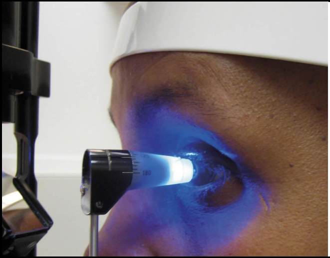
Beyond Estimation: Next Steps for Precision Medicine

Michael Wallace, University of Waterloo

Glaucoma: One Disease, Many Treatments

Glaucoma: group of eye diseases associated with elevated intraocular pressure (IOP).

IOP can be measured in various ways.



Glaucoma: One Disease, Many Treatments

Treatment options attempt to lower IOP, they include:

- Lifestyle changes.
- Eye drops (numerous options).
- Surgery.

Glaucoma: One Disease, Many Treatments

Treatment decisions are made based on various factors:

- Current and past IOP.
- Current and past treatments.
- Concerns over side effects.
- Broader risk factors.
- Other characteristics (such as age).

Glaucoma: One Disease, Many Treatments

Treatment decisions are made based on various factors:

- Current and past IOP.
- Current and past treatments.
- Concerns over side effects.
- Broader risk factors.
- Other characteristics (such as age).

Precision Medicine: tailoring treatment decisions to patient-level characteristics.

- Dynamic treatment regimes (DTRs) ‘formalize’ personalized treatment:

  “Patient presents with historic IOP of 13 and is taking Azarga. If current IOP is 15 or higher, add Alphagan, otherwise continue with only Azarga.”

- Dynamic treatment regimes (DTRs) 'formalize' personalized treatment:

  "Patient presents with historic IOP of 13 and is taking Azarga. If current IOP is 15 or higher, add Alphagan, otherwise continue with only Azarga."

- How do we choose the best DTR?

  Should our IOP cut-off be 13, 15, 20?

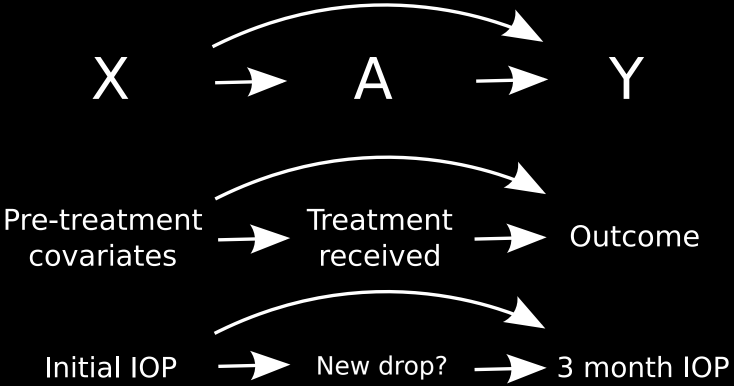
- What makes this difficult?

We typically work with data from observational studies.

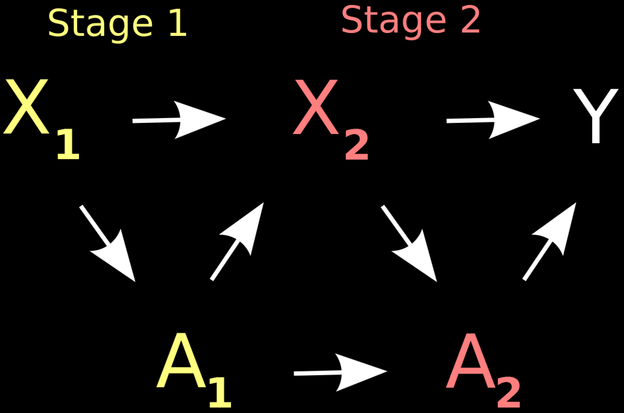
Patient	Observed IOP	Drop added?	IOP at 3 months
1	16	No	15
2	20	Yes	16
3	21	Yes	17
4	16	Yes	16
5	15	No	18
...	...	...	...

We typically work with data from observational studies.

Patient	Observed IOP X	Drop added? A	IOP at 3 months Y
1	16	No	15
2	20	Yes	16
3	21	Yes	17
4	16	Yes	16
5	15	No	18
...	...	...	...



DTR: treatment A^{opt} that optimizes $E[Y|X, A^{opt}]$



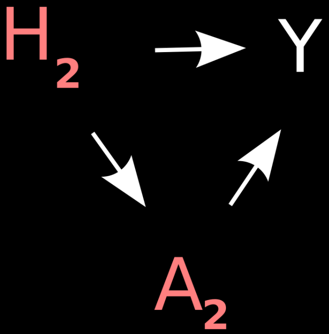
More often,
we work
in stages.

DTR: treatment sequence A_1^{opt}, A_2^{opt}

Identifying the best treatment regime: multi-stage

$$H_2 = (X_1, A_1, X_2)$$

Stage 2



Single Stage Analysis

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

Patient	Observed IOP X	Drop added? A	IOP at 3 months Y
1	16	No	15
2	20	Yes	16
3	21	Yes	17
4	16	Yes	16
5	15	No	18
...	...	...	...

Lots of methods available:

Q-learning

MSMs

G-estimation

IPTW

dWOLS

OWL

A-learning

etc...

Lots of methods available:

Q-learning

MSMs

G-estimation

IPTW

dWOLS

OWL

A-learning

etc...

