Submitted to the Annals of Applied Statistics

BAYESIAN SENSITIVITY ANALYSIS FOR CAUSAL EFFECTS FROM 2×2 TABLES IN THE PRESENCE OF UNMEASURED CONFOUNDING WITH APPLICATION TO PRESIDENTIAL CAMPAIGN VISITS*

By Luke Keele† and Kevin M. Quinn‡

Georgetown University† and University of California, Berkeley‡

Presidents often campaign on behalf of candidates during elections. Do these campaign visits increase the probability that the candidate will win? While one might attempt to answer this question by adjusting for observed covariates, such an approach is plagued by serious data limitations. In this paper we pursue a different approach. Namely, we ask: what, if anything, should one infer about the causal effect of a presidential campaign visit using a simple cross-tabulation of the data? We take a Bayesian approach to this problem and show that if one is willing to use substantive information to make some (possibly weak) assumptions about the nature of the unmeasured confounding, sharp posterior estimates of causal effects are easy to calculate. Using data from the 2002 midterm elections, we find that, under a reasonable set of assumptions, a presidential campaign visit on the behalf of congressional candidates helped those candidates win elections.

1. Introduction. In many social science settings, randomized experiments are simply not possible while the assumptions required for other causal identification strategies are often implausible. Moreover, many questions of scientific and policy relevance fall prey to this problem. Nonetheless, this does not mean that one can learn nothing about causal effects of interest. To provide a better sense of this, in the next subsection, we describe a particular substantive question from political science that we will focus on throughout this paper.

1.1. *Presidential Campaigning for Co-Partisans.* The powers of U.S. presidents are largely informal. The veto and power of appointment are rare

*We thank Adam Glynn, Dan Ho, Gary King, and seminar participants at Penn State University and the Northeast Methods Meeting for helpful comments and conversations. Kevin Quinn gratefully acknowledges support from the U.S. National Science Foundation (grant BCS 05-27513) and the Institute for Quantitative Social Science at Harvard University.

Keywords and phrases: causal inference, Bayesian statistics, partial identification, sensitivity analysis

instances of formal powers given the president in the Constitution. Presidential power arises from the more informal power to persuade Congress (Neustadt 1960). One powerful means of persuasion is to make Members of Congress indebted to the president. One method that presidents have at their disposal to develop such indebtedness is the Congressional campaign visit. Here, the president makes a personal visit on behalf of someone running for the House or the Senate. Campaign visits by presidents are likely to be more effective for candidates to the House, since these elections are often low information affairs (Jacobson 2003).

Political scientists have long attempted to estimate the causal effect of a presidential campaign visit on the probability that a candidate wins(Herrnson and Morris 2007; Sellers and Denton 2006; Cohen, Krassa and Hamman 1991; Keele, Fogarty and Stimson 2004). Particular attention has been paid to the 2002 midterm election, when George W. Bush campaigned extensively for Republican candidates. A number of media reports inferred that a campaign visit that year was highly effective in helping candidates win (Keele, Fogarty and Stimson 2004). Of course, these analyses must rely on observational data, and the resulting estimates of causal effects are surely subject to bias from confounding. To increase the credibility of estimated statistical associations as causal effects, analyst adjust for observed covariates thought to be confounders. As we outline below, there is little reason to think this assumption is plausible in this setting.

1.2. *Overview.* Many researchers might say that little about causal effects can be inferred from such data. While there are certainly aspects of truth to the position that unmeasured confounding simply cannot be overcome in many cases, it is generally possible to learn some things about the size of the causal effects of interest regardless of the nature and degree of the unmeasured confounding. For instance, Manski (1990) derived bounds for the average treatment effect under very general assumptions and showed that, in the case of binary treatment and outcome, the width of these bounds is 1. Since the average treatment effect can take values between -1 and 1, Manski's bounding interval always includes an estimate of zero for the treatment effect. However, auxiliary assumptions can narrow this bounding interval Manski (2003). See Robins (1989); Manski (1993, 2003); Imai and Yamamoto (2008) and Balke and Pearl (1997) for related work on the partial identification of treatment effects.

In a similar spirit, this paper further explores what can be learned about treatment effects under arbitrary forms of unmeasured confounding. More specifically, we ask what can one infer about the causal effect of a bi-

nary treatment on a binary outcome from a 2×2 cross-tabulation of non-experimental data? We reduce the case of arbitrary unmeasured confounding with binary treatment and binary outcome to its most basic, yet still general, form that retains a readily interpretable parameterization of the key quantities. We then show how a Bayesian prior distribution can be placed over the four free parameters that govern the type and extent of the unmeasured confounding. If one is willing and able to use background knowledge to make some (possibly weak) assumptions about the nature of the unmeasured confounding, sharp posterior estimates of partially identified causal effects are easy to calculate. As such, in the spirit of Manski we focus on estimates of bounds for causal effects using Bayesian priors. Since these assumptions are formalized within the Bayesian framework, subjective uncertainty about causal effects is calculated in a logically coherent manner. The end result is a procedure that allows researchers to make probability statements about the likely size of causal effects based on the evidence in a 2×2 (or $2 \times 2 \times K$) table regardless of the sample size and the amount of unmeasured confounding.

Manski-type large sample bounds are a special case of the more general sensitivity analysis that we develop. Ichino, Mealli and Nannicini (2008, p. 320) explicitly demonstrate the link between Manski bounds and sensitivity analysis of the type we propose. While large sample bounds generally consider worst case scenarios, we allow for investigating the sensitivity of conclusions over a range of priors deemed reasonable by the analyst. Our work is also similar to other forms of sensitivity analysis that place bounds on causal quantities. For example, one early method of sensitivity analysis outlined by Cornfield et al. (1959) developed bounds for causal quantities with binary responses but abstracted away from issues of sampling variability. Rosenbaum and Rubin (1983) and Rosenbaum (1987, 2002) developed sharp bounds on causal quantities for any type of response using randomization distributions. See Imbens (2003) and Ding and VanderWeele (2016a); Ding and Vanderweele (2016b) for more recent studies that further extend the Cornfield framework. One primary distinction between our work and the extant literature is our use of the Bayesian inferential framework.

1.3. *Notation and Causal Model.* Next, we present our notation and place our analysis in the formal statistical framework of causal inference based on potential outcomes (Holland 1986a). For each Congressional district $i = 1, 2, \ldots, 435$, we define two potential outcomes $Y_i(1)$ and $Y_i(0) \in \{0, 1\}$. $Y_i(1)$ denotes a potential electoral victory by the candidate in district i when the president campaigns for that candidate through a personal visit to that district. In contrast, $Y_i(0)$ represents a potential victory by the can-

didate in district i when the president does not campaign on the candidate's behalf. We use the indicator variable $D_i \in \{0, 1\}$ to denote the treatment status in district i. In our application, D_i is equal to 1 if the president campaigned for candidate in district i through a personal visit and 0 otherwise. Under this causal model, the observed outcome Y_i is a function of the treatment variable and the potential outcomes: $Y_i = Y_i(1)D_i + (1 - D_i)Y_i(0)$. In our notation, we use upper-case letters to distinguish a random variable from its realization. The fundamental problem of causal inference is that for a single unit at most only one of the potential outcomes can be revealed (Holland 1986a).

This framework implicitly assumes that there is no interference among units: the potential outcomes of one unit do not depend on the treatment status of other units (Cox 1958; Rubin 1990). In our application, this assumption implies that the potential electoral victory status of a candidate in one district does not depend on whether the president campaigned for a candidate in another district. This assumption is reasonable given that presidents campaign for specific candidates and a visit is unlikely to help other candidates that did not specifically receive a presidential campaign visit. Under this framework, the individual level causal effect is defined as a contrast in potential outcomes: $Y_i(1) - Y_i(0)$. Rather than focusing on unit-level causal effects, we will concern ourselves with aggregate effects within some collection of units. The causal quantity that we focus on is the average treatment effect (ATE)

$$\begin{aligned} ATE &\equiv \mathbb{E}[Y(1) - Y(0)] \\ &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \end{aligned}$$

where the expectation is taken over all units in the study (congressional districts in our running example). Note that with a binary outcome ATE can also be written as

$$ATE = \Pr[Y(1) = 1] - \Pr([Y(0) = 1]. \tag{1}$$

To give the observed ATE a causal interpretation depends on counterfactual quantities. While we can easily calculate estimates of $\mathbb{E}[Y = y|D = d]$ and/or $\Pr(Y = y|D = d)$ from the observed joint distribution of Y and D and use those in place of counterfactual quantities in the equations above, these counterfactual quantities can only be estimated consistently from the joint distribution of the data if untestable causal assumptions are maintained (Rubin 1978; Holland 1986b; Robins 1986; Pearl 1995, 2000).

