NBER WORKING PAPER SERIES

OVERCONFIDENCE AND TECHNOLOGY ADOPTION IN HEALTH CARE

Diego A. Comin
Jonathan S. Skinner
Douglas O. Staiger

Working Paper 30345
http://www.nber.org/papers/w30345

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2022

We are grateful for financial support from the National Institute on Aging (P01-AG19783 and U01-AG046830). In addition, we are indebted to Kimon Bekelis, Lauren Gilstrap, Peter Groeneveld, Edward Norton, Greg Roth, and to seminar participants for very helpful comments and suggestions. Weiping Zhou provided essential programming and analysis. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w30345.ack

Overconfidence and Technology Adoption in Health Care
Diego A. Comin, Jonathan S. Skinner, and Douglas O. Staiger
NBER Working Paper No. 30345
August 2022
JEL No. I1,O3,O33

ABSTRACT

Variation in technology adoption is a key driver of differences in productivity. Previous studies sought to explain variations in technology adoption by heterogeneity in profitability, costs of adoption, or other factors. Less is known about how adoption is affected by bias in the perceived skill to implement the technology. We develop a Bayesian framework in which the use of the technology depends on perceived skill, while the outcomes from using it depend on actual skill. We study the determinants of adoption in the case of implantable cardiac defibrillators (ICDs) for which we document large differences across hospitals in the rate of adoption between 2002-2006, and a strong reversal from 2006-2013. We find that perception bias explains two-thirds of the cross-hospital variation in ICD use. A dynamic version of the model with learning about bias predicts accurately the subsequent decline in ICD use between 2006-2013. These results suggest an important role for misperception in explaining the wide variation in the adoption of new technologies.

Diego A. Comin
Dartmouth College
Economics Department
6106 Rockefeller Hall, Room 327
Hanover, NH 03755
and CEPR
and also NBER
diego.comin@dartmouth.edu

Douglas O. Staiger
Dartmouth College
Department of Economics
HB6106, 301 Rockefeller Hall
Hanover, NH 03755-3514
and NBER
douglas.staiger@dartmouth.edu

Jonathan S. Skinner
Department of Economics
6106 Rockefeller Hall
Dartmouth College
Hanover, NH 03755
and NBER
jonathan.skinner@dartmouth.edu

1. Introduction

Technology is the key determinant of long-run living standards, and a better understanding of the determinants of technology adoption is central to improving productivity (e.g., Comin and Hobijn, 2004, 2009, 2010; Skinner and Staiger, 2015, Comin and Mestieri, 2018). Concerns about lagging adoption rates that slow productivity growth have largely centered around the role of heterogeneity across potential adopters in the value of the new technologies (Griliches, 1957; Comin and Hobijn, 2007; Caselli and Coleman, 2006; Chandra and Staiger, 2007; Jovanovic and Nyarko, 1996; Suri, 2011), and in the cost of implementing, using or managing the new technologies (Caselli, 2004; Skinner and Staiger, 2007, Foster and Rosenzweig, 1995; Conley and Udry, 2010; Duflo et al., 2011; Rogers, 2010).

In contrast, much less attention has been devoted to perception biases, or gaps between beliefs and reality in the value or the costs of new technologies, as potential sources of the large variation observed in technology adoption patterns. Overconfidence (and underconfidence) may have important effects on adoption patterns as ultimately, it is adopters' expectations that determine whether to adopt or not.[1] Even in the case where all potential adopters have identical abilities, access to technology, and face identical prices, differences across adopters in their perceptions about these variables will generate variation in the timing or intensity of adoption. Variation in adoption due to misperceptions leads to allocative inefficiency in the use of the technology; the pervasiveness of misperceptions therefore has important implications for welfare and productivity.

The neglect of perception biases in the context of technology adoption may be due to the inherent difficulty of separately identifying perceived and actual net values of a new technology for adopters. The main goal of this paper is to develop a framework that uses variation in technology adoption and outcomes conditional on adoption to disentangle two structural parameters: the true value of the technology for the adopter, and the adopter's misperception of its true value. We present and apply the framework in the context of a specific technology, implantable cardioverter defibrillators (ICDs). These are expensive medical devices whose

[1] Two related literatures are (a) CEO overconfidence in supporting innovation (Galasso and Simcoe, 2011), and (b) an agent's inability to implement the optimal adoption decision due to time-inconsistency and lack of access to commitment devices (Duflo, Kremer, and Robinson, 2008). Time-inconsistence is unlikely to be an issue in our context as doctors do not bear the economic cost of adopting the technology.

purpose for patients with congestive heart failure (CHF) is to reduces the risk of sudden cardiac arrest, and thus extend survival. We use a 100% fee-for-service cohort of Medicare enrollees aged 65 and over with CHF for each year from 2002-2013 to document the following: (1) a rapid average diffusion between 2002 and 2006 in the use of ICDs for patients with CHF, but with wide variation in growth across regions; and (2) a decline in the average ICD rate after 2006 with regions with higher initial (2006) ICD use rates more likely to scale back on their use.

We then match to the Medicare data a comprehensive national 2006-13 ICD registry with detailed clinical information about every patient receiving an ICD. We used the combined data to find that (3) hospitals with the highest rate of ICD use exhibited a larger fraction of patients outside clinical guidelines; (4) there was wide variability in ICD mortality rates across hospitals; (5) hospitals with larger ICDs use rate exhibited higher conditional (risk-adjusted) mortality; and (6) average mortality rates showed only a modest decline between 2006-13.

In our theoretical framework, Bayesian physicians decide whether patients should receive an ICD implant based on their imperfect assessment of both the patient's type and their own skill in implanting ICDs. A key prediction of the model is that the ICD use rate depends on the *perceived* level of physician skill, while the outcome from the implant, which in our case is the conditional mortality rate, depends on the *true* skill of the physician. These predictions provide an empirical strategy to identify the physician's true and perceived skill from data on ICD use and conditional mortality; a hospital adopting ICDs rapidly despite worse patient outcomes is consistent with the hospital's physicians being overconfident about their skill. (While there may be more than one ICD-capable physician in a given hospital, we are limited empirically to hospital-level measures of utilization and outcomes.)

We implement this empirical strategy in our data and identify, for each hospital and year, the level of true and perceived skill and consequently the gap between the two which we label physician bias. We find that most (but not all) physicians are overconfident, in the sense that their perceived skill exceeds their actual skill. The structural model, in addition to helping to identify these parameters, allows for counterfactual analyses. These imply that variation in overconfidence explains two-thirds of hospital variation in the ICD use rate. In turn, misperception leads to less appropriate patients receiving ICDs, which attenuates the net benefits of the new technology by 40 percent.

We extend the model to explore the nature of physician biases and the drivers of its evolution. We consider three possible sources of bias: (i) an individual doctor's overconfidence, (ii) a common bias across all doctors in the perceived net value of ICD use in the relevant population, and (iii) a fixed private financial benefit for the doctor to implant ICDs (sometimes referred to as supplier-induced demand). We allow doctors to learn individually about their true skill in implanting ICDs, and collectively about the population-wide net value of ICDs. The model yields an intuitive framework to identify these three possible sources of physician bias. For ICDs during this period, private benefits did not change over time and so introduce a time-invariant hospital-specific intercept. Learning about common bias affects all hospitals equally and introduces a time-specific intercept. In contrast, learning about true skill introduces mean reversion in physician bias. That is, overconfidence attenuates over time leading to a reduction in the average level and dispersion of ICD use across hospitals.

