Investigating mediation when counterfactuals are not metaphysical: Does sunlight exposure mediate the effect of eye-glasses on cataracts?

Brian Egleston
Fox Chase Cancer Center

Collaborators:
Daniel Scharfstein, Beatriz Munoz, Sheila West
Johns Hopkins University

Public Health Goals

- Cataracts are a major source of vision loss in older persons.

- Promoting the use of eye-glasses when people are outdoors might reduce the incidence of cataracts through their reduction in the amount of sunlight that reaches the eye.

Salisbury Eye Evaluation

- Population based study of approximately 2,500 older adults in Salisbury, Maryland.
- Participants asked about their lifetime glasses use, jobs, and leisure activities.
- The current study uses recalled eye-glasses use and sun exposure and presenting cortical cataract data.

Data Structure and Notation

- Z indicates glasses use (1 = use, 0 = no use)
- $Y(z)$ indicates potential cataract outcome (1=cataract, 0 otherwise) under $Z = z$.
- $M(z)$ is potential ocular UV exposure in Maryland Sun Years (MSYs) under $Z = z$.
- Full set of potential outcomes:

$$\{M(0), Y(0), M(1), Y(1)\}$$

Data Structure and Notation

- X is a vector of confounding covariates: Age, Type of job in 30s, Race, Sex, Diabetic status, Education level

- $M = M(Z), Y = Y(Z)$ are observed UV and cataract outcomes.

- Observed data:

$$\{Z, X, M, Y\}$$

Mediation

- A mediator is the causal mechanism linking an exposure to an outcome.

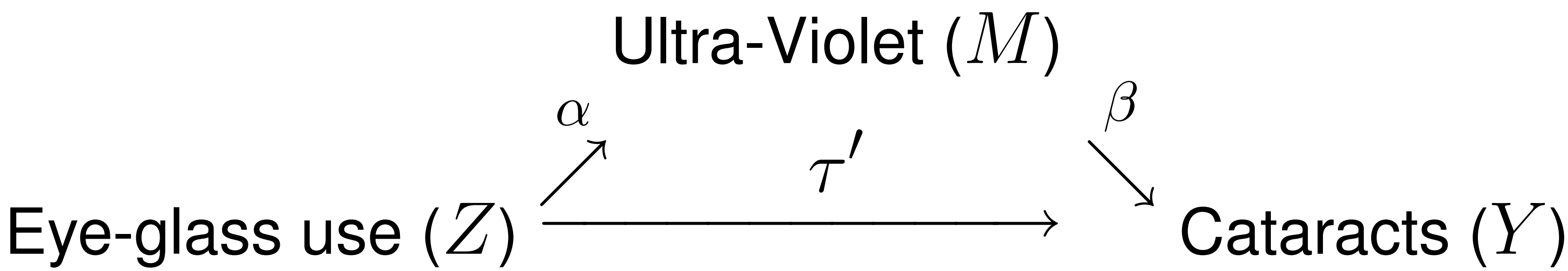
- Causal hypothesis:

Eye glasses $\longrightarrow$ UV exposure $\longrightarrow$ Cataracts

$\uparrow$ $\uparrow$ $\uparrow$

Age 31 Ages 31-34 Age 65+

Traditional Model



$$Y = \gamma_1 + \tau Z + \epsilon_1 \tag{1}$$

$$M = \gamma_2 + \alpha Z + \epsilon_2 \tag{2}$$

$$Y = \gamma_3 + \tau' Z + \beta M + \epsilon_3 \tag{3}$$

- For Y as a continuous measure, cov(ϵ_2, ϵ_3)=0
- Total effect of Z on Y is τ, direct effect is τ'.

Traditional Model

- Under Baron and Kenny, a measure of the indirect (mediated) effect is $\alpha\beta$.

  Total effect = $\tau = \tau' + \alpha\beta$.

- Stringent assumptions are necessary to give causal meaning to τ' and $\alpha\beta$.

“Controlled” Effects

- Effects on outcomes after manipulating Z, $M(Z)$
  $\rightarrow Y(z, m)$
- Exchangeability assumptions needed to identify controlled effects under randomization.
- $Y(z, m) \perp M(z) | Z = z$, which implies,
  $$E[Y(z, m)] = E[Y(z, m) | M(z) = m, Z = z]$$
- Then, $E[Y(1, m)] - E[Y(0, m)] = \tau'$
- Are controlled effects meaningful?

“Natural” Effects

- Proposed by Robins and Pearl.

$$\begin{aligned}\text{Total Effect} &= E[Y(1)] - E[Y(0)] \\ &= \underbrace{E[Y(1, M(1))] - E[Y(0, M(1))]}_{\text{Direct Effect}} \\ &+ \underbrace{E[Y(0, M(1))] - E[Y(0, M(0))]}_{\text{Indirect (Mediated) Effect}}\end{aligned}$$

“Natural” Effects

- An assumption is needed to identify natural mediational effects in addition to the assumption necessary for controlled effects.
- One assumption: $Y(1, m) - Y(0, m) = B$ is a random variable that does not depend on m.
- Natural effects have become the reference for assigning cause to mediational, surrogate marker, and indirect effect models (e.g. Taylor *et al.*, 2005)

“Natural” Effects

- Are natural effects meaningful?
- How could one ever experimentally observe $Y(1, M(0))$?
- We would need to observe UV exposure in 30s when a person does not wear glasses, then go back in time and assign glasses but exposure under no glasses.

Proposed Causal Estimand

$$RR(p,m) = \frac{P[Y(1) = 1 | P = p, M(0) = m]}{P[Y(0) = 1 | P = p, M(0) = m]}$$

- $P = M(1)/M(0)$. P is the proportion of potential UV that reaches eyes under glasses.
- Relative risk of cataracts with glasses use within strata based on baseline exposure and shielding effect of glasses.

Proposed Causal Estimand

- In a case of complete mediation we would expect that $RR(1, m) = 1$ for all m.

- If glasses use does not change an individual's UV exposure then glasses should not be associated with cataracts.

Proposed Causal Estimand

- In a case of mediation, we would expect that,

$$1 \geq RR(p, m) > RR(p', m) \text{ if } p > p'$$

- The more that glasses prevent UV exposure, the more they prevent cataracts.

