STAT/BIOST 572
Final Presentation

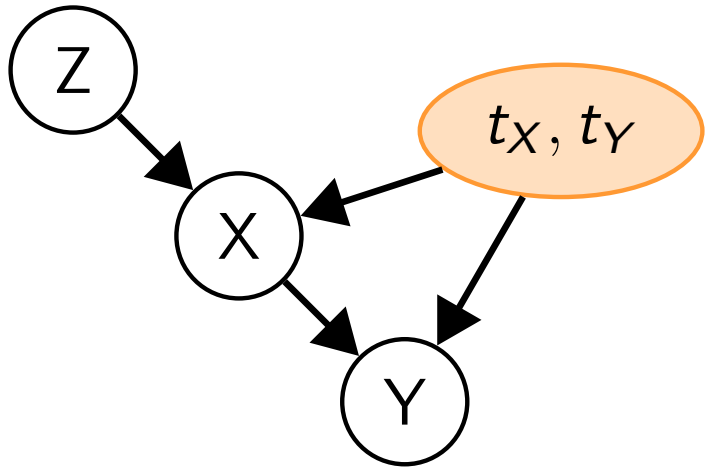
Transparent Parameterizations of Models for Potential Outcomes [Richardson et al., 2011]

Wen Wei Loh

May 15, 2012

Recap: Potential Outcomes under IV model

- Observed
  - Z = Assignment to treatment (Instrument)
  - X = Receipt/Exposure to treatment
  - Y = Response
- **I**nstrumental **V**ariable : Effect of Z on Y is only through X
- Unobserved (due to imperfect compliance)
  - t_X = Underlying compliance "type"
  - t_Y = Underlying response "type"



Recap: t_X and t_Y

Table: Compliance types (t_X) based on potential outcomes

$X_{z=0}$	$X_{z=1}$	Compliance Type t_X	
0	0	NT	Never Taker
1	0	DE	Defier
0	1	CO	Complier
1	1	AT	Always Taker

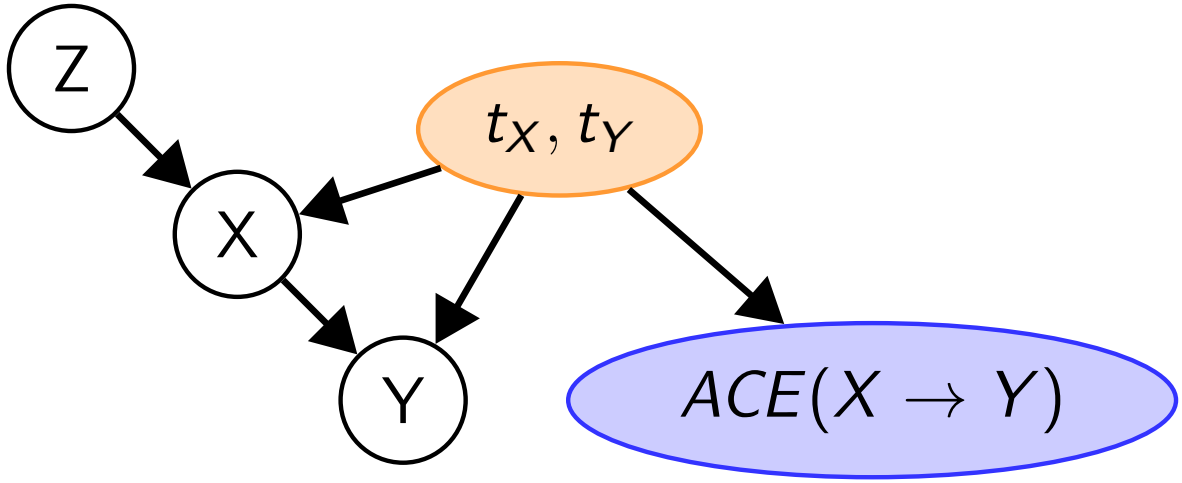
Table: Response types (t_Y) based on potential outcomes under Exclusion Restriction [Angrist et al., 1996]

$Y_{0\cdot}$	$Y_{1\cdot}$	Response Type t_Y	
0	0	NR	Never Recover
1	0	HU	Hurt
0	1	HE	Helped
1	1	AR	Always Recover

Measuring the effects of causes

- Causal estimand of interest: Average Causal Effect (ACE)
  - Marginal effect over the population
  - $ACE(X \to Y) = \mathrm{E}[Y_{x=1\cdot} - Y_{x=0\cdot}] = p(HE) - p(HU)$

    $= \sum_{t_X} \Big[p(t_X, HE) - p(t_X, HU) \Big]$
  - Depends on parameters that are not fully-identified e.g. $p(CO, HE)$.



Where are the partially-identified parameters?

$p(t_x, t_y)$		$p(y, x \mid \mathbf{z=1})$			
		y_0, x_0	y_1, x_0	y_0, x_1	y_1, x_1
$p(y, x \mid \mathbf{z=0})$	y_0, x_0	$p(NT, NR)$			$p(CO, HE)$
	y_1, x_0				$p(CO, AR)$
	y_0, x_1				
	y_1, x_1				

Example: Type $(t_x = CO, t_y = HE)$ is observed in $p(y = 0, x = 0 \mid z = 0)$, and $p(y = 1, x = 1 \mid z = 1)$.

But . . . type $(t_x = \underline{NT}, t_y = \underline{NR})$ is also observed in $p(y = 0, x = 0 \mid z = 0)$.

And . . . type $(t_x = CO, t_y = \underline{AR})$ is observed in $p(y = 1, x = 1 \mid z = 1)$

Where are the partially-identified parameters?

		$p(y, x \mid \mathbf{z=1})$			
$p(t_x, t_y)$		y_0, x_0	y_1, x_0	y_0, x_1	y_1, x_1
$p(y, x \mid \mathbf{z=0})$	y_0, x_0	$p(NT, NR)+$ $p(NT, HE)$		$p(CO, NR)$	$p(CO, HE)$
	y_1, x_0		$p(NT, AR)+$ $p(NT, HU)$	$p(CO, HU)$	$p(CO, AR)$
	y_0, x_1	$p(DE, NR)$	$p(DE, HU)$	$p(AT, NR)+$ $p(AT, HU)$	
	y_1, x_1	$p(DE, HE)$	$p(DE, AR)$		$p(AT, AR)+$ $p(AT, HE)$

Table: Two-way table for binary Instrumental Variable model

Recap: ACE bounds

- Pearl [2000] proposed bounds on ACE:
  - minimum and maximum ACE, based on the set of distributions of $p(t_X, t_Y)$ compatible with observed data
- Illustrated with data from a double-blind placebo-controlled randomized trial [1]
- No DEfiers and no Always Takers since $Z = 0 \Rightarrow X = 0$

z	x	y	count	z	x	y	count
0	0	0	158	1	0	0	52
0	0	1	14	1	0	1	12
0	1	0	*0*	1	1	0	23
0	1	1	*0*	1	1	1	78