- If only one treatment decision:

$$\underbrace{E[Y|X, A]}_{\substack{\text{Expected outcome} \\ \text{(to be maximized)}}}$$

- If only one treatment decision:

$$\underbrace{E[Y|X, A]}_{\substack{\text{Expected outcome} \\ \text{(to be maximized)}}}$$

- We might propose the following model

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \text{IOP} + A(\psi_0 + \psi_1 \text{IOP})$$

"Treat $(A = 1)$ if $\psi_0 + \psi_1 \text{IOP} > 0$"

- If only one treatment decision:

$$\underbrace{E[Y|X, A]}_{\substack{\text{Expected outcome} \\ \text{(to be maximized)}}}$$

- We might propose the following model

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \mathrm{IOP} + A(\psi_0 + \psi_1 \mathrm{IOP})$$

"Treat $(A = 1)$ if $\psi_0 + \psi_1 \mathrm{IOP} > 0$"

- More generally, split outcome into two components:

$$\underbrace{E[Y|X, A; \beta, \psi]}_{\substack{\text{Expected outcome} \\ \text{(to be maximized)}}} = \overbrace{G(X; \beta)}^{\substack{\text{Impact of patient history} \\ \text{in the absence of treatment}}} + \underbrace{\gamma(X, A; \psi)}_{\substack{\text{Impact of treatment} \\ \text{on outcome}}}$$

- Simplifies focus: find A^{opt} that maximizes $\gamma(X, A; \psi)$.

- Suppose the true outcome model is:

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \mathrm{IOP} + \beta_2 \mathrm{IOP}^2 + A(\psi_0 + \psi_1 \mathrm{IOP})$$

- Suppose the true outcome model is:

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \mathrm{IOP} + \beta_2 \mathrm{IOP}^2 + A(\psi_0 + \psi_1 \mathrm{IOP})$$

- But we propose:

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \mathrm{IOP} + A(\psi_0 + \psi_1 \mathrm{IOP})$$

$$E[Y|X, A; \beta, \psi] = G(X; \beta) + \gamma(X, A; \psi)$$

- Three models to specify:
  1. Blip model: $\gamma(X, A; \psi)$.
  2. Treatment-free model: $G(X; \beta)$.
  3. Treatment model: $P(A = 1|X; \alpha)$.

$$E[Y|X, A; \beta, \psi] = G(X; \beta) + \gamma(X, A; \psi)$$

- Three models to specify:
  1. Blip model: $\gamma(X, A; \psi)$.
  2. Treatment-free model: $G(X; \beta)$.
  3. Treatment model: $P(A = 1|X; \alpha)$.
- Estimate ψ via WOLS of Y on covariates in blip and treatment-free models, with weights $w = |A - P(A = 1|X; \hat{\alpha})|$.

- Suppose the true outcome model is:

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \text{IOP} + \beta_2 \text{IOP}^2 + A(\psi_0 + \psi_1 \text{IOP})$$

- Suppose the true outcome model is:

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \mathrm{IOP} + \beta_2 \mathrm{IOP}^2 + A(\psi_0 + \psi_1 \mathrm{IOP})$$

- But we propose:

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \mathrm{IOP} + A(\psi_0 + \psi_1 \mathrm{IOP})$$

- Suppose the true outcome model is:

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \text{IOP} + \beta_2 \text{IOP}^2 + A(\psi_0 + \psi_1 \text{IOP})$$

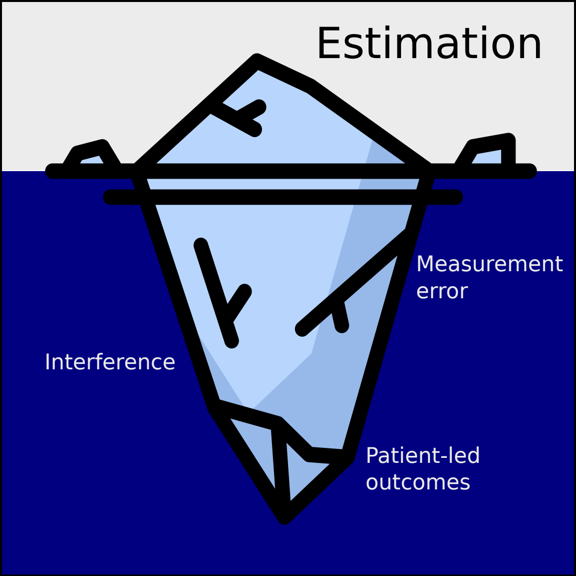
- But we propose:

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \text{IOP} + A(\psi_0 + \psi_1 \text{IOP})$$

- A weighted regression with weights $w = |A - P(A = 1|X; \hat{\alpha})|$ will still yield consistent estimators of ψ_0, ψ_1.

- Suppose the true outcome model is:
$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \text{IOP} + \beta_2 \text{IOP}^2 + A(\psi_0 + \psi_1 \text{IOP})$$
- But we propose:
$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 \text{IOP} + A(\psi_0 + \psi_1 \text{IOP})$$
- A weighted regression with weights $w = |A - P(A = 1|X; \hat{\alpha})|$ will still yield consistent estimators of ψ_0, ψ_1.
- The estimators are "doubly robust": consistent if at least one of the treatment-free or treatment components is correctly specified.
- The blip must always be correct.

So all we have to do is specify some models, estimate the model parameters, then choose the treatment that maximizes the expected outcome - easy!