If one is willing to assume that treatment assignment is ignorable, i.e. $Y(1) \perp\!\!\!\perp D$ and $Y(0) \perp\!\!\!\perp D$, and that the potential outcomes are well-defined as written,[1] then it is possible to consistently estimate the counterfactual quantities in Equation 1 with simple sample averages. In a well-run randomized controlled experiment, ignorability of treatment assignment and SUTVA are likely to hold because of the design of the experiment.

However, with observational data, ignorability of treatment assignment is unlikely to hold. Consequently, analysts of observational data often invoke the assumption of *conditionally* ignorable treatment assignment. Under conditional ignorability of treatment assignment the claim is that there exists a collection of pretreatment variables $\mathbf{U}$ such that treatment assignment is conditionally ignorable given $\mathbf{U}$. Formally, conditional ignorability of treatment assignment assumes that $Y(1) \perp\!\!\!\perp D|\mathbf{U}$ and $Y(0) \perp\!\!\!\perp D|\mathbf{U}$. Conditional ignorability of treatment assignment is generally considered to be a strong assumption, since the analyst must assume that simple ignorability of treatment assignment holds within each level of $\mathbf{U}$. Moreover, this assumption is not testable. Next, we review the data from our empirical application and these assumptions within the context of our application.

1.4. *Data and Elementary Analyses.* Our data is from the 2002 midterm election in the United States. These data were first reported in Keele, Fogarty and Stimson (2004). The outcome is measured as a binary indicator for whether the Republican candidate won the election or not and was constructed from *Congressional Quarterly's Politics in America* (2001). The treatment indicator is whether George W. Bush campaigned on behalf of U.S. House candidate through a personal appearance in that members district. Keele, Fogarty and Stimson (2004) used Lexis-Nexus state level AP reports from the 2002 election cycle to determine whether President George W. Bush personally campaigned for the Republican candidate between Labor Day and the election in November. The observed data are summarized as a 2×2 cross-tabulation in Table 1. The cross-tabulation excludes races where one candidate ran unopposed, leaving us with 348 races with two candidates.

We see that the president campaigned on behalf of 21 different Republican candidates. Of the 348 races with two candidates, the president only selected approximately six percent for a campaign visit. A naive analysis that assumes there is no confounding would clearly conclude that presidential visits are effective in helping Republican candidates win elections. The

[1]This latter assumption is sometimes referred to as the stable unit treatment value assumption or SUTVA.

	$Y_i = 0$ Republican Loses	$Y_i = 1$ Republican Wins
No Visit $D_i = 0$	164	163
Visit $D_i = 1$	3	18

TABLE 1
Observed Data For Presidential Visits in 2002. *The units of analysis are the 348 Republican candidates running in two-party races for the U.S. House of Representatives in 2002.*

proportion of candidates that won when the president visited was 0.86 while the proportion of candidates that won without a presidential visit was 0.50. Using these observed data quantities, we can calculate the average treatment effect at 0.36 with a large sample 95% confidence interval [0.17, 0.54]. Of course, confounding is likely as a president may strategically select candidates for visits based on their ability to win an election. Given the likelihood of confounding, one strategy is to use a set of covariates to model the treatment assignment mechanism (Rubin 2008). In this context, however, there are a limited number of observable covariates that we can credibly use to model the assignment mechanism. Moreover, much of the data available to model the assignment mechanism are simply descriptive characteristics of the Congressional district such as the level of education or income in the district. The actual decision-making process of presidents is unobserved, there is little reason to think much progress can be made by modeling the assignment mechanism. Thus while modeling the assignment mechanism is often a reasonable strategy for estimating causal effects, in this context we cannot credibly model for the assignment mechanism well enough to make conditional ignorability a reasonable assumption. The remainder of this paper is devoted to the question of what can be inferred about the ATE from the data in Table 1 by making reasonable assumptions about the patterns of confounding that are present.

The paper proceeds as follows. In Section 2 we introduce the necessary terminology and notation and demonstrate how inferences can be constructed from a 2×2 table with general unmeasured confounding. Section 3 shows how causal quantities of interest such as the average treatment effect can be written in terms of the model parameters from Section 2. Large sample nonparametric bounds on these causal quantities are also derived in this section. These bounds coincide with those of Manski (1990) although the derivation is slightly different. Section 4 discusses the choice of prior distribution for the

$Y_i(0)$	$Y_i(1)$	Z_i	
0	0	0	Never Succeed
0	1	1	Helped
1	0	2	Hurt
1	1	3	Always Succeed

TABLE 2

Possible Patterns of Potential Outcomes and the Coarsest General Confounding Variable. $Y_i(0)$ *is the potential outcome for unit i when* D_i *is set to 0 (no visit).* $Y_i(1)$ *is the potential outcome for unit i when* D_i *is set to 1 (presidential visit). A unit i for which* $Z_i = 0$ *has a value of* U_i *that causes it to always have* $Y_i = 0$ *regardless of the (counterfactual) value of* D_i*. We say these units are "never succeeders". If* $Z_i = 1$ *we say that unit i is "helped" by treatment because its potential outcome under* $D_i = 1$ *is equal to 1 (success) while its potential outcome under* $D_i = 0$ *is 0 (failure). If unit i has* $Z_i = 2$ *we say that i is "hurt" by treatment because its potential outcome under* $D - i = 1$ *is equal to 0 (failure) while its potential outcome under* $D_i = 0$ *is 1 (success). Finally, if* $Z_i = 3$ *we say that i is an "always succeeder" because its value of* U_i *is such that* Y_i *will always equal 1 regardless of the (counterfactual) value of* D_i*.*

model parameters and then describes a simple posterior sampling algorithm that does not require Markov chain Monte Carlo. In Section 5 we revisit the example data in Table 1. Here we see how defensible prior beliefs can be operationalized in a prior distribution over the model parameters and what this implies for inferences about a possible presidential visit treatment effect. The final section concludes.

2. The Probability Model under Unobserved Confounding. Here we assume that some $\mathbf{U}$ exists such that $Y(1) \perp\!\!\!\perp D|\mathbf{U}$ and $Y(0) \perp\!\!\!\perp D|\mathbf{U}$, i.e. that a presidential visit is conditionally ignorable given $\mathbf{U}$. Importantly, we do not assume that we can measure $\mathbf{U}$ or that we even know what variables it includes. While $\mathbf{U}$ may be extremely complicated, the binary nature of both treatment and outcome implies that the domain of $\mathbf{U}$ can be partitioned into four equivalence classes depending on the pattern of potential outcomes associated with each point in the domain of $\mathbf{U}$ (Angrist, Imbens and Rubin 1996; Balke and Pearl 1997; Chickering and Pearl 1997). We introduce a new categorical variable Z_i that labels these equivalence classes. The values of Z_i along with the associated patterns of potential outcomes are presented in Table 2.

If the joint distribution of Z, D, and Y (P_{DYZ}) were observed, one could write the probabilities of the various potential outcomes as:

$$\Pr(Y(d) = y) = \sum_{z=0}^{3} \Pr(Y = y|D = d, Z = z) \Pr(Z = z).$$

The probabilities on the right-hand-side of the equation above can be calculated directly from P_{DYZ}. If P_{DYZ} is not directly observed but can be consistently estimated, then one can construct consistent estimates of the potential outcome probabilities via the plug-in principle.

In our application, as is common in the social sciences, P_{DYZ} is unknown and Z_i is not observable for any i. Without data on Z_i, it is impossible to consistently estimate P_{DYZ}. Nevertheless, there is some information about Z_i in the observed (D, Y) data sampled from P_{DYZ}. The goal of this paper is to show how this information can be combined with subjective background knowledge to yield causal inferences from 2×2 tables even when the confounding variables in $\mathbf{U}$ are not measured.