The estimates show that there is significant learning about pure skill, with an average half-life in hospital's misperceptions about their true skill of about 3 years. Estimates suggest that about one-third of the average initial physician bias is due to the doctor's private benefit (e.g., supplier-induced demand) while the remaining two-thirds is due to physician's being initially overconfident about their skill in performing ICDs. Using the estimates of the learning model over the cross-section of hospitals during the first year, we conduct an out-of-sample forecasting exercise that provides the levels of perceived skill in each hospital in 2013 predicted by the learning model. This exercise shows that learning predicts a reduction in the average ICD use rate across hospitals equal to two-thirds of the observed decline in the data. Furthermore, doctors learning about true skill more than accounts for the reduction in cross-hospital variance in ICD use observed in the data. Across hospitals, the out-of-sample model forecast of the change in ICD use is strongly correlated (0.68) with the actual (2006-2013) change.

Our work builds on earlier papers by Chandra and Staiger (2010, 2020) that separately identify procedure skill from bias in determining the treatment of heart attacks, and by Abaluck et al. (2016) and Mullainathan and Obermeyer (2022) that use similar methods to study inefficiency in testing in the emergency room. Like our paper, these papers use both variation in patient outcomes and rates of procedure/testing use across hospitals to separately identify skill from bias. Our paper differs from these earlier papers in three regards. First, our study focuses on misperception in the adoption and diffusion of a new technology – or more specifically, the use of an existing

technology – ICDs – for an entirely new population of people with congestive heart failure that was only approved in the mid-2000s. Second, whereas earlier papers used a more reduced-form methodology, we develop a structural empirical model related to Olley and Pakes (1996) who identify firm-level total factor productivity (TFP) in the presence of endogenous inputs choice and selection by inferring firm's expected TFP from the firm's investment behavior. Though in a very different setting, our framework shares with Olley and Pakes the insight that agents' expectations about key unobserved parameters (skill in our case, TFP in theirs) can be identified by using the model structure and observed choices (ICD use in our case, investment in theirs). Finally, our approach provides an empirical framework for the analysis of data generated by clinical registries where the sample is typically limited to only those who received the treatment.

Other papers have considered instead variation across physicians in diagnostic expertise (the ability to choose the right patients for treatment). Currie and McLeod (2017) study the role of diagnostic expertise in the use of Cesarean procedures (C-sections), where the ability to diagnose appropriate patients varies across physicians; patients of physicians with better diagnostic abilities experience better outcomes. Similarly, Chan et al. (2022) consider diagnostic expertise in the context of radiologists diagnosing pneumonia. In contrast, we model heterogeneity across doctors in their skill at treating all patients and their biases about their perceived skill. All these papers find evidence of productive inefficiency in diagnosis that is consistent with (but not unique to) our model of misperceptions in treatment choice, but none consider bias in the perception of procedural skill (e.g., the ability to perform the procedure, as in Birkmeyer et al., 2013) and the associated misallocation of patients to procedures, nor do they consider the potential for dynamic learning.

More generally, the literature has studied dynamics in the use of medical procedures and outcomes as dependent on two distinct mechanisms. One strand has emphasized the role of clinical learning-by-doing (e.g., Jovanovic and Nyarko, 1995; Gong, 2017). The other has focused on the general perception about the net value of new technologies or procedures, documenting both cases where the general perception was initially too optimistic (Jupiter and Burke, 2013), and too pessimistic (Currie, MacLeod and Van Parys, 2015; Wu and David, 2022) relative to its long-run level. Our framework allows both for variation over time in true skill and in the economy-wide perception of the net value of the technology. However, we do not find that they played a significant role in the evolution of ICD use. Instead, the strong mean reversion in hospital-level overconfidence and the simultaneous declines in cross-hospital variation in ICD use points to the

importance of physicians learning about bias and overconfidence and revising their practice accordingly.

Finally, our framework and findings are related to various strands of the broader literature on overconfidence. Overconfidence has received considerable attention in psychology (e.g., Moore and Healy, 2008), finance (Barber and Odean, 2001; Glaser and Weber, 2007; O'Neill, Pouder; and Buchholtz, 1998), CEO decisions regarding the exercising of stock options, firm investments, and innovation (Galasso and Simcoe, 2011; Malmendier and Tate, 2015), market entry in industrial organization (Camerer and Lovallo, 1999), and in health care regarding diagnostic error or beliefs not supported by clinical evidence (Berner and Graber, 2008; Cutler et al. 2019). To our knowledge, perception bias has not been explored in the context of technology adoption and diffusion.

The rest of the paper is organized as follows. Section 2 describes how we have assembled our data set and documents the key facts about the evolution of the use of ICDs and the outcomes after an ICD implant across US hospitals. Section 3 presents our model. Section 4 contains our analysis, while Section 5 concludes.

2. Implantable Cardioverter Defibrillators (ICDs)

Congestive heart failure (CHF) is a common illness among the elderly with an estimated prevalence of 6.2 million people in the U.S (Virani, 2020). It is very different from acute myocardial infarction (heart attack). While heart attacks are sudden medical emergencies treated (often successfully) with a variety of medical interventions, CHF is a chronic illness whose progression can only be slowed by appropriate medical management. The severity (and hence progression) of CHF has been categorized by New York Heart Association to range from Class I (the least severe) through to Class IV (the most severe), at which point the annual mortality rate is as high as 20-50 percent (Ahmed et al., 2006).

An important risk facing CHF patients is a sudden cardiac arrest, which occurs when the heart suddenly stops functioning, typically because of arrhythmia, or irregular heart rhythm. This causes rapid and unsynchronized heartbeat, leading to little or no blood being pumped from the heart, and absence of a heartbeat (van Reys, 2014). Implantable cardioverter defibrillators (ICDs) are small electronic devices that are surgically implanted in the pectoral region of the chest and connected with wire "leads" to key locations of the heart. These leads serve two functions. The

first is to monitor the rhythm and detect tachycardia (irregular or weak heart beats), and the second is, when necessary, to shock the heart with a strong electrical current, effectively "rebooting" the conduction system. (TV series and movies often show physicians using paddles to administer electrical shocks;[2] ICDs are internal automated versions.) Over time ICDs have become more effective and entailed fewer complications as the size of the ICD shrunk, and the sophistication of the computer programs designed to detect arrhythmias improved.

Initially, ICDs were developed in the 1980s and 1990s for people who had already experienced and survived a cardiac arrest but were at risk of experiencing another one. As ICDs became more compact and reliable, attention turned to the larger group of people with congestive heart failure (CHF) also at risk of cardiac arrest but who had not yet experienced the life-threatening event; for these patients the ICD is deemed "preventive." A large 2005 randomized trial, SCD-HeFT, found substantial mortality benefits of up to 7 percentage point increases in survival 5 years after the procedure (Bardy et al., 2005). Soon after the SCD-HeFT trial, ICDs were allowed by Medicare in the U.S. to be used as a preventive device for patients with weakened hearts (congestive heart failure, or CHF) who had not yet experienced a cardiac arrest, thus expanding dramatically the population of those eligible for ICDs. In this case, "adoption" is the expanded use of an existing technology for an entirely new population, rather than a brand-new technology. We use the Medicare claims data linked to a Centers for Medicare and Medicaid Services (CMS) clinical registry of every ICD implanted during 2006-13 with detailed information on key clinical variables that characterize both appropriateness for treatment, and subsequent risk of mortality.