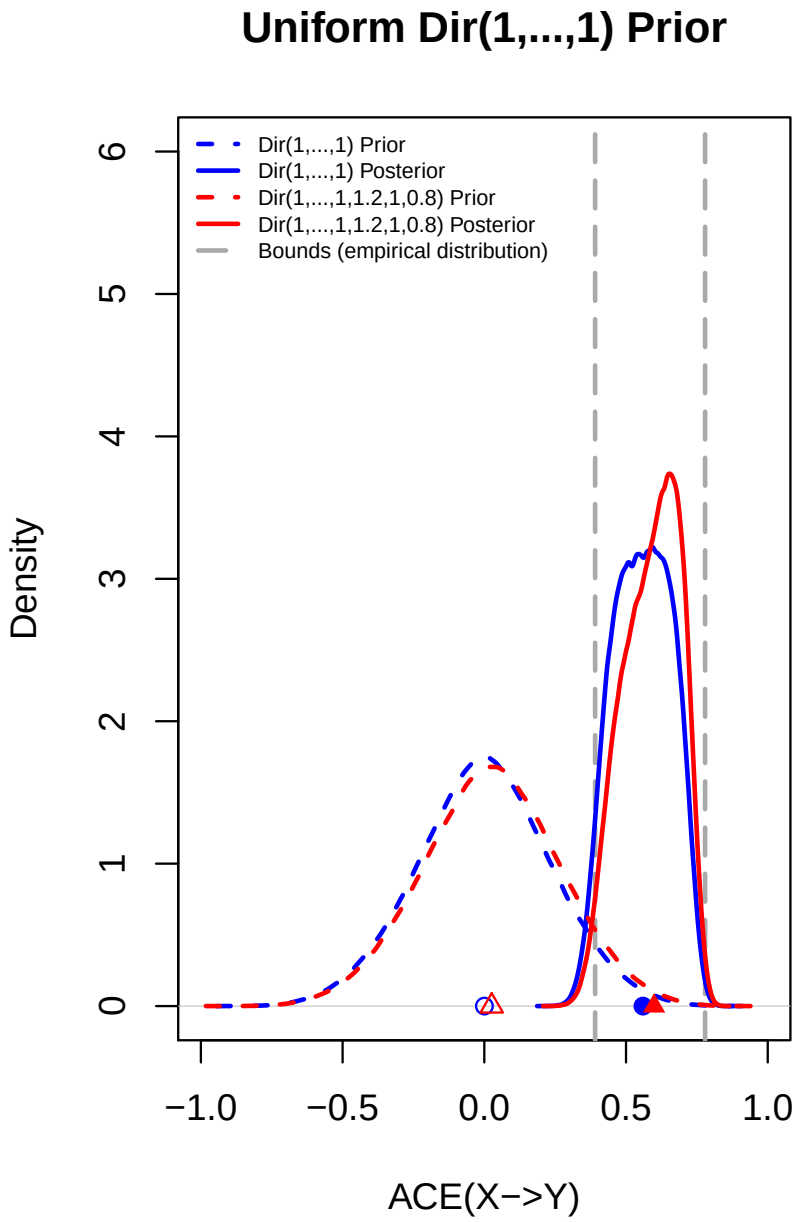
Table: Lipid data; there are two structural zeros

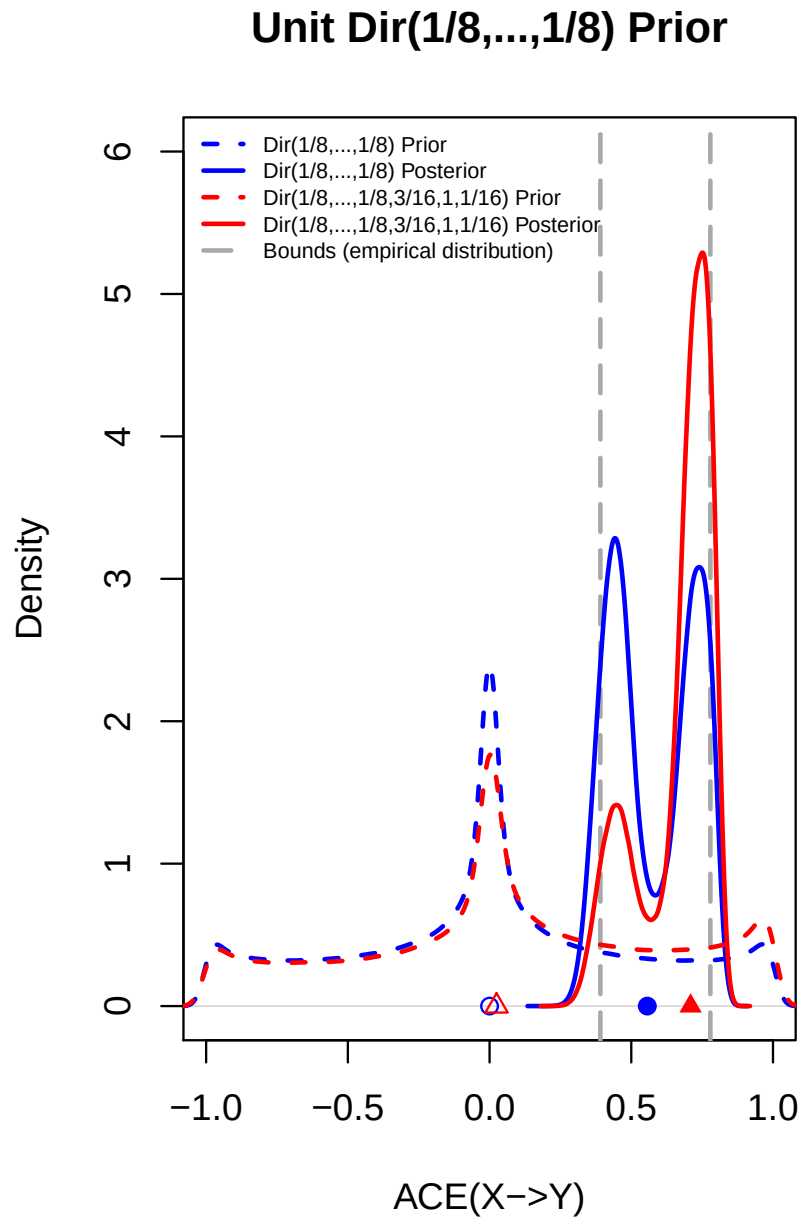
[1] Compliance was originally a continuous measure (proportion of the time during the trial that the patient took the medication), which was dichotomized by Pearl [2000]

Recap: Prior sensitivity

- Prior distribution over potential outcomes $p(t_X, t_Y)$
  - Reflects our beliefs about the proportion of individuals in population that possess characteristics corresponding to (t_X, t_Y)
  - Uniform: Dir $(1, \ldots, 1) \rightarrow$ Dir $(1, \ldots, 1.2, 1, 0.8)$
  - Unit : Dir $(\frac{1}{8}, \ldots, \frac{1}{8}) \rightarrow$ Dir $(1, \ldots, \frac{3}{16}, \frac{1}{8}, \frac{1}{16})$
  - Such perturbation should not have large effect on posterior ACE
- Gibbs sampling to sample (t_X, t_Y) from resulting posterior
- Find $ACE(X \rightarrow Y) = p(Helped \mid data) - p(Hurt \mid data)$

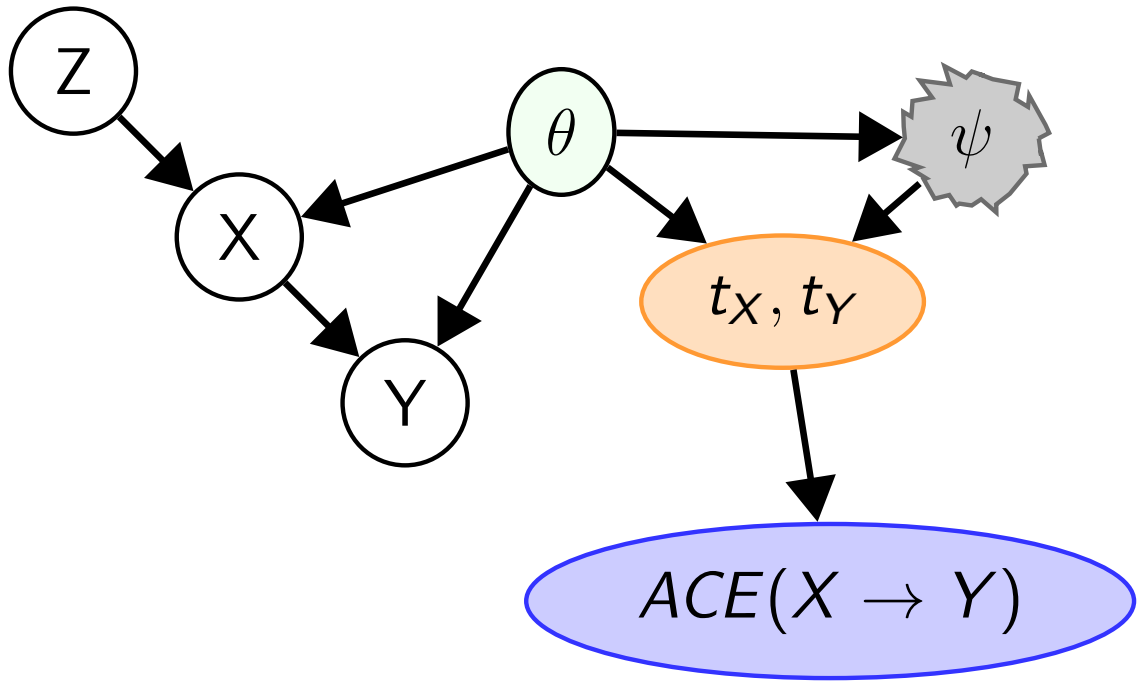
Recap: Prior sensitivity





Transparent Parametrization

- Re-parameterize $p(t_X, t_Y)$ into $f(\theta, \psi)$
  - $\theta =$ identifiable parameter
    i.e. estimable from observed (X, Y, Z)
  - $\psi =$ non-identifiable parameter
    i.e. given θ*, there is no information in the likelihood/data concerning this parameter*