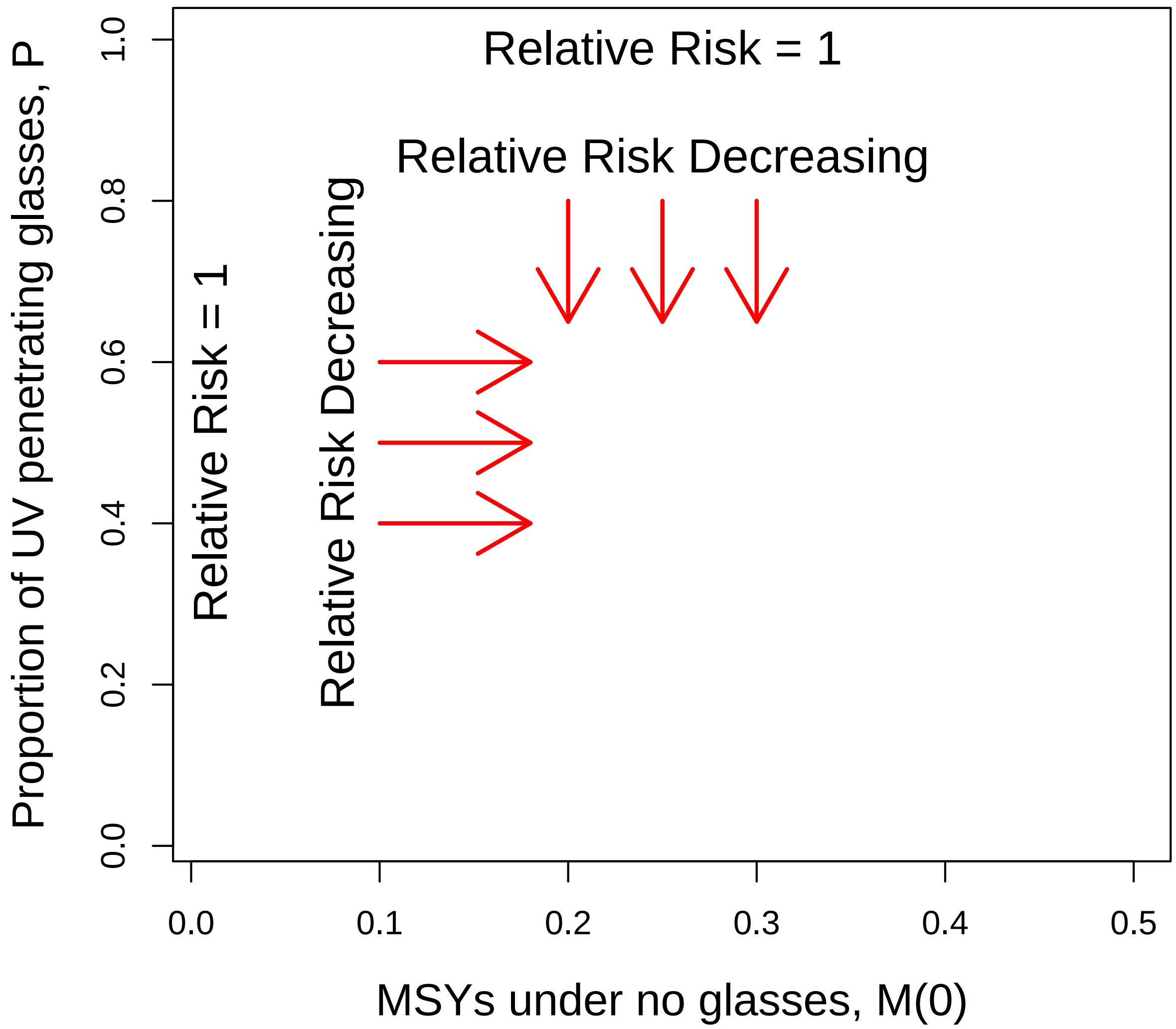
- This monotonicity might be broken if the principal stratum defined by $\{p, m\}$ includes individuals who are very different from the principal stratum defined by $\{p', m\}$.

Local Causal Inference

- Strata are likely similar within neighborhoods of p for given $M(0)$.

- After controlling for $M(0)$, those with very different values of P might have different characteristics, but we did not expect this to be the case a priori.

Hypothesized RR(p,m)



Non-Metaphysical Counterfactual

- $M(0)$ observable on everyone (Duncan *et al.*, 1997)

$$M = \sum_{s=1}^{12} G(s)R(s) \sum_{t=5}^{18} F(t,s)H(t,s)T_{hats}(t,s)T_{eye}(t,s)$$

M	=	Total UV exposure
s	=	Month
t	=	Hour of day
$G(s)$	=	Geographic correction factor
$R(s)$	=	Ocular ambient exposure ratio
$F(t,s)$	=	Fraction of time spent outdoors
$H(t,s)$	=	Global ambient exposure
$T_{hats}(t,s)$	=	Percent of UV penetrating hats
$T_{eye}(t,s)$	=	Percent UV penetrating glasses; Set to 1 to identify M(0)

Identification of Estimand

Assumption 1: Stable Unit Treatment Value

- An individual's potential outcomes are unrelated to glasses use of other study participants and there are only two well-defined treatment arms.

Assumption 2:

$$Z \perp \{Y(0), Y(1), M(1)\} \mid M(0), X$$

- This is an observational study equivalent of the randomization assumption in randomized trials.

Identification of Estimand

Assumption 3: $Y(0) \perp M(1) \mid Z, M(0), X$

- If we already know someone's glasses use status, baseline UV exposure and set of confounding covariates, knowing UV exposure that would occur when a person wears glasses gives us no additional information about baseline cataract outcomes.

Identification of Estimand

- For a neighborhood dp of $P = p$,

$$P[Y(1) = 1 | P \in dp, M(0)] = E\left[\frac{P[Y = 1 | Z = 1, P \in dp, M(0), X] P[P \in dp \, | Z = 1, M(0), X]}{E[P[P \in dp \, | Z = 1, M(0), X]]} \middle| M(0)\right]$$

$$P[Y(0) = 1 | P \in dp, M(0)] = E\left[\frac{P[Y = 1 | Z = 0, M(0), X] P[P \in dp \, | Z = 1, M(0), X]}{E[P[P \in dp \, | Z = 1, M(0), X]]} \middle| M(0)\right]$$

- Use assumption 2 for first equality, assumptions 2 and 3 for second.

Models

- Models of primary interest:

$$\begin{aligned} \text{logit } P[Y(0)=1|P,M(0)] &= g_0(P,M(0);\boldsymbol{\beta}_0^*) \\ \text{logit } P[Y(1)=1|P,M(0)] &= g_1(P,M(0);\boldsymbol{\beta}_1^*) \end{aligned}$$

- Propensity model used for assumption 2:

$$\text{logit } P[Z=1|M(0),X] = h(M(0),X;\boldsymbol{\gamma}^*)$$

Models

- Beta regression of P since we do not observe $M(1)$ on those who did not wear glasses.

$$\begin{aligned} \text{logit } E[P|M(0), X] &= k(M(0), X; \boldsymbol{\eta}^*) \\ E[P|M(0), X] &= \mu(M(0), X; \boldsymbol{\eta}^*) \\ Var[P|M(0), X] &= \\ &\frac{\mu(M(0), X; \boldsymbol{\eta}^*)(1 - \mu(M(0), X; \boldsymbol{\eta}^*))}{1 + \phi^*} \end{aligned}$$