The SCD-HeFT trial included only the intermediate Class II and Class III CHF patients with low "ejection fractions" or the heart's ability to pump blood to the rest of the body.[3] The reason why the trial was limited to only these two groups was the consensus that for Class I (the least serious) CHF patients, the risks outweighed potential benefits given the rarity of sudden cardiac arrest in this group versus the risks of broken leads or infections, while for the more severe Class IV patients, the heart is so weakened that it can no longer sustain pumping, no matter how

[2] An example from CSI: New York: https://www.youtube.com/watch?v=_lJCDrYxK9A

[3] As well, the ejection fraction should be 35% or less in patients with Class II or III Heart Failure. Despite the rarity of older patients in the randomized trials, there are no guidelines that recommend against the use of ICDs because of age.

many times it reboots. For Class IV patients, ICDs can lead to a series of successive and painful shocks, sometimes delaying an otherwise peaceful demise as the ICD continues to go off until the batteries are drained (Friedrich and Bohm, 2007). Despite these guidelines, a small fraction of ICD procedures were done for those with either Class I or Class IV patients, or for those who had been diagnosed with CHF only recently, and thus have not yet tried medical management. In our analysis, we adjust for these different characteristics, but do not address the more complex problem of whether higher-quality physicians should be more or less likely to follow guidelines.[4]

Finally, to understand the growth and subsequent reduction in the use of ICDs, it is important to rule out the development of a new technology that might have led to a shift away from ICDs. While during the 2000s, there was increased adherence to guideline-directed drug prescriptions (e.g., Roth et al., 2016), there was no innovation or breakthrough developed to reduce mortality among CHF patients that would have caused physicians to replace ICDs during the period of analysis (Kolata, 2017).

2.1 Measuring ICD use in the Medicare Population

To study the evolution of ICD use, we use a 100% fee-for-service cohort of Medicare enrollees age 65 and over diagnosed with CHF (based on Hierarchical Condition Category) for each year from 2002 to 2013. These are created both to illustrate the wide differences in the diffusion of ICDs for CHF patients, and to derive regional utilization rates that can then be assigned to hospitals (as described below). We include only new ICD implantations during 2002-13 and thus exclude replacement ICDs because of failed batteries or other reasons.[5] For these cohorts of CHF patients, we estimate rates of ICD use at the hospital referral region (HRR) level, of which there are 306 in the U.S.[6] These utilization measures are based on the residence of the patient; if a resident of the Memphis HRR received their ICD in Atlanta, the ICD would be assigned to the Memphis HRR rather than to Atlanta.

[4] In the context of our model below, it is possible that higher-skill physicians could still gain good outcomes even for out-of-guideline patients, although the evidence for this is weak (e.g., Abaluck et al., 2020). During the period of analysis, CMS cracked down on hospitals billing for out-of-guideline patients.

[5] We begin the analysis using the claims data in 2002, when the sample of Part B claims data relevant for analysis is 20% of all fee-for-service enrollees; the sample rises to 40% in 2003-05 and becomes 100% thereafter. We use CPT 33249 rather than in-hospital DRG codes to measure incidence.

[6] HRRs were first developed by the Dartmouth Atlas Project in the 1990s to create regions based on the migration patterns of individuals to their hospitals. Thus, HRR boundaries will often follow (e.g.) interstate highways and cross state lines. Each HRR includes a major tertiary hospital that performs neurosurgery and cardiac surgery. We use HRRs rather than the smaller hospital service areas (HSAs) for better sample precision.

In addition to restricting the sample to CHF patients, we also include a variety of other risk adjusters to control for more severe CHF or other comorbidities that could affect the likelihood of ICD placement. For this reason (and for consistency with the theoretical model), we used a year-specific probit risk-adjustment model with HRR-level fixed effects. We include as risk-adjusters individual five-year age brackets (with a category of 85+ for older patients), sex, race/ethnicity (black, white, and other), and dual eligibility with Medicaid (an individual indicator of serious illness, poverty, or both). At the ZIP code level we included poverty rates and income (from the 2010 Census) and at the county level smoking, obesity, and diabetes based on Behavioral Risk Factor Surveillance System (BRFSS) data; these latter health behavior measures are highly predictive of regional mortality rates (Wennberg et al., 2014). Risk-adjusted population-based rates of ICD use for each HRR for each year were calculated as the predicted ICD rate from this probit for an average CHF patient.[7]

The regression estimates are presented in Appendix Table B.1 for three selected years (2002, 2006, and 2013). In all years, the regressions indicate that the county health indicators are not strongly predictive of receiving an ICD.[8] While individual characteristics such as age are important predictors at the individual level, average age across HRRs conditional on Medicare enrollment varies little so that the correlation between raw and fully-adjusted ICD rates at the HRR level is 0.96. Finally, the regression coefficients suggest a narrowing in racial disparities for the use of ICDs; by 2013 there are no meaningful differences in risk-adjusted ICD use across the three racial/ethnic groups.

In Figure 1, we present risk-adjusted rates of newly implanted ICDs per 100 CHF patients by HRR between 2002-13 for the U.S., and for selected regions, with an emphasis on the regions adopting most rapidly. At the national level (shown as a dashed line in Figure 1), ICD use among this sample nearly doubled, from 0.8 percent in 2003 to 1.4 percent in 2005, before a gradual decline to 1.0 percent in 2013.

[7] The "average" CHF patient is defined as a patient with average mortality risk (rather than a patient with average characteristics) to ensure in this nonlinear model that the average predicted rate in a given year was equal to the actual average. See Appendix A for more details.

[8] Recall that the sample is for people who have already been diagnosed with CHF, while the rates of health behaviors are for the entire populations. Similar results were found when we included the entire sample of Medicare enrollees, including (in addition to the variables above) a dummy variable for diagnosed CHF.

As suggested by Figure 1, there was widespread variability in rates of diffusion. Two of the most rapid adopters were Terre Haute IN, which rose 4-fold in two years, from 0.8 percent in 2003 to 3.3 percent in 2005, and Wilmington NC, which appears to have risen by a similar amount (although the 2002 rate is suppressed under CMS rules because fewer than 11 ICDs were reported in the sample). By contrast, many larger metropolitan regions exhibited much smaller increases; Los Angeles and Providence RI, for example, remained consistently below the national average (Figure 1). While both Terre Haute and Wilmington also experienced a decline in later years, other regions such as Harlingen TX exhibited steady growth during the entire period.

Figure 2 provides a map for the entire U.S. of 2006 ICD utilization rates by HRR. This figure confirms the geographic disparity in the use of ICDs across the entire U.S., with a 10-fold difference between Olympia WA (0.28 percent of CHF patients) and Terre Haute (2.9 percent in 2006).[9] Furthermore, there is considerable variation even within states. Figure 3 plots the change in hospital-level ICD use rate between 2002 and 2005 (x-axis), and between 2006 and 2013 (y-axis). The population-weighted cross-hospital correlation between initial and subsequent change in ICD rates is negative (-.38, $p < .001$), showing that the hospitals that adopted ICDs more intensively also experienced the most subsequent rapid decline.

