Online Appendix

Mastering the Art of Cookbook Medicine:
Machine Learning, Randomized Trials, and Misallocation
Jason Abaluck, Leila Agha, David Chan, Daniel Singer, Diana Zhu

A.1 Bayesian Model of Decision-Making

In this appendix, we describe in greater detail the Bayesian model of decision-making specified in Equation (10), which we restate here:

$$W_i = \mathbf{1}\left\{\beta^s \tilde{\tau}^s_{i,g} + \beta^b \tilde{\tau}^b_{i,g} + f(X_i) + v_i > 0\right\},$$

focusing on the Bayesian posterior beliefs about treatment effects: $\tilde{\tau}^s_{i,g}$ and $\tilde{\tau}^b_{i,g}$. Recall that g denotes the adoption status of the physician. Adoption status may change the informativeness of physician beliefs about treatment effects.

A.1.1 Component Treatment Effects, Signals, and Beliefs

Treatment effects and physician beliefs depend on patient characteristics, which we may orthogonalize into components $k \in \mathcal{K}$. We can conceptualize each principal component as implying additional (orthogonal) information about treatment effects. Specifically, assume that treatment effects are normally distributed and comprise component treatment effects that are also normally distributed:

$$\tau^o_{i,k} \sim N\left(\overline{\tau}^o_k, \sigma^2_{\tau(o),k}\right), \tag{A.1}$$

for outcome o and each $k \in \mathcal{K}$. We assume that physicians know the moments of each component treatment effect $\left(\overline{\tau}^o_k, \sigma^2_{\tau(o),k}\right)_{k=1}^{K}$.[28]

For each component k, physicians receive a noisy signal of the underlying treatment effect, $\dot{\tau}^o_{i,k}$:

$$\dot{\tau}^o_{i,g,k} = \tau^o_{i,k} + \epsilon^o_{i,g,k}, \tag{A.2}$$

where $\epsilon^o_{i,g,k}$ is a normally distributed noise term with variance $\sigma^2_{\epsilon(o),g,k}$, or $\epsilon^o_{i,g,k} \sim N\left(0, \sigma^2_{\epsilon(o),g,k}\right)$. Note the dependence of signals on g. This models the possibility that adoption status may change the quality of information that physicians receive about treatment effects.

Given prior beliefs and the noisy signals, physicians form posterior beliefs, $\tilde{\tau}^o_{i,k}$. Specifically,

$$\tilde{\tau}^o_{i,g,k} = \lambda^o_{g,k}\dot{\tau}^o_{i,g,k} + (1-\lambda^o_{g,k})\overline{\tau}^o_k, \tag{A.3}$$

[28] Our model in Equation (10) allows for potentially non-Bayesian beliefs that can shift decision-making via $f(X_i)$ and v_i. In order to study the effect of information in a Bayesian framework, we compartmentalize the two components of the model and consider the first component, described in this appendix, as Bayesian.

where $\lambda^o_{g,k} = \frac{\sigma^2_{\tau(o),k}}{\sigma^2_{\tau(o),k}+\sigma^2_{\epsilon(o),g,k}}$ is the signal-to-noise ratio of the kth component.

A.1.2 Regression Interpretation

The relationship between posterior beliefs and signals in Equation (A.3) can be interpreted as a regression of posterior beliefs on signals. This relationship may also be interpreted as a regression of posterior beliefs on true treatment effects, since the noise component of signals is orthogonal to treatment effects:

$$\begin{aligned}\tilde{\tau}^o_{i,g,k} &= \lambda^o_{g,k}\dot{\tau}^o_{i,g,k} + (1-\lambda^o_{g,k})\overline{\tau}^o_k \\ &= \lambda^o_{g,k}\tau^o_{i,k} + (1-\lambda^o_{g,k})\overline{\tau}^o_k + \lambda^o_{g,k}\epsilon^o_{i,g,k},\end{aligned}$$

where the second line uses the definition of the signal in Equation (A.2). In other words, a unit increase in the treatment effect $\tau^o_{i,k}$ should increase posterior beliefs by $\lambda^o_{g,k}$.

We may use this framework to consider the relationship between overall treatment effects, overall signals, and overall posterior beliefs, aggregated across components $k \in \mathcal{K}$. These overall objects are, respectively, $\tau^o_i \equiv \sum_{k\in\mathcal{K}}\tau^o_{i,k}$; $\dot{\tau}^o_{i,g} = \sum_{k\in\mathcal{K}}\dot{\tau}^o_{i,g,k}$; and $\tilde{\tau}^o_{i,g} = \sum_{k\in\mathcal{K}}\tilde{\tau}^o_{i,g,k}$. Substituting the definition of the component signals from Equation (A.3), we may also state the overall posterior belief as

$$\tilde{\tau}^o_{i,g} = \sum_{k\in\mathcal{K}}\left(\lambda^o_{g,k}\dot{\tau}^o_{i,g,k} + (1-\lambda^o_{g,k})\overline{\tau}^o_k\right). \tag{A.4}$$

We now consider the overall signal-to-noise ratio in a regression predicting the overall posterior belief using the signal:

$$\tilde{\tau}^o_{i,g} = \lambda^o_g\dot{\tau}^o_{i,g} + (1-\lambda^o_g)\overline{\tau}^o. \tag{A.5}$$

Using Equation (A.4) for $\tilde{\tau}^o_{i,g}$ and the definition of the overall signal for $\dot{\tau}^o_{i,g}$, the coefficient λ^o_g in this regression is

$$\begin{aligned}\lambda^o_g &= \frac{\operatorname{Cov}\left(\tilde{\tau}^o_{i,g},\dot{\tau}^o_{i,g}\right)}{\operatorname{Var}\left(\dot{\tau}^o_{i,g}\right)} = \frac{\sum_{k\in\mathcal{K}}\lambda^o_{g,k}\operatorname{Var}\left(\dot{\tau}^o_{i,g,k}\right)}{\sum_{k\in\mathcal{K}}\operatorname{Var}\left(\dot{\tau}^o_{i,g,k}\right)} &\text{(A.6)} \\ &= \frac{\sum_{k\in\mathcal{K}}\sigma^2_{\tau(o),g,k}}{\sum_{k\in\mathcal{K}}\left(\sigma^2_{\tau(o),g,k}+\sigma^2_{\epsilon(o),g,k}\right)}. &\text{(A.7)}\end{aligned}$$

Equation (A.6) reveals that the overall signal-to-noise ratio, λ^o_g, can be thought of as a variance-weighted average of the component signal-to-noise ratios, $\lambda^o_{g,k}$. Equation (A.7) shows that a posterior belief formed directly from the aggregate signal, as in Equation (A.5), will have the same signal-to-noise ratio as a posterior belief aggregated from component posterior beliefs, as in Equation (A.4).

A.1.3 CHADS$_2$ and Residual Treatment Effects

We are now in a position to state posterior beliefs as in Equations (11) and (12). For strokes, we can separate the set of components $\mathcal{K}_c$ that predict CHADS$_2$-related treatment effects and $\mathcal{K} \setminus \mathcal{K}_c$ components that predict residual treatment effects. We expect that the component posterior beliefs related to the CHADS$_2$ score should increase in informativeness. That is, we expect that $\lambda^s_{g,k}$ should increase with $g = \text{post}$, for $k \in \mathcal{K}_c$. We first define the two components of stroke treatment effects: $\tau_i^{s(c)} \equiv \sum_{k \in \mathcal{K}_c} \tau^s_{i,k}$, and $\tau_i^{s(r)} \equiv \sum_{k \notin \mathcal{K}_c} \tau^s_{i,k}$. Restating Equation (11) as

$$\tilde{\tau}^s_{i,g} = \lambda_g^{s(c)} \tau_i^{s(c)} + \lambda_g^{s(r)} \tau_i^{s(r)} + \mu^s_g + \upsilon^s_{i,g},$$

we can then interpret the signal-to-noise coefficients in the equation as follows:

$$\begin{aligned} \lambda_g^{s(c)} &= \frac{\sum_{k \in \mathcal{K}_c} \sigma^2_{\tau(s),g,k}}{\sum_{k \in \mathcal{K}_c} \left(\sigma^2_{\tau(s),g,k} + \sigma^2_{\epsilon(s),g,k} \right)}; \\ \lambda_g^{s(r)} &= \frac{\sum_{k \notin \mathcal{K}_c} \sigma^2_{\tau(s),g,k}}{\sum_{k \notin \mathcal{K}_c} \left(\sigma^2_{\tau(s),g,k} + \sigma^2_{\epsilon(s),g,k} \right)}. \end{aligned}$$