2.1. *Likelihood.* We adopt a Bayesian approach to make inferences about the form of P_{DYZ} without information on Z_i. The main reason for taking a Bayesian approach in this paper is that it allows us to incorporate background knowledge about the (potentially unobserved) confounder Z_i in a principled fashion (Kadane and Wolfson 1998; Western and Jackman 1994; Gill and Walker 2005). Our work is not the first Bayesian approach to partial identification. See Gustafson (2010); Gustafson et al. (2010); Moon and Schorfheide (2012); Richardson, Evans and Robins (2011); McCandless, Gustafson and Levy (2007) for other examples of partial identification within a Bayesian framework. Among prior work, our approach is most similar to Ding and Dasgupta (2016) and Jin and Rubin (2008) in that we separate identifiable parameters from unidentifiable parameters and introduce prior information on the unidentifiable parameters to study how conclusions are altered. We begin by discussing the likelihood function and then discuss our choice of prior distribution along with the resulting posterior distribution.

Let $\mathcal{Z}_i$ denote the set of possible values Z_i could take given the observed data on unit i. More formally,

$$\mathcal{Z}_i = \begin{cases} \{0,1\} & \text{if } d_i = 0,\ y_i = 0 \\ \{2,3\} & \text{if } d_i = 0,\ y_i = 1 \\ \{0,2\} & \text{if } d_i = 1,\ y_i = 0 \\ \{1,3\} & \text{if } d_i = 1,\ y_i = 1 \end{cases}$$

We can then write the likelihood as:

$$
\begin{aligned}
p(\mathbf{d},\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\psi}) &= \prod_{i=1}^{n} \sum_{z_i \in \mathcal{Z}_i} p(d_i, y_i, z_i|\boldsymbol{\theta},\boldsymbol{\psi}) \\
&= \prod_{i=1}^{n} p(d_i, y_i|\boldsymbol{\theta}) \left\{ \sum_{z_i \in \mathcal{Z}_i} p(z_i|d_i, y_i, \boldsymbol{\psi}) \right\} \\
&= \prod_{i=1}^{n} \theta_{00}^{\mathbb{I}(d_i=0,y_i=0)} \theta_{01}^{\mathbb{I}(d_i=0,y_i=1)} \theta_{10}^{\mathbb{I}(d_i=1,y_i=0)} \theta_{11}^{\mathbb{I}(d_i=1,y_i=1)} \times \\
&\qquad \left\{ \sum_{z_i \in \mathcal{Z}_i} \psi_{00}^{\mathbb{I}(d_i=0,y_i=0,z_i=1)} (1-\psi_{00})^{\mathbb{I}(d_i=0,y_i=0,z_i=0)} \times \right. \\
&\qquad\quad \psi_{01}^{\mathbb{I}(d_i=0,y_i=1,z_i=3)} (1-\psi_{01})^{\mathbb{I}(d_i=0,y_i=1,z_i=2)} \times \\
&\qquad\quad \psi_{10}^{\mathbb{I}(d_i=1,y_i=0,z_i=2)} (1-\psi_{10})^{\mathbb{I}(d_i=1,y_i=0,z_i=0)} \times \\
&\qquad\quad \left. \psi_{11}^{\mathbb{I}(d_i=1,y_i=1,z_i=3)} (1-\psi_{11})^{\mathbb{I}(d_i=1,y_i=1,z_i=1)} \right\} \\
&= \theta_{00}^{C_{00+}} \theta_{01}^{C_{00+}} \theta_{10}^{C_{10+}} \theta_{11}^{C_{11+}} \qquad (2)
\end{aligned}
$$

where $\mathbb{I}(\cdot)$ is the indicator function, $C_{dy+} = \sum_{i=1}^{n} \mathbb{I}(d_i = d, y_i = y)$, $\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11} \geq 0$, $\theta_{00} + \theta_{01} + \theta_{10} + \theta_{11} = 1$, and $\psi_{00}, \psi_{01}, \psi_{10}, \psi_{11} \in [0, 1]$.

While this model for (D, Y, Z) might seem to contain a large number of parameters, there are two key sets of parameters θ_{dy} and ψ_{dy} for $d = 0, 1$ and $y = 0, 1$. The θ parameters govern a multinomial distribution for the distribution of (D, Y) after Z has been marginalized out of P_{DYZ}. The ψ parameters govern the conditional distribution of Z given D and Y. Note that because of the definition of Z (see Table 2) only 2 values of Z are logically possible given any admissible (D_i, Y_i) pair. The distribution of Z given $D = d$ and $Y = y$ is thus Bernoulli with parameter ψ_{dy}. Here, ψ_{01} gives the probability that $Z = 3$ given $D = 0$ and $Y_i = 1$ while $(1-\psi_{01})$ gives the probability that $Z = 2$ given $D = 0$ and $Y = 1$. The other conditional distributions for Z given $D = d$ and $Y = y$ are similarly parameterized. Table 6 in the appendix provides a complete summary of the parameters and their intuitive meanings.

2.2. *Prior and Posterior.* Bayesian inference centers on the posterior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ given the observed data. The posterior distribution is given (up to proportionality) by:

$$p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}, \mathbf{d}, \mathbf{z}) \propto p(\mathbf{y}, \mathbf{d}, \mathbf{z}|\boldsymbol{\theta}, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \boldsymbol{\psi})$$

We defined the likelihood, $p(\mathbf{y}, \mathbf{d}, \mathbf{z}|\boldsymbol{\theta}, \boldsymbol{\psi})$, in the previous section, and we now discuss specification of the prior distribution $p(\boldsymbol{\theta}, \boldsymbol{\psi})$. A natural choice for the joint prior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ is to assume that $\boldsymbol{\theta}$, $\psi_{00}, \psi_{01}, \psi_{10}$, and ψ_{11} are mutually independent a priori and that $\boldsymbol{\theta} \sim \mathcal{D}irichlet(a_{00}, a_{01}, a_{10}, a_{11})$, $\psi_{dy} \sim \mathcal{B}eta(b_{dy}, c_{dy})$, for $d = 0, 1$ and $y = 0, 1$. This is the conjugate prior distribution for this model. This prior specification will allow us to think of the hyper-parameters a_{dy}, b_{dy}, and c_{dy} for $d = 0, 1$ and $y = 0, 1$ as additional "pseudo-observations." This makes the prior distributions more easily interpretable, which is important for the current application where inferences are dependent on the prior.

Combining this prior with the likelihood in Equation 2 gives us the following posterior density (up to proportionality):

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{d}, \mathbf{y}) \propto\ & \theta_{00}^{C_{00+}+a_{00}-1}\theta_{01}^{C_{01+}+a_{01}-1}\theta_{10}^{C_{10+}+a_{10}-1}\theta_{11}^{C_{11+}+a_{11}-1}\times \\ & \psi_{00}^{b_{00}-1}(1-\psi_{00})^{c_{00}-1}\psi_{01}^{b_{01}-1}(1-\psi_{01})^{c_{01}-1}\times \\ & \psi_{10}^{b_{10}-1}(1-\psi_{10})^{c_{10}-1}\psi_{11}^{b_{11}-1}(1-\psi_{11})^{c_{11}-1} \end{aligned} \tag{3}$$

Note that the only information about $\boldsymbol{\psi}$ is coming from the prior distribution. This implies that inferences that depend on $\boldsymbol{\psi}$ will be dependent on one's choice of prior for $\boldsymbol{\psi}$.