Estimation

- Maximum likelihood estimates can be used for $\boldsymbol{\beta}_1^*$, $\boldsymbol{\gamma}^*$, η^*, and ϕ^*.
- Unbiased estimating equation for $\boldsymbol{\beta}_0^*$:

$$
\begin{aligned}
&\boldsymbol{U}_{\boldsymbol{\beta}_0}(O^\dagger; \boldsymbol{\psi}^*) \\
&= E\left[\left.\frac{(1-Z)g_0'(P, M(0); \boldsymbol{\beta}_0^*)\left(Y - \text{ expit}\left\{g_0(P, M(0); \boldsymbol{\beta}_0^*)\right\}\right)}{(1 - \text{expit}\left\{h(M(0), X; \boldsymbol{\gamma}^*)\right)\}}\right| O^\dagger\right] \\
&= \frac{\int_0^1 (1-Z)g_0'(p, M(0); \boldsymbol{\beta}_0^*)Y(0)f(p|M(0), Z=1, X; \boldsymbol{\eta}^*, \phi^*)dp -}{(1 - \text{expit}\left\{h(M(0), X; \boldsymbol{\gamma}^*)\right)\}} \\
&\quad \frac{\int_0^1 (1-Z)g_0'(p, M(0); \boldsymbol{\beta}_0^*)\text{expit}\left\{g_0(p, M(0); \boldsymbol{\beta}_0^*)\right\}f(p|M(0), Z=1, X; \boldsymbol{\eta}^*, \phi^*)dp}{(1 - \text{expit}\left\{h(M(0), X; \boldsymbol{\gamma}^*)\right)\}}
\end{aligned}
$$

where $O^\dagger = \{Z, X, M(0), M, Y\}$.

Estimation

$$\widehat{P}[Y(z) = 1 | P = p, M(0) = m] = \frac{\exp\{g_z(p, m; \widehat{\boldsymbol{\beta}}_z)\}}{1 + \exp\{g_z(p, m; \widehat{\boldsymbol{\beta}}_z)\}}$$

$$\widehat{RR}(p, m) = \frac{\widehat{P}[Y(1) = 1 | P = p, M(0) = m]}{\widehat{P}[Y(0) = 1 | P = p, M(0) = m]}$$

Large Sample Theory

- Stack the score equations and $\boldsymbol{U}_{\boldsymbol{\beta}_0}(O^\dagger;\boldsymbol{\psi}^*)$.

$$\boldsymbol{U}(O^\dagger;\boldsymbol{\psi}) =$$

$$\left[\boldsymbol{U}_{\boldsymbol{\beta}_0}(O^\dagger;\boldsymbol{\psi})', \boldsymbol{U}_{\boldsymbol{\beta}_1}(O^\dagger;\boldsymbol{\psi})', \boldsymbol{U}_{\boldsymbol{\gamma}}(O^\dagger;\boldsymbol{\psi})', \boldsymbol{U}_{\boldsymbol{\eta}}(O^\dagger;\boldsymbol{\psi})', \boldsymbol{U}_{\phi}(O^\dagger;\boldsymbol{\psi})\right]'$$

- Under mild regularity conditions (Huber, 1964),

$$\sqrt{n}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) \xrightarrow{D} Normal(0, \Sigma^*)$$

$$\Sigma^* = E\left[\frac{\partial \boldsymbol{U}(O^\dagger;\boldsymbol{\psi}^*)}{\partial \boldsymbol{\psi}}\right]^{-1} E\left[\boldsymbol{U}(O^\dagger;\boldsymbol{\psi}^*)\boldsymbol{U}(O^\dagger;\boldsymbol{\psi}^*)'\right] E\left[\frac{\partial \boldsymbol{U}(O^\dagger;\boldsymbol{\psi}^*)}{\partial \boldsymbol{\psi}}\right]^{-1'}$$

- By the δ-method,

$$\sqrt{n}(RR(p,m;\widehat{\boldsymbol{\psi}}) - RR(p,m;\boldsymbol{\psi}^*)) \xrightarrow{D} N\left(0, \frac{\partial RR(p,m;\boldsymbol{\psi}^*)}{\partial \boldsymbol{\psi}} \Sigma^* \frac{\partial RR(p,m;\boldsymbol{\psi}^*)'}{\partial \boldsymbol{\psi}}\right)$$

Analysis

Table 1: Characteristics of sample

Variable	No Eye-glass Use	Eye-glass Use
Number of participants	830 (42%)	1125 (58%)
Cortical cataracts	16.1%	11.6%
Sun exposure if glasses worn, M(1)	-	0.06
Sun exposure if glasses never worn, M(0)	.17 (.11)	.16 (.11)
Age	73.5 (5.0)	72.7 (4.8)
Diabetic	17.4%	17.2%
Male	54.6%	39.9%
Black	30.7%	22.1%
Not high school graduate	58.0%	45.6%
Job characteristics		
Worked over water	1.7%	1.2%
Worked outside on land	41.1%	28.5%
Worked inside	38.9%	44.2%
Worked as homemaker	18.3%	26.1%