Transparent Parametrization

Some notation:

- $\gamma_{t_x}^{x=i} = Pr(Y_{x=i} = 1 \mid t_x) = Pr(y = 1 \mid x = i, t_x)$
- e.g. Probability of a COmplier recovering ($y = 1$) when treatment is not received ($x = 0$) is:
  $\gamma_{CO}^{x=0} = Pr(Y_{x=0} = 1 \mid CO) = Pr(y = 1 \mid x = 0, CO)$.
- Probability does *not* depend on treatment assignment Z (i.e. Instrumental Variable under Exclusion Restriction)
- π_x = Probability that a subject is of compliance type t_x

Transparent Parametrization

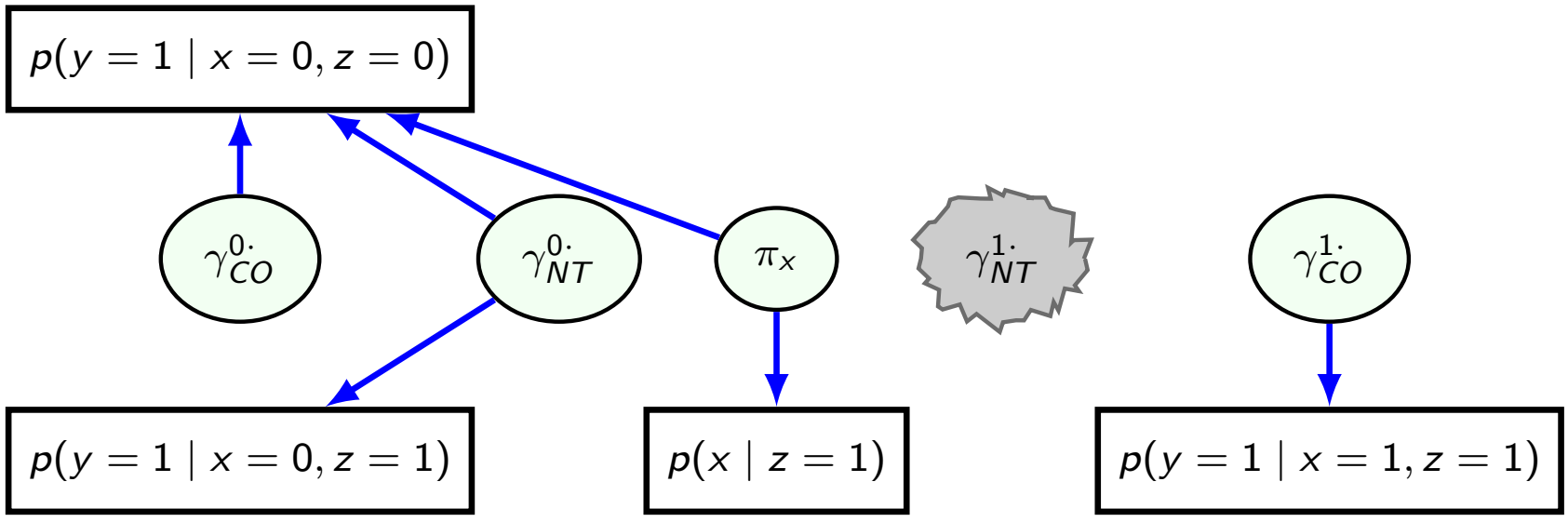
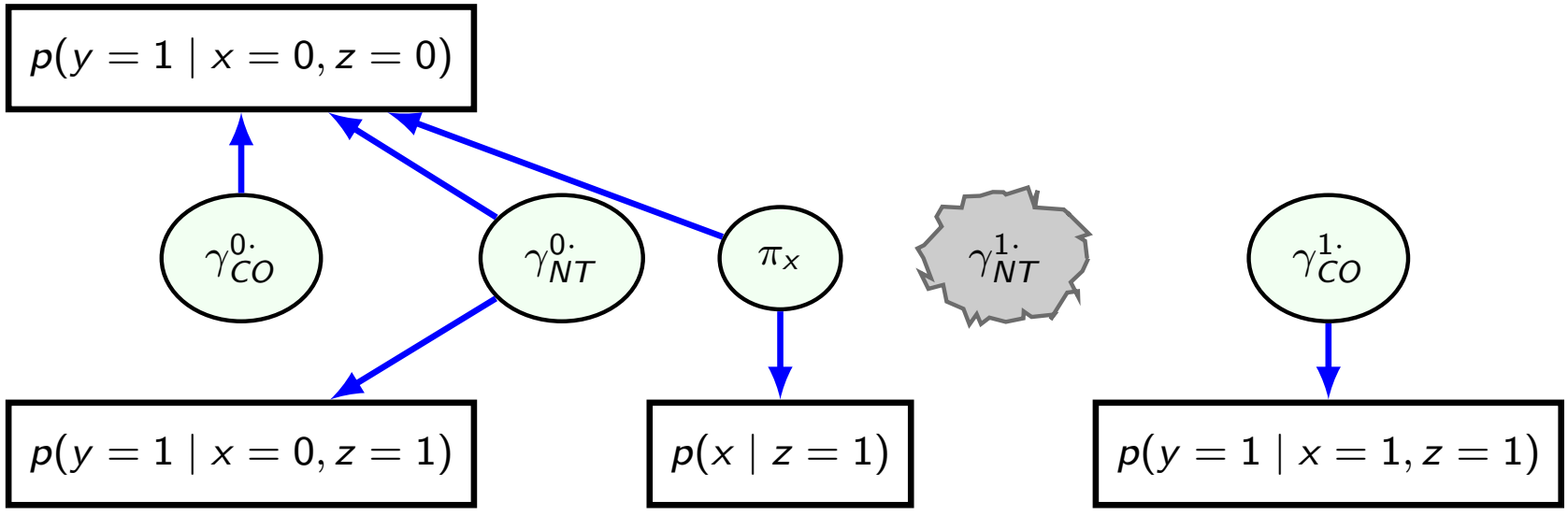


Figure: Transparent parametrization of the simple IV model with no Always Takers or DEfiers (from the Lipid data).
Oval nodes are (unknown) parameters in the model.
Rectangular nodes are observable probabilities from the data, and are *deterministic* functions of their parents (oval nodes).