If we conceptualize the posterior belief as directly formed from $\dot{\tau}_i^{s(c)} \equiv \tau_i^{s(c)} + \sum_{k \in \mathcal{K}_c} \epsilon^s_{i,g,k}$ and $\dot{\tau}_i^{s(r)} \equiv \tau_i^{s(r)} + \sum_{k \notin \mathcal{K}_c} \epsilon^s_{i,g,k}$, then we can interpret the constant, μ^s_g, and error term, $\upsilon^s_{i,g}$ as

$$\begin{aligned} \mu^s_g &= \sum_{k \in \mathcal{K}} \left(1 - \mathbf{1}(k \in \mathcal{K}_c) \lambda_g^{s(c)} - \mathbf{1}(k \notin \mathcal{K}_c) \lambda_g^{s(r)} \right) \overline{\tau}^s_k; \\ \upsilon^s_{i,g} &= \sum_{k \in \mathcal{K}} \left(\mathbf{1}(k \in \mathcal{K}_c) \lambda_g^{s(c)} + \mathbf{1}(k \notin \mathcal{K}_c) \lambda_g^{s(r)} \right) \epsilon^s_{i,g,k}. \end{aligned}$$

Unlike $\lambda_g^{s(c)}$ and $\lambda_g^{s(r)}$, μ^s_g and $\mathrm{Var}\left(\upsilon^s_{i,g}\right)$ are not exactly invariant to the level of aggregation with which posterior beliefs are formed.[29] Nevertheless, regardless of this level of aggregation, qualitative interpretations are unchanged: μ^s_g is a function of the signal-to-noise ratio and prior beliefs, and $\upsilon^s_{i,g}$ is a function of signal-to-noise ratio and noise. If $\lambda^s_{g,k} = 1$ for all $k \in \mathcal{K}$, there is no noise, and $\upsilon^s_{i,g} = 0$. At the other extreme, if $\lambda^s_{g,k} = 0$ for all $k \in \mathcal{K}$, there is no meaningful signal. In this case, physicians will ignore all $\dot{\tau}^s_{g,k}$, and we will also have $\upsilon^s_{i,g} = 0$.

For bleeding events, we are only interested in overall treatment effects: $\tau_i^b \equiv \sum_{k \in \mathcal{K}} \tau^b_{i,k}$. Restating Equation (12) as

$$\tilde{\tau}^b_{i,g} = \lambda^b_g \tau^b_i + \mu^b_g + \upsilon^b_{i,g},$$

[29] For μ^s_g to be invariant, we require λ^s_g to be a different weighted average of $\lambda^s_{g,k}$, with weights proportional to $\overline{\tau}^s_k$ rather than $\mathrm{Var}\left(\dot{\tau}^s_{i,g,k}\right)$. For $\mathrm{Var}\left(\upsilon^s_{i,g}\right)$ to be invariant, we require $(\lambda^s_g)^2$ to be a weighted average of $(\lambda^s_{g,k})^2$, with weights proportional to $\mathrm{Var}\left(\epsilon^s_{i,g,k}\right)$ rather than $\mathrm{Var}\left(\dot{\tau}^s_{i,g,k}\right)$.

we interpret the signal-to-noise coefficient as

$$\lambda_g^b = \frac{\sum_{k\in\mathcal{K}} \sigma^2_{\tau(b),g,k}}{\sum_{k\in\mathcal{K}} \left(\sigma^2_{\tau(b),g,k} + \sigma^2_{\epsilon(b),g,k}\right)}.$$

If we conceptualize the posterior belief as directly formed from $\dot{\tau}_i^b \equiv \tau_i^b + \sum_{k\in\mathcal{K}_c} \epsilon^b_{i,g,k}$, then we can similarly interpret the constant, μ_g^b, as a function of λ_g^b and prior beliefs, or $\mu_g^b = \left(1-\lambda_g^b\right)\sum_{k\in\mathcal{K}} \overline{\tau}_k^s$. The error term, $\upsilon^b_{i,g}$, is similarly a function of λ_g^b and noise, or $\upsilon^b_{i,g} = \lambda_g^b \sum_{k\in\mathcal{K}} \epsilon^s_{i,g,k}$.

A.2 Adoption of Guideline Revealing $\left(\tau^{s(c)}(x), \tau^{s(r)}(x)\right)$

In this appendix, we discuss how we evaluate a counterfactual scenario in which physicians adopt a guideline that contains information on both $\tau^{s(c)}(x)$ and $\tau^{s(r)}(x)$. We assume that physicians who have adopted this guideline will have equal information on $\hat{\tau}^{s(c)}(x)$ and $\hat{\tau}^{s(r)}(x)$, which implies that $\lambda^{s(r)}_{\text{post}} = \lambda^{s(c)}_{\text{post}}$. We assume the same distraction effects as the CHADS$_2$ score, so that α^b_{post} remains unchanged. To evaluate counterfactual outcomes, we need to know the effect of this policy on $\sigma_{\varepsilon,\text{post}}$. Although this object is not identified in the data, we proceed with a bounding exercise.

We first note that adoption of the CHADS$_2$ score did little to change $\sigma_{\varepsilon,\text{post}}$. This suggest that, although the CHADS$_2$ score had a significant impact on physicians' information on $\tau^{s(c)}(x)$ (i.e., doubling $\lambda^{s(c)}_{\text{post}}$ relative to $\lambda^{s(c)}_{\text{pre}}$), uncertainty about $\tau^{s(c)}(x)$ accounts for very little in the variation in treatment decisions across patients with the same treatment effects. We therefore expect very little impact of the comprehensive guideline that also informs physicians of $\tau^{s(r)}(x)$.

As a baseline assumption, we consider no additional effect of the comprehensive guideline on $\sigma_{\varepsilon,\text{post}}$. We also consider a lower bound on $\sigma_{\varepsilon,\text{post}}$, which corresponds to an upper bound on the welfare improvement from the comprehensive guideline. In this lower bound, we consider the effect of the comprehensive guideline reducing the variance of $\beta^s \lambda_g^{s(r)} \sum_{k\notin\mathcal{K}_c} \epsilon^s_{i,g,k}$ to 0. We note that

$$\begin{aligned}
\text{Var}\left(\beta^s \lambda_g^{s(r)} \sum_{k\notin\mathcal{K}_c} \epsilon^s_{i,g,k}\right) &= \left(\beta^s \lambda_g^{s(r)}\right)^2 \sum_{k\notin\mathcal{K}_c} \sigma^2_{\epsilon(s),g,k} \\
&= \left(\frac{\alpha_g^{s(r)}}{\lambda_g^{s(r)}}\right)^2 \left(\lambda_g^{s(r)}\right)^2 \sum_{k\notin\mathcal{K}_c} \sigma^2_{\epsilon(s),g,k} \\
&\leq \left(2\alpha_g^{s(r)}\right)^2 \frac{1}{4} \sum_{k\notin\mathcal{K}_c} \sigma^2_{\tau(s),g,k} \\
&= \left(\alpha_g^{s(r)}\right)^2 \text{Var}\left(\tau^{s(r)}(x)\right).
\end{aligned}$$

The inequality comes from the fact that $\left(\lambda_g^{s(r)}\right)^2 \sum_{k\notin\mathcal{K}_c} \sigma^2_{\epsilon(s),g,k}$ is largest when $\lambda_g^{s(r)} = \frac{1}{2}$. This occurs when $\sum_{k\notin\mathcal{K}_c} \sigma^2_{\epsilon(s),g,k} = \sum_{k\notin\mathcal{K}_c} \sigma^2_{\tau(s),g,k}$. To evaluate this bound, we use the empirical variance of $\hat{\tau}^{s(r)}_{BLP}(x) \approx 0.00013$ for $\text{Var}\left(\tau^{s(r)}(x)\right)$.

