$, which are taken from the literature.
Hyperparameters $(\mu_{RR}, \tau_{RR}^2), (\mu_{p0}, \tau_{p0}^2)$ and $(\mu_{p1}, \tau_{p1}^2)$

What prior distribution would describe the typical magnitude confounding from a single missing binary variable?

$U =$ engaging in 3 or more preventative medical practices

\[
\begin{align*}
\mu_{RR} &= -0.62 \text{ and } \tau_{RR} = 0.27 \\
\mu_{p0} &= -0.476 \text{ and } \tau_{p0} = 0.014 \\
\mu_{p1} &= -0.285 \text{ and } \tau_{p1} = 0.0423
\end{align*}
\]

Based on Dormuth al. (2007).

$\rightarrow$ Statin users are more likely to engage in preventative medical practices, and this reduces fracture risk. OR $\approx 1.5$ or 0.67.
Update in 3 blocks. Two blocks are conditionally conjugate.

The posterior:

\[
P(\theta, RR, p_1, p_0, \mu, \tau^2 | y, \sigma) \propto \prod_{j=1}^{k} \exp \left[ -\frac{1}{2\sigma_i^2} \{ y_i - (\theta_j + \Omega(RR_j, p_{1j}, p_{0j})) \}^2 \right] 
\]

\[
P(\theta_i | \mu, \tau)P(RR_j)P(p_{1j})P(p_{0j}) \times 
\]

\[
P(\mu)P(\tau^2). 
\]

Three Blocks:

\[
[\theta \mid RR, p_1, p_0, \mu, \tau^2 \mid y, \sigma] \quad [\mu, \tau^2 \mid \theta] \quad [RR, p_1, p_0 \mid y, \theta], 
\]
\[ [\theta_j | RR_j, p_{1j}, p_{0j}, \mu, \tau^2, \gamma_j, \sigma_j] \sim N \left( \frac{\tau^2 \{ y_i - \Omega(RR_j, p_{1j}, p_{0j}) \} + \sigma_j^2 \mu}{\sigma_j^2 + \tau^2}, \frac{\sigma_j^2 \tau^2}{\sigma_j^2 + \tau^2} \right), \]

\[ [\mu, \tau^2 | \theta] \]

→ Usual meta-analysis conditionally conjugate.

\[ [RR, p_1, p_0 | y, \theta] \]

→ Complicated dependence on \( \Omega \) requires Metropolis steps.
Guess what: Adjustment for a typical missing confounder has no impact on the results.
Suppose that statins do not reduce fracture risk, and the protective association is entirely due to a single unmeasured confounder.

What are the likely characteristics of such a confounder?

We can answer this question by locking the bias-corrected pooled RR to be 1.0 and then looking at the posterior distribution that is induced over the bias parameters.
Posterior mean of bias parameters is (-1.65, 1.65). ORs and RRs roughly 5 or 1/5. Tremendous confounding is required to destroy association between statins and fractures.
Results

This investigation reveals that adjustment for a single missing variable would not remove the discrepancy between observational and randomized estimates.

This is important because much of the debate in the literature centers on the role of variables covariates that may or may not have been measured in different studies.
Conclusions & Going Forward

What are other hypotheses for the failure of epidemiology in the context of statins and fractures?

- **Multiple** correlated unmeasured unconfounders (and extremely hard/impossible statistical problem).
- Publication bias
- Selection bias because of prevalent users of statins.

McCandless (2013) Statins and fracture risk: Can we quantify the healthy-user effect? *Epidemiology (in press)*