3. Causal Quantities of Interest. Next we discuss how causal quantities such as the ATE can be partially identified based on beliefs about $\boldsymbol{\psi}$. Typically causal quantities are calculated directly from the data based on an assumption of conditional unconfoundedness. In our analysis of whether presidential campaign visits help candidates, unconfoundness, conditional or otherwise, is unrealistic. In our approach, based on the probability model specified in Section 2, we calculate causal quantities that depend on the parameter $\boldsymbol{\psi}$. We denote these as *sensitivity analysis quantities* since the inference is based on prior beliefs about the unobserved distribution of Z_i. Sensitivity analysis quantities depend on the distribution of Z_i through the

following set of equations:

$$\begin{aligned}
\Pr_s(Y(0)=0) &= \sum_{z=0}^{3} \Pr(Y=0|D=0,Z=z)\Pr(Z=z) \\
&= \Pr(Z=0)+\Pr(Z=1) \\
&= \theta_{10}(1-\psi_{10})+\theta_{11}(1-\psi_{11})+\theta_{00} \\
\Pr_s(Y(0)=1) &= \sum_{z=0}^{3} \Pr(Y=1|D=0,Z=z)\Pr(Z=z) \\
&= \Pr(Z=2)+\Pr(Z=3) \\
&= \theta_{10}\psi_{10}+\theta_{11}\psi_{11}+\theta_{01} \\
\Pr_s(Y(1)=0) &= \sum_{z=0}^{3} \Pr(Y=0|D=1,Z=z)\Pr(Z=z) \\
&= \Pr(Z=0)+\Pr(Z=2) \\
&= \theta_{00}(1-\psi_{00})+\theta_{01}(1-\psi_{01})+\theta_{10} \\
\Pr_s(Y(1)=1) &= \sum_{z=0}^{3} \Pr(Y=1|D=1,Z=z)\Pr(Z=z) \\
&= \Pr(Z=1)+\Pr(Z=3) \\
&= \theta_{00}\psi_{00}+\theta_{01}\psi_{01}+\theta_{11}.
\end{aligned}$$

If $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ were known, the probabilities above would be the true probabilities of the potential outcomes. This would hold regardless of the forms of confounding present in the data. Of course, $\boldsymbol{\psi}$ is never known (and typically not identified) so the sensitivity analysis quantities will depend on one's prior beliefs about $\boldsymbol{\psi}$. We denote sensitivity analysis quantities using subscript s. For example, ATE_s denotes the average treatment effect calculated using prior beliefs about $\boldsymbol{\psi}$.

3.1. *Average Treatment Effects.* Next, we describe how causal quantities can be calculated based on values of $\boldsymbol{\psi}$. Here, we focus on the ATE. While we focus on the ATE, it is also possible to define average treatment effects on the treated (ATT) or just the control group (ATC) and calculate bounds and sensitivity analysis distributions for these estimands as well. Moreover, we can also calculate sensitivity analysis quantities based on the relative risk

as well. The sensitivity analysis ATE for a binary outcome is defined as:

$$\begin{aligned} ATE_s &= \Pr_s(Y(1)=1) - \Pr_s(Y(0)=1) \\ &= (\theta_{00}\psi_{00} + \theta_{01}\psi_{01} + \theta_{11}) - (\theta_{10}\psi_{10} + \theta_{11}\psi_{11} + \theta_{01}) \end{aligned} \tag{4}$$

Manski (1990) derived nonparametric bounds for the average treatment effect that will contain the true average treatment effect with probability 1 as sample size goes to infinity. Here, we show how these bounds can be calculated as a function $\boldsymbol{\psi}$. Inspection of Equation 4 reveals that the minimum value of ATE_s will occur when $\psi_{00} = 0, \psi_{01} = 0, \psi_{10} = 1$, and $\psi_{11} = 1$. Similarly, the maximum value of ATE_s will occur when $\psi_{00} = 1, \psi_{01} = 1, \psi_{10} = 0$, and $\psi_{11} = 0$. Substituting these values into the expression for ATE_s and recognizing that $ATE_s = ATE$ we see that:

$$ATE \in [-(\theta_{10} + \theta_{01}), \quad (\theta_{00} + \theta_{11})] \tag{5}$$

Note that this interval will always include 0. Further, since $\sum_d \sum_y \theta_{dy} = 1$, the width of this interval will always be 1 (see also Manski (1990)).

3.1. *Average Treatment Effects.* We next describe a Bayesian method for obtaining sensitivity analysis quantities. Here, we focus on obtaining sensitivity analysis quantities for the ATE, but sensitivity analysis quantities are easily obtained for estimands such as the average treatment effect of the treated as well. A Bayesian approach requires one to specify a prior distribution for $(\boldsymbol{\theta}, \boldsymbol{\psi})$. In Section 2 we argued that independent Dirichlet and Beta distributions made sense in terms of interpretability. In the remainder of this section we discuss how the parameters governing these prior distributions can be chosen and how one can summarize the resulting posterior distribution to make inferences about causal quantities of interest such as the ATE under general but unobservable patterns of confounding.

4.1. *Choosing a Prior Distribution.* It is worth emphasizing that, unlike Bayesian inference for parameters which are point identified, the impact of the choice of the prior for $\boldsymbol{\psi}$ on the posterior distribution for $\boldsymbol{\psi}$ and functionals of that posterior distribution will not diminish as n gets large if Z is completely unobserved. In fact, since no new information about $\boldsymbol{\psi}$ is arriving as n gets large, the marginal posterior for $\boldsymbol{\psi}$ will always be equal to the prior for $\boldsymbol{\psi}$. Consequently, the choice of a particular prior should be justified using substantive background knowledge. As a result, the analyst should report numerous sensitivity analysis quantities in which multiple reasonable priors are used.

Each ψ_{dy} represents the conditional probability of one of the two possible configurations of potential outcomes among units in which we observe $D = d$ and $Y = y$.[2] Thus the $\mathcal{B}eta(b_{dy}, c_{dy})$ prior for ψ_{dy} can be thought of as a statement of belief that $b_{dy} - 1$ of the C_{dy+} units have one potential outcome profile while $c_{dy} - 1$ of the C_{dy+} units have the other possible potential outcome profile. If $b_{dy} + c_{dy} = C_{dy+} + 2$ then the information in the prior is equivalent to the information that would be in the sample data in the ideal case in which the potential outcome patterns are observed for units with $D = d$ and $Y = y$. If $b_{dy} + c_{dy} < C_{dy+} + 2$ then there is less information in the prior than this ideal situation and if $b_{dy} + c_{dy} > C_{dy+} + 2$ then the prior is adding more information than one could ever get directly from the sample data. In the appendix, we provide a full elaboration of the relationship between the b_{dy} and c_{dy} parameters and potential outcomes. Here, we provide two useful heuristics for prior selection based on two models of treatment response and selection that should be widely applicable in many social science settings. We use both of these heuristics in the analysis of the presidential visit data from 2002. The first heuristic we consider is the possibility of a monotonic treatment effect (Manski 1997). Under such a complete version of monotonicity, we would assume

$$Y_i(1) \geq Y_i(0) \text{ or } Y_i(1) \leq Y_i(0) \; \forall \; i = 1, \cdots, n.$$

In words, we would assume that treatment either has no effect or moves outcomes in the same direction for all units. In the context here, a positive monotone treatment response assumption implies that a presidential campaign visit does not hurt the election chances of any candidates. Our approach weakens this assumption by assuming that treatment has a monotonic effect for most, but not all, units. Put another way, our assumption is that very few candidates are hurt by a presidential visit. If one believes that such a generally positive monotonic treatment effect is reasonable then one could set $b_{01} \gg c_{01}$ and $b_{10} \ll c_{10}$. Conversely, one could set $b_{00} \ll c_{00}$ and $b_{11} \gg c_{11}$ to operationalize a generally negative monotonic treatment effect (very few units are helped by the active treatment).

Setting the parameters on the prior distributions as follows $b_{01} \to \infty$, $c_{01} = 0$ and $b_{10} = 0$, $c_{10} \to \infty$ operationalizes Manski's version of the positive monotone treatment response assumption where no units are hurt by the treatment. Manski's version of monotone treatment response is, in fact, a limiting case of monotonic treatment effects. We need not assume that treatment is uniformly positive for all units. We can vary the fraction

[2]See the appendix for a full elaboration of how each ψ_{dy} parameter relates to a particular set of potential outcomes.

of units helped or hurt to assess sensitivity of the resulting estimates to the monotonicity assumption. Monotone treatment response is a fairly weak assumption in this context. In general, it seems unlikely that a campaign visit would hurt a candidate's vote share even if the visit did not help the candidate's election chances.