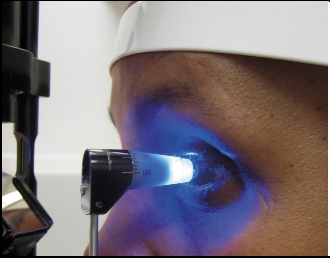

Estimation
Measurement error
Interference
Patient-led outcomes

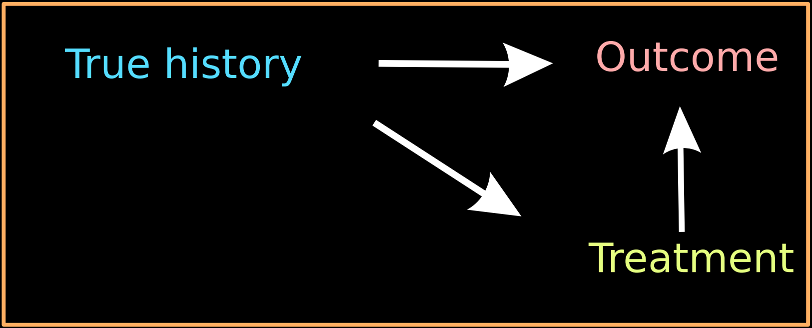
Measurement error

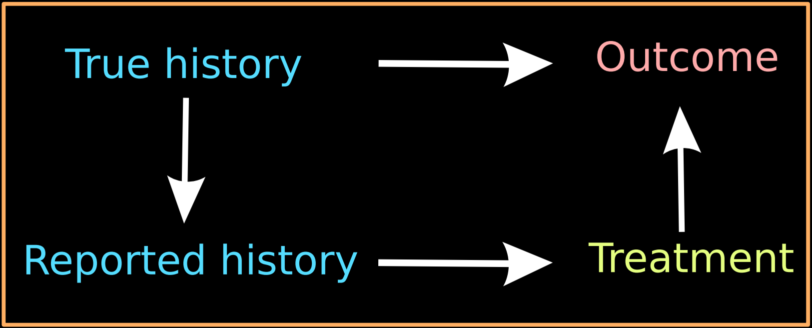
Target measurement: ‘average’ IOP.

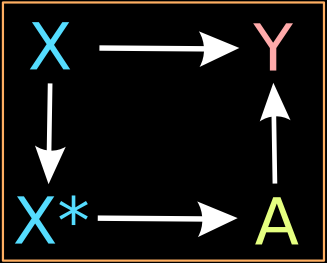
Observed measurement: 1-3 in-clinic readings within <5 minutes.

Some patients have access to more regular at-home tonometry.




True history
Outcome
Treatment

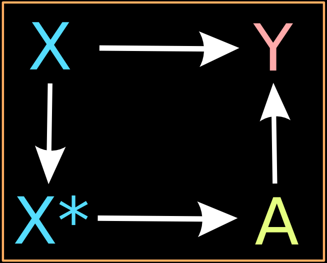

True history
Outcome
Reported history
Treatment



Estimation: suppose the true outcome model is:

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 X + \beta_2 X^2 + A(\psi_0 + \psi_1 X)$$

If we only observe X^*. What happens? What can we do about it?



Estimation: suppose the true outcome model is:

$$E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 X + \beta_2 X^2 + A(\psi_0 + \psi_1 X)$$

If we only observe X^*. What happens? What can we do about it?

Solution: 'correct' for the measurement error using additional data.

Assume: classical additive measurement error:

$$\text{Observed} = \text{True} + \text{Error}$$
$$X^* = X + U$$

$U \sim N(0, \sigma_u^2)$; $Y \perp X^* | X$

Assume: replicate measurements available on at least some patients.

Patient	First IOP measurement	Second IOP measurement
1	16	15
2	20	16
3	21	17
4	16	16
5	15	18
...	...	...

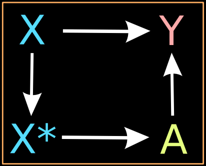
Simple correction method: Regression Calibration.

Principle:

1. Use additional data to estimate $E[X|X^*, A] = X_{rc}$.
2. Replace X with X_{rc} and carry out a standard analysis.
3. Adjust the resulting standard errors to account for the estimation in step 1.

- Suppose the true outcome model is:

$$E[Y|\cdot] = \beta_0 + \beta_1 X + \beta_2 X^2 + A(\psi_0 + \psi_1 X)$$

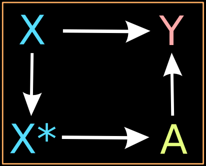


- Suppose the true outcome model is:

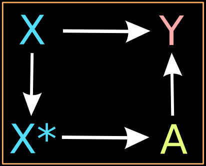
$$E[Y|\cdot] = \beta_0 + \beta_1 X + \beta_2 X^2 + A(\psi_0 + \psi_1 X)$$

- If we have RC estimates X_{rc} then we could fit

$$E[Y|\cdot] = \beta_0 + \beta_1 X_{rc} + \beta_2 X_{rc}^2 + A(\psi_0 + \psi_1 X_{rc})$$



- Suppose the true outcome model is:
$$E[Y|\cdot] = \beta_0 + \beta_1 X + \beta_2 X^2 + A(\psi_0 + \psi_1 X)$$
- If we have RC estimates X_{rc} then we could fit
$$E[Y|\cdot] = \beta_0 + \beta_1 X_{rc} + \beta_2 X_{rc}^2 + A(\psi_0 + \psi_1 X_{rc})$$
- But we might mis-specify the model as
$$E[Y|\cdot] = \beta_0 + \beta_1 X_{rc} + A(\psi_0 + \psi_1 X_{rc})$$
where A depends on X^*.