A.3 Optimal Treatment as Function of $\left(\tau^s(x), \tau^b(x)\right)$

In this appendix, we will examine the optimal ranking of patients as a function of $\left(\tau^s(x), \tau^b(x)\right)$. We will use this ranking to characterize the optimal guideline under strict adherence, in which physicians follow the ranking of patients specified by the guideline. Intuitively, we find that the optimal guideline should rank patients in terms of the ratio of stroke treatment effects per bleed treatment effect. In other words, patients should be ranked by $|\tau^s(x)|/\tau^b(x)$, and patients with the highest value of $|\tau^s(x)|/\tau^b(x)$.

In order to see this intuition, we may first consider a social welfare function of the form

$$U_i = \overline{\beta}^s Y_i^s + \overline{\beta}^b Y_i^b, \tag{A.8}$$

where $\overline{\beta}^s$ and $\overline{\beta}^b$ are social preference weights with respect to stroke and bleeds, and Y_i^s and Y_i^b are realize stroke and bleed outcomes. Using notation from Section 5.1, we note that $Y_i^o = W_i Y_i^s(1) + (1 - W_i) Y_i^s(0)$, for $o \in \{s, b\}$. If the planner knows conditional treatment effects, $\left(\tau^s(x), \tau^b(x)\right)$, defined in Equations (2) and (3) as

$$\begin{aligned} \tau^s(x) &\equiv E\left[Y_i^s(1) - Y_i^s(0) \middle| X_i = x\right]; \\ \tau^b(x) &\equiv E\left[Y_i^b(1) - Y_i^b(0) \middle| X_i = x\right], \end{aligned}$$

she may maximize $E[U_i]$ with the following decision rule:

$$W_i = \mathbf{1}\{\overline{\beta}^s \tau^s(X_i) + \overline{\beta}^b \tau^b(X_i) > 0\}. \tag{A.9}$$

This decision rule is equivalent to a rule that treats if and only if $|\tau^s(x)|/\tau^b(x) > \overline{\beta}^b/\overline{\beta}^s$.

Note the similarity between the decision rule implied by Equation (A.9) and our model of physician decision-making in Equation (10). Physician decisions W_i will maximize U_i if and only if $\overline{\beta}^b/\overline{\beta}^s = \beta^b/\beta^s$ and $f(X_i) + \nu_i = 0$ for all i. The first condition reflects alignment in physician preferences and social preference weights. The second condition implies that physicians perfectly observe $\left(\tau^s(x), \tau^b(x)\right)$ and have no other decision-making considerations.

The decision rule in Equation (A.9) can be generalized to any arbitrary social preference ratio $\pi = \overline{\beta}^b/\overline{\beta}^s$. Suppose we consider optimal decisions under two preference ratios, π and $\pi' < \pi$. All patients who should be treated under the preference ratio π should also be treated under the preference ratio π'. No patients who are not treated under the preference ratio π' should be treated under the preference ratio π. Therefore, there are only incremental patients who should be treated under π' and not under π, and not vice versa. These patients have a ratio of treatment effects $|\tau^s(x)|/\tau^b(x) \in (\pi', \pi]$. This equivalence between patient rankings and utility maximization is noted in Vytlacil (2002) and Chan et al. (2019).

A.4 Predicting Physician Treatment Decisions

Observable variation in treatment effects, patient age, and time trends explain a relatively small fraction of the total variation in treatment decisions. In this section, we explore other factors that might drive physician treatment decisions. Specifically, we consider the following additional variables, which may influence physicians' treatment decisions. None of these variables are available in the AFI database, and so estimated treatment effects are not a direct function of these variables.

1. *Variables related to patient's ability to comply with warfarin monitoring*: drug abuse, depression, psychoses, number of years of military service. Appropriate management of patients on warfarin requires blood work repeated at regular intervals (typically every 2-4 weeks) to ensure the dosing is appropriate. Optimal dosing can depend on a patient's diet and other medications, and may need to be adjusted from time to time as those factors change. If the warfarin dosage is too low, the patient will not reap the benefits of anticoagulation for stroke reduction; if the dosage is too high, the patient will be at elevated risk of bleeds. These variables included here are related to the likelihood that the patient can comply with the monitoring regimen.

2. *Variables included in the HAS-BLED score to predict bleeding risk if anticoagulated*: liver disease, renal failure, alcohol abuse, history of bleeds. These variables are included in the HAS-BLED score, which is a predictive risk score that aims to inform physicians of the risk of induced bleed, if the patient is anticoagulated. The HAS-BLED score incorporates three variables that we have already included into our predictions of bleed treatment effect heterogeneity, including age, hypertension, and stroke history; we do not consider these variables separately here, since included bleed treatment effects may already depend on these variables. The HAS-BLED also includes a measure of a measure of unstable or high INRs among treated patients, which is not observed prior to treatment, and so not included here. Finally, HAS-BLED score also includes medication usage that predisposes patients to bleeding, such as aspirin or NSAIDS. Unfortunately, we do not consistently observe the use of these medications because they are widely available over the counter, without a prescription.

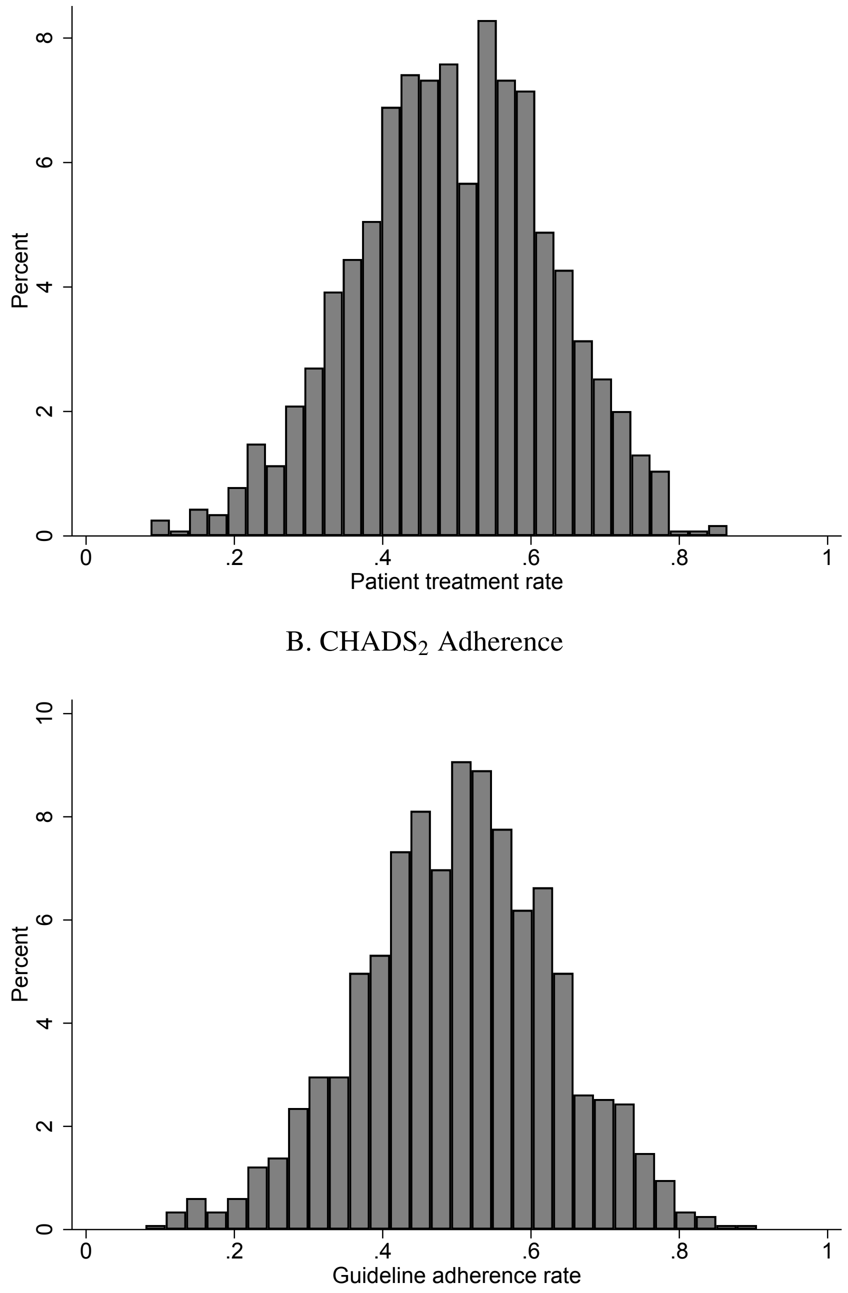
3. *Variables related to frailty and fall risk:* neurologic disorder (including Parkinson's Disease), fall risk (neuropathy, muscle weakness, dizziness), vision problems, arthritis, head injury, fracture. Frailty and fall risk are frequently cited clinical explanations for not prescribing warfarin to patients with high $CHADS_2$ scores. Patients with high fall risk may be more likely to suffer intracranial bleeds if they are taking warfarin.