The next heuristic we consider is that of treatment selection (Manski 1995). Under positive selection, we assume that treated units are more likely to be helped than control units. This implies that the president selects candidates for a visit based on whether the visit is likely to increase the chances of winning that election. Formally, we write a positive selection assumption in the following way:

$$\Pr(Z = 1|D = 1) > \Pr(Z = 1|D = 0).$$

In words, the treatment selection assumption expresses the belief that the probability of randomly selecting a helped unit among the treated units is greater than the probability of randomly selecting a helped unit among the control units. Next, we write this assumption in terms of the model parameters that govern the distribution of Z given D and Y. Some elementary calculations allow us to write:

$$\Pr(Z = 1|D = 1) = (1 - \psi_{11}) \Pr(Y = 1|D = 1)$$

and

$$\Pr(Z = 1|D = 0) = \psi_{00} \Pr(Y = 0|D = 0)$$

Next we substitute in the MLEs for $\Pr(Y = 1|D = 1)$ and $\Pr(Y = 0|D = 0)$, which allows us to express the positive treatment selection assumption for the data in Table 1 as:

$$(1 - \psi_{11}) \left(\frac{18}{21}\right) > \psi_{00} \left(\frac{164}{327}\right). \tag{6}$$

To operationalize positive selection effects through the prior, one could set values of b_{00}, c_{00}, b_{11}, and c_{11} such that Inequality (6) is likely to be satisfied. The limiting case is that in which Inequality (6) is always satisfied a priori, while weaker versions of positive selection are obtained by setting the priors so that prior probability that Inequality (6) holds is between 0.5 and 1. Is selection a reasonable assumption in this context? The president is unlikely to waste time campaigning for candidates that have little chance of winning nor is the president likely to campaign much for candidates that will win easily. Thus the president is likely to avoid never succeed and always succeed

types and instead attempt to identify candidates that will be helped. This is consistent with our positive selection assumption.

We can also combine the monotonicity and selection assumptions and estimates effects under the assumption of monotone treatment selection (MTS), which should further sharpen the inference. While, we cannot verify that either assumption holds individually or in combination, there is no reason to think that the presence of one assumption decreases the likelihood of the other assumption. While researchers may choose different prior values, we think these two heuristics are widely applicable to many empirical settings. It is important to note that any particular choice or prior over $\boldsymbol{\psi}$ will have implications for the fractions of helped, hurt, never succeed, and always succeed units in the sample. Alternative prior assumptions will yield alternative inferences. Researchers need to consider the implications of any priors that are used.

4.2. *Posterior Inference.* The posterior distributions for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ discussed in Section 2 can all be sampled using simple independent Monte Carlo sampling. Unlike many Bayesian methods for partial identification (Gustafson 2010; Gustafson et al. 2010), Markov chain Monte Carlo methods are unnecessary here. To produce a Monte Carlo sample of size m from the distribution with density given (up to proportionality) by Equation 3 we can use Algorithm 4.1.

Algorithm 4.1: Posterior Sampling Unobserved Z $(\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, m)$

for $j \leftarrow 1$ **to** m

do
$$\begin{cases} \boldsymbol{\theta}^{(j)} \leftarrow rdirichlet(C_{00+} + a_{00},\ C_{01+} + a_{01},\ C_{10+} + a_{10},\ C_{11+} + a_{11}) \\ \psi_{00}^{(j)} \leftarrow rbeta(b_{00},\ c_{00}) \\ \psi_{01}^{(j)} \leftarrow rbeta(b_{01},\ c_{01}) \\ \psi_{10}^{(j)} \leftarrow rbeta(b_{10},\ c_{10}) \\ \psi_{11}^{(j)} \leftarrow rbeta(b_{11},\ c_{11}) \end{cases}$$

return $(\{\boldsymbol{\theta}^{(j)}\}_{j=1}^{m},\ \{\psi_{00}^{(j)}\}_{j=1}^{m},\ \{\psi_{01}^{(j)}\}_{j=1}^{m},\ \{\psi_{10}^{(j)}\}_{j=1}^{m},\ \{\psi_{11}^{(j)}\}_{j=1}^{m})$

Here $rdirichlet(d, e, f, g)$ is a function that returns a pseudo-random draw from a $\mathcal{D}irichlet(d, e, f, g)$ distribution and $rbeta(d, e)$ is a function that returns a pseudo-random draw from a $\mathcal{B}eta(d, e)$ distribution.

Once a sample $\{\boldsymbol{\theta}^{(j)}, \boldsymbol{\psi}^{(j)}\}_{j=1}^{m}$ from the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\psi})$

has been drawn, these draws are plugged into the formulas for the causal quantity of interest. A sample from the posterior distribution of the sensitivity analysis average treatment effect ($\{ATE_s^{(j)}\}_{j=1}^m$) can be constructed by taking the jth sample to be

$$ATE_s^{(j)} = \left(\theta_{00}^{(j)}\psi_{00}^{(j)} + \theta_{01}^{(j)}\psi_{01}^{(j)} + \theta_{11}^{(j)}\right) - \left(\theta_{10}^{(j)}\psi_{10}^{(j)} + \theta_{11}^{(j)}\psi_{11}^{(j)} + \theta_{01}^{(j)}\right)$$

for $j = 1, \ldots, m$. Samples from the posterior distributions of other causal quantities of interest follow analogously. Once we obtain a sample from the posterior distribution of interest, we can summarize the distribution by calculating density estimates, highest posterior density regions (the smallest region that contains a pre-specified amount of the posterior mass), the probability that a quantity of interest is greater than 0, etc. using the sampled parameter values. See Gelman et al. (2003) and Gill (2007) for discussions of how posterior samples can be summarized.

4.3. *Extensions to Cases with Observed Confounders.* We can also adjustment for measured confounders in our approach. Next, we outline the simplest case where there is a single measured discrete confounder. We also briefly discuss continuous confounders and multiple confounders. First, we assume there is a single measured confounder, X, which is discrete with K categories. In this case, it is appropriate to perform the simple 2×2 analyses described above separately within each of the $k = 0, \ldots, K-1$ levels of X and then weight by $\phi_k \equiv \Pr(X = k)$. More formally, we can write the likelihood for the observed data marginalized over the unobserved Z variables as:

$$\begin{aligned} p(\mathbf{x}, \mathbf{d}, \mathbf{y}|\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\psi}) &= \prod_{i=1}^{n} \sum_{z_i \in \mathcal{Z}_i} p(x_i, d_i, y_i, z_i|\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\psi}) \\ &= \prod_{i=1}^{n} p(x_i, d_i, y_i|\boldsymbol{\phi}, \boldsymbol{\theta}) \left\{ \sum_{z_i \in \mathcal{Z}_i} p(z_i|x_i, d_i, y_i, \boldsymbol{\psi}) \right\} \\ &= \prod_{i=1}^{n} p(x_i|\boldsymbol{\phi}) p(d_i, y_i|x_i, \boldsymbol{\theta}) \left\{ \sum_{z_i \in \mathcal{Z}_i} p(z_i|x_i, d_i, y_i, \boldsymbol{\psi}) \right\} \end{aligned}$$

where the $p(x_i|\boldsymbol{\phi})$ term is a multinomial mass function with probability vector $\boldsymbol{\phi}$ and multinomial sample size 1, $p(d_i, y_i|x_i, \boldsymbol{\theta})$ is a multinomial mass function that depends on the value of x_i with probabilities $\theta_{00|x_i}, \theta_{01|x_i}, \theta_{10|x_i}, \theta_{11|x_i}$

and multinomial sample size 1, and the term involving the sum over Z consists of $4K$ Bernoulli mass functions with parameters $\psi_{00|x_i}, \psi_{01|x_i}, \psi_{10|x_i}$ and $\psi_{11|x_i}$. Note that everything is analogous to the 2×2 analysis except for the conditioning on x_i throughout.