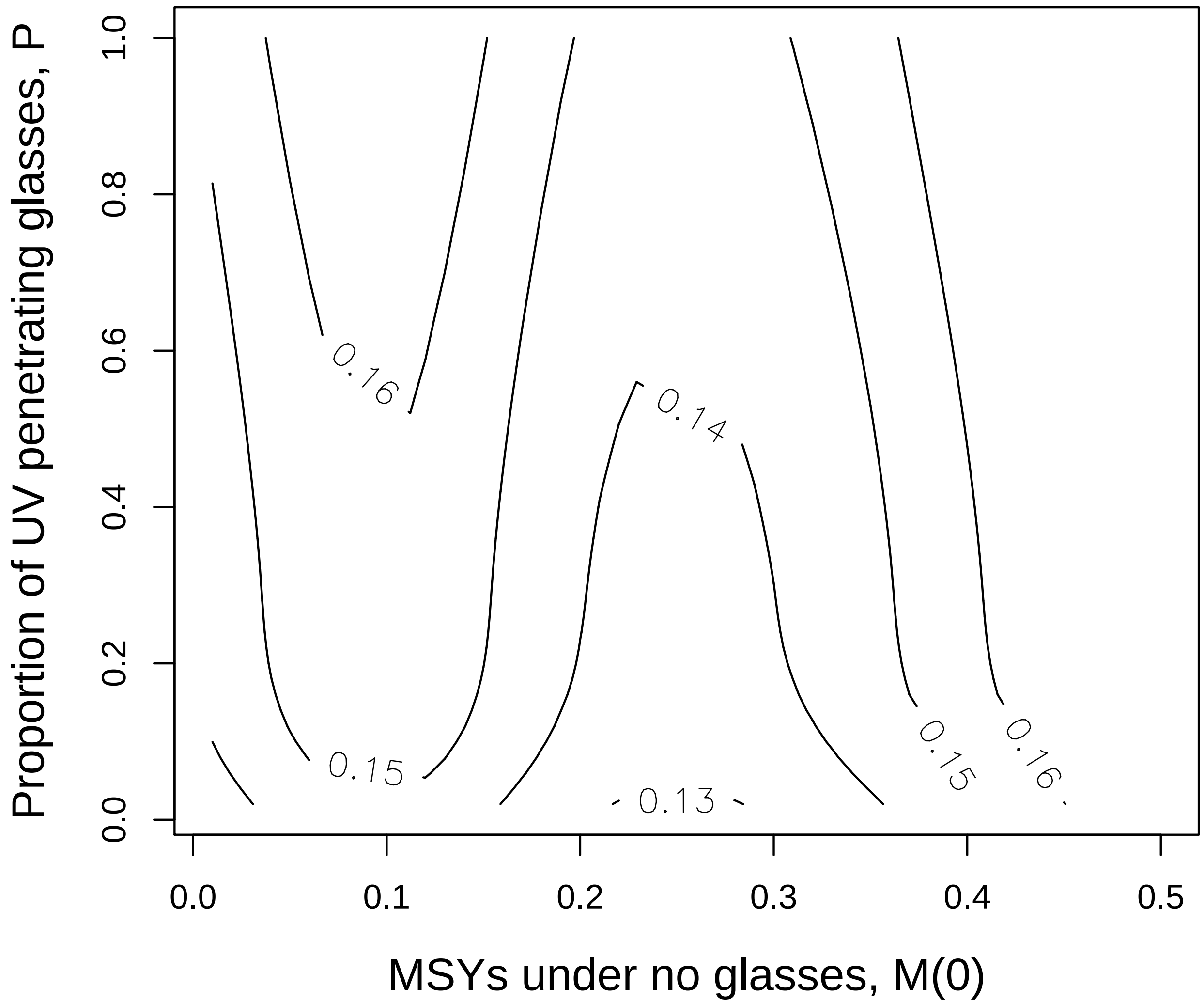
Analysis: Baron and Kenny's Method

Table 2: Logistic models of cataract development (coefficients as odds ratios).

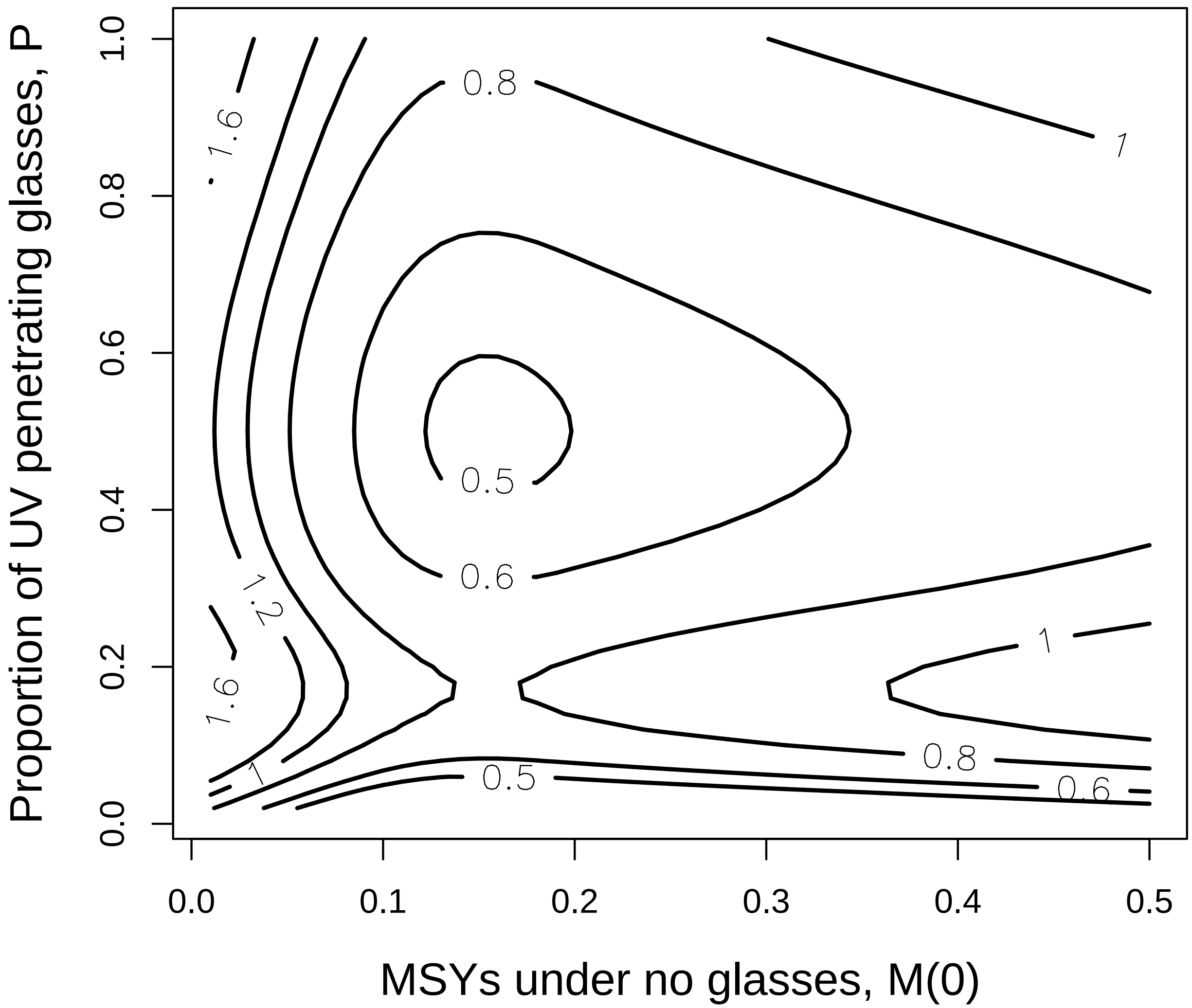
Variable	Model 1	95% CI	Model 2	95% CI
Cataract Models				
Age	1.17	(1.07, 1.28)	1.17	(1.07, 1.28)
Age spline term	0.89	(0.78, 1.03)	0.89	(0.78, 1.03)
Diabetic	1.43	(1.02, 2.00)	1.43	(1.02, 2.00)
Male	0.64	(0.45, 0.92)	0.63	(0.44, 0.91)
Black	4.23	(3.13, 5.72)	4.22	(3.12, 5.71)
Not high school grad	1.10	(0.81, 1.48)	1.09	(0.81, 1.48)
Worked over water	Reference		Reference	
Worked outside	0.50	(0.20, 1.27)	0.52	(0.20, 1.32)
Worked inside	0.64	(0.25, 1.66)	0.70	(0.26, 1.90)
Worked as homemaker	0.54	(0.19, 1.51)	0.57	(0.20, 1.61)
Glasses	0.74	(0.56, 0.99)	0.78	(0.57, 1.09)
UV			1.80	(0.30, 10.76)

