Improving estimation of total effects when direct effects are provided in a meta-analysis

Colleen Chan

Yale University

Collaborators: Tonatiuh Barrientos Gutierrez[1], Dalia Camacho García Formentí[1], Rodrigo Barran[1], Rosana Torres Alvarez[1], Dalia Stern-Solodkin[1], Donna Spiegelman[2]

[1]Instituto Nacional de Salud Pública [2]Yale School of Public Health

JSM 2020

August 6, 2020

Motivation

Suppose we are interested in estimating the effect of sugar-sweetened beverages (SSB's) on incidence of type II diabetes.

Much research has been conducted investingating this question, each studying different populations at different time periods and using different controls.

Goal: Aggregate this information together to get a pooled estimate

Background

Meta-Analysis

- Combines information across multiple studies that aim to estimate the same quantity to produce a pooled estimate with smaller error
- Let $\widehat{T}_s$ be the estimated effect from study s.
  Common weighting scheme: inverse-variance weighted average

$$\sum_s w_s \widehat{T}_s \Big/ \sum_s w_s \quad \text{where the weight } w_s = \frac{1}{\mathrm{Var}(\widehat{T}_s)}$$

- Two models for assigning weights:
  - Fixed effects:

$$w_s = \frac{1}{\mathrm{Var}(\widehat{T}_s)}$$

  - Random effects:

$$w_s = \frac{1}{\mathrm{Var}(\widehat{T}_s) + \tau^2}$$ [1]

[1] Der Simonian and Lair 1986

Mediation Proportion

$$\begin{tikzcd} X \arrow[rr] \arrow[dr] & & Y \\ & M \arrow[ur] & \end{tikzcd}$$

X is an exposure, M is a mediator, Y is an outcome

The direct effect (DE) can be represented as that flowing through $X \to Y$. The indirect effect (IE) can be represented as that flowing through $X \to M \to Y$.

The total effect (TE) is the sum of the indirect effect (IE) and the direct effect (DE).

The mediation proportion is the proportion of the total effect that is mediated, or the indirect effect divided by the total effect.

Computing the Mediation Proportion (Difference Method)

$$
\begin{tikzcd}
X \arrow[rr] \arrow[dr] & & Y \\
& M \arrow[ur] &
\end{tikzcd}
$$

Fit the following regressions, where g is a link function.

With adjustment for mediator:

$$\mathbb{E}[Y|X, M, \mathbf{W}] = g^{-1}[\beta_0 + \beta_D X + \beta_I M + \beta_1' \mathbf{W}]$$

Without adjustment for mediator:

$$\mathbb{E}[Y|X, \mathbf{W}] = g^{-1}[\beta_0^* + \beta_T X + \beta_1'^* \mathbf{W}]$$

If both of the equations hold simultaneously, we says we have *g-linkability*. Under g-linkability, $DE = \beta_D$, $IE = \beta_T - \beta_D$.

Computing the Mediation Proportion (Difference Method)

From previous slide:

$$\mathbb{E}[Y|X, M, \mathbf{W}] = g^{-1}[\beta_0 + \beta_D X + \beta_I M + \beta_1' \mathbf{W}]$$
$$\mathbb{E}[Y|X, \mathbf{W}] = g^{-1}[\beta_0^* + \beta_T X + \beta_1'^* \mathbf{W}]$$

The mediation proportion $\widehat{p}$ can then be estimated as

$$\widehat{p} = \frac{\widehat{\beta}_I}{\widehat{\beta}_T} = \frac{\widehat{\beta}_T - \widehat{\beta}_D}{\widehat{\beta}_T}$$

Under g-linkability, this estimator is consistent for p.

Common link functions: identity, log, logit

Traditional Meta-Analysis Approach for Estimating Total Effects

Studies can be classified into three groups:

- Studies that only report direct effect (S_D)
- Studies that only report total effect (S_T)
- Studies that report both total and direct effects (S_B)

If we want to estimate a summary total effect, studies in S_D should not be included.