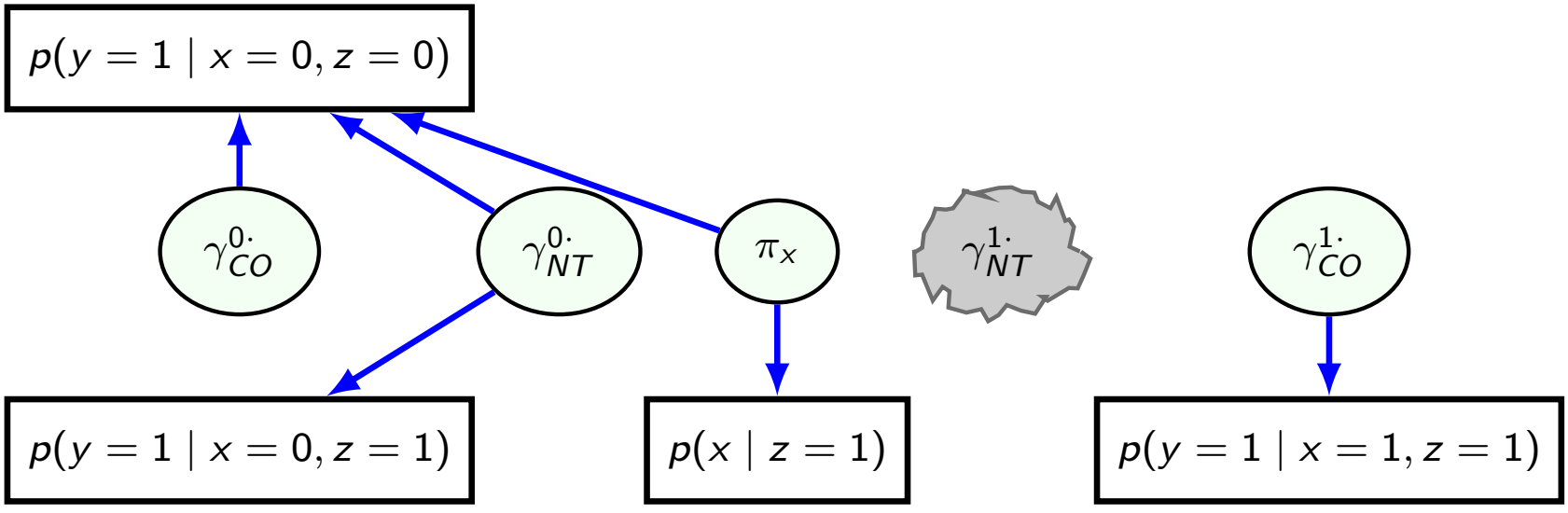
Transparent Parametrization



(a) Characterize set of distributions π_X compatible with $p(x \mid z)$:

$$\pi_{DE} = \pi_{AT} = 0$$
$$p(x = 1 \mid z = 1) = \pi_{CO}$$
$$p(x = 0 \mid z = 1) = \pi_{NT}$$

Transparent Parametrization



(b) For fixed marginal distribution π_X, describe set of values of $\gamma^{x=i}_{t_x}$ compatible with *observed* $p(y \mid x, z)$.

$$p(y = 1 \mid x = 1, z = 1) = \gamma^{1\cdot}_{CO}$$
$$p(y = 1 \mid x = 0, z = 1) = \gamma^{0\cdot}_{NT}$$
$$p(y = 1 \mid x = 0, z = 0) = \gamma^{0\cdot}_{NT} \cdot \pi_{NT} + \gamma^{0\cdot}_{CO} \cdot \pi_{CO}$$

Transparent Parametrization

(c) Identify the parameters

$$\pi_{CO} = p(x = 1 \mid z = 1)$$
$$\pi_{NT} = p(x = 0 \mid z = 1)$$
$$\gamma^{1\cdot}_{CO} = p(y = 1 \mid x = 1, z = 1)$$
$$\gamma^{0\cdot}_{NT} = p(y = 1 \mid x = 0, z = 1)$$
$$\gamma^{0\cdot}_{CO} = \frac{p(y = 1 \mid x = 0, z = 0) - \gamma^{0\cdot}_{NT} \cdot \pi_{NT}}{\pi_{CO}}$$
$$= \frac{p(y = 1, x = 0 \mid z = 0) - p(y = 1, x = 0 \mid z = 1)}{p(x = 1 \mid z = 1)}$$

(d) Find restrictions on observed $p(y, x \mid z)$

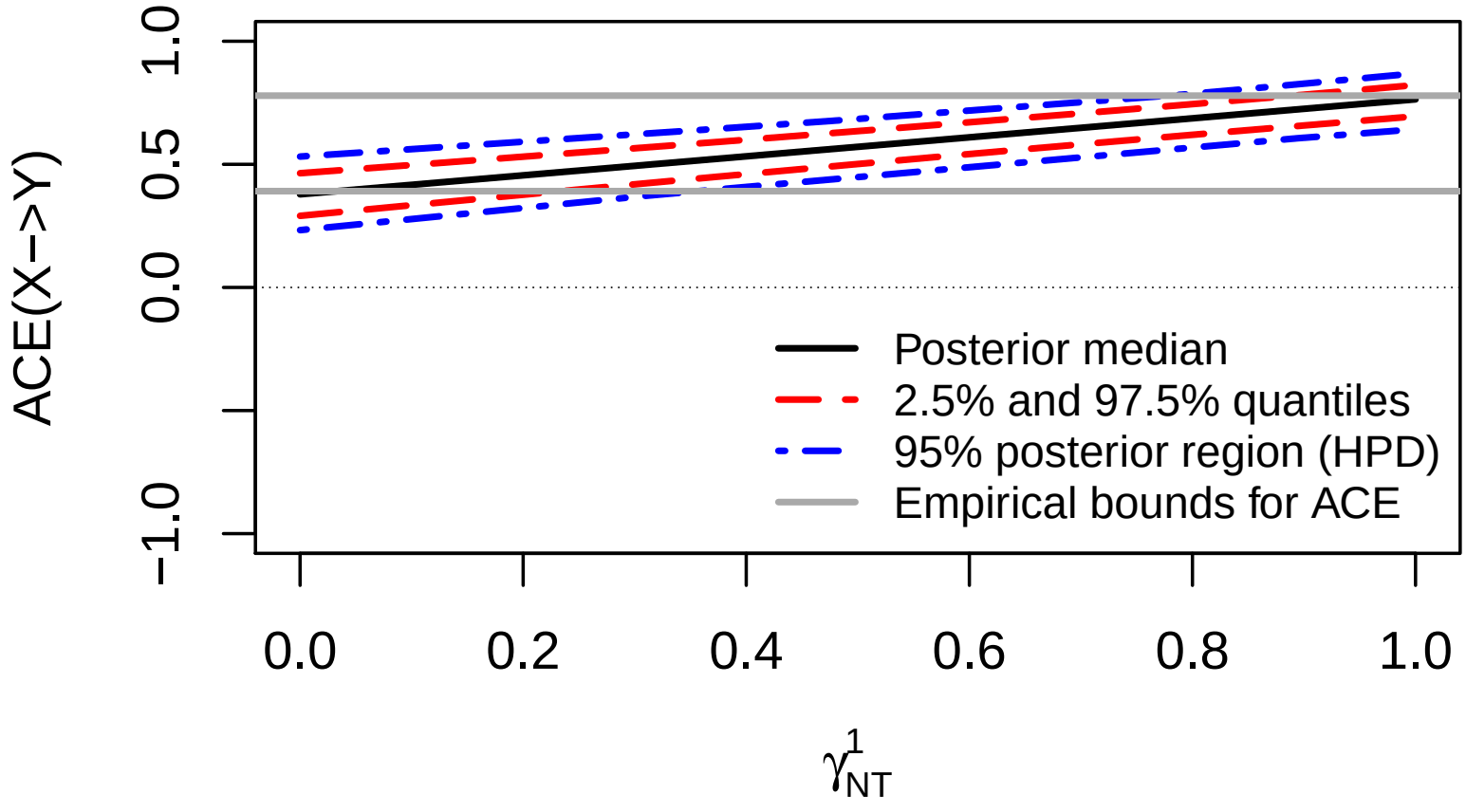
$$\gamma^{0\cdot}_{CO} \geq 0 \Rightarrow p(y = 1, x = 0 \mid z = 0) \geq p(y = 1, x = 0 \mid z = 1)$$
$$\gamma^{0\cdot}_{CO} \leq 1 \Rightarrow p(y = 0, x = 0 \mid z = 0) \geq p(y = 0, x = 0 \mid z = 1)$$