One hypothesis for why ICD use declined gradually following an initially high rate (as we observe in Figures 1 and 3) is a "stock-flow" model; the stock or backlog of patients newly eligible for an ICD could have led to an uptick in utilization for 2006, followed by a dissipation in the backlog with subsequent utilization rates more closely matched the steady-state flow of newly eligible patients. As discussed in more detail below, the ICD registry data includes the duration of the CHF prior to the ICD, but there was no evidence of a diminishing backlog.[10]

While population-based rates of ICD utilization are drawn from HRRs, we seek to estimate our model at the level of the hospital that performs the ICD.[11] We do this by assigning to each

[9] One might be concerned with small-sample bias in these relatively small HRRs, but the patterns show a strong temporal trend; high rates in 2006 are not an outlier relative to previous or subsequent years. Instead, the large variation in smaller areas is most likely because just one or two physicians are the key decision-makers for ICD implantations.

[10] In 2006, 72 percent of those receiving an ICD had had CHF for at least 9 months; by 2013, the fraction of ICD recipients with long-term CHF had risen to 82 percent.

[11] Measures of ICD intensity in utilization requires both a numerator (the number of ICDs implanted in a given year) and a denominator (the number of potential patients). While regions are well suited to calculate both numerator and denominator (e.g., as done in the HRR-level analysis above), calculating the denominator of a given hospital, particularly in a city with multiple hospitals, is difficult.

patient their HRR-level utilization measure. For example, if a hospital in the Boston area draws from the Boston, Providence, and Portland ME HRRs for their ICD patients, the hospital-specific rate of ICD utilization will be a weighted average of those three HRR rates; this is shown in a schematic in Figure 4a.

2.2. Variation in Health Outcome Following ICD Implantation

In the study of technology diffusion, it is rare to measure accurately the performance of adopters after adopting the technology. When CMS approved the use of ICDs for preventive purposes, it was done with the understanding that hospitals would send detailed clinical information about the patient to CMS. We use this 100% registry, linked to the Medicare denominator file for people age 65+, during 2006-13, which allows us to calculate mortality rates based on Medicare denominator files available through 2015. The registry includes detailed information on each patient that includes whether the ICD was for patients with CHF, their risk class (I through IV) as well as ejection fraction and many other clinically relevant factors such as ventricular tachycardia, family history of cardiac arrest and the length of time diagnosed with CHF; importantly we know the hospital performing the procedure.[12] These data are far more detailed than what could ever be recovered from Medicare billing claims.

We link the data to the Medicare denominator file to measure one-year and two-year mortality rates. We view mortality as the apposite outcome because ICDs provide no other benefit to patients other than to "reboot" the heart in the case of sudden cardiac arrest. To estimate mortality rates, we focus on a relatively homogenous group of CHF patients who have never previously had an ICD implanted. Ideally, we would like to measure true treatment effects; the benefit of an ICD relative to the status quo of medical management for CHF. However, because registry data is lacking for patients not receiving an ICD, our estimates are specific to mortality rates only among those treated.[13] As well, we focus solely on hospital-level measures of outcomes; we recognize that some hospitals may include more than one cardiologist or electrophysiologist

[12] One complexity associated with identifying hospitals is that in some cases, the hospital was not identified; only the NPI for the provider who performed the procedure. We are grateful to Andrea Austin for providing a crosswalk from ICD-capable providers to the hospital where they performed the plurality of procedures.

[13] The modeling in Section 3 and analysis in Section 3 derive predictions based on these estimates using the registry data, supplemented by treatment effects of ICDs estimated based on the SCD-HeFT randomized trial (Bardy et al., 2005).

who performs the procedures but identifying pure physician effects from the hospital-level team that both implants and maintains the ICD is problematic.

Table 1 provides summary statistics of the ICD sample (N = 238,059). The average age among the Medicare enrollees (all of whom are 65+) is 74.9, and just 28 percent are female. We also include summary statistics for additional covariates from the registry, including the ejection fraction, prior cardiac arrest, family history, prior heart attack, and other variables. Hospital-level risk-adjusted mortality is modeled using the following hierarchical structure:

$$M_{jit} = \Psi_{it} + X_{jit}\beta + \zeta_{jit} \quad (1)$$

$$\text{where } \Psi_{it} = ICD\_rate_{it}\Gamma + \theta_i + v_{it} \quad (2)$$

The first equation is at the patient level, where mortality (M_{jit}) for patient j treated at hospital i in year t is a binary variable that depends on characteristic of the patient (X_{jit}), and a hospital effect (Ψ_{it}). At the hospital level, the hospital effect in turn depends on the hospital-level risk-adjusted utilization rate of ICDs in that year (ICD_rate_{it}) plus a random hospital effect (θ_i) and a random hospital-year effect (v_{it}). We allow the mortality hospital effect Ψ_{it} to depend on the utilization rate of ICDs to estimate a reduced-form correlation between ICD aggressiveness and risk-adjusted mortality. As well, we are particularly interested in the variance of Ψ_{it} and its covariance with the hospital's ICD utilization, which depends both on the predictable characteristics of the hospital, Var($ICD\_rate_{it}\Gamma$), and the provider-specific error term Var(θ_i).

Our preferred specification is a hierarchical random-effects model, which provides estimates of the key parameters ($\Gamma, Var(\theta_i)$) and also estimates of the individual hospital effects, $\hat{\Psi}_{it} = ICD\_rate_{it}\hat{\Gamma} + \hat{\theta}_i + \hat{v}_{it}$, where we use best linear unbiased predictions for the hospital and hospital-year random effects (e.g., use Empirical Bayes to "shrink" the estimate of the provider residual towards the fitted value $ICD\_rate_{it}\hat{\Gamma}$ depending on the sample size of the provider). We focus on random-effects models, but in sensitivity analyses we also consider least-squares regressions and models with provider-level fixed effects. Because we wish to estimate the hospital-specific effect on mortality of ICD relative to medical management, in some specifications we add hospital-level controls (to X_{jit}) to proxy for quality of medical management such as patient volume and the use of guideline-consistent medical treatment for CHF patients.

The benefits inherent in ICD implantation arise only after several years (Bardy et al., 2005) so we estimate the conditional hospital effect (Ψ_{it}) on both 1-year and 2-year mortality using the random-effects model (Equations 1 and 2). Figure 5 shows the distribution across hospitals in conditional 2-year mortality rates. The standard deviation of the 2-year conditional mortality rate across hospitals is 3.0 percentage points (relative to a mean of 22 percentage points), with mortality rates in high-mortality hospitals nearly twice as high as those in low-mortality ones.[14] Similar variation is observed in 1-year conditional mortality rates, with a standard deviation of 2.2 percentage points relative to a mean of 12 percent. Finally, Figure 6 presents the evolution of 1-year and 2-year mortality over time for the U.S., with both showing slight declines between 2006 and 2013 of less than half a percentage point. Thus, in the early years when ICD use was higher there was also higher mortality rates among those patients receiving an ICD.

2.3 The Association of ICD Diffusion with ICD Patient Characteristics

The types of patients being treated with ICDs in the top quartile of hospitals with the highest ICD utilization appear different from patients treated in the bottom quartile of hospitals with the lowest ICD utilization (Table 2). Relative to hospitals with low ICD utilization, patients treated in hospitals with high ICD utilization were more likely to be inappropriate for ICD based on guidelines, more likely to be at high ex-ante risk (>20%) of dying within 1 and 2 years, and more likely to have Class IV CHF severity (with all differences significant at the .001 level). Each measure is an indication that counsels against the use of ICDs, suggesting that hospitals with high utilization rates are drawing from a less appropriate distribution of patient characteristics.