Analysis

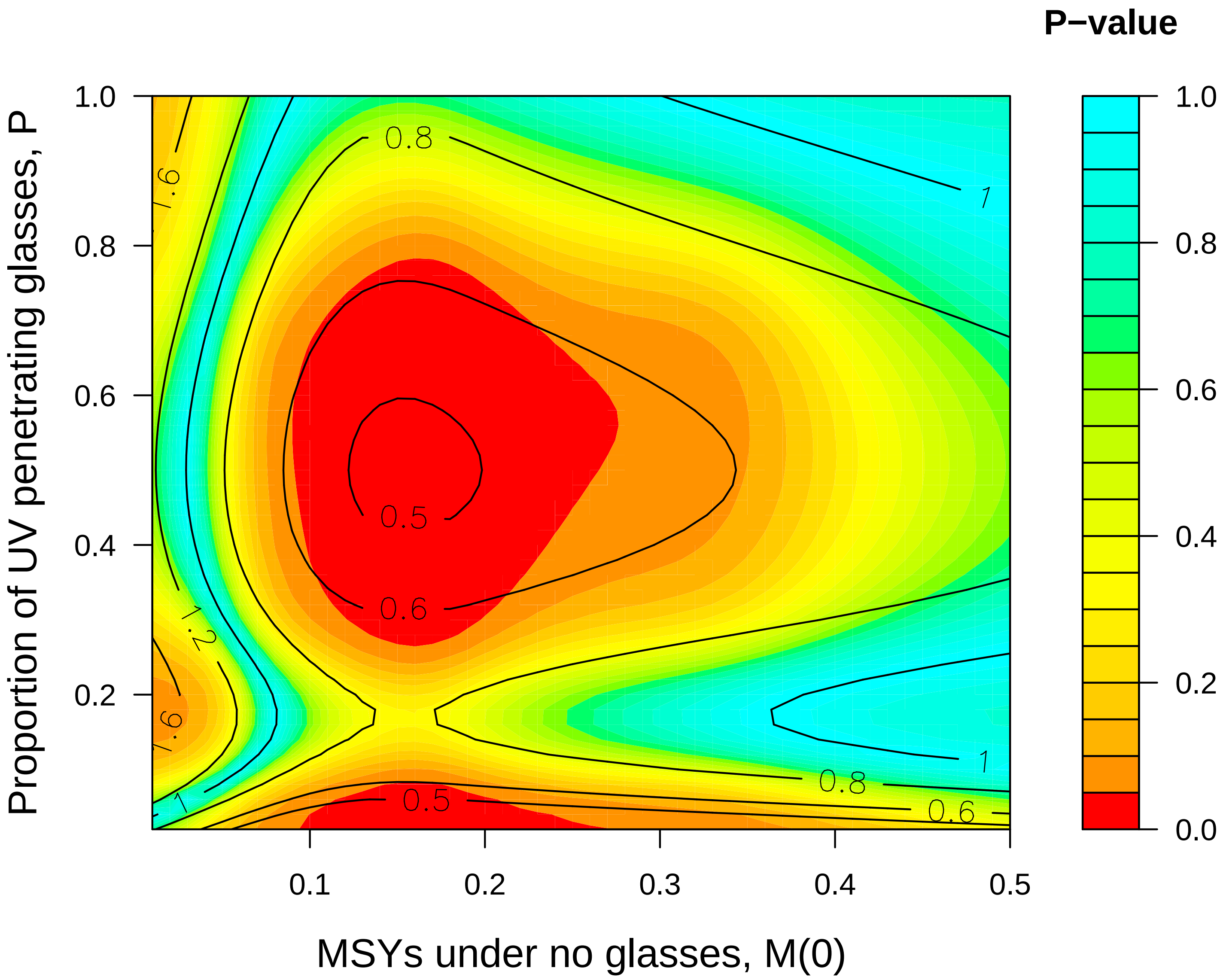
Figure 1: Estimates of $P[Y(0) = 1 \mid P = p, M(0) = m]$: Probabilities of developing cataracts under no glasses within strata.