- Suppose the true outcome model is:
$$E[Y|\cdot] = \beta_0 + \beta_1 X + \beta_2 X^2 + A(\psi_0 + \psi_1 X)$$
- If we have RC estimates X_{rc} then we could fit
$$E[Y|\cdot] = \beta_0 + \beta_1 X_{rc} + \beta_2 X_{rc}^2 + A(\psi_0 + \psi_1 X_{rc})$$
- But we might mis-specify the model as
$$E[Y|\cdot] = \beta_0 + \beta_1 X_{rc} + A(\psi_0 + \psi_1 X_{rc})$$
where A depends on X^*.
- Establish (approximate) covariate balance in X_{rc} by regressing A on X_{rc}.

$$\begin{array}{ccc} X & \longrightarrow & Y \\ \downarrow & & \uparrow \\ X^* & \longrightarrow & A \end{array}$$

Suppose we conclude that our treatment rule should be:

"If 3-month average IOP > 15 add secondary drop, otherwise, maintain current treatment regime."

Suppose we conclude that our treatment rule should be:

"If 3-month average IOP > 15 add secondary drop, otherwise, maintain current treatment regime."

I go to the clinic and my IOP measurement is 16. Then what?

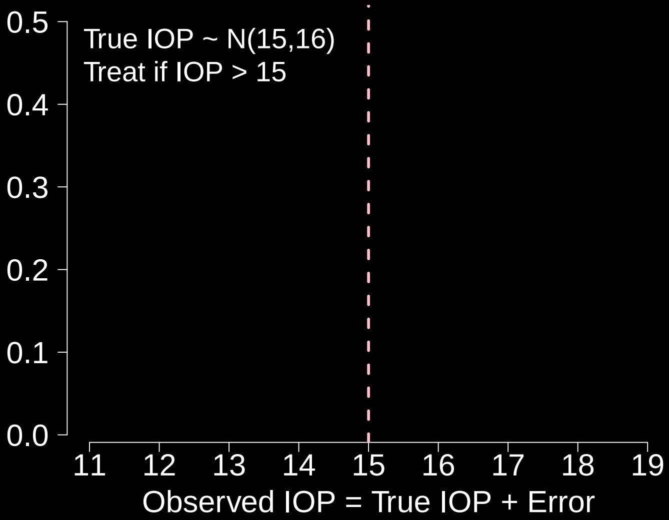
Suppose we conclude that our treatment rule should be:

"If 3-month average IOP > 15 add secondary drop, otherwise, maintain current treatment regime."

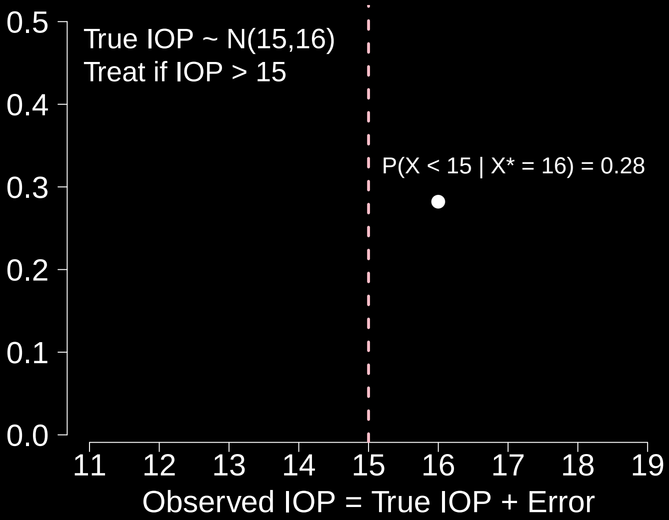
I go to the clinic and my IOP measurement is 16. Then what?

What is $P(X \leq 15 | X^* = 16)$?

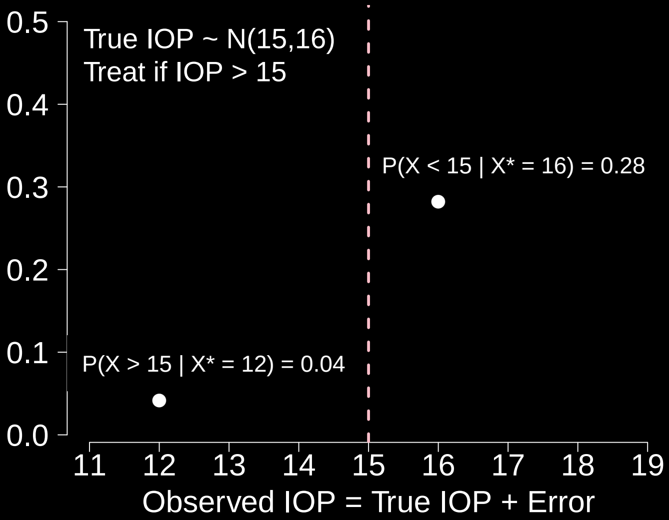
In some settings, results fairly intuitive:



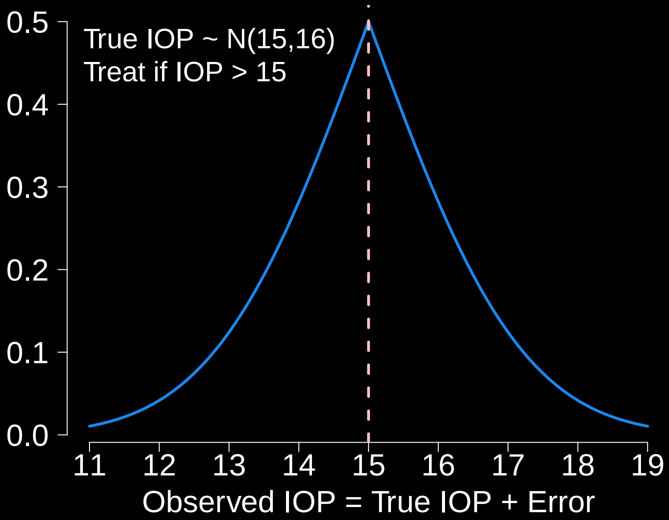
In some settings, results fairly intuitive:



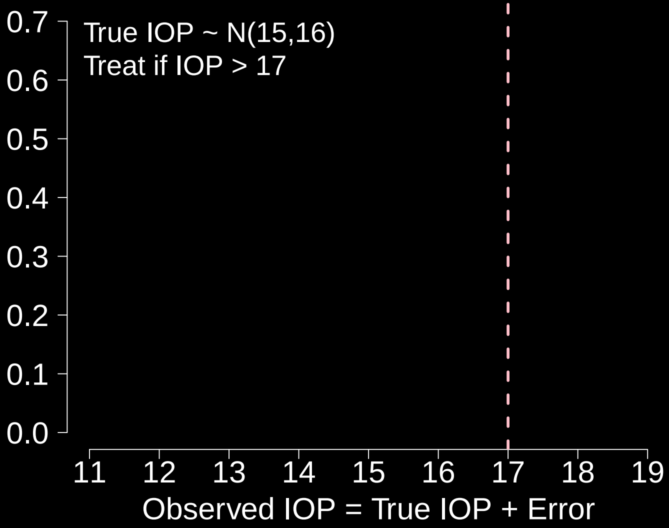
In some settings, results fairly intuitive:



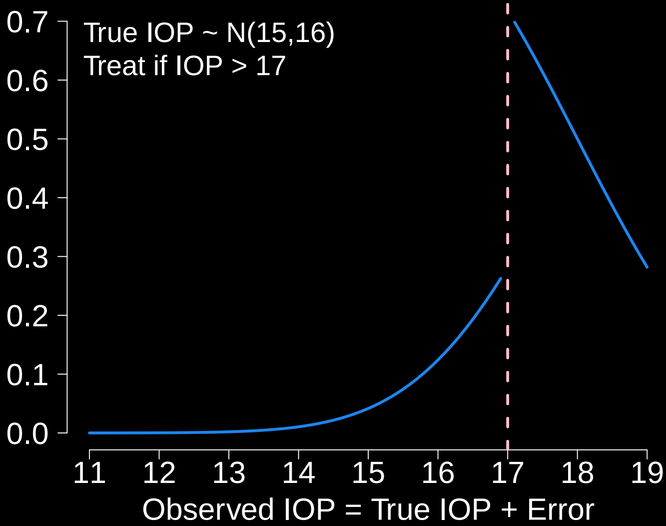
In some settings, results fairly intuitive:



In others, perhaps more of a surprise (to some):



In others, perhaps more of a surprise (to some):



“Isn’t this just a prediction problem?”

Data availability will vary by study and by variable:

Scenario	Analysis	Application
1: “We never observe the truth.”	X^*	X^*

“Isn’t this just a prediction problem?”

What if error-free data are possible, but expensive?

Scenario	Analysis	Application
1: “We never observe the truth.”	X^*	X^*
2: “Past data are error-prone, but future data may not be.”	X^*	X
3: “Past data are not error-prone, but future data may be.”	X	X^*

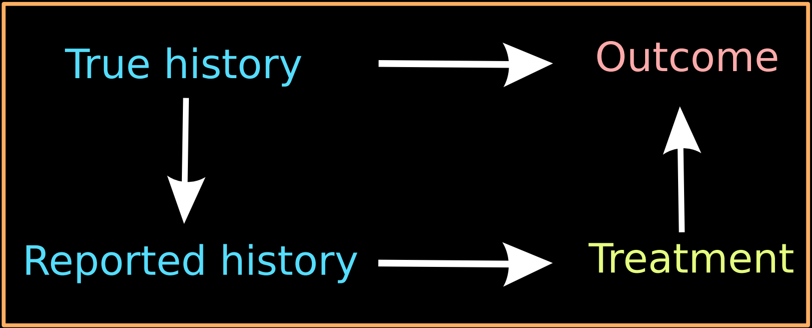
Only Scenario 4 is well-studied.

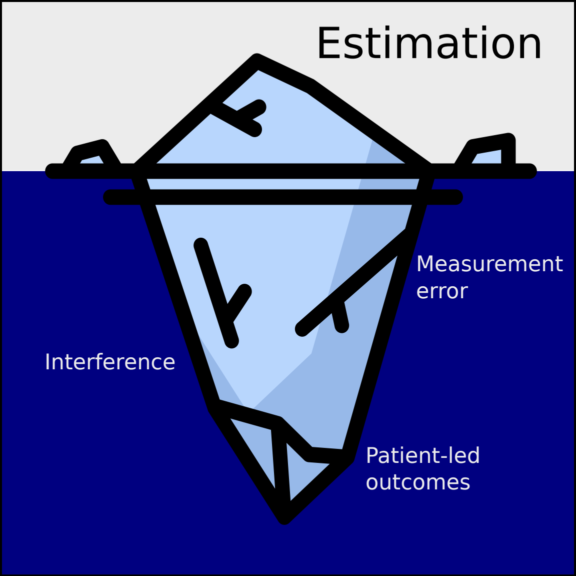
Scenario	Analysis	Application
1: "We never observe the truth."	X^*	X^*
2: "Past data are error-prone, but future data may not be."	X^*	X
3: "Past data are not error-prone, but future data may be."	X	X^*
4: "We always observe the truth."	X	X

"Isn't this just a prediction problem?"