The summary total effect (fixed effects) would be estimated as

$$\widehat{\widehat{\beta}}_T = \sum_{s \in S_T \cup S_B} w_{T_s} \widehat{\beta}_{T_s} \Bigg/ \sum_{s \in S_T \cup S_B} w_{T_s} \quad \text{where } w_{T_s} = \frac{1}{\mathrm{Var}(\widehat{\beta}_{T_s})}.$$

What can go wrong

Studies can be classified into three groups:

- Studies that only report direct effect (S_D)
- Studies that only report total effect (S_T)
- Studies that report both total and direct effects (S_B)

Common confusion:

- Using the direct effect instead of the total effect when both are provided (study is in S_B)
- Using the direct effect from a study in S_D when the study should not be incorporated in the analysis

Incorporating the direct effect estimate would lead to an underestimate of the total effect.

An Example

Malik et al. 2018 uses the direct effect estimate in place of the total effect estimate.

Ref.	Population (cases)	Age range (years)	Duration (years)	Dietary assessment method	Outcome	Results	Adjustment for potential confounders
Montonen et al., 2007 (14)	2,360 adults, Finnish Mobile Clinic Health Examination, Finland (177)	40–69	12	Diet history	Type 2 diabetes*	RR (95% CI) between extreme quartiles of median SSB intake (0 vs. 143 g/day): $\boxed{1.67}$ (0.98–2.87); P_{trend} = 0.01	Age, sex, $\boxed{\text{BMI}}$, energy intake, smoking, geographic area, physical activity, family history of diabetes, prudent dietary score, and conservative pattern score
Paynter et al., 2006 (15)	12,204 adults ARIC study, U.S. (718 men, 719 women)	45–64	9	FFQ	Type 2 diabetes†	Men: RR (95% CI) between extreme quartiles of SSB intake (<1 8-oz serving/day vs. ≥2 8-oz servings/day): 1.09 (0.89–1.33); P_{trend} = 0.68. Women: RR (95% CI) between extreme quintiles of SSB intake: 1.17 (0.94–1.46); P_{trend} = 0.05	Race, age
Schulze et al., 2004 (16)	91,249 women NHS II, U.S. (741)	24–44	8	133-item FFQ	Type 2 diabetes†	RR (95% CI) between extreme quartiles of SSB intake (<1 serving/month vs. ≥1 serving/day: 1.83 (1.42, 2.36); P_{trend} = <0.001	Age, alcohol intake, physical activity, family history of diabetes, smoking, postmenopausal hormone use, oral contraceptive use, cereal fiber, magnesium, trans fat, ratio of polyunsaturated to saturated fat, diet soft drinks, fruit juice, fruit punch

Methods

Estimating the Summary Mediation Proportion

1. For each study in S_B:
   1. Compute mediation proportion for each study
   $$\widehat{p}_s = 1 - \frac{\widehat{\beta}_{D_s}}{\widehat{\beta}_{T_s}}.$$
   2. Compute the variance of the mediation proportion using the multivariate delta method
   $$\mathrm{Var}(\widehat{p}_s) \approx \frac{(\widehat{\beta}_{D_s})^2\ \mathrm{Var}(\widehat{\beta}_{T_s})}{(\widehat{\beta}_{T_s})^4} + \frac{\mathrm{Var}(\widehat{\beta}_{D_s})}{(\widehat{\beta}_{T_s})^2} - \frac{2\widehat{\beta}_{D_s}\mathrm{Cov}(\widehat{\beta}_{D_s}, \widehat{\beta}_{T_s})}{(\widehat{\beta}_{T_s})^3}.$$
   Use data duplication algorithm[2] to find the correlation between total effect and direct effect estimate (requires individual level data). Otherwise, use a sensitivity analysis approach.

[2] Nevo, Liao, and Spiegelman 2017

Estimating the Summary Mediation Proportion (cont.)