Adjustment for X also requires that the formulas for the sensitivity analysis probabilities of the various potential outcomes be rewritten to account for this adjustment.

$$\begin{aligned}
\Pr_s(Y(0)=0) &= \sum_{x=0}^{K-1}\sum_{z=0}^{3} \Pr(Y=0|D=0,X=x,Z=z)\Pr(Z=z|X=x)\Pr(X=x) \\
&= \sum_{x=0}^{K-1} [\theta_{10|x}(1-\psi_{10|x}) + \theta_{11|x}(1-\psi_{11|x}) + \theta_{00|x}]\phi_x \\
\Pr_s(Y(0)=1) &= \sum_{x=0}^{K-1}\sum_{z=0}^{3} \Pr(Y=1|D=0,X=x,Z=z)\Pr(Z=z|X=x)\Pr(X=x) \\
&= \sum_{x=0}^{K-1} [\theta_{10|x}\psi_{10|x} + \theta_{11|x}\psi_{11|x} + \theta_{01|x}]\phi_x \\
\Pr_s(Y(1)=0) &= \sum_{x=0}^{K-1}\sum_{z=0}^{3} \Pr(Y=0|D=1,X=x,Z=z)\Pr(Z=z|X=x)\Pr(X=x) \\
&= \sum_{x=0}^{K-1} [\theta_{00|x}(1-\psi_{00|x}) + \theta_{01|x}(1-\psi_{01|x}) + \theta_{10|x}]\phi_x \\
\Pr_s(Y(1)=1) &= \sum_{x=0}^{K-1}\sum_{z=0}^{3} \Pr(Y=1|D=1,X=x,Z=z)\Pr(Z=z|X=x)\Pr(X=x) \\
&= \sum_{x=0}^{K-1} [\theta_{00|x}\psi_{00|x} + \theta_{01|x}\psi_{01|x} + \theta_{11|x}]\phi_x
\end{aligned}$$

It is straightforward to show that conditioning on an observed, pre-treatment covariate X as above does not change the large sample bounds on the ATE, ATT, or ATC. For example, consider the case of the ATE. As before, $ATE_s = \Pr_s(Y(1)=1) - \Pr_s(Y(0)=1)$. Substituting, we get

$$ATE_s = \sum_{x=0}^{K-1} [\theta_{00|x}\psi_{00|x} + \theta_{01|x}\psi_{01|x} + \theta_{11|x}]\phi_x - \sum_{x=0}^{K-1} [\theta_{10|x}\psi_{10|x} + \theta_{11|x}\psi_{11|x} + \theta_{01|x}]\phi_x.$$

Assuming that $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are fixed, this quantity takes its maximum value when $\psi_{00|x} = \psi_{01|x} = 1$ and $\psi_{10|x} = \psi_{11|x} = 0$ for all x. Substituting these values into the expression for ATE_s and simplifying we get $\max[ATE_s] = \sum_{x=0}^{K-1} \theta_{00|x}\phi_x + \sum_{x=0}^{K-1} \theta_{11|x}\phi_x = \theta_{00} + \theta_{11}$, which is equivalent to the upper bound in Expression 5. Similar logic can be used to get the lower bound as well as the bounds on the ATT and ATC. Thus, like the bounds in (Mealli and Pacini 2013), conditioning on a pre-treatment covariate does not narrow our proposed bounds. Note that in scenarios where one is willing to make additional causal assumptions, conditioning on additional variables (oftentimes post-treatment variables) can narrow the bounds on causal effects of interest. For instance, see Grilli and Mealli (2008) and Glynn and Quinn (2011).

As such, the primary utility of conditioning on a pre-treatment covariate is that it may be easier to specify substantively plausible prior distributions for $\boldsymbol{\psi}$ conditional on X. Extensions to the case with multiple or continuous X are straightforward using a model for the conditional probability of treatment given X. While the use of covariates is possible, we see one key advantage of our approach is that credible inferences are possible in the absence of credible observed confounders.

5. Sensitivity Analysis Quantities for Presidential Campaigning. We begin the analysis with the estimation of naive treatment effect of a presidential campaign visit in 2002. For this estimate to be a valid causal effect, we must assume that presidential visits were as-if randomly assigned across Congressional Districts. An assumption that is clearly implausible. As we noted earlier, the naive estimate of the average treatment effect estimate is approximately 0.36, with a 95% confidence interval of 0.17 and 0.54. In contrast, Manski's no-assumption bounds on the treatment effect are also reported in Table 3 are -0.477 and 0.523. Next, we report results from an analysis that assumes independent, uniform priors on $\psi_{00}, \psi_{01}, \psi_{10}$, and ψ_{11}. We also assume, as we do throughout, that $\boldsymbol{\theta} \sim \mathcal{D}irichlet(0.25, 0.25, 0.25, 0.25)$. The resulting estimates and 95% credible intervals are in Table 3. Under this prior assumption, the endpoints of the 95% highest posterior density region are now -0.342 and 0.387, with a posterior mean of 0.023. Unlike in many applications of Bayesian inference, the use of uniform priors make little sense, since the goal is to use substantive information to reason about the nature of confounding. The uniform priors on $\boldsymbol{\psi}$ used here are not motivated by substantive knowledge and we do not attempt to defend them as such. Nonetheless, we report the results in Table 3 as a reference point for our other analyses. Our goal is to produce an inference under assumptions

that are more realistic than the naive estimate of the causal effect but also to sharpen the inference over the no assumption bounds using credible but weak assumptions.

TABLE 3

Estimated Treatment Effects for Presidential Visits. Point estimates (posterior means) are on the first row. Interval estimates are on the second row. The intervals in columns 1 and 3 are 95% highest posterior density intervals.

Naive Estimate	No-Assumption Bounds	Uniform Prior
0.36		0.023
[0.189, 0.492]	[-0.477, 0.523]	[-0.342, 0.387]

Next, we use a set of priors that assume varying levels of a monotonic treatment effect. Here, we make assumptions about the fraction of units helped by the treatment. The monotonicity assumption seems reasonable; given that we expect that a visit from George W. Bush is unlikely to hurt the vote shares of few if any Republican candidates in 2002. Please see the Appendix for a full discussion of the prior distributions used in this analysis. The first row of Table 4 contains estimates for the effect of a presidential visit under three different levels of monotonicity. The strongest monotonicity assumption used here assumes that about 1% (with 95% prior interval of [0%, 2%]) of the units that could have been hurt by a visit from President Bush were hurt units ($Z = 2$). This is nearly equivalent to Manski's(1997) definition of monotone treatment response (no unit is hurt by a visit from President Bush). The weakest monotonicity assumption used here assumes that about 10% (with 95% prior interval of [2%, 23%]) of the units that could have been hurt by a visit from President Bush were hurt units ($Z = 2$). Our intermediate monotonicity assumptions assumes that about 5% (with 95% prior interval of [1%, 12%]) of the units that could have been hurt by a visit from President Bush were hurt units ($Z = 2$).

Under the weakest monotonicity assumption, the estimated treatment effect is 0.213, however the 95% highest posterior density (HPD) interval includes zero and thus we cannot conclude that visits were effective. Strengthening the monotonicity assumption to the intermediate level increases the point estimate to nearly 0.24 and the associated 95% HPD interval no longer includes 0. Under the strongest monotonicity assumption, the ATE estimate is 0.257 with a 95% HPD interval of $[0.025, 0.489]$. This a sizable effect but still more than 25% smaller than the naive treatment effect estimate. It appears that so long as we are willing to assume that the effect of the treatment was moderately or strongly monotonic presidential visits were an effective

campaign strategy.

TABLE 4
Estimated Treatment Effects for Presidential Visits Under Monotonicity and Selection Assumptions. Point estimates are posterior means and interval estimates are 95% highest posterior density intervals.

Monotonicity		
Weak	Medium	Strong
0.213 [-0.024, 0.455]	0.238 [0.005, 0.472]	0.257 [0.025, 0.489]

Selection		
Weak	Medium	Strong
0.163 [-0.107, 0.436]	-0.025 [-0.304, 0.248]	-0.165 [-0.426, 0.082]

Next, we use a set of priors that assume that candidates in the treatment group where chosen to maximize the outcome: wins by Republican candidates. Thus we assume that Bush chose to campaign for candidates that he thought would be most helped by a visit. Given that presidents have many constraints on their time, the selection assumption is quite plausible. We vary the strength of this assumption and estimate the treatment effect under weak, medium, and strong selection assumptions. Under each of these three versions of the selection assumption we assume that ψ_{11} has a uniform prior ($b_{11} = c_{11} = 1$) and we vary the prior on ψ_{00}.[3] Under weak selection we assume that the prior mean of ψ_{00} is 0.8 with a 95% prior interval of $[0.52, 0.97]$. This corresponds to

$$\Pr\left((1-\psi_{11})\left(\frac{18}{21}\right) > \psi_{00}\left(\frac{164}{327}\right)\right) \approx 0.53.$$

Under this configuration of the prior for the treatment selection assumption, this inequality holds approximately 53% of the time, which implies that selection only weakly holds in the population of candidates. Alternative, under this assumption we are expressing the belief that, the probability of randomly selecting a helped candidate among the treated units is greater