4. *Elixhauser comorbidities that are not in the AFI database:* HIV/AIDS, deficiency anemia, hypothyroidism, tumor, metastasis, lymphoma, obesity, weight loss, paralysis, pulmonary circulation disorders, ulcer, valvular disease. These are additional patient characteristics that have been shown to predict health care spending and mortality.

5. *Physician characteristics:* doctor specialty code (cardiology, internal medicine, primary care). This specialty coding variable indicates the doctor's training and role at the VHA.

Controlling for these variables in our model estimation does not materially change the conclusions of our analysis. Figure A.4 reports the results of regressions that permute the control variable sets to cover every possible combination of the above list. In Panel A, we find a similar increase in sensitivity to the CHADS_2-component of stroke treatment effects after guideline adoption in each model, regardless of the set of included controls. In Panel B, we show that the unexplained variance in treatment propensity does not change substantially, even after we control for these detailed patient and physician characteristics.

Figure A.1: Distribution of Physician Treatment Decisions

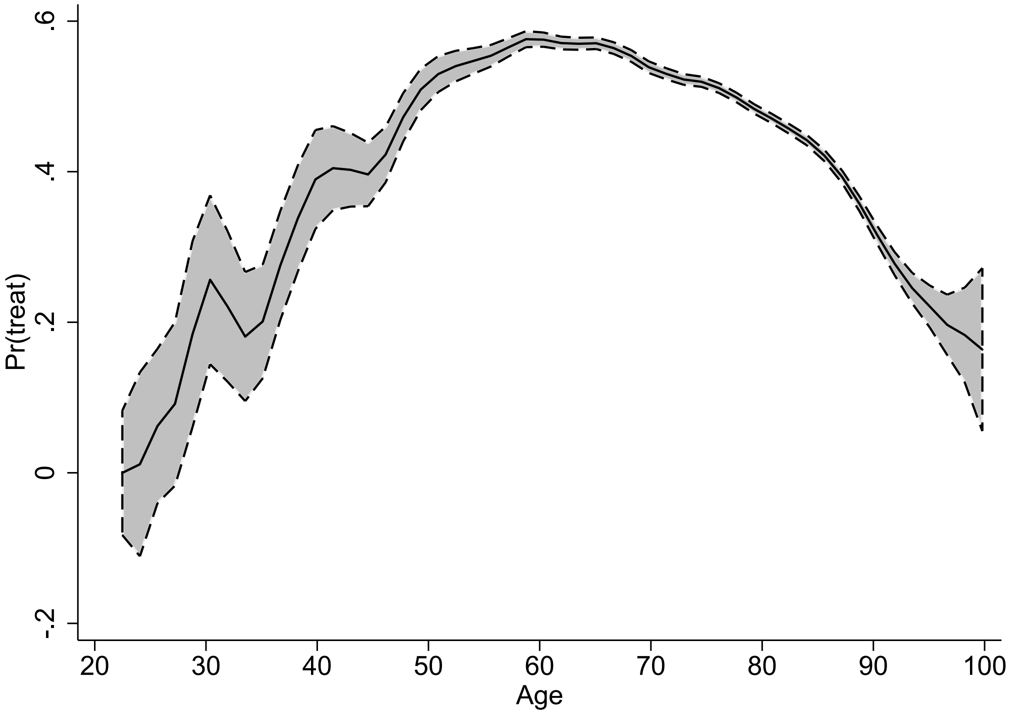


A. Treatment

B. CHADS$_2$ Adherence

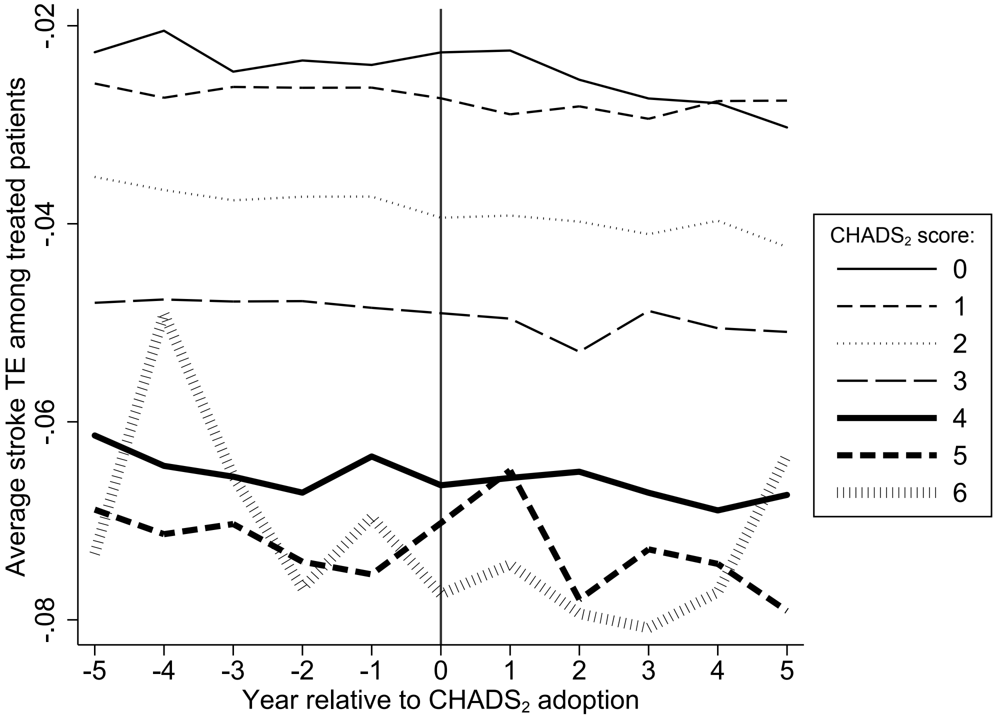
Notes: These figures show the distribution of treatment rates and CHADS$_2$ adherence rates across physicians. They cover the subsample of 1,146 physicians treating at least 30 patients in our final analysis sample. This covers 50,426 patients treated by the higher volume doctors, or a little less than half of the VHA sample defined in Table 1. Panel A shows the distribution of treatment rates. Panel B shows the distribution of CHADS$_2$ adherence rates. We define CHADS$_2$-adherent anticoagulation decisions as follows: No anticoagulation for patients with a CHADS$_2$ of 0 and anticoagulation for patients with a CHADS$_2$ score greater than or equal 2; we omit patients with a CHADS$_2$ score of 1 from this calculation, since the 2008 ACCP guideline allowed for either anticoagulation or aspirin for these patients.