3. Compute the summary mediation proportion using a normalized inverse variance weighted average

$$\widehat{\overline{p}} = \sum_{s \in S_B} w_{p_s} \widehat{p}_s \Big/ \sum_{s=1}^{S_B} w_{p_s} \quad \text{where} \quad w_{p_s} = \frac{1}{\mathrm{Var}(\widehat{p}_s)}.$$

4. Compute the variance of the summary mediation proportion

$$\mathrm{Var}(\widehat{\overline{p}}) = \left(\sum_{s \in S_B} w_{p_s} \right)^{-1}.$$

Estimating the Summary Total Effect

1. For each study in S_D:
   1. Back-calculate the total effect

   $$\widehat{\beta}_{T_s} = \widehat{\beta}_{D_s} / (1 - \widehat{\overline{p}})$$

   2. Compute the variance of the back-calculated total effect using the multivariate delta method

   $$\mathrm{Var}(\widehat{\beta}_{T_s}) \approx \frac{\widehat{\beta}^2_{T_s} \mathrm{Var}(\widehat{\overline{p}}) + \mathrm{Var}(\widehat{\beta}_{D_s})}{(1 - \widehat{\overline{p}})^2}$$

Estimating the Summary Total Effect (cont.)

3. Compute the summary total effect using an inverse variance weighted average

$$\widehat{\overline{\beta}}_T = \sum_{s \in S} w_{T_s} \widehat{\beta}_{T_s} \Big/ \sum_{s \in S} w_{T_s} \quad \text{where} \quad w_{T_s} = \frac{1}{\mathrm{Var}(\widehat{\beta}_{T_s})}$$

where $S = S_D \cup S_T \cup S_B$.

4. Compute the variance of the summary total effect

$$\mathrm{Var}(\widehat{\overline{\beta}}_T) = \left(\sum_{s \in S} w_{T_s} \right)^{-1}.$$

Random Effects Model

The variance of each study is the sum of the within-studies variance and between studies variance.

$$\mathrm{Var}(\widehat{\beta}_{T_s})^* = \mathrm{Var}(\widehat{\beta}_{T_s}) + \tau^2.$$

The between-studies variance is

$$\tau^2 = (Q - df)/C$$

where $Q = \sum w_{T_s}\widehat{\beta}_s^2 - \frac{\left(\sum w_{T_s}\widehat{\beta}_s\right)^2}{\sum w_{T_s}}$ is the total variance, df is the number of studies - 1, and C is a normalizing constant.

Take an inverse-variance weighted average as before

$$\widehat{\overline{\beta}}^*_T = \sum_{s\in S} w^*_{T_s}\widehat{\beta}_{T_s} \Big/ \sum_{s\in S} w^*_{T_s} \quad \text{where } w^*_{T_s} = \frac{1}{\mathrm{Var}(\widehat{\beta}_{T_s})^*}$$

Illustrative Example

Illustrative Example

Interested in estimating the total effect of drinking one sugar sweetened beverage (SSB)/day on incidence of type II diabetes.

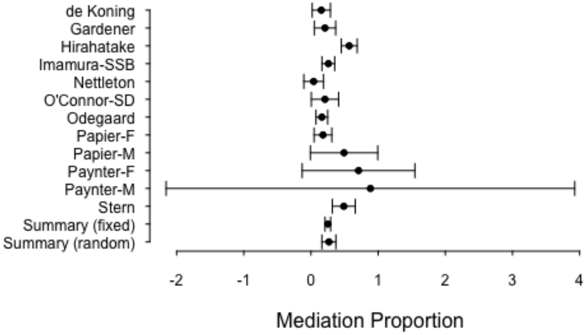
25 independent studies total.

S_D (5)	S_B (12)	S_T (8)
Auerbach	de Koning	Bazzano
Fresan	Gardener	Imamura (FJ)
Huang	Hirahatake	O'Connor (FJ)
Löfvenborg	Imamura (SSB)	O'Connor (SMB)
Motonen	Nettleton	O'Connor (STC)
	O'Connor (SD)	Palmer
	Odegaard	Sakurai
	Papier (F)	Schulze
	Papier (M)	
	Paynter (F)	
	Paynter (M)	
	Stern	

Abbreviations: FJ = fruit juices, SD = soft drinks, SMB = sweetened milk beverages, STC = sweetened tea or coffee, SSB = sugar-sweetened beverages, F = females, M = males