Monte Carlo rejection sampling

1. Draw from uniform Dirichlet priors for $p(y, x \mid z)$; $z = \{0, 1\}$
2. Update the conjugate Dirichlet posterior using a multinomial likelihood, and simulate from the posterior
3. Truncate using restrictions (d) on previous slide
   (discard simulations that violate the inequalities)
4. Estimate identifiable parameters using (c) on previous slide
5. Estimate causal effect as function of parameters:

$$
\begin{aligned}
\text{e.g. } ACE &= \mathrm{E}[Y_{x=1\cdot} - Y_{x=0\cdot}] \\
&= \sum_{t_X} p(t_X) \cdot \mathrm{E}[Y_{x=1\cdot} - Y_{x=0\cdot} \mid t_X] \\
&= \sum_{t_X} p(t_X) \cdot (\gamma_{t_X}^{1\cdot} - \gamma_{t_X}^{0\cdot}) \\
&= p(CO) \cdot (\gamma_{CO}^{1\cdot} - \gamma_{CO}^{0\cdot}) + p(NT) \cdot (\gamma_{NT}^{1\cdot} - \gamma_{NT}^{0\cdot}) \\
&= \text{function of } \textit{wholly} \text{ unidentified parameter } \gamma_{NT}^{1\cdot}
\end{aligned}
$$

Posterior ACE as a function of $\gamma^{1\cdot}_{NT}$



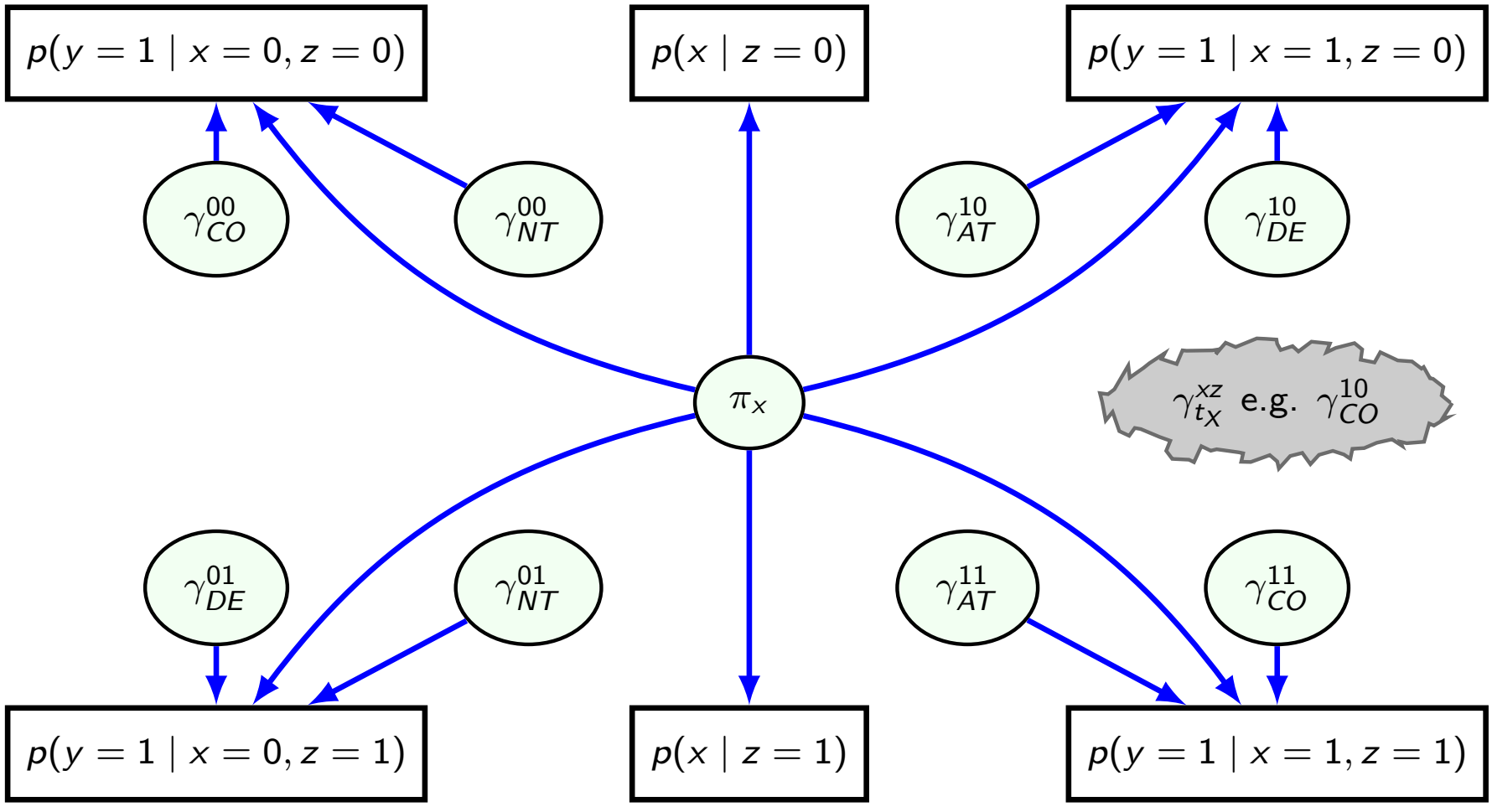
General Framework

- Remove:
  - Assumption of Exclusion Restriction [2]
  - Restrictions on state space (t_X, t_Y)
    i.e. don't rule out possibility of Always Takers or DEfiers
- Consider the following causal estimands as well:
  - $ACDE_{t_X}(x) = \mathrm{E}[Y_{x,z=1} - Y_{x,z=0} \mid t_X]$
    Average Controlled Direct Effect [3] , due to exposure x, for response type t_X
  - $ITT_{t_X} = \mathrm{E}[Y_{X_{z=1},z=1} - Y_{X_{z=0},z=0} \mid t_X]$
    Intent-To-Treat Effect for response type t_X

[2] Effect of Z on Y is only through X

[3] vs Average Causal Effect $ACE(X \to Y) = \mathrm{E}[Y_{x=1\cdot} - Y_{x=0\cdot}]$

Saturated Model: No assumptions or restrictions on (t_X, t_Y)



Average Controlled Direct Effects: $ACDE_{t_X}(x)$

Deriving the bounds for NT and AT:

$$1 - \frac{p(y=0, x=0 \mid z=1) + p(y=1, x=0 \mid z=0)}{p(x=0 \mid z=0) - p(x=1 \mid z=1)} \leq ACDE_{NT}(x=0) \leq \frac{p(y=0, x=0 \mid z=0) + p(y=1, x=0 \mid z=1)}{p(x=0 \mid z=0) - p(x=1 \mid z=1)} - 1$$

$$1 - \frac{p(y=0, x=1 \mid z=1) + p(y=1, x=1 \mid z=0)}{p(x=1 \mid z=1) - p(x=0 \mid z=0)} \leq ACDE_{AT}(x=1) \leq \frac{p(y=0, x=1 \mid z=0) + p(y=1, x=1 \mid z=1)}{p(x=1 \mid z=1) - p(x=0 \mid z=0)} - 1$$

Intent-To-Treat Effect: ITT_{t_X}

Under additional exclusion assumptions for COmpliers [4], ITT_{CO} is also the **COmplier Average Causal Effect** of X on Y, with bounds:

$$1 - \frac{p(y=0, x=1 \mid z=1) + p(y=1, x=0 \mid z=0)}{p(x=1 \mid z=1) - p(x=1 \mid z=0)}$$

$$\leq ITT_{CO} \leq$$

$$\frac{p(y=0, x=0 \mid z=0) + p(y=1, x=1 \mid z=1)}{p(x=1 \mid z=1) - p(x=1 \mid z=0)} - 1$$

[4] $\gamma^{00}_{CO} = \gamma^{01}_{CO} = \gamma^{0\cdot}_{CO}$ and $\gamma^{11}_{CO} = \gamma^{10}_{CO} = \gamma^{1\cdot}_{CO}$

Motivating example: Flu vaccine data

- $Z = 1$: Patient's physician was asked to *remind* patients to obtain flu shots
- $X = 1$: Patient actually *received* a flu shot
- $Y = 1$: Patient was *not* hospitalized

z	x	y	count	z	x	y	count
0	0	0	99	1	0	0	84
0	0	1	1027	1	0	1	935
0	1	0	30	1	1	0	31
0	1	1	233	1	1	1	422