Analysis: Relative Risk

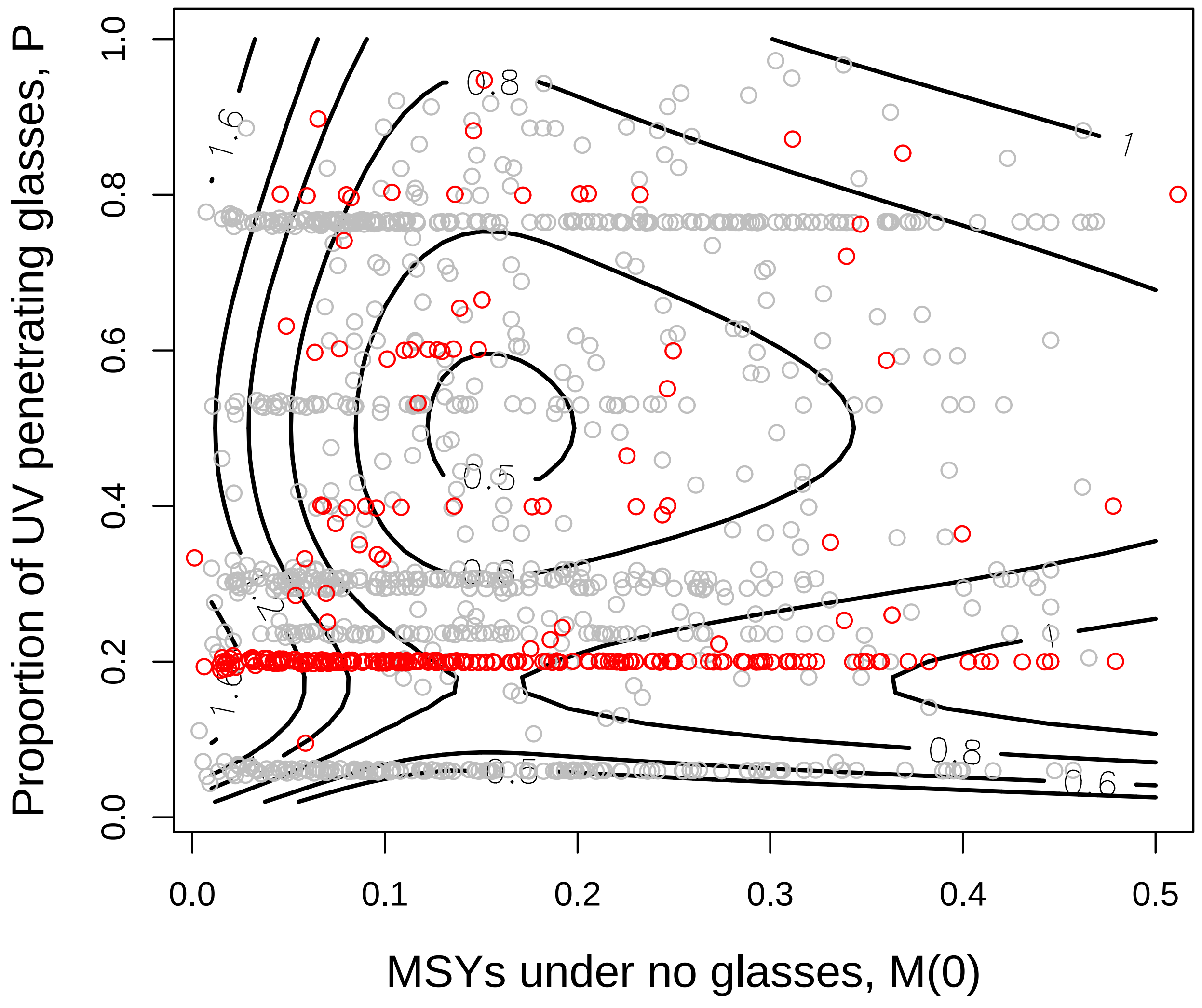


Analysis: Relative Risk



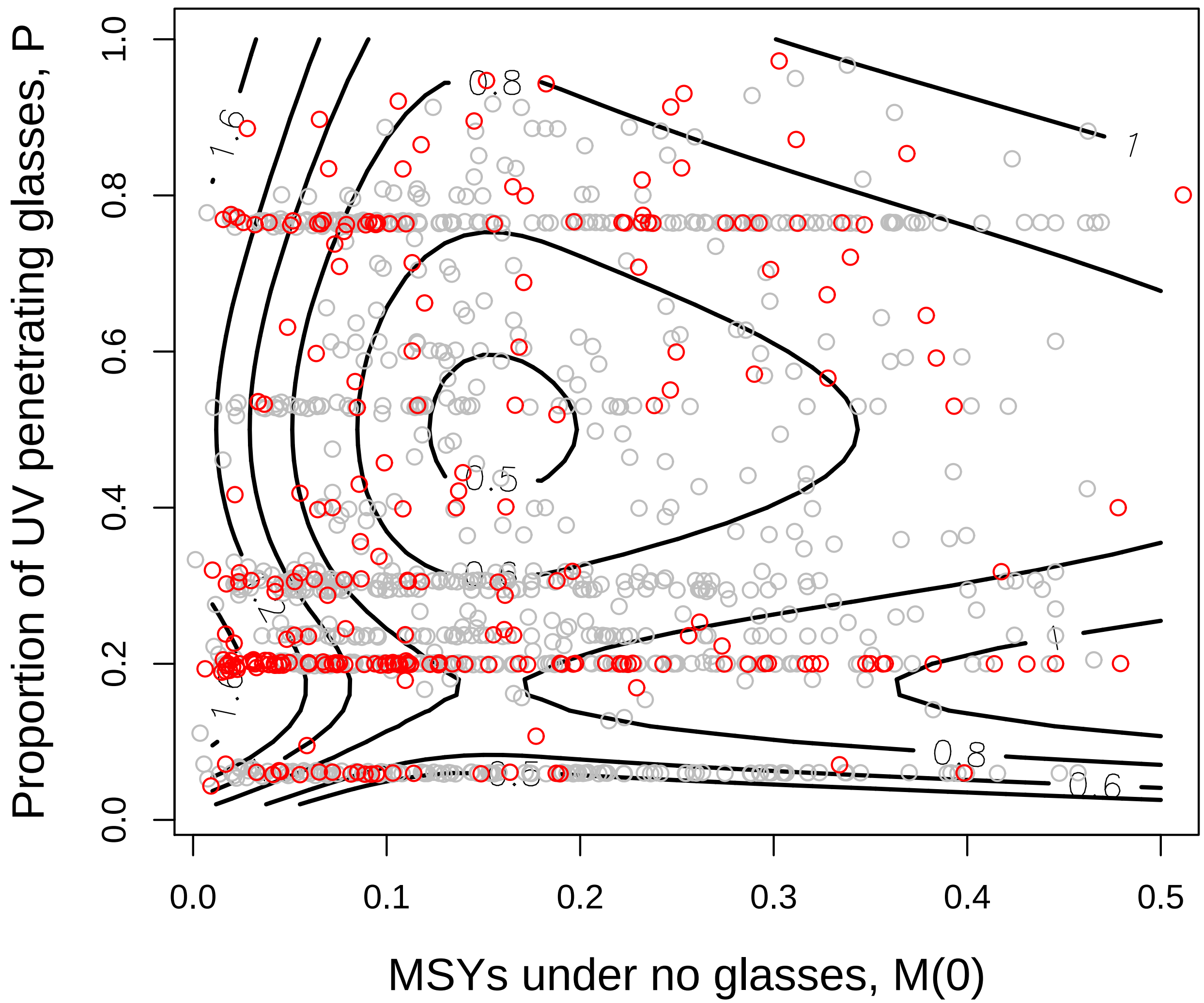
Analysis: Relative Risk

Figure 2: RR(p,m) and P vs. M(0) among glasses wearers; Red=Non-sunglasses users



Analysis: Relative Risk

Figure 3: RR(p,m) and P vs. M(0) among glasses wearers; Red=Black



Discussion

- The RR is approximately 1 when P=1.
- The RR decreases as P decreases, suggesting a protective effect of glasses.
- The decrease in the RR is not monotone; this might be due to differences in principal strata. Sunglass users have higher values of P.
- These results are consistent with mediation.

Discussion

- The traditional method of analysis provided only marginal evidence of mediation.

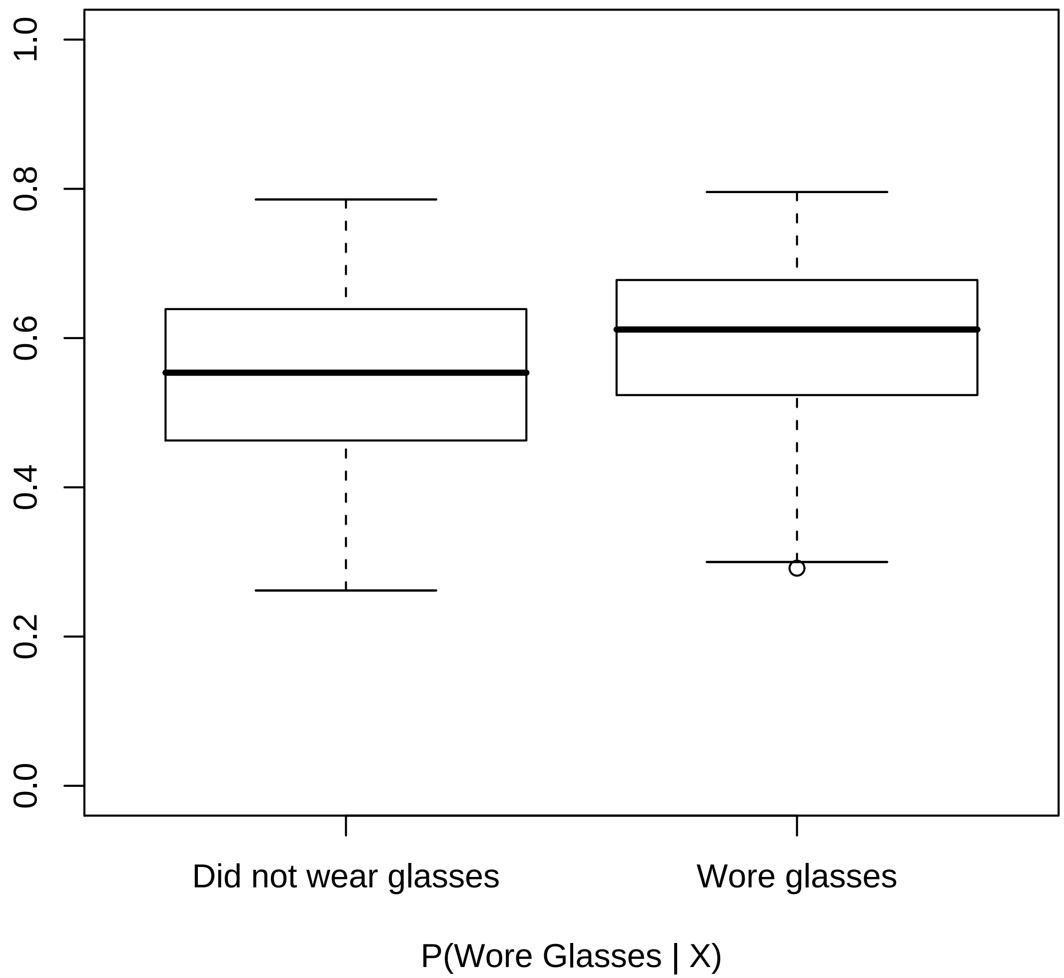
- Our causal estimand provides a richer analysis.