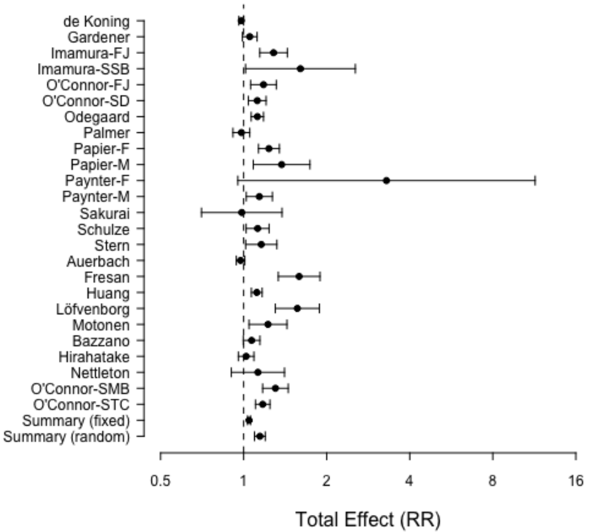
Mediation Proportion

From `metamediate` package (coming soon!):



Fixed effects: 0.251 (0.209, 0.292)
Random effects: 0.267 (0.162, 0.372)

Total Effect



Total Effect

Method	# Studies	RR (fixed)	RR (random)
New	25	1.044 (1.031, 1.056)	1.143 (1.093, 1.194)
New, w/o Auerbach	24	1.090 (1.072, 1.107)	1.152 (1.091, 1.217)
Naive	25	1.031 (1.021, 1.042)	1.1608 (1.065, 1.265)
Naive, w/o Auerbach	24	1.087 (1.070, 1.104)	1.177 (1.043, 1.328)
Traditional	20	1.085 (1.067, 1.103)	1.159 (1.019, 1.317)
Malik et. al	9	–	1.26 (1.12, 1.41)

Conclusion

Conclusion

- Presented methodological overview for estimating total and direct effects
- Introduced a new, more efficient estimator for the summary total effect in a meta-analysis
- Reperform an analysis of SSB consumption on incidence of type II diabetes using new method

Thank you

Back-up Slides

Bias Analysis

Q: How does the bias change as a function of parameters when we use direct effects as total effects for studies in S_D?

Classify studies into two groups:

- S_P: studies where the total effect is provided (Note $S_P = S_T \cup S_B$)
- S_D: studies where only the direct effect is provided

Assume each group has a true total effect and each study has the same weight (variance). Let π be the proportion of studies where the total effect is provided.

$=$

Then the naive summary total effect estimator is

$$\widehat{\overline{\beta}}_{T,\text{naive}} = \pi \widehat{\overline{\beta}}_{T_P} + (1-\pi)\widehat{\overline{\beta}}_D.$$

Bias Analysis

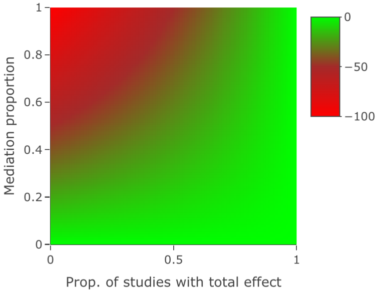
The unbiased summary total effect estimator is

$$\widehat{\overline{\beta}}_T = \pi \widehat{\overline{\beta}}_{T_P} + (1-\pi)\frac{\widehat{\overline{\beta}}_D}{1-\widehat{\overline{p}}}$$

which converges in probability to $\overline{\beta}_T$.

The relative bias of the naive estimator simplifies to:

$$(\pi - 1)\overline{p}$$



Confidence Intervals and P-values

A $(1-\alpha)\%$ level confidence interval for $\widehat{\overline{\beta}}_T$ can be obtained as

$$\widehat{\overline{\beta}}_T \pm z_{1-\frac{\alpha}{2}} \sqrt{\mathrm{Var}(\widehat{\overline{\beta}}_T)}.$$

The p-value for the one-sided test can be estimated as

$$p = 1 - \Phi\left(\widehat{\overline{\beta}}_T) \Big/ \sqrt{\mathrm{Var}(\widehat{\overline{\beta}}_T)}\right).$$

For a two-sided test, the p-value would be

$$p = 2\left[1 - \Phi\left(\widehat{\overline{\beta}}_T) \Big/ \sqrt{\mathrm{Var}(\widehat{\overline{\beta}}_T)}\right)\right].$$