Figure A.2: Treatment Probability by Patient Age



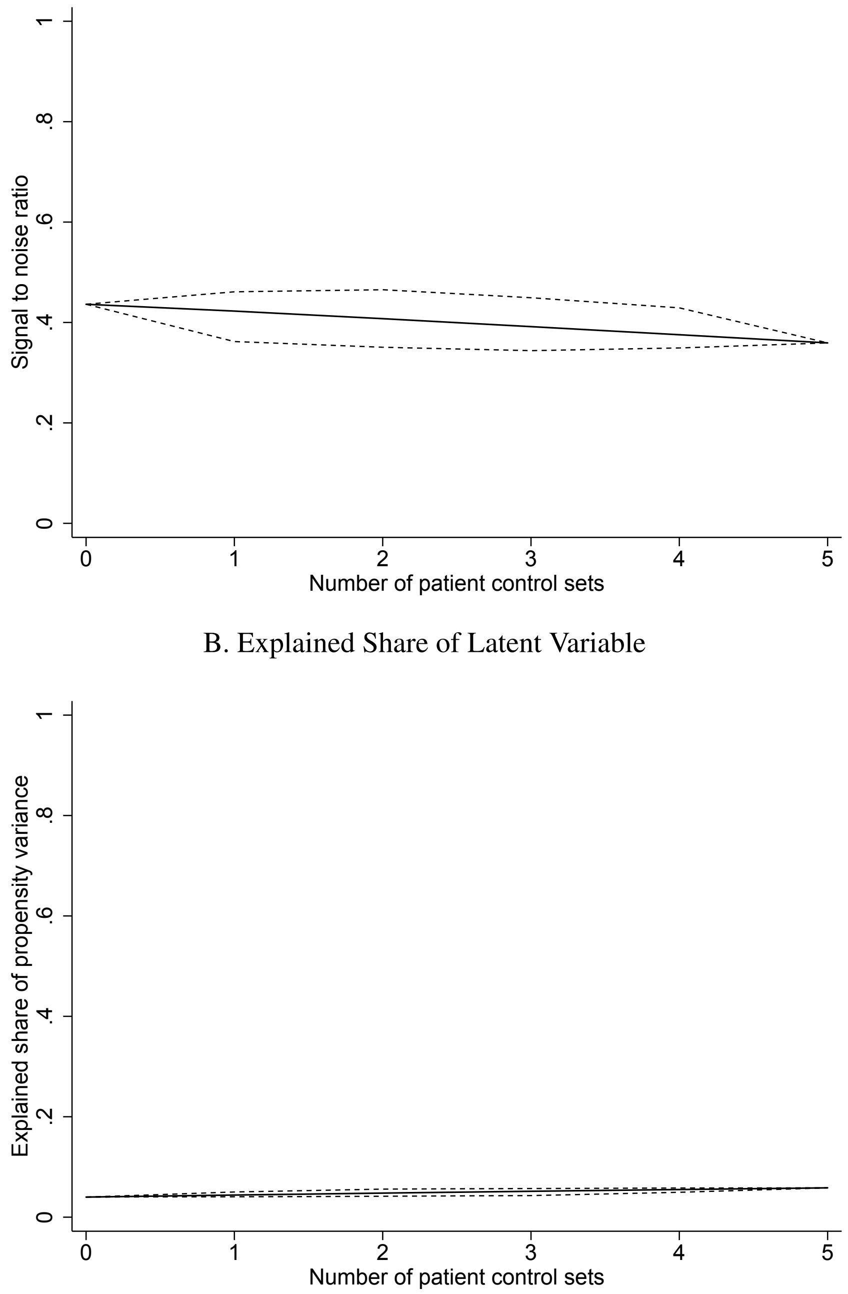
Notes: This figure shows the probability of anticoagulation as a function of patient age in the VHA sample. The curve fits the observed data with a kernel weighted local polynomial; the shaded area represents the 95% confidence interval.

Figure A.3: Stroke Treatment Effects Among Treated Patients



Notes: This figure displays time trends in the average stroke treatment effect among treated patients, separately by CHADS$_2$ score. The physician's first CHADS$_2$ mention occurs on the first day of year 0. Stroke treatment effects are predicted using causal forest rules trained and validated in the AFI database; the causal forest rules are then applied to patient characteristics in the VHA data. Adoption status and treatment decisions are measured in the VHA data.

Figure A.4: Stability of the Structural Results

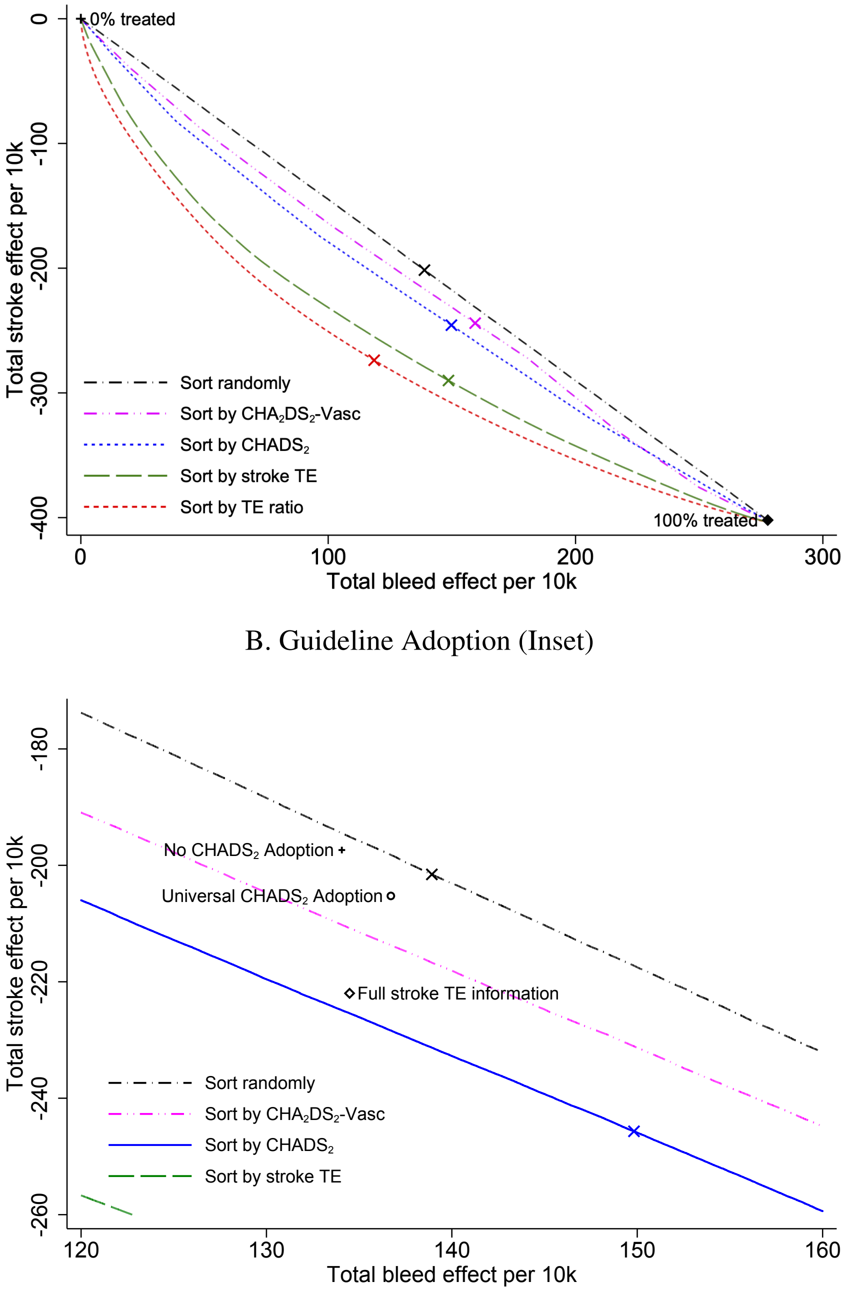


A. Increase in Signal-to-Noise Ratio, $\lambda_{\text{pre}}^{s(c)}/\lambda_{\text{post}}^{s(c)}$

B. Explained Share of Latent Variable

Notes: These graphs illustrate how key results of our structural model vary as we include various sets of control variables in its estimation. Panel A examines the increase in the informativeness of physicians' beliefs about CHADS$_2$-related stroke treatment effects with CHADS$_2$ adoption, or $\lambda_{\text{pre}}^{s(c)}/\lambda_{\text{post}}^{s(c)}$. Panel B examines the proportion of variance in the latent variable that we can explain with observable characteristics (i.e., the complement of the share explained by $\sigma_{\varepsilon,g}^2$). In each panel, we include varying sets of patient characteristics in $f(X_i)$ in our structural model stated in Equation (13). We estimate the baseline specification, shown in Column 1 of Table 4. The solid line shows the mean value of the statistic among specifications with the indicated number of control sets; the top (bottom) dashed line shows the maximum (minimum) of the statistic. The control variables are detailed in Appendix Section A.4.