Table: Flu vaccine data previously analyzed by Hirano et al. [2000], but now ignoring the baseline covariates

Motivating example: Flu vaccine data

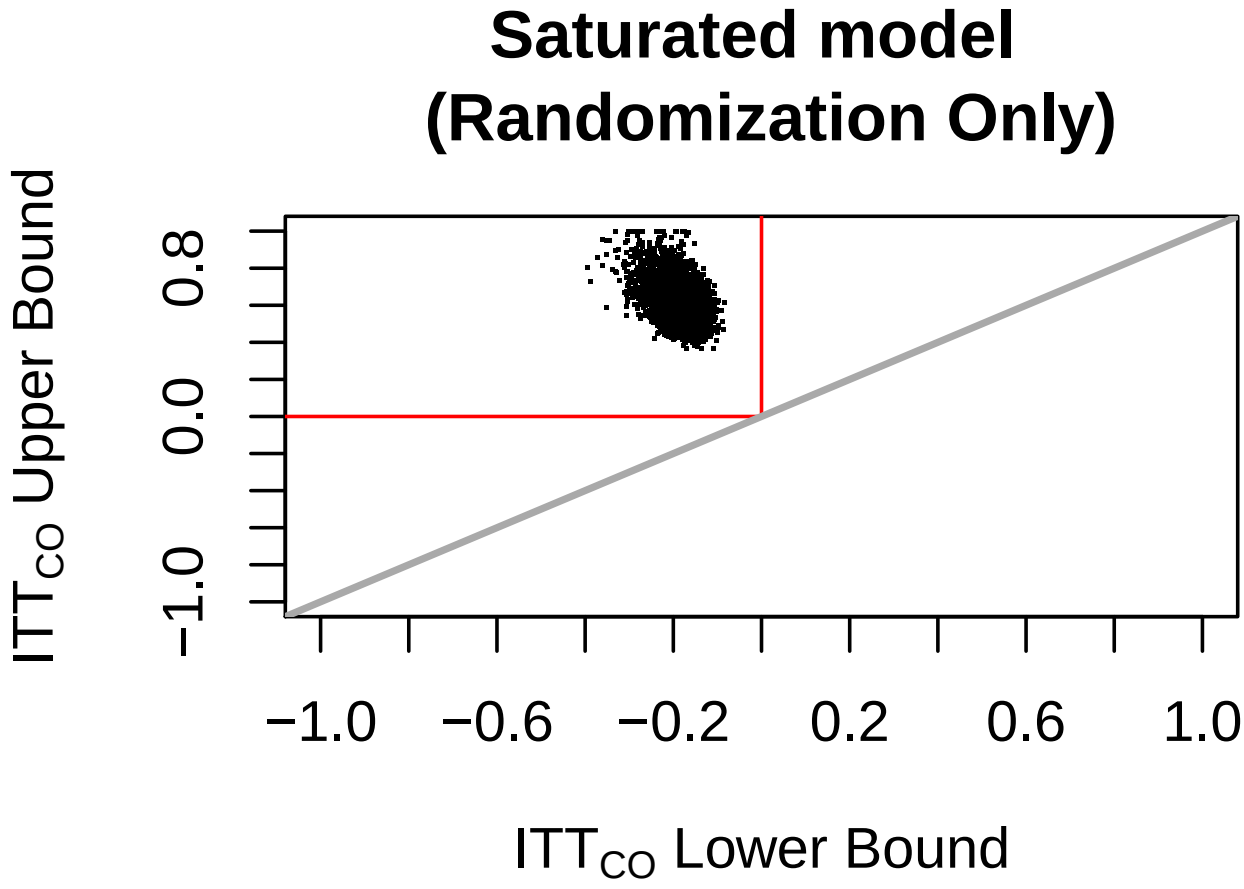
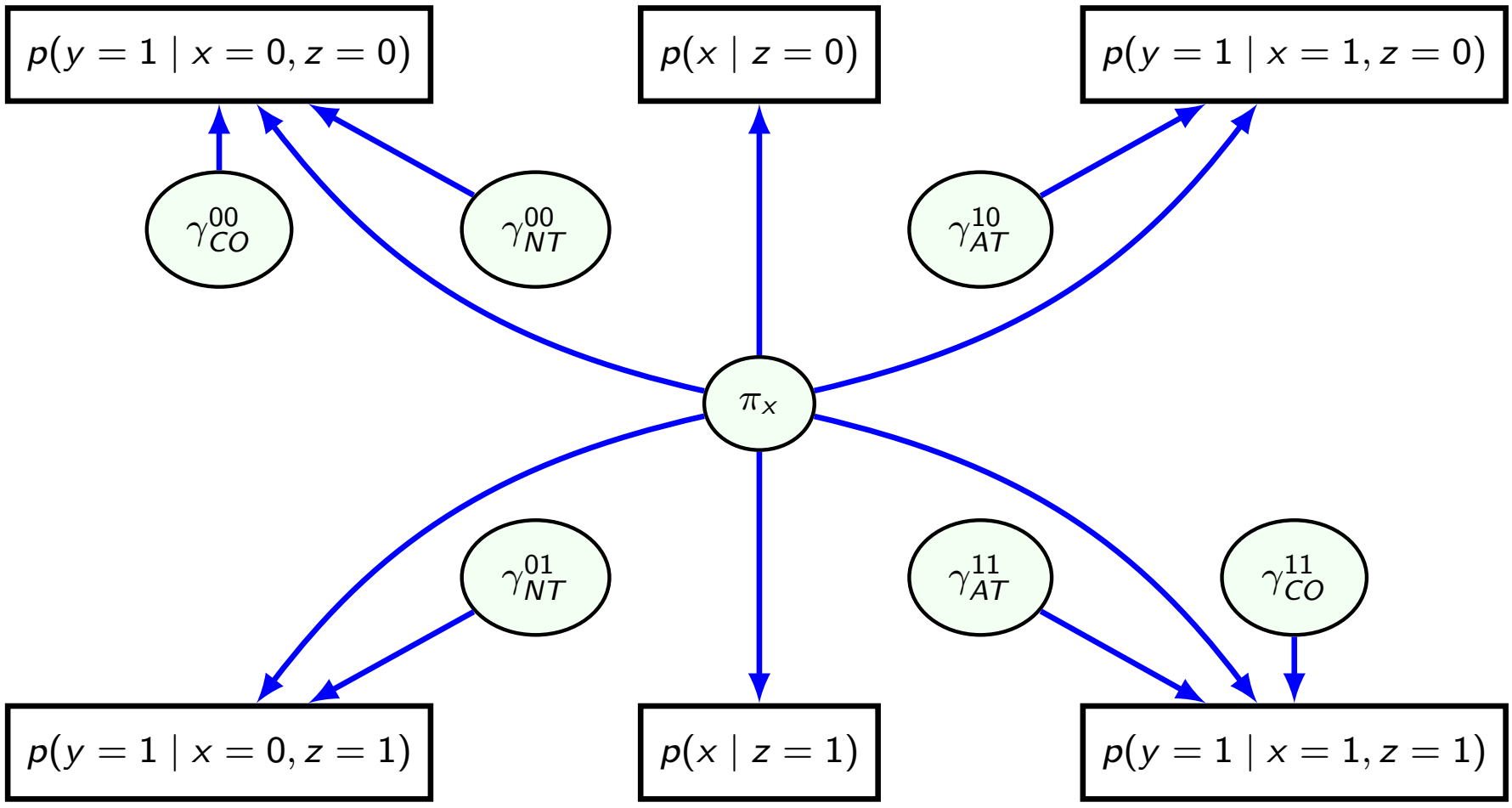


Figure: Posterior distribution over upper and lower bounds on ITT_{CO}, using Monte Carlo rejection sampling