[3]Note that with two free parameters in Inequality 6 there is no single way to operationalize positive selection via the prior. For instance, one could also fix the prior for ψ_{00} and vary the prior for ψ_{11} (or vary both simultaneously).

than the probability of randomly selecting a helped candidate among the control units. However, that probability just barely holds at approximately 0.53. Under medium selection we assume that the prior mean of ψ_{00} is 0.4 with a 95% prior interval of $[0.14, 0.70]$. This corresponds to

$$\Pr\left((1-\psi_{11})\left(\frac{18}{21}\right) > \psi_{00}\left(\frac{164}{327}\right)\right) \approx 0.76.$$

In this second version of the prior the inequality holds approximately 75% of the time, which implies that selection is more likely in the population of candidates. Finally, under strong selection we assume that the prior mean of ψ_{00} is 0.1 with a 95% prior interval of $[0.003, 0.34]$. This corresponds to

$$\Pr\left((1-\psi_{11})\left(\frac{18}{21}\right) > \psi_{00}\left(\frac{164}{327}\right)\right) \approx 0.94.$$

Now the inequality holds approximately 95% of the time in the population of candidates.

The estimates under selection are in the second row of Table 4. Here, we see that positive selection pushes the naive estimate toward it's lower bound— with stronger forms of positive selection moving the estimate closer to the lower Manski bound. The reason the selection assumption moves the estimates toward zero is that fraction of treated candidates is quite small. Under all three levels of positive selection, the 95% HPD interval includes 0. Thus under various degrees of positive selection (but no explicit assumption of monotonic treatment response) we cannot rule out the possibility that a campaign visit by President Bush had no average effect on the election chances of Republican candidates.

To this point we have seen that two seemingly plausible types of assumptions (monotonic treatment response and positive selection) push the naive inference in opposite directions. What should one conclude if one believes that both monotonic treatment response and positive selection are plausible? To answer that question, we use priors that assume both selection and monotonicity of varying levels. If we again vary each assumption at three levels, that creates nine different prior combinations. Table 5 contains the nine different estimates of the treatment effect along with the associated 95% HPD intervals. The combination of the monotonicity and selection assumptions is sufficient to result in informative inferences in six out of the nine combinations. Only when positive selection is strong does the 95% credible interval contain zero. Under the other six combinations the treatment effect varies from 0.166 to 0.398.

TABLE 5
Estimated Treatment Effects for Presidential Visits Under Combinations of Monotonicity and Selection Assumptions. Point estimates are posterior means and interval estimates are 95% highest posterior density intervals.

		Monotonicity		
		Weak	Medium	Strong
Selection	Weak	0.355 [0.219, 0.479]	0.379 [0.254, 0.494]	0.398 [0.270, 0.511]
	Medium	0.166 [0.022, 0.318]	0.190 [0.052, 0.335]	0.209 [0.074, 0.351]
	Strong	0.026 [-0.073, 0.143]	0.049 [-0.030, 0.150]	0.068 [-0.001, 0.163]

What conclusion should we draw about the effectiveness of presidential campaign visits? Our conclusions here must be evaluated against the plausibility of the assumptions. The selection assumption is the most likely of the two assumptions to hold. There is every reason to believe that presidents' select which candidates to campaign for based some hope of being effective. It is unlikely that Presidents would spend much time campaigning for candidates that have little chance of winning. With only the selection assumption in place, we cannot rule out the possibility that the average effect of a Presidential visit is 0. However, If we believe that treatment response is also monotone for most candidates there does appear to be a positive treatment effect. Consider the effect under strong selection and strong monotonicity. While the 95% HPD intervals just barely contain zero, the point estimate is 0.068. If we take that as the most credible set of assumptions, presidential campaigning would appear to be effective, but the estimate is much reduced compared to the naive estimate of 0.36.

As we noted above, identification via a conditional ignorability assumption is possible, but we argued that this assumption is unrealistic in this context given that we think there are unobservable factors related to treatment. However, we estimated treatment effects under the conditional ignorability assumption to compare to the estimates from our Bayesian partial identification approach. To estimate the treatment effect of a presidential campaign visit under a conditional ignorability assumption, we use a marginal structural model (Robins, Hernan and Brumback 2000). In the model, we use

stabilized weights (Cole and Hernán 2008). We use the same set of covariates used in the original analysis in Keele, Fogarty and Stimson (2004). These covariates include the vote share received by George Bush in the 2000 election, an indicator for whether there was a Senate race in the state in 2002, an indicator for whether the race did not have an incumbents running, an indicator for whether Congressional Quarterly designated the race as competitive in the Spring of 2002, percentage of residents that live in urban areas, the percentage of residents that are African American, the percentage of Hispanic residents, the percentage of residents with a high school degree, the percentage of residents with a college degree, and median income.

Under the conditional ignorability assumption, the estimated treatment effect is 0.24, with a 95% confidence interval of -0.18 and 0.67. While the magnitude of this effect is reduced compared to the naive estimate, this estimate is notably larger than most of the estimates under the assumptions in Table 5. Moreover, inspection of the data revealed clear violations of the positivity assumption which implies that there must be a nonzero probability of treatment for every level and combination of confounders (Cole and Hernán 2008). Thus we have multiple reasons to doubt an approach based on modeling the assignment mechanism with covariates.

6. Conclusion. In this article we have illustrated how to conduct a form of sensitivity analysis under general patterns of unobserved confounding. In many social science applications interventions cannot be randomized, and the assumption of that all confounders are observed is implausible. For example, in our application, there is little reason to think we observe all the necessary covariates that would make congressional districts visited by the president comparable to the congressional districts that do not receive a presidential campaign visit. The development of methods of sensitivity analyses for situations in which unmeasured confounding is present, as is done in this paper, serves to shift empirical social science research away from the all too typical enterprise of defending indefensible causal assumptions to the practice of honestly stating the range of assumptions that are consistent with a size of causal effect. Using Bayesian methods, analysts can present a range of estimates under different patterns of confounding. In the analysis presented here, we find some evidence for a positive effect of Presidential visits on election success under some ranges of plausible assumptions. However, under other plausible assumptions the null of no average effect cannot be rejected. Generally, we find that estimate form our Bayesian partial identification approach are much reduced compared to both the naive estimate and an estimate based on a conditional ignorability assumption.

References.

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* **91** 444-455.

Balke, A. and Pearl, J. (1997). Bounds on Treatment Effects From Studies With Imperfect Compliance. *Journal of the American Statistical Association* **92** 1171-1176.

Chickering, D. M. and Pearl, J. (1997). A Clinician's Tool for Analyzing Non-Compliance. *Computing Science and Statistics* **29** 424-431.

Cohen, J. E., Krassa, M. A. and Hamman, J. A. (1991). The Impact of Presidential Campaigning on Midterm U.S. Senate Elections. *American Political Science Review* **85** 165-178.

Cole, S. R. and Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American journal of epidemiology* **168** 656–664.

Cornfield, J., Haenszel, W., Hammond, E., Lilienfeld, A., Shimkin, M. and Wynder, E. (1959). Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions. *Journal of National Cancer Institute* **22** 173-203.

Cox, D. R. (1958). *Planning of Experiments*. Wiley, New York.

Ding, P. and Dasgupta, T. (2016). A Potential Tale of Two-by-Two Tables From Completely Randomized Experiments. *Journal of the American Statistical Association* **111** 157–168.

Ding, P. and VanderWeele, T. J. (2016a). Sensitivity analysis without assumptions. *Epidemiology (Cambridge, Mass.)* **27** 368.

Ding, P. and Vanderweele, T. J. (2016b). Sharp sensitivity bounds for mediation under unmeasured mediator-outcome confounding. *Biometrika* **103** 483–490.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall, London.

Gill, J. (2007). *Bayesian Methods: A Social and Behavioral Science Approach*, second ed. Chapman & Hall/CRC, Boca Raton, FL.

Gill, J. and Walker, L. D. (2005). Elicited Priors for Bayesian Model Specifications in Political Science Research. *The Journal of Politics* **67** 841-872.

Glynn, A. N. and Quinn, K. M. (2011). Why Process Matters for Causal Inference. *Political Analysis* **19** 273-286.