2.4 The Correlation Between ICD Diffusion and Mortality

A key moment to understand the determinants of ICD use is the relationship between ICD adoption and conditional mortality. To this end, we combine the utilization data (Section 2.1) and the outcome estimates (Section 2.2) to compute the correlation between conditional (risk-adjusted) mortality and ICD utilization. Rather than report mortality rates at the hospital level we instead aggregate them back to the more precisely estimated HRR level as shown schematically in Figure 4b.

[14] Recall that these estimates are derived from the random-effects model and are therefore already shrunken towards the mean; a fixed-effects model would have exhibited even more variability.

Figure 7 shows a positive correlation between the average (2006-13) ICD utilization rate, and the fully risk-adjusted 2-year mortality at the HRR level. The graph also identifies several regions of interest. For example, some HRRs exhibit both low risk-adjusted ICD mortality rates and low use of ICDs (Minneapolis-St. Paul, Syracuse NY, Owensville KY) while others exhibit high rates of ICD use and high ICD mortality (e.g., Miami, Terra Haute IN, and Munster, IN). That Munster is an outlier may be explained in part by a cardiologist there who was later sued for malpractice (Creswell, 2015).

Table 3 displays estimates of the hospital-level relationship between mortality and ICD utilization from OLS, random effect, and fixed-effect models, limited to just two-year mortality; full regression results (with one-year and two-year mortality) are reported in Appendix Table B.2. As shown in Table 3, there is a consistent positive association between hospital ICD utilization and risk-adjusted mortality rates, which is statistically significant in both the OLS and random-effects model, suggesting in the reduced form that patients in regions with the most rapid diffusion experience worse outcomes. The point estimates are positive but smaller and less precise in the fixed-effect model – the association between ICD utilization and conditional mortality comes primarily from between-hospital rather than within-hospital variation.

As sensitivity analysis, Table 3 also includes a regression specification including the log of annual volume for all ICD performed at the hospital for the over-65 population (including non-CHF patients), to adjust for the conventional finding that higher-volume hospitals yield better outcomes (Freeman et al., 2010). The coefficients on these variables are as expected; an increase in log-volume of 1 is associated with a 1.4 percentage-point decline in 2-year mortality in the random-effects model (Column 4 of Table 2). We also include the fraction of ICD-appropriate patients receiving high-quality medical management, which is the alternative treatment for CHF patients not receiving an ICD.[15] Including both the volume and medical management variables in the mortality regression leads to a larger positive coefficient for ICD use in the random-effects regression (Column 4 in Table 3).

To sum up, we have established several empirical patterns in the ICD data: (a) rapid average diffusion between 2002 and 2006 in the use of ICDs for patients with congestive heart

[15] The HRR level data, from Roth et al. (2016), measured the fraction of patients receiving guideline-directed medical treatments prior to their receiving an ICD.

failure, but with wide variation in growth across regions; (b) a reversion to the mean with regard to utilization, in the sense that regions with the most rapid growth in ICDs were most likely to scale back on their use; (c) hospitals with the highest ICD use were treating less appropriate patients; (d) there was wide variability in ICD mortality rates across hospitals; (e) between 2006 and 2013 the average mortality rate one- and two-year after an ICD implant has declined; and (f) a positive correlation between ICD utilization and conditional mortality across hospitals. We turn next to developing a model that can explain these empirical patterns.

3. The Model

The model builds on an optimizing Bayesian framework where both physicians and patients are heterogeneous and health outcomes are uncertain. Patients differ in the potential benefits from an ICD implant, and physicians can only observe patient type imperfectly. Additionally, physicians and their teams differ in their ability in implanting ICDs, and they may have biased perceptions about their true ability.

As noted above, while our model is couched in terms of a physician's decision, our data is at the level of the hospital. This is because ICD procedures are typically team efforts; nurses, anesthesiologists, cardiologists, electrophysiologists, and technicians contribute at various stages to better or worse outcomes. For many hospitals, there is only one primary ICD-capable physician, in which case this assumption is innocuous; for larger hospitals we will be blending the choices of two or more physicians.[16]

3.1 Static setting

We begin with the decision faced by a physician, indexed by i.[17] There is a continuum of patient types j that differ in their potential values of the ICD implant v_j and of the alternative treatments w_j. The utility of a patient with type j that receives an ICD implant by a physician with skill level a_i is $v_j + a_i$; while her utility after receiving an alternative treatment is w_j. Without loss

[16] We also assume that the ICD-capable physician makes the final decision about which patients to choose. The networks of primary care physicians and how they "feed" patients to the ICD-capable hospitals may also affect choices of patients; see for example Moen et al. (2018).

[17] In our analysis, we omit demand-side factors – e.g., patient preferences unrelated to health – that lead to systematically overusing or underusing the procedure. While patient preferences may be important at the individual level, there is less evidence that such preferences can explain a large fraction of such variation across hospitals/regions (Cutler et al., 2019).

of generality, we interpret v_j and w_j as risk-adjusted values, by which we mean (a non-linear monotonic transformation of) outcomes for patient j (e.g., years of survival or mortality) after controlling for patient characteristics observable to the econometrician. This definition is symmetric to how we have controlled for risk factors in section 2 to construct hospital-level measures of risk adjusted mortality.

We assume that v_j and w_j are independent and distributed normally.[18] That is, $v_j \sim N(\bar{v}, \sigma_v^2)$ and $w_j \sim N(\bar{w}, \sigma_w^2)$. Let μ_j denote the difference between the patient's potential value from receiving an ICD implant and her value from alternative treatments; $\mu_j \equiv v_j - w_j$. It follows from the distributional assumptions made above that:

$$\mu_j \sim N(\bar{\mu}, \sigma_\mu^2), \tag{3}$$

where $\bar{\mu} = \bar{v} - \bar{w}$ is the population mean of μ_j, and $\sigma_\mu^2 = \sigma_v^2 + \sigma_w^2$.

The Physician's information structure and priors. Physicians do not directly observe μ_j, just an imperfect signal, s_j.

$$s_j = \mu_j + \varepsilon_j \tag{4}$$

where ε_j is normal with mean 0 and variance σ_ε^2.

Physician may have a biased perception of their true skill, a_i; a_i^p denotes the provider's perceived skill. The gap between the perceived and true skill is the misperception bias, o_i.[19] If $a_i^p = a_i$ the physician is unbiased in her assessed skill; $a_i^p > a_i$ corresponds to being overconfident (or overly optimistic), while $a_i^p < a_i$ denotes underconfidence.[20]

Treatment decision. For the time being, we assume physicians obtain no private benefit or cost from implanting ICDs. Physician *i* will implant an ICD to patient *j* if the patient's expected value from implanting an ICD given the signal s_j is greater than the expected value from alternative treatment. That is, the physician treats if:

$$E\left[v_j - w_j + a_i | s_j\right] \geq 0. \tag{5}$$

[18] We have conducted the analysis allowing for correlation between patient types when treated and untreated and have found that the results are robust to this extension.

[19] For simplicity, we assume that o_i and a_i are uncorrelated. We have extended the analysis to allowing for correlation across doctors in these parameters and the results are completely consistent with those from the uncorrelated case.