Figure A.5: Counterfactual Outcomes with CHA_2DS_2-VASc Guideline

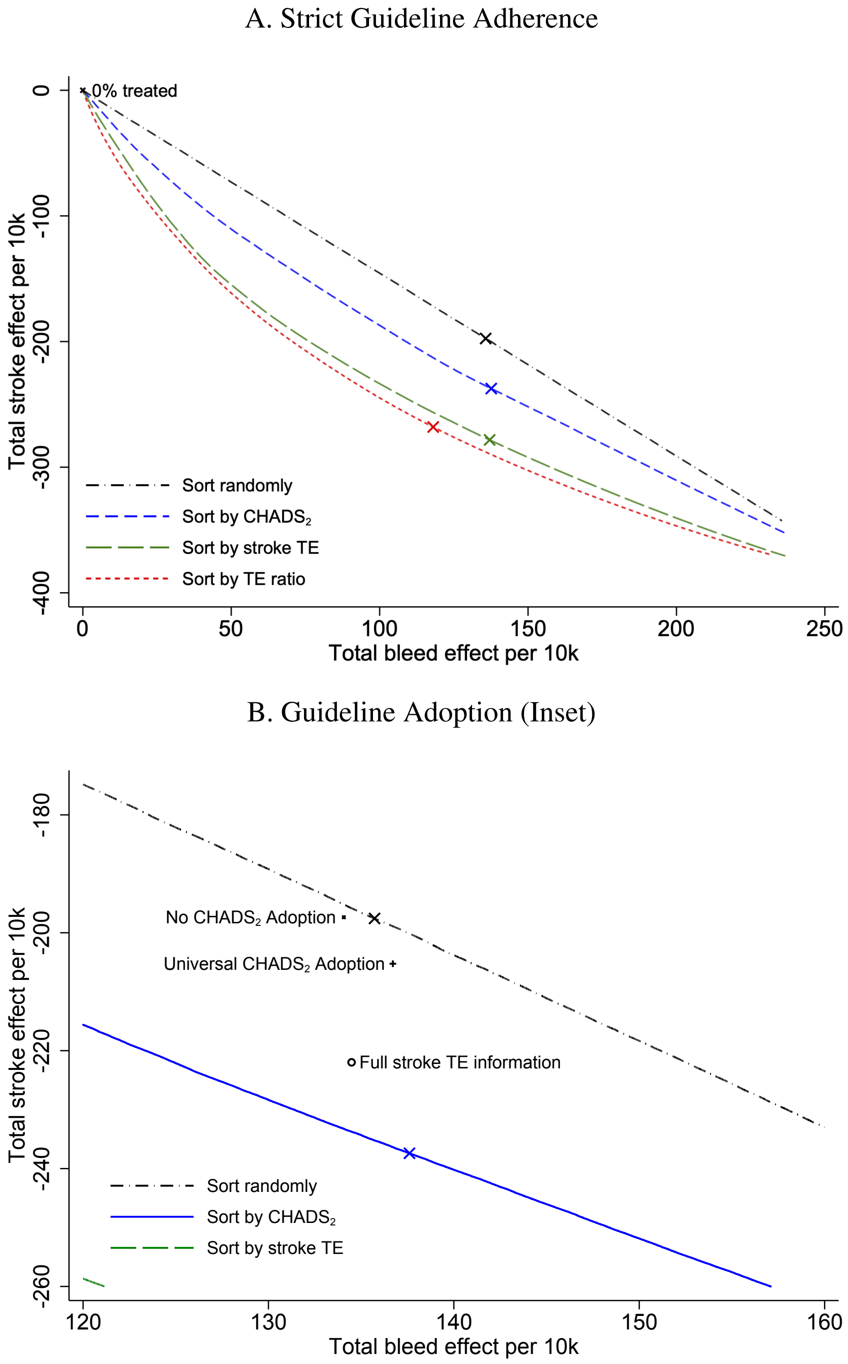


A. Strict Guideline Adherence

B. Guideline Adoption (Inset)

Notes: Relative to Figure 6, this figure includes an additional set of counterfactual outcomes under strict adherence to the CHA_2DS_2-VASc guideline. Like other counterfactuals of strict adherence, strict adherence to this guideline implies ranking patients by their CHA_2DS_2-VASc score. The CHA_2DS_2-VASc score assigns one point for congestive heart failure, hypertension, age 65-74 years, female sex, vascular disease, and diabetes; it assigns two points for age 75 years or older, and for stroke, transient ischemic attack, or thromboembolism. Details for this figure are otherwise described in the notes for Figure 6.