Correcting for measurement error is <u>at worst</u> competitive with an analysis that ignores it completely.

Scenario	Analysis	Application
1: "We never observe the truth."	X^*	X^*
2: "We always observe the truth."	X	X
3: "Past data are error-prone, but future data may not be."	X^*	X
4: "Past data are not error-prone, but future data may be."	X	X^*


True history
Outcome
Reported history
Treatment

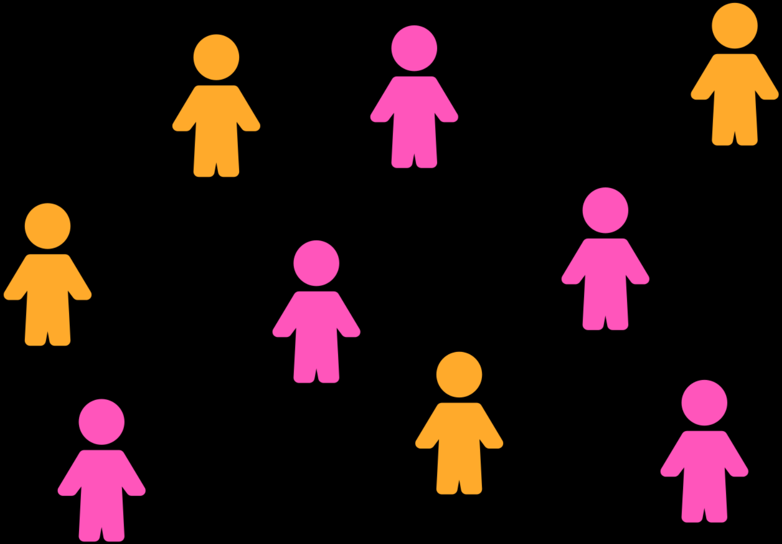

Estimation
Measurement error
Interference
Patient-led outcomes

No man is an Island, entire of itself; every man is a piece of the Continent, a part of the main; if a clod be washed away by the sea, Europe is the less, as well as if a promontory were, as well as if a manor of thy friends or of thine own were; any man's death diminishes me, because I am involved in Mankind; And therefore never send to know for whom the bell tolls; It tolls for thee.

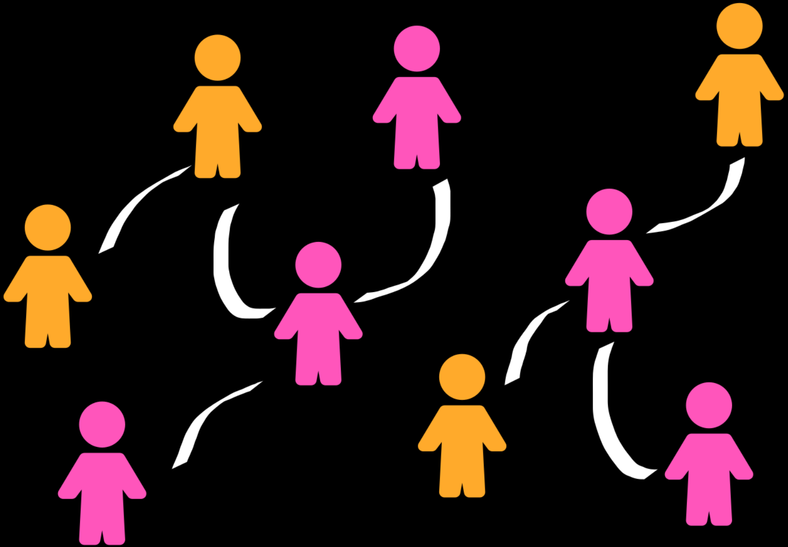
(John Donne)

Interference: another patient's treatment doesn't affect my outcome.

Common assumption: no interference (or ‘spillover’):

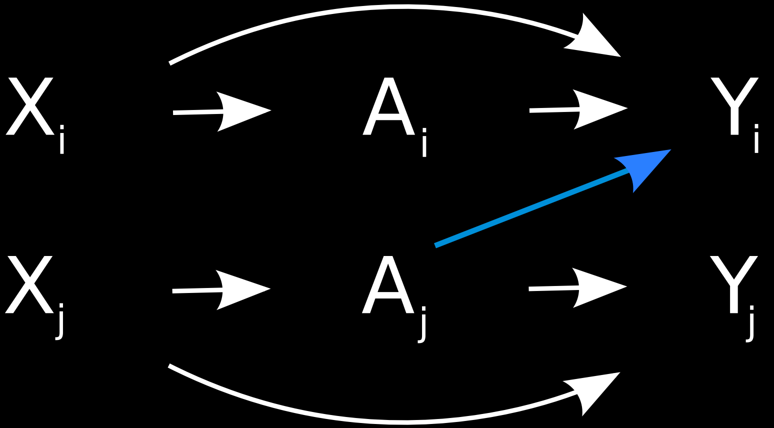


Common assumption: no interference (or 'spillover'):



Interference

$$\begin{tikzcd}
X_i \arrow[r] \arrow[rr, bend left] & A_i \arrow[r] & Y_i \\
X_j \arrow[r] \arrow[rr, bend right] & A_j \arrow[r] & Y_j
\end{tikzcd}$$