[20] Our definition of overconfidence is similar to Malmendier and Tate's (2015) in the context of CEOs: "We define overconfidence as the overestimation of the value a manager believes he or she can create." (p. 46).

Given the information structure, the posterior distribution of the patient's net type conditional on s_j is

$$\mu_j \mid s_j \sim N\left(\bar{\mu}_j, \frac{\sigma_\mu^2 \sigma_\varepsilon^2}{\sigma_\mu^2 + \sigma_\varepsilon^2}\right) \tag{6}$$

where the posterior mean, $\bar{\mu}_j$, is

$$\bar{\mu}_j = (1-\alpha)\bar{\mu} + \alpha s_j, \tag{7}$$

with $\alpha = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2}$. Recognizing that $E\left[v_j - w_j | s_j\right] = \bar{\mu}_j$, we can express condition (5), as

$$\overbrace{(1-\alpha)\bar{\mu} + \alpha s_j + a_i}^{Unbiased\ net\ benefit\ of\ ICD} + \overbrace{\tilde{o}_i}^{Physician\ bias} \geq 0 \tag{8}$$

Expression (8) decomposes the net benefits from implanting an ICD perceived by the physician into two components. The first component – the unbiased net benefit – is the net benefit received by the patient if she is implanted an ICD. The second term – physician bias – captures the additional net benefit that the physician believes the implant brings the patient due to the physician bias in her perceived skill.

Isolating s_j from (8), it follows that the provider implants an ICD if she receives a signal s_j greater than a threshold $s(a_i^p)$ defined by:

$$s_j \geq s(a_i^p) \equiv \frac{1-\alpha}{\alpha}\bar{\mu} - \frac{a_i^p}{\alpha} \tag{9}$$

where $a_i^p = a_i + o_i$.

ICD usage. Given the decision rule (9) and the distribution of types and signals, the probability of implanting an ICD for a physician with perceived skill a_i^p is:

$$\Pr(ICD = 1) = \int_{s(a_i^p)}^{\infty} f(s)\, ds, \tag{10}$$

where $s\left(a_i^p\right)$ is defined by equation (9) and *f(.)* is the pdf of the signal s_j. That is, a normal distribution with mean $\bar{\mu}$, and variance $\sigma_\mu^2 + \sigma_\varepsilon^2$.

Proposition 1 (Determinants of diffusion). Ceteris paribus, the use of ICDs increases with perceived skilled, a_i^p.

Proof:

$\frac{\partial \Pr(ICD=1)}{\partial a_i^p} = -f\left(s(a_i^p)\right) * \frac{\partial s(.)}{\partial a_i^p} > 0$ because, from expression (9), $\frac{\partial s(.)}{\partial a_i^p} < 0$. □

Intuitively, the threshold signal required to implant an ICD decreases with perceived skill. Note that expression (10) shows that the only physician-specific parameter that affects the ICD use rate is the perceived skilled of the physician, a_i^p. Therefore, for a given skill, the use of ICDs increases with overconfidence, o_i; similarly, for a given level of o_i, higher (true) skill induces a greater use of ICDs. Proposition 1 is the basis for the identification of perceived skill, a_i^p. Given the population parameters that define $\bar{\mu}$, α and the distribution of signals, we use the observed ICD use rates to infer the physician's perceived skill level.

Outcomes. The outcome we measure in our dataset is the mortality rate conditional on an ICD implant. To use this information, we need to interpret what death means in our model. Naturally, the event of death (in the near term) represents a very low ex-post value for the patient. It also seems reasonable that a death that occurs further in the future is preferred to one occurring earlier. Accordingly, we interpret the death of the patient within *x* years as an ex-post utility below a threshold $\underline{\kappa}_x$, where $\underline{\kappa}_x$ is increasing in *x*.

The *x*-years mortality rate conditional on an ICD implant for a physician with perceived skill, a_i^p, and actual skill, a_i, is:

$$Pr\left(v_j + a_i \leq \underline{\kappa}_x \middle| ICD = 1\right) = \frac{\Pr\left(v_j \leq \underline{\kappa}_x - a_i \cap\ ICD = 1\right)}{\Pr\left(\ ICD = 1 | a_i^p\right)} =$$

$$\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\varepsilon(\varepsilon')\, f_\omega(\omega')\left(\int_{s(a_i^p)+\omega'+\varepsilon'}^{\underline{\kappa}_x - a_i} f_v(v')dv'\right)d\omega' d\varepsilon'}{\int_{s(a_i^p)}^{\infty} f_s(s')ds'} \qquad (11)$$

where $f_\varepsilon(.)$ is the pdf for ε, $f_v(.)$ and $f_\omega(.)$ are the pdf for patient's type v_j and ω_j, and $f_s(.)$ is the pdf for the signal *s*. The following proposition characterizes the impact of true and perceived skill on conditional mortality.

Proposition 2 (Determinants of mortality conditional on ICD implant). (i) The probability of death conditional on implanting an ICD increases with the physician's misperception, o_i. (ii) Skill has an ambiguous effect on the mortality rate conditional on receiving the ICD. However, (iii) conditional on a level of perceived skill, a_i^p, the probability of death after an ICD implant declines with true skill, a_i.

Proof: The proofs are as follows:

(i)

$$\frac{\partial \Pr(v_j + a_i \leq \underline{\kappa}_x | \, ICD = 1)}{\partial o_i} =$$

$$\left[1 - \Pr(v_j \leq \underline{\kappa}_x \mid ICD = 1, a_i^p, a_i\right]\left(-\frac{\partial s}{\partial o_i}\right)\left[\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\varepsilon(\varepsilon') f_\omega(\omega') f_v(s(a_i^p) - \varepsilon') d\omega' d\varepsilon'}{\int_{-\infty}^{\infty} f_\varepsilon(\varepsilon') \int_{-\infty}^{\infty} f_\omega(\omega') \left(\int_{s(a_i^p)+\omega'-\varepsilon'}^{\infty} f_v(v') dv'\right) d\omega' d\varepsilon'}\right] > 0 \qquad (12)$$

Both the first and third terms are positive, but the key is the middle expression; that when overconfidence rises, the "hurdle" point at which the physician does the procedure declines, thus expanding the number of patients for which the net benefit is negative.

(ii)

$$\frac{\partial \Pr(v_j + a_i \leq \underline{\kappa}_x | \, ICD = 1)}{\partial a_i} = -\left[\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\varepsilon(\varepsilon') f_\omega(\omega') f_v(\underline{\kappa}_x - a_i) d\omega' d\varepsilon'}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\varepsilon(\varepsilon') f_\omega(\omega') \left(\int_{s(a_i^p)-\varepsilon}^{\infty} f_v(v') d\, v'\right) d\omega' d\varepsilon'}\right]$$

$$+ \frac{\left[1 - \Pr(v_j \leq \underline{\kappa}_x \mid ICD = 1, a_i^p, a_i\right]}{\alpha}\left[\frac{\int_{-\infty}^{\infty} f_\varepsilon(\varepsilon') f_\omega(\omega') f_v(s(a_i^p - \varepsilon') d\varepsilon'}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\varepsilon(\varepsilon') f_\omega(\omega') \left(\int_{s(a_i^p)-\varepsilon}^{\infty} f_v(v') dv'\right) d\omega' d\varepsilon'}\right] \qquad (13)$$

Expression (13) shows that skill affects mortality by improving the outcomes for patients who would have been treated anyway (first term), but also by bringing in more patients with net benefit, but whose underlying mortality probability could be higher as well (second term). As a result, the net effect of skill on conditional mortality is ambiguous.