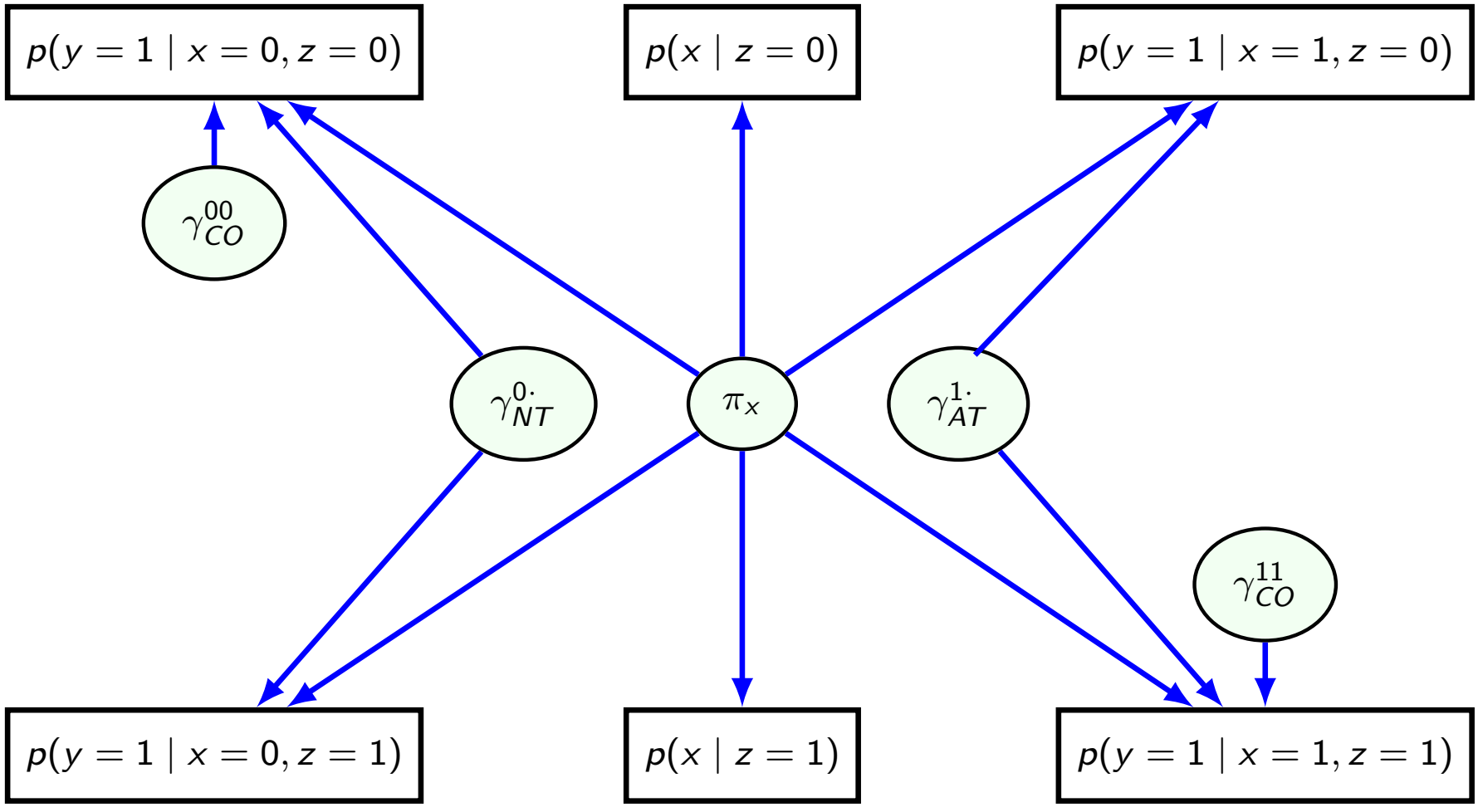
. . . And Then There Were Three . . .

- Now consider models with combinations where the following three assumptions hold:
  - MON_X : Monotonic Compliance
    $\Rightarrow$ No DEfiers
  - EX_{NT} : Stochastic Exclusion for Never-Takers under non-exposure
    $\Rightarrow \gamma^{0\cdot}_{NT} = \gamma^{00}_{NT} = \gamma^{01}_{NT}$
  - EX_{AT} : Stochastic Exclusion for Always-Takers under exposure
    $\Rightarrow \gamma^{1\cdot}_{AT} = \gamma^{11}_{AT} = \gamma^{10}_{AT}$

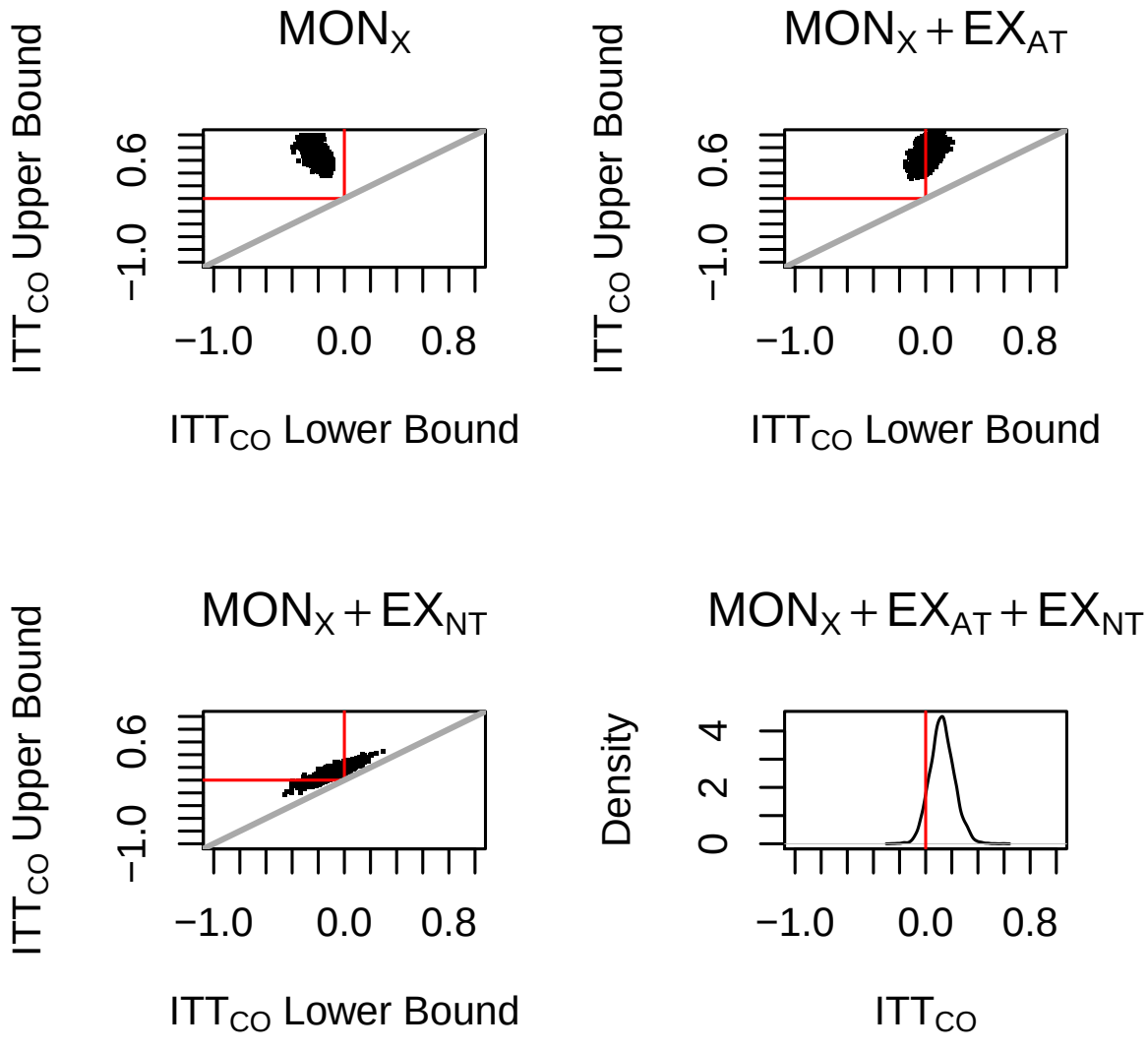
MON_X


$p(y = 1 \mid x = 0, z = 0)$
$p(x \mid z = 0)$
$p(y = 1 \mid x = 1, z = 0)$
γ^{00}_{CO}
γ^{00}_{NT}
γ^{10}_{AT}
π_x
γ^{01}_{NT}
γ^{11}_{AT}
γ^{11}_{CO}
$p(y = 1 \mid x = 0, z = 1)$
$p(x \mid z = 1)$
$p(y = 1 \mid x = 1, z = 1)$