X_i
A_i
Y_i
X_j
A_j
Y_j

- No interference:

$$\underbrace{E[Y|X,A]}_{\substack{\text{Expected outcome}\\\text{(to be maximized)}}}$$

- We might propose the following model

$$E[Y_i|X_i, A_i; \beta, \psi] = \beta_0 + \beta_1 X_i + A_i(\psi_0 + \psi_1 X_i)$$

"Treat ($A_i = 1$) if $\psi_0 + \psi_1 X_i > 0$"

With interference: two ideas:

- (1) Add interaction terms:
  $E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 X_i + \beta_2 X_j + A_i(\psi_0 + \psi_1 X_i + \psi_2 A_j)$
- Then dWOLS proceeds as usual: "Treat ($A_i = 1$) if $\psi_0 + \psi_1 X_i + \psi_2 A_j > 0$"

With interference: two ideas:

- (1) Add interaction terms:
  $E[Y|X, A; \beta, \psi] = \beta_0 + \beta_1 X_i + \beta_2 X_j + A_i(\psi_0 + \psi_1 X_i + \psi_2 A_j)$
- Then dWOLS proceeds as usual: "Treat ($A_i = 1$) if $\psi_0 + \psi_1 X_i + \psi_2 A_j > 0$"
- (2) Use 'network propensity weights': for each individual i, apply weights based on the probability their neighbour is treated.

Estimation is only half the battle.

Example: Alex and Blake share a household. The following table summarizes the ranking of the four possible combinations of a binary treatment:

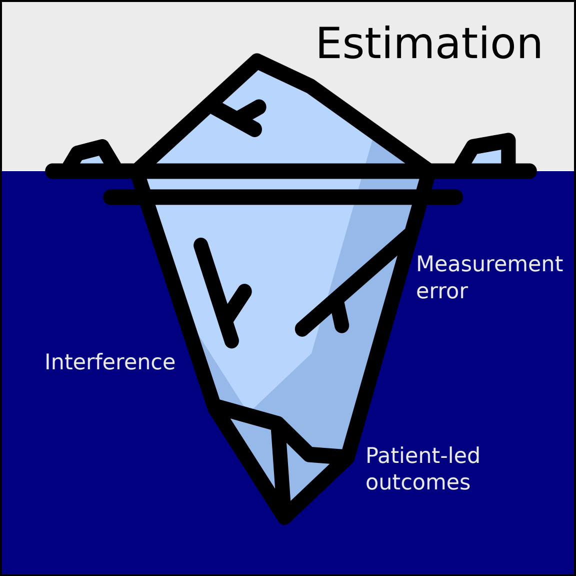
Rank	Alex	Blake
1	o	-
2	-	o
3	o	o
4	-	-

Treatment: o
No treatment: -

- In what order do we prescribe treatments to patients?
- What if resources are limited?
- How do we balance the needs of the individual against the needs of the population?

Rank	Alex	Blake
1	o	-
2	-	o
3	o	o
4	-	-

Treatment: o
No treatment: -


Estimation
Measurement error
Interference
Patient-led outcomes

Recall the treatment options for glaucoma:

- Lifestyle changes.
- Eye drops (numerous options).
- Surgery.

Treatments have various side effects, including:

- ‘Minor’: eyelash growth, iris discoloration.
- ‘Major’: additional vision loss.

Prior research tells us the probability of side effects, and the 'average' IOP decrease following each treatment.

Suppose you had the swallowing predicted effects:

Treatment	Iris discoloration	Eyelash growth	Vision loss	E[IOP change]
Drop A	80%	40%	1%	2
Drop B	90%	30%	1%	2

Which treatment do you choose? What influences your decision?

How do we elicit this information from patients?

Prior research tells us the probability of side effects, and the 'average' IOP decrease following each treatment.

Suppose you had the swallowing predicted effects:

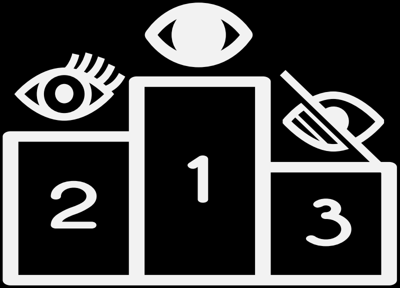
Treatment	Iris discoloration	Eyelash growth	Vision loss	E[IOP change]
Drop A	80%	40%	1%	2
Drop B	90%	30%	1%	2
Surgery	1%	0%	5%	3

Which treatment do you choose? What influences your decision?

How do we elicit this information from patients?

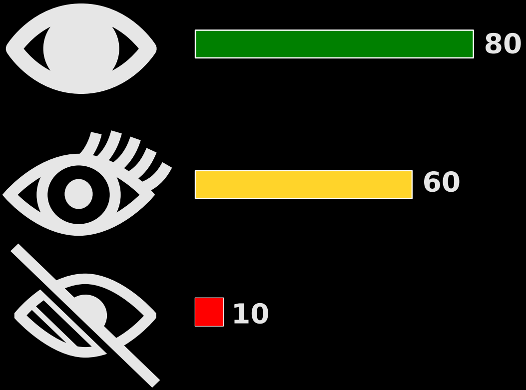
Patients can provide information to varying degrees:

- Ranking: "I would prefer iris discoloration to eyelash growth, and prefer both to additional vision loss."