Discussion

- This work presents how one might develop, identify, and estimate a scientifically meaningful causal estimand.

- The results suggest that encouraging people to wear eyeglasses in mid-life can reduce cataracts later in life.

Figure 4: Boxplots of Propensity of Wearing Glasses



Analysis: Relative Risk

Figure 5: RR(p,m) and P vs. M(0) among glasses wearers; Red=Cataracts

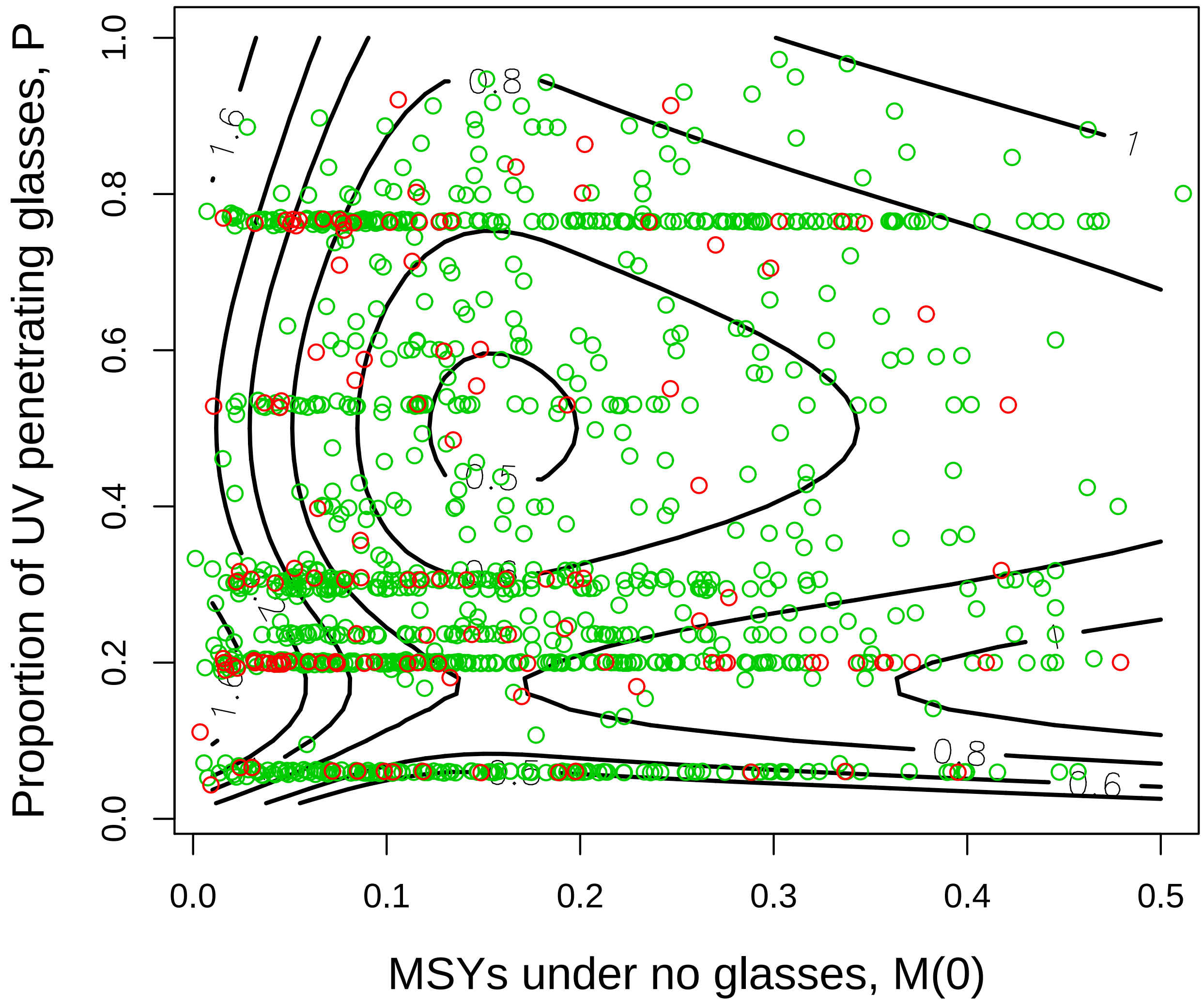


Table 3: Characteristics of sample after weighting by estimated probability of observed glasses use.

Variable	No Eye-glass Use	Eye-glass Use
Number of participants	830 (42%)	1125 (58%)
Cortical cataracts	15.0%	12.2%
Sun exposure if glasses never worn, M(0)	.17	.17
Age	73.0	73.0
Diabetic	17.3%	17.4%
Male	46.3%	46.1%
Black	25.5%	25.5%
Not high school graduate	51.1%	50.9%
Job characteristics		
Worked over water	1.4%	1.4%
Worked outside on land	33.4%	33.4%
Worked inside	42.2%	42.3%
Worked as homemaker	23.1%	22.9%

Figure 6: Histogram of $M(0)$.