$MON_X + EX_{AT} + EX_{NT}$



ITT_{CO} under other models



Conclusions

- The problem of non-compliance in causal models
  - Instrumental variables are a well-established method of estimating (population) causal effects
  - However, the resulting posteriors may be highly sensitive to the specification of the prior distribution over compliance types

Conclusions

- Transparent parameterizations
  - Re-parameterize the model such that the complete parameter vector may be divided into point-identified and entirely non-identified subvectors
  - Work out the distribution of the observed data implied by the transparent model
  - Use Monte-Carlo rejection sampling to draw from the conjugate posterior
  - Obtain point estimates or develop bounds on causal measures [5] in terms of identified parameter(s)

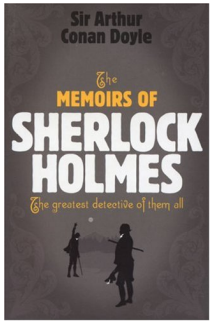
[5] regardless of chosen scale

Covered in the paper, but future work for me

- Incorporating covariates
  - Examine causal effects in sub-populations defined by baseline covariates $\mathbf{V}$
  - For discrete covariates with small numbers of levels, repeat analysis within each level of $\mathbf{V}$
  - For continuous covariates, re-parametrize each of the set of distributions, and construct (multivariate) generalized linear models for $p(y, x \mid z)$ as a function of $\mathbf{V}$.
- Robustness to model mis-specification of nuisance models
  - Posterior ITT estimate centered on MLE of ITT asymptotically
  - MLE generally inconsistent under the ITT null hypothesis [6]
  - Results in high type I error

[6] That treatment assignment Z and response Y are independent

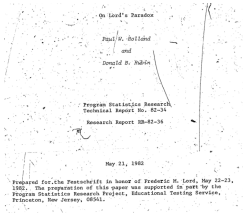
Finally . . .



Results without causes *are much more impressive.*

S. Holmes, 1893
The Adventure of the Stockbroker's Clerk

The scientific and practical interpretations of the results . . . dramatically different for descriptive and causal *questions.*


On Lord's Paradox

Paul W. Holland
and
Donald B. Rubin

Program Statistics Research
Technical Report No. 82-34

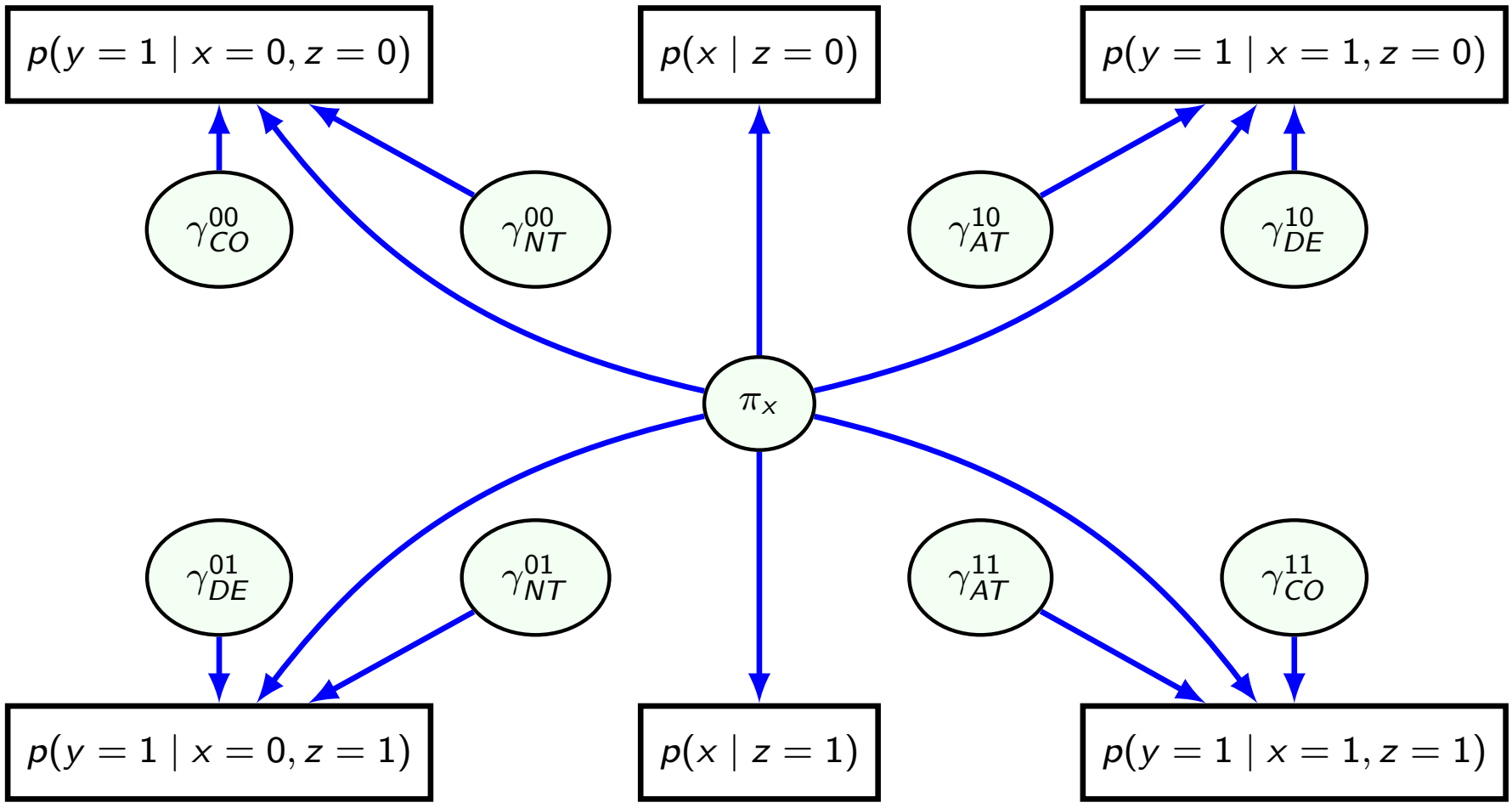
Research Report RR-82-36

May 21, 1982

Prepared for the Festschrift in honor of Frederic M. Lord, May 22-23, 1982. The preparation of this paper was supported in part by the Program Statistics Research Project, Educational Testing Service, Princeton, New Jersey, 08541.

P.W. Holland and D.B. Rubin, 1982
On Lord's paradox; Prepared for the Festschrift in honor of Frederic M. Lord

Thank You


$p(y = 1 \mid x = 0, z = 0)$
$p(x \mid z = 0)$
$p(y = 1 \mid x = 1, z = 0)$
γ^{00}_{CO}
γ^{00}_{NT}
γ^{10}_{AT}
γ^{10}_{DE}
π_x
γ^{01}_{DE}
γ^{01}_{NT}
γ^{11}_{AT}
γ^{11}_{CO}
$p(y = 1 \mid x = 0, z = 1)$
$p(x \mid z = 1)$
$p(y = 1 \mid x = 1, z = 1)$

References

J.D. Angrist, G.W. Imbens, and D.B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996.

Keisuke Hirano, Guido W. Imbens, Donald B. Rubin, and Xiao-Hua Zhou. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, 1(1):69–88, 2000.

J. Pearl. *Causality: models, reasoning, and inference*, volume 47. Cambridge Univ Press, 2000.

T.S. Richardson, R.J. Evans, and J.M. Robins. Transparent parameterizations of models for potential outcomes. *Bayesian Statistics*, 9:569–610, 2011.