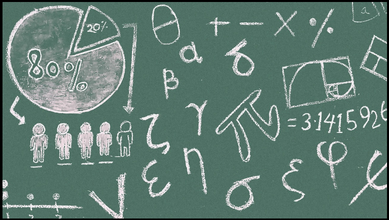
Patients can provide information to varying degrees:

- Weighting: "I would prefer iris discoloration twice as much as eyelash growth, and ten times as much as additional vision loss."



Patients can provide information to varying degrees:

- Utility functions: "I want to maximize this complicated function of all the possible outcomes."



Patients can provide information to varying degrees:

- Ranking: "I would prefer iris discoloration to eyelash growth, and prefer both to additional vision loss."
- Weighting: "I would prefer iris discoloration twice as much as eyelash growth, and ten times as much as additional vision loss."
- Utility functions: "I want to maximize this function of the possible outcomes."

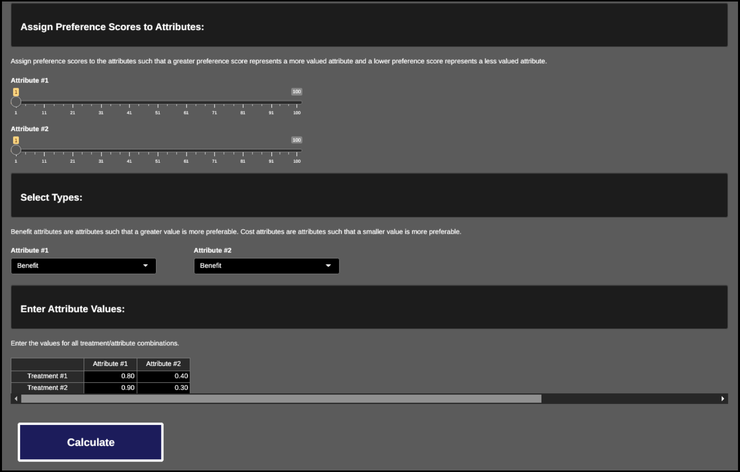
Methods from *multi-attribute decision making* can be applied to recommend treatment options.

We can also consider non-probabilistic properties of treatments, such as their cost.

Treatment	Cost	Iris discoloration	Eyelash growth	Vision loss	IOP change
Drop A	$	80%	40%	1%	2
Drop B	$$	90%	30%	1%	2
Surgery	$$$	1%	0%	5%	3

Such methods can be implemented via web-based applications:

https://shiny.math.uwaterloo.ca/sas/mwallace/mapp/madm/



Such methods can be implemented via web-based applications:

`https://shiny.math.uwaterloo.ca/sas/mwallace/mapp/madm/`

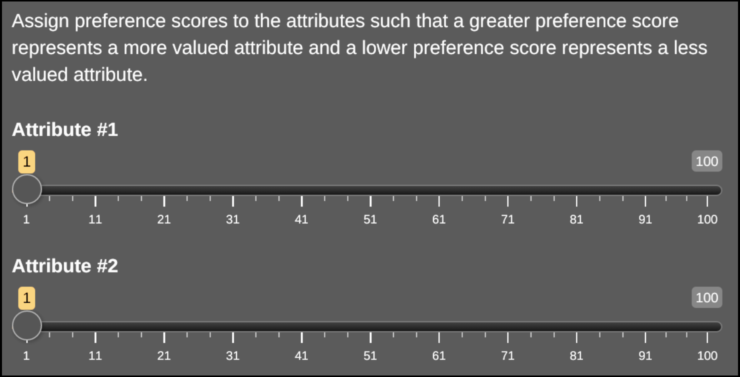
Enter Attribute Values:

Enter the values for all treatment/attribute combinations.

	Attribute #1	Attribute #2
Treatment #1	0.80	0.40
Treatment #2	0.90	0.30

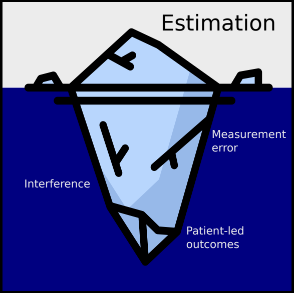
Such methods can be implemented via web-based applications:

`https://shiny.math.uwaterloo.ca/sas/mwallace/mapp/madm/`



So where are we now?

- DTRs an important tool in precision medicine.
- Lots of estimation methods available, ease of implementation, and avoiding ‘black boxes’ a big challenge.
- Beyond estimation: lots of interesting, open, practical problems to work on.
  - Measurement error.
  - Interference.
  - Patient-led outcomes.
  - Model assessment.
  - Treatment quality.
  - etc. etc. etc.



Acknowledgments

Dylan Spicker: Measurement Error

Cong Jiang: Interference and Networks

Grace Tompkins: Patient-Led Outcomes

Marzieh Rizi: Interference and Networks

NSERC CRSNG

CIHR IRSC Canadian Institutes of Health Research Instituts de recherche en santé du Canada

- **dWOLS**: M. P. Wallace and E. E. M. Moodie (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics* **71(3)** 636-644.
- **Precision Medicine and Measurement Error**: D. Spicker and M. P. Wallace (2020). Measurement error and precision medicine: error-prone tailoring covariates in dynamic treatment regimes. *Statistics in Medicine* **39(26)** https://doi.org/10.1002/sim.8690
- **Precision Medicine and Interference**: C. Jiang, M. E. Thompson, M. P. Wallace (2022). Dynamic treatment regimes with interference. *Canadian Journal of Statistics. In press*

michael.wallace@uwaterloo.ca @statacake

https://shiny.math.uwaterloo.ca/sas/mwallace/mapp/madm/