Figure A.6: Counterfactual Outcomes, Fixed Treated Age Distribution

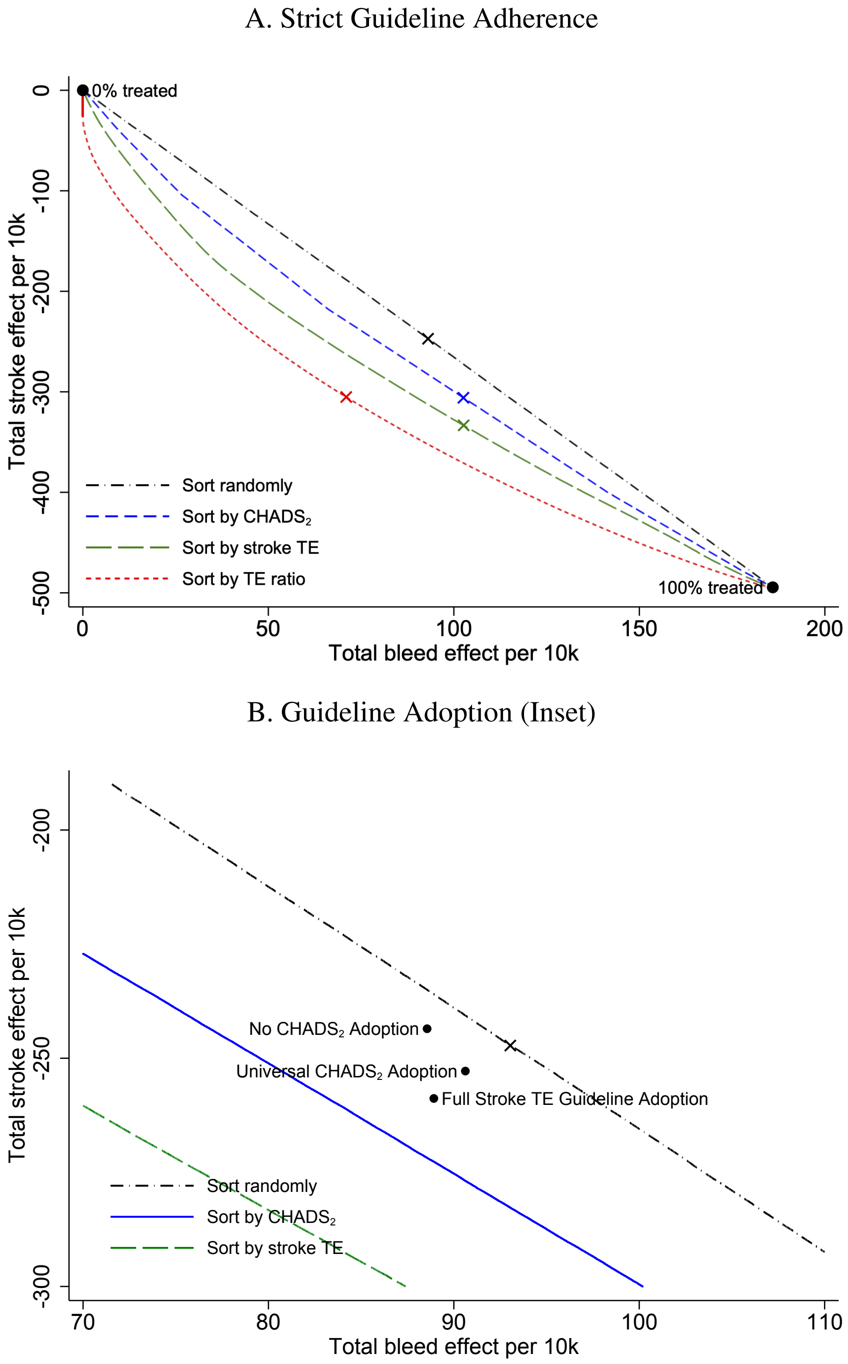


A. Strict Guideline Adherence

B. Guideline Adoption (Inset)

Notes: Relative to Figure 6, this figure shows counterfactual outcomes for patient rankings that hold fixed the age distribution of treated patients at every point along the curve. The fraction of treated patients in each 5-year age bin matches the fraction observed treated in our sample. Within each age group, patients are ranked according to scores in each guideline. In order to maintain a fixed age distribution of treated patients, it is not possible to treat 100% of patients. This is reflected by curves for counterfactual outcomes not reaching the same bottom-right corner of Figure 6. Only counterfactual outcomes for strict adherence and for random sorting are changed in this figure; outcomes for adoption scenarios are unchanged from Figure 6. For more details, see notes to Figure 6.

Figure A.7: Counterfactual Outcomes, Unadjusted Treatment Effects



A. Strict Guideline Adherence

B. Guideline Adoption (Inset)

Notes: Relative to Figure 6, this figure shows counterfactual outcomes that apply the raw stroke and bleed treatment effects from the causal forest instead of the best linear predictor (BLP) adjusted estimates. For this exercise, we bound stroke treatment effects above by 0, trimming the treatment effects for 0.06% of the sample with wrong-signed predictions of stroke treatment effects. We bound bleed treatment effects from below by 0, trimming the treatment effects for 6% of the sample with wrong-signed bleed treatment effects. While wrong-signed bleed treatment effects are more common when we do not apply the BLP adjustment, the point estimates are very close to zero, with the largest magnitude wrong-signed bleed treatment effect being −0.003. For more details, see notes to Figure 6.

Table A.1: AFI Database Covariates

Age
Sex
Height
Weight
Race
Smoker
Systolic blood pressure
Diastolic blood pressure
History of angina
History of congestive heart failure
History of diabetes
History of hypertension
History of myocardial infarction
History of TIA or stroke
Worst prior event: TIA or stroke
Time from last stroke or TIA

Notes: This table lists all the covariates from the AFI database that are used in our causal forest implementation. We set the treatment variable to be 1 for patients treated with warfarin and 0 for patients patients on control or ASA therapy (aspirin). Observations are dropped for those on low warfarin and low warfarin plus ASA therapy. Patients with especially relevant disease histories missing are also excluded from the sample. These disease histories include Transient Ischemic Attack (TIA), stroke, diabetes and hypertension. We regrouped the variable "Race" so that it equals 1 when the patient is white and 0 otherwise. We constructed two binary variables related to smoking history based on the variable "Smoker" in the AFI database, which takes 3 values. The new variables, current smoker and never smoker indicates whether the patient is a current smoker and whether he or she has ever smoked. Otherwise, missing variables are replaced by their mean and mode.

Table A.2: Variable Importance

Stroke Causal Forest	Stroke Regression Forest	Bleed Causal Forest	Bleed Regression Forest
Stroke Risk (−)	Past Stroke or TIA (+)	Bleed Risk (+)	Age (+)
Age (−)	CHADS_2 (+)	Age (+)	Race (−)
Systolic Blood Pressure (−)	White (−)	Race (+)	
Hemoglobin (+)	Systolic Blood Pressure (+)		
Past Stroke or TIA (−)	Age (+)		
Height (+)	Hemoglobin (−)		
CHADS_2 (−)	Height (−)		
White (−)	Past Diabetes (+)		
Past Angina (−)	Past Angina (+)		
Past Diabetes (−)	Past MI (+)		
Past MI (−)			

Notes: Each column of this table shows the important variables for each forest in descending order of importance. Only variables selected with the LASSO procedure are included for the forests. For bleeds, LASSO only selected age and race. Risks computed from regression forest using only the control sample are then used as an input into causal forests. The +/− signs following each variable indicates the sign of its coefficient in a bivariate linear regression with the causal forest output as the dependent variable.

Table A.3: Balance Table

Patient Characteristics	Control Group Mean	Treatment Group Mean	Coefficient
Age	70.9	70.7	0.134
			(0.269)
Congestive Heart Failure	0.27	0.32	-0.004
			(0.012)
Age above 65	0.76	0.77	0.003
			(0.012)
History of Hypertension	0.47	0.50	-0.007
			(0.014)
History of Stroke	0.19	0.15	-0.007
			(0.008)
History of Diabetes	0.15	0.15	-0.007
			(0.010)
Male	0.64	0.66	-0.015
			(0.013)

Notes: This table shows the unadjusted means of each patient characteristics in the treatment and control group. The last column shows results of a regression of each patient characteristics on trial fixed effects and treatment indicator in AFI database. Standard errors are shown in parentheses.

Table A.4: Causal Forest BLP Validation Regressions

	Dependent Variable		
	Stroke		Bleed
	(1)	(2)	(3)
Treatment, W_i	−0.043	−0.041	0.024
	(0.007)	(0.007)	(0.005)
Treatment effect interactions			
$W_i \times \hat{\tau}^{o}_{-j}(X_i)$	1.027		1.149
	(0.242)		(0.324)
$W_i \times \hat{\tau}^{s(c)}_{-j}(X_i)$		0.763	
		(0.340)	
$W_i \times \hat{\tau}^{s(r)}_{-j}(X_i)$		1.301	
		(0.348)	
Outcome mean	0.071	0.071	0.030
Observations	6,707	6,707	6,707
Trial fixed effects	Yes	Yes	Yes
Predicted outcome controls			
$\hat{Y}^{o}_{-j,1}(X_i)$, control group	Yes	Yes	Yes
$\hat{Y}^{o}_{-j,2}(X_i)$, control and treatment groups	Yes	Yes	Yes

Notes: This table reports the coefficients of best linear predictor (BLP) validation regressions of stroke and bleed outcomes on treatment W_i and interactions with treatment effects. Treatment effects are demeaned, so that the coefficient on the treatment indicator W_i reflects the average treatment effect. All specifications control for trial fixed effects and for regression forest predictions of the outcome, $\hat{Y}^{o}_{-j,1}(X_i)$, that are estimated in the control groups of leave-out trials as and analogous predictors $\hat{Y}^{o}_{-j,2}(X_i)$, that are estimated in both control and treatment groups of leave-out trials. Columns 1 and 3 corresponds to Equation (6), which interacts treatment with the full treatment effect, or $\hat{\tau}^{o}_{-j}(X_i)$; Column 2 corresponds to Equation (8), which interacts treatment with the CHADS$_2$ and the residual components of stroke treatment effects, or $\hat{\tau}^{s(c)}_{-j}(X_i)$ and $\hat{\tau}^{s(r)}_{-j}(X_i)$. Standard errors are shown in parentheses.