(iii)

$$\frac{\partial \Pr(v_j + a_i \leq \underline{\kappa}_x \mid ICD = 1, a_i^p)}{\partial a_i} = -\left[\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\varepsilon(\varepsilon') f_\omega(\omega') f_v(\underline{\kappa}_x - a_i) d\omega' d\varepsilon'}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\varepsilon(\varepsilon') f_\omega(\omega') \left(\int_{s(a_i^p)-\varepsilon}^{\infty} f_v(v') dv'\right) d\omega' d\varepsilon'}\right] < 0$$

□

While utilization is affected only by perceived skill, the average mortality of physician *i'* patients with an ICD implant, conditional on population parameters, $\underline{\kappa}_x$, and perceived skill a_i^p, is determined by the true skill level of physician, a_i. Therefore, conditional on those parameters, we can identify the level of true skill in a hospital by inverting expression (11) after replacing in the left-hand-side the conditional mortality rate observed in the data. In summary, whether a given

patient receives an ICD is affected only by the physician's *perceived* skill, but the likelihood of death is further affected by the physician's *true* skill. This insight is the basis for our identification of the model.

A (slight) generalization. So far, the only source of physician bias we have considered is misperception of true skill. However, there may be other sources of bias that influence the decision to implant ICDs; we consider two here. The first bias reflects the possibility of misperception by all physicians of the average net value of an ICD. That is, instead of all physicians understanding that the true population mean of the treatment effect is $\bar{\mu}$, they may instead believe that $\hat{\mu} \neq \bar{\mu}$ is the correct value. Since information about the value of ICDs comes largely from professional journals and widely-reported trials, it is natural to model this bias as common for all physicians. The second bias we consider stems from the possibility that physician may derive a private benefit or cost, B_i, from implanting ICDs. Private benefits may be positive if a physician receives incentive payments from manufacturers or derives a high utility from the fees from implanting ICDs, and negative if she experiences high costs (psychic or otherwise) from performing the procedure (Cutler et al., 2019).

In this general setting, a physician implants an ICD if:

$$\overbrace{(1-\alpha)\bar{\mu} + \alpha s_j + a_i}^{Unbiased\ net\ benefit\ of\ ICD} + \overbrace{o_i + (1-\alpha)(\hat{\mu} - \bar{\mu}) + B_i}^{Generalized\ physician\ bias} \geq 0 \qquad (14)$$

This decision rule is isomorphic to expression (8) in the baseline model once we reinterpret the physician bias so that, in addition to including the physician's overconfidence (o_i), it also includes the common bias in the assessment of the net value of ICDs in population ($(1-\alpha)(\hat{\mu} - \bar{\mu})$) and the physician's private benefit from implanting the ICD (B_i). We denote this generalized physician bias by O_i (to highlight the symmetry with the bias in the baseline case, o_i). Importantly, the insights advanced above about how we can use the model to identify true and perceived skill extend naturally to the general setting. To see this, re-define perceived skill, A_i^p, as

$$A_i^p \equiv a_i + O_i = a_i^p + (1-\alpha)(\hat{\mu} - \bar{\mu}) + B_i \qquad (15)$$

which reflects true skill plus the generalized physician bias. As in the baseline, perceived skill, A_i^p, can be identified by inverting expression (10) using as threshold signal $s(A_i^p)$ defined as

$$s(A_i^p) \equiv -\frac{(1-\alpha)}{\alpha}\bar{\mu} - \frac{A_i^p}{\alpha} \qquad (16)$$

Furthermore, conditional mortality is unaffected by this generalization because true skill is unchanged; conditional on the rate of ICD implants, mortality only depends on true skill.

3.2 Dynamics of physician bias

While the cross-sectional variation across hospitals in conditional mortality and ICD use is sufficient to identify physician skill (a_i) and generalized physician bias (O_i) in our static model, it does not separately identify the three possible sources for this bias (overconfidence, common bias, and private benefit). Additionally, because the model is static, it does not predict how bias will evolve over time. To address both issues, we extend our model to a dynamic setting; while true skill and private benefits do not change, we do allow for the evolution of both perception bias and the common evaluation across all providers of the ICD's average benefit.

We assume that in every period, physicians receive two unbiased signals. The first is $s_a = a_i + \varepsilon_{a_i}$, which provides information about the physician's true skill but contains noise, ε_{a_i} which has a zero mean and is distributed normally. The physician knows the precision (or inverse of the variance) of signal s_a that we denote by ρ_a. The second, s_μ, is a common signal to all physician about the true average value of ICDs in the population, perhaps drawn from clinical journals or other sources; $s_\mu = \bar{\mu} + \varepsilon_\mu$, again where ε_μ is zero-mean and normally distributed, with known precision ρ_μ.

Beliefs about a_i and $\bar{\mu}$ are random variables that we respectively denote as $\tilde{a}_i$ and $\tilde{\bar{\mu}}$.[21] We assume that the prior distribution of $\tilde{a}_i$ is normal with mean a_i^p and precision $\tau > 0$. Providers' common prior about the mean net value of ICDs in the population, $\tilde{\bar{\mu}}$, is distributed normal with mean $\hat{\mu}$ and precision $\tau_\mu > 0$.

The following Lemma characterizes the posterior distributions of $\tilde{a}_i$ and $\tilde{\bar{\mu}}$

Lemma 1 (Posterior distribution of skill) The posterior distribution of $\tilde{a}_i$ is normal with mean $a_i^{p\prime}$ and precision $\tau + \rho_a$, where:

$$a_i^{p'} = \frac{\tau a_i^p + \rho_a s_a}{\tau + \rho_a}. \qquad (17)$$

The posterior distribution of $\tilde{\bar{\mu}}$ is normal with mean $\hat{\mu}'$ and precision $\tau_\mu + \rho_\mu$ where:

[21] We use "~" to denote random variables.

$$\hat{\mu}' = \frac{\tau_\mu \hat{\mu} + \rho_\mu s_\mu}{\tau_\mu + \rho_\mu} \tag{18}$$

Proof: See De Groot (1971), page 167. □

At any given period *t*, the posterior distribution of a random variable constitutes the prior distribution at *t+1*. Using Lemma 1, and the definition of A_{it}^p, we can derive the following expression for its law of motion:

$$\Delta A_{it+1}^p = A_{it+1}^p - A_{it}^p = -\delta_a\left(a_{it}^p - a_i\right) - \delta_\mu(1-\alpha)(\hat{\mu}_t - \bar{\mu}) + \varepsilon_t \tag{19}$$

where $\delta_a = \frac{\rho_a}{\tau+\rho_a}$, $\delta_\mu = \frac{\rho_\mu}{\tau_\mu+\rho_\mu}$ and $\varepsilon_t = \delta_a \varepsilon_{a_{it}} + \delta_\mu \varepsilon_{\mu t}$.