Grilli, L. and Mealli, F. (2008). Nonparametric Bounds on the Causal Effect of University Studies on Job Opportunities Using Principal Stratification. *Journal of Educational and Behavioral Statistics* **33** 111-130.

Gustafson, P. (2010). Bayesian inference for partially identified models. *The international journal of biostatistics* **6**.

Gustafson, P., McCandless, L., Levy, A. and Richardson, S. (2010). Simplified Bayesian Sensitivity analysis for mismeasured and unobserved confounders. *Biometrics* **66** 1129–1137.

Herrnson, P. S. and Morris, I. L. (2007). Presidential Campaigning in the 2002 Congressional Elections. *Legislative Studies Quarterly* **32** 629-648.

Holland, P. W. (1986a). Statistics and Causal Inference. *Journal of the American Statistical Association* **81** 945-960.

Holland, P. W. (1986b). Statistics and Causal Inference. *Journal of the American Statistical Association* **81** 945-960.

Ichino, A., Mealli, F. and Nannicini, T. (2008). From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and Their Sensitivity? *Journal of Applied Econometrics* **23** 305-327.

Imai, K. and Yamamoto, T. (2008). Causal Inference with Measurement Error: Nonparametric Identification and Sensitivity Analyses of a Field Experiment on Democratic Deliberations. Paper presented at the 2008 Summer Political Methodology Meeting.

Imbens, G. W. (2003). Sensitivity to Exogeneity Assumptions in Program Evaluation. *The American Economic Review Papers and Proceedings* **93** 126-132.

Jacobson, G. C. (2003). *The Politics of Congressional Elections.*, 6th edition ed. Longman, New York.

Jin, H. and Rubin, D. B. (2008). Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association* **103** 101–111.

Kadane, J. B. and Wolfson, L. J. (1998). Experiences in Elicitation. *The Statistician* **47** 3-19.

Keele, L. J., Fogarty, B. and Stimson, J. A. (2004). The Impact of Presidential Visits in the 2002 Congressional Elections. *PS: Political Science & Politics* **34** 971-986.

Manski, C. F. (1990). Nonparametric Bounds on Treatment Effects. *The American Economic Review Papers and Proceedings* **80** 319-323.

Manski, C. F. (1993). Identification Problems in the Social Sciences. *Sociological Methodology* **23** 1-56.

Manski, C. F. (1995). *Identification Problems in the Social Sciences*. Harvard University Press, Cambridge, MA.

Manski, C. F. (1997). Monotone Treatment Response. *Econometrica* **65** 1311-1334.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer, New York.

McCandless, L. C., Gustafson, P. and Levy, A. (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in medicine* **26** 2331–2347.

Mealli, F. and Pacini, B. (2013). Using Secondary Outcomes to Sharpen Inference in Randomized Experiments With Noncompliance. *Journal of the American Statistical Association* **108** 1120-1131.

Moon, H. R. and Schorfheide, F. (2012). Bayesian and frequentist inference in partially identified models. *Econometrica* **80** 755–782.

Neustadt, R. E. (1960). *Presidential Power*. Wiley, New York, NY.

Nutting, B. and Stern, A. H. (2001). *CQ's Politics in America: 2002 the 107th Congress*. Congressional Quarterly Press, Washington, D.C.

Pearl, J. (1995). Causal Diagrams for Empirical Research. *Biometrika* **82** 669-710.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York.

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

Robins, J. M. (1986). A new aproach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modeling* **7** 1393-1512.

Robins, J. M. (1989). The Analysis of Randomized and Nonrandomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies. In *Health Service Research Methodology: A Focus on AIDS* (L. Sechreset, H. Freeman and A. Mulley, eds.) U.S. Public Health Service, Washington DC.

Robins, J. M., Hernan, M. A. and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.

Rosenbaum, P. R. (1987). Sensitivity Analysis For Certain Permutation Inferences in Matched Observational Studies. *Biometrika* **74** 13-26.

Rosenbaum, P. R. (2002). *Observational Studies*, 2nd ed. Springer, New York, NY.

Rosenbaum, P. R. and Rubin, D. B. (1983). Assessing Sensitivity to an Unobserved Covariate in an Observational Study With Binary Outcome. *Journal of The Royal Statistical Society Series B* **45** 212-218.

Rubin, D. B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics* **6** 34-58.

Rubin, D. B. (1990). Formal Modes of Statistical Inference For Causal Effects. *Journal of Statistical Planning* **25** 279-292.

Rubin, D. B. (2008). For Objective Causal Inference, Design Trumps Analysis. *The Annals of Applied Statistics* **2** 808-840.

Sellers, P. J. and Denton, L. M. (2006). Presidential Visits and Midterm Senate Elections. *Presidential Studies Quarterly* **36** 410-432.

Western, B. and Jackman, S. (1994). Bayesian Inference for Comparative Research. *American Political Science Review* **88** 412-423.

Appendices

APPENDIX A: PRIOR SELECTION AND PATTERNS OF CONFOUNDING

Table 6 provides an overview of the various parameters in the model. It is by placing prior distributions on the $\boldsymbol{\psi}$ parameters that we use subjective knowledge of possible patterns of confounding into the model.

Parameter θ_{dy}	Probability $\Pr(D_i = d, Y_i = y)$	Interpretation Probability D_i is equal to d and Y_i is equal to y
ψ_{00}	$\Pr(Z_i = 1 \mid D_i = 0, Y_i = 0)$	Probability i would be helped by treatment given i not treated and i failed
$1 - \psi_{00}$	$\Pr(Z_i = 0 \mid D_i = 0, Y_i = 0)$	Probability i would never succeed given i not treated and i failed
ψ_{01}	$\Pr(Z_i = 3 \mid D_i = 0, Y_i = 1)$	Probability i would always succeed given i not treated and i succeeded
$1 - \psi_{01}$	$\Pr(Z_i = 2 \mid D_i = 0, Y_i = 1)$	Probability i would be hurt by treatment given i not treated and i succeeded
ψ_{10}	$\Pr(Z_i = 2 \mid D_i = 1, Y_i = 0)$	Probability i was hurt by treatment given i treated and i failed
$1 - \psi_{10}$	$\Pr(Z_i = 0 \mid D_i = 1, Y_i = 0)$	Probability i would never succeed given i treated and i failed
ψ_{11}	$\Pr(Z_i = 3 \mid D_i = 1, Y_i = 1)$	Probability i would always succeed given i treated and i succeeded
$1 - \psi_{11}$	$\Pr(Z_i = 1 \mid D_i = 1, Y_i = 1)$	Probability i was helped by treatment given i treated and i succeeded

TABLE 6
Interpretation of Parameters in the Model for (D,Y,Z). *The i indices denote a randomly selected unit.*

APPENDIX B: PRIOR SELECTION IN THE APPLICATION

McCourt School of Public Policy
& Dept of Government
304 Old North, Georgetown University
37th & O St, NW, Washington, D.C. 20057
E-mail: lk681@georgetown.edu

UC Berkeley School of Law
University of California, Berkeley
490 Simon #7200,
Berkeley, CA 94720
E-mail: kquinn@law.berkeley.edu

TABLE 7
Prior Parameter Values Used in Analysis

	b_{00}	c_{00}	b_{01}	c_{01}	b_{10}	c_{10}	b_{11}	c_{11}
weak mono.	1	1	27	3	3	27	1	1
med. mono.	1	1	57	3	3	57	1	1
strong mono.	1	1	396	4	4	396	1	1
weak select.	8	2	1	1	1	1	1	1
med. select.	4	6	1	1	1	1	1	1
strong select.	1	9	1	1	1	1	1	1
weak mono. & weak select.	8	2	27	3	3	27	1	1
med. mono. & weak select.	8	2	57	3	3	57	1	1
strong mono. & weak select.	8	2	396	4	4	396	1	1
weak mono. & med. select.	4	6	27	3	3	27	1	1
med. mono. & med. select.	4	6	57	3	3	57	1	1
strong mono. & med. select.	4	6	396	4	4	396	1	1
weak mono. & strong select.	1	9	27	3	3	27	1	1
med. mono. & strong select.	1	9	57	3	3	57	1	1
strong mono. & strong select.	1	9	396	4	4	396	1	1