Table A.5: Patient Characteristics in the VHA Data and Across Trials

	VHA Data	AFI Database Trial							
		AFASAK1	BAATAF	CAFA	SPINAF	SPAF2	AFASAK2	EAFT Group 1	PATAF Group 1
Age	74.0	73.0	67.9	68.0	67.9	70.3	73.7	70.7	70.5
Stroke treatment effect, $\hat{\tau}^s_{BLP}$	−0.05	−0.05	−0.03	−0.03	−0.03	−0.04	−0.04	−0.08	−0.03
Bleed treatment effect, $\hat{\tau}^b_{BLP}$	0.03	0.03	0.02	0.02	0.02	0.02	0.03	0.02	0.02
Congestive heart failure	0.16	0.52	0.26	0.22	0.31	0.21	0.70	0.09	0.00
Hypertension	0.51	0.32	0.51	0.39	0.59	0.53	0.44	0.44	0.36
Age ≥ 65	0.84	0.84	0.67	0.68	0.69	0.73	0.90	0.81	0.79
Diabetes	0.15	0.08	0.15	0.12	0.19	0.16	0.12	0.13	0.17
Stroke history	0.18	0.06	0.03	0.03	0.08	0.06	0.05	0.76	0.00
Male	0.99	0.54	0.72	0.75	1.00	0.70	0.61	0.59	0.46
Number of patients	113,270	1,007	420	378	571	1,100	339	668	272

Note: This table shows the mean of each listed patient characteristic in the VHA data, as well as each of the eight trials with both control and treatment arms in the AFI database. There are a total of 10 trials in the AFI database. In three of the trials, patients are divided into eligible versus ineligible groups for anticoagulation and then randomized within each group. In the estimation of causal forest, we treat these groups as separate trials. AFASAK1: Atrial Fibrillation, Aspirin, and Anticoagulation Study 1; BAATAF: Boston Area Anticoagulation Trial for Atrial Fibrillation; CAFA: Canadian Atrial Fibrillation Anticoagulation; SPINAF: Stroke Prevention in Non-rheumatic Atrial Fibrillation; SPAF2: Stroke Prevention in atrial Fibrillation; AFASAK2: Atrial Fibrillation, Aspirin, and Anticoagulation Study 1; EAFT Group 1: European Atrial Fibrillation Trial; PAATAF Group 1: Primary Prevention of Arterial Thromboembolism in Atrial Fibrillation.

Table A.6: Probit Estimates: Additional Results

	Dependent Variable: Anticoagulant Prescription		
	(1)	(2)	(3)
Adoption status intercepts			
Post-adoption intercept, μ_{post}	−0.296***	−0.274***	−0.232***
	(0.046)	(0.048)	(0.052)
Never-adopter intercept, μ_{never}	0.015	0.016	0.023
	(0.038)	(0.038)	(0.041)
Standard deviation of error term			
Post-adoption, $\ln\left(\sigma_{\varepsilon,\text{post}}\right)$	0.015	0.032	0.115
	(0.077)	(0.080)	(0.0868)
Never-adopter, $\ln\left(\sigma_{\varepsilon,\text{never}}\right)$	0.012	0.013	0.0232
	(0.055)	(0.055)	(0.0576)
Year fixed effects, age spline controls	Yes	Yes	Yes
Differential trends on treatment effects	No	Yes	No
Treatment effects interacted with year identities	No	No	Yes
Number of observations	113,270	113,270	113,270

Notes: This table shows additional results from the specifications reported in Table 4. The probit models allow that $\sigma_{\varepsilon,g}$ may vary with adoption status (post-adoption and non-adopter) and year fixed effects. See notes for Table 4 for additional details. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.7: Probit Estimates: Average Marginal Effects

	Dependent Variable: Anticoagulant Prescription		
	(1)	(2)	(3)
CHADS$_2$-related stroke treatment effect, $\hat{\tau}^{s(c)}_{BLP}(x)$			
Prior to adoption, $\alpha^{s(c)}_{\text{pre}}$	−1.861***	−1.916***	−2.429***
	(0.210)	(0.228)	(0.387)
Post-adoption difference, $\alpha^{s(c)}_{\text{post}} - \alpha^{s(c)}_{\text{pre}}$	−2.404***	−2.369***	−1.903***
	(0.373)	(0.381)	(0.496)
Never-adopter difference, $\alpha^{s(c)}_{\text{never}} - \alpha^{s(c)}_{\text{pre}}$	0.372	0.355	0.440
	(0.282)	(0.294)	(0.506)
Residual stroke treatment effect, $\hat{\tau}^{s(r)}_{BLP}(x)$			
Prior to adoption, $\alpha^{s(r)}_{\text{pre}}$	0.241	0.177	0.153
	(0.121)	(0.128)	(0.698)
Post-adoption difference, $\alpha^{s(r)}_{\text{post}} - \alpha^{s(r)}_{\text{pre}}$	−0.366*	−0.305	−0.288
	(0.188)	(0.191)	(0.283)
Never-adopter difference, $\alpha^{s(r)}_{\text{never}} - \alpha^{s(r)}_{\text{pre}}$	0.017	0.056	0.142
	(0.160)	(0.165)	(0.347)
Bleed treatment effect, $\hat{\tau}^{b}_{BLP}(x)$			
Prior to adoption, α^{b}_{pre}	−1.294***	−1.226***	−1.111**
	(0.321)	(0.353)	(0.568)
Post-adoption difference, $\alpha^{s(r)}_{\text{post}} - \alpha^{s(r)}_{\text{pre}}$	0.974*	0.875	0.700
	(0.517)	(0.541)	(0.695)
Never-adopter difference, $\alpha^{b}_{\text{never}} - \alpha^{b}_{\text{pre}}$	0.421	0.357	0.113
	(0.416)	(0.436)	(0.739)
Year fixed effects, age spline controls	Yes	Yes	Yes
Differential trends on treatment effects	No	Yes	No
Treatment effects interacted with year identities	No	No	Yes
Number of observations	113,270	113,270	113,270

Notes: This table reports average marginal effects from the specifications reported in Table 4. See notes for Table 4 for additional details. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.8: Counterfactual Outcomes, Fixed Treated Age Distribution

	Treated Patients	Strokes Prevented	Bleeds Induced
A: Benchmarks (repeated from Table 5)			
Status quo	49.8%	199	134
Randomly assigned treatment	49.8%	196	135
B: Strict Guideline Adherence, Fixed Treated Age Distribution			
Strict CHADS_2 adherence	48.9%	234	134
Strict adherence to stroke TE guideline	49.1%	275	134
Strict adherence to TE-ratio guideline	55.6%	287	134

Notes: This table reports counterfactual outcomes that hold the fixed the age distribution of treated patients. The fraction of treated patients in each 5-year age bin matches the fraction observed treated in our sample. We also hold the overall percentage of treated patients fixed at 49.8%. Within each age group, patients are treated according to rankings implied by the noted guideline. Figure A.6 shows counterfactual outcomes varying the overall percentage of treated patients. For more details, see notes to Table 5 and Figure A.6.

Table A.9: Counterfactual Outcomes, Unadjusted Treatment Effects

	Treated Patients	Strokes Prevented	Bleeds Induced
A: Benchmarks			
Observed treatment choices	49.8%	245	89
Randomly assigned treatment	49.8%	246	93
B: Guideline Adoption			
No CHADS$_2$ adoption	49.7%	244	89
Universal CHADS$_2$ adoption	50.2%	252	91
Universal adoption of stroke TE guideline	49.8%	256	89
C: Strict Guideline Adherence, Fixed Bleeds Induced			
Strict CHADS$_2$ adherence	43.1%	273	89
Strict adherence to stroke TE guideline	43.4%	304	89
Strict adherence to TE-ratio guideline	58.5%	344	89

Notes: This table reports counterfactual outcomes that apply the raw stroke and bleed treatment effects from the causal forest rather than the best linear predictor (BLP) adjusted estimates. For this exercise, we bound stroke treatment effects above by 0, trimming the treatment effects for 0.06% of the sample with wrong-signed predictions of stroke treatment effects. We bound bleed treatment effects from below by 0, trimming the treatment effects for 6% of the sample with wrong-signed bleed treatment effects. While wrong-signed bleed treatment effects are more common when we do not apply the BLP adjustment, the point estimates are very close to zero, with the largest magnitude wrong-signed bleed treatment effect being −0.003. For more details, see notes to Table 5 and Figure A.7.