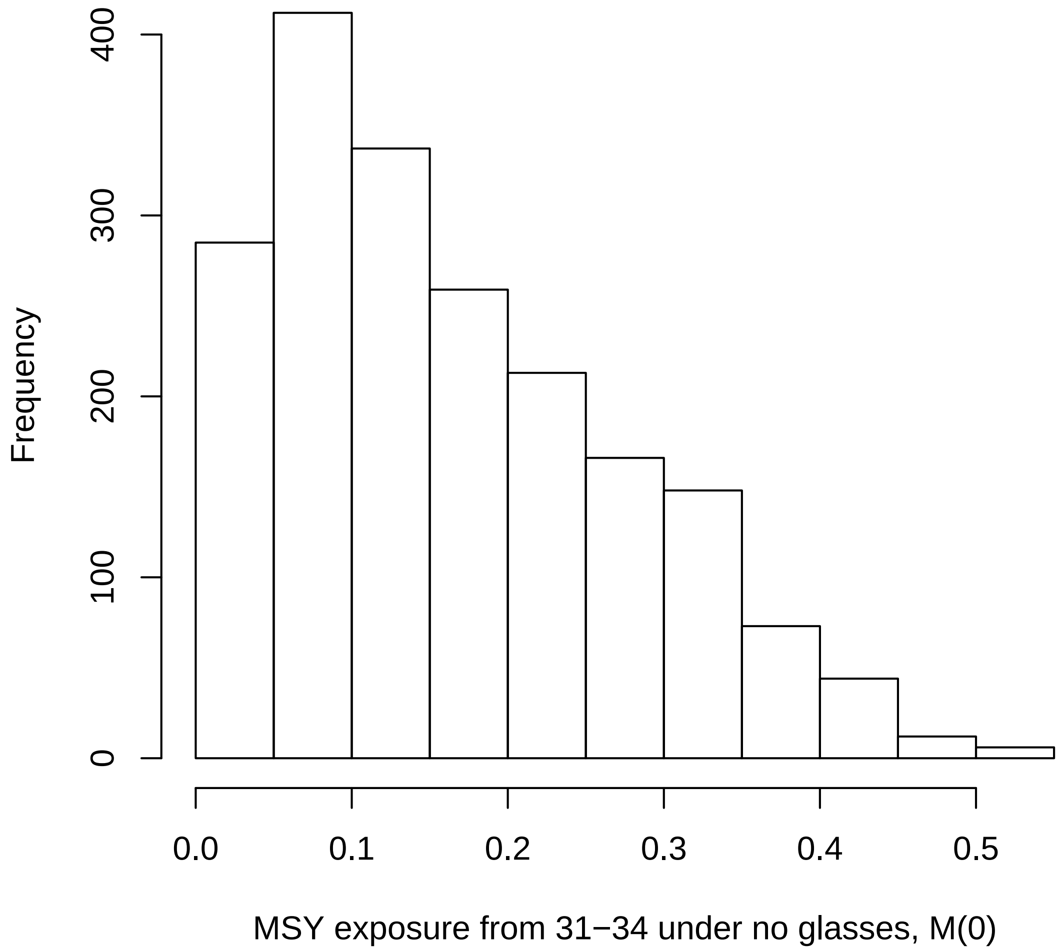


Figure 7: Scatterplot of M(1) vs. M(0) among participants who wore glasses; jitter added.

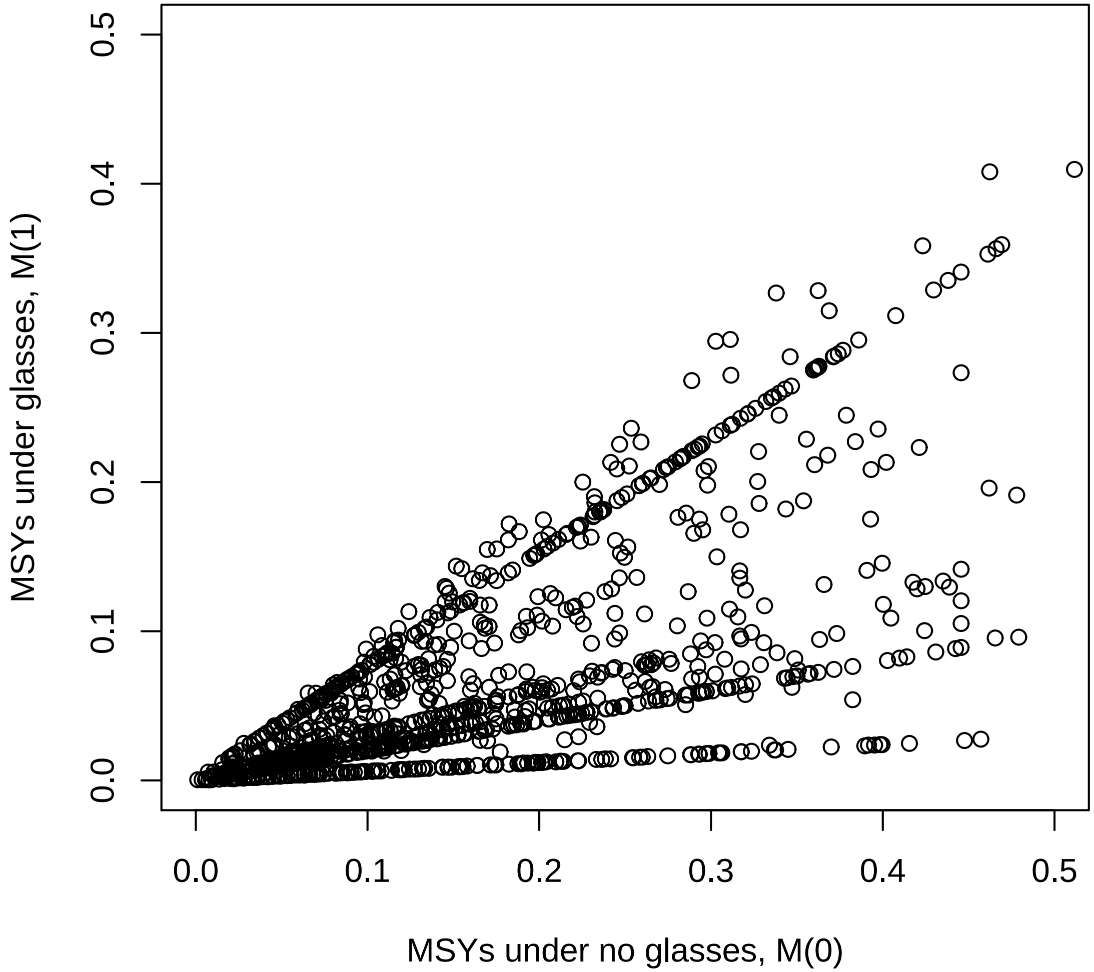


Table 4: Results from logistic model of outdoor glasses use at age 31.

Variable	Estimate	95% CI
Intercept	2.82	(-1.15, 6.80)
Age	-0.03	(-0.09, 0.02)
Age spline term	0.00	(-0.09, 0.09)
Diabetic	0.12	(-0.13, 0.36)
Male	-0.53	(-0.78, -0.29)
Black	0.39	(0.16, 0.61)
Not high school grad	-0.39	(-0.58, -0.19)
Worked over water	Reference	
Worked outside	-0.44	(-1.24, 0.35)
Worked inside	-0.25	(-1.10, 0.59)
Worked as homemaker	-0.26	(-1.11, 0.60)
UV	3.96	(-1.72, 9.63)
UV cubic spline term 1	-24.70	(-52.11, 2.72)
UV cubic spline term 2	59.02	(-3.02, 121.07)

Table 5: Results from Beta regression of P=M(1)/M(0)

Variable	Estimate	95% CI
Intercept	-0.56	(-2.98, 1.86)
Age	-0.01	(-0.04, 0.03)
Age spline term	0.02	(-0.04, 0.08)
Diabetic	-0.02	(-0.17, 0.13)
Male	0.17	(0.02, 0.33)
Black	-0.09	(-0.24, 0.06)
Not high school grad	0.02	(-0.10, 0.14)
Worked over water	Reference	
Worked outside	-0.08	(-0.62, 0.47)
Worked inside	0.18	(-0.38, 0.75)
Worked as homemaker	0.02	(-0.55, 0.58)
UV	4.40	(0.87, 7.93)
UV cubic spline term 1	-10.61	(-27.69, 6.47)
UV cubic spline term 2	19.72	(-18.94, 58.38)
ϕ	3.12	(2.89, 3.35)

Analysis: Relative Risk