In the special case we have used as baseline (i.e., $\hat{\mu}_t = \bar{\mu}$, and $B_i = 0$), $A_{it}^p = a_{it}^p$ and equation (19) becomes:

$$\Delta a_{it+1}^p = -\delta_a\left(a_{it}^p - a_i\right) + \delta_a \varepsilon_{a_{it}} \tag{20}$$

Or equivalently, since a_i is fixed,

$$\Delta o_{it+1} = -\delta_a o_{it} + \delta_a \varepsilon_{a_{it}} \tag{21}$$

Since $\delta_a > 0$, equation (21) demonstrates that learning about true skill induces mean-reversion in overconfidence (i.e., hospitals with greater overconfidence tend to experience larger subsequent declines). The mean reversion in overconfidence generally induces mean reversion in perceived skill,[22] leading to a reduction over time in cross-hospital variation in perceived skill and utilization rates if the initial variance is above its long-run level.[23]

Going back to the general case, we note that $a_{it}^p = A_{it}^p - B_i - (1-\alpha)(\hat{\mu}_t - \bar{\mu})$ which allows us to express the mean reversion in perceived skill as:

$$-\delta_a\left(a_{it}^p - a_i\right) = \delta_a B_i - \delta_a\left(A_{it}^p - a_i\right) + \delta_a(1-\alpha)(\hat{\mu}_t - \bar{\mu}) \tag{22}$$

Plugging (22) in (19), and expressing $(\hat{\mu}_t - \bar{\mu})$ as $\left(\hat{\mu}_t - \bar{\hat{\mu}}\right) + \left(\bar{\hat{\mu}} - \bar{\mu}\right)$ where $\bar{\hat{\mu}}$ is the average level of the perceived net value of ICDs ($\hat{\mu}_t$), we obtain the following expression for the evolution of A_{it}^p:

[22] The exception would be when true skill and overconfidence are sufficiently negatively correlated.

[23] The long-run variance of overconfidence is $\frac{\alpha_a^2 \sigma_\varepsilon^2}{1-(1-\alpha_a)^2}$.

$$\Delta A_{it+1}^{p} = \Delta O_{it+1} = \overbrace{\delta_a \widetilde{B_i}}^{Intercept} - \overbrace{\delta_a\left(A_{it}^{p} - a_i\right)}^{Mean\ reversion} - \overbrace{\left(\delta_\mu - \delta_a\right)(1-\alpha)(\hat{\mu}_t - \bar{\mu})}^{Time\ effect} + \varepsilon_{t+1} \quad (23)$$

Equation (23) yields an intuitive framework to identify the three possible sources of the generalized overconfidence bias: pure overconfidence, common bias, and private benefit. Private benefits do not change over time and introduce a time-invariant hospital-specific intercept. Learning about common bias affects all hospitals equally and introduces a time-specific intercept. In contrast, learning about true skill introduces a negative relationship between (generalized) overconfidence and the change in perceived skill. While both forms of learning (true skill and net value of ICDs) can generate a negative drift in (generalized) perceived skill, only learning about true skills can reduce the cross-hospital variance in (generalized) perceived skill.

4. Analysis

In this section, we use our model to study the empirical regularities in Section 2. First, we identify the hospital-levels of true and perceived skill (a_i and A_{it}^{p}), with the difference between perceived and true skill capturing generalized overconfidence ($O_{it} = A_{it}^{p} - a_i$). Second, we analyze the relevance of each of them in explaining the patterns of variation in ICD use and conditional mortality across hospitals. Third, we use our model and estimates to evaluate the welfare effects arising from overconfidence and misperception more generally. Finally, we estimate dynamic models of overconfidence to study the role of learning for the evolution of perceived skill and its impact on the evolution of the cross-hospital distribution in ICD use and conditional mortality.

4.1 Identification of parameters

We identify the model parameters in two stages. First, we use information on moments from the entire distribution of hospitals to calibrate parameters that are common to all hospitals and that we denote as aggregate parameters. Once we have calibrated the aggregate parameters, hospital-level information on ICD use and conditional mortality in each year are used to identify the hospital-year levels of true and perceived skill.

Calibrating the aggregate parameters. Without loss of generality, we normalize the average skill in population, $\bar{a}_i$, and the average utility of a patient with heart failure in the absence

of ICD treatments, $\bar{w}$, to 0. These parameters are isomorphic to $\bar{v}$ in the ICD use equation (12), and to $\underline{\kappa_x}$ in the mortality rate equation (11). We also normalize the variance of the signals noise, σ_ε^2, to 1, as the relevant moment for the ICD implant decision is the noise to signal ratio ($\sigma_\varepsilon^2/(\sigma_v^2 + \sigma_{w,}^2)$).

An important aspect of the calibration is to bridge the conceptual gap between the units in the model (i.e., utility) and in the outcomes we observe (i.e., mortality or years of survival). Using the logic of a probit specification, we calibrate the threshold $\underline{\kappa_2}$ to match the unconditional mortality for patients with congestive heart failure (CHF) who do not receive an ICD. We know the 2-year average mortality rate for treated CHF patients from the registry data, but calculate average mortality for untreated patients by subtracting the estimated treatment effects from the landmark randomized trial (Bardy et al., 2005) -- a 3 percentage-point reduction in mortality at two years -- from the mortality rate for treated patients.

These normalizations leave us with 6 parameters to calibrate: the average levels of perception bias and the value of ICDs in the population ($\bar{O}$ and $\bar{v}$), the variances of the patient values with and without treatment (σ_v^2, σ_w^2), and the variances of true skill and provider bias (σ_a^2, σ_O^2). To identify these parameters, we have six moments. Five of them are computed from the moments in our data: the mean and variance of ICD use rate across hospitals,[24] the mean and variance of conditional mortality across hospitals, and the cross-hospital correlation between the ICD use rate and the conditional mortality rates. Additionally, we use information on the ratio of the variance in life expectancy for treated and untreated patients during the ICD trials, equal to 0.89 (Bardy et al. 2005), to calibrate the relative variance of the distribution of utility values (σ_v^2/σ_w^2).[25] Using the variance ratio in mortality is a reasonable assumption to the utility-based variance ratio given that the sole purpose of ICDs is to reduce the risk of sudden cardiac arrest. We use this restriction as the sixth moment to calibrate the six aggregate parameters.

[24] We target the rate of ICD use among eligible patients from Al-Khatib et al. (2012) which is 39.9 percent. They measure the implementation of ICD among patients hospitalized for heart failure during 2005-09 (as in our empirical analysis) that are eligible for ICD implantation. This seems the most natural denominator to consider in our context given that, despite the small share of potentially eligible patients with HF, over 90% of the ICDs in the registry data are administered among eligible patients. We use the ratio of the average ICD use rate among eligible patients (from in Al-Khatib et al. (2012)) and among patients with HF in our data to scale accordingly the standard deviation of the utilization rate.

[25] The variance of life expectancy was calculated for patients randomly assigned into the treatment and control groups over the 5-year trial period based on Figure 1 in Bardy et al. (2005). The variance of life expectancy for treated patients was 2.48 years, while the variance for untreated patients was 2.80 years; the ratio is 0.89.

With six moments, the six parameters of the model are just identified (although with non-linearities the model may not fit the moments exactly). Intuitively, the variance parameters are identified in the following way. The variance of the hospital distributions of skill and overconfidence are inferred from the cross-hospital variances and correlation of the ICD use rate and conditional mortalities. The variance of the ICD use rate across hospitals depends on the variance of perceived skill ($\sigma_a^2 + \sigma_O^2$). The correlation across hospitals of the ICD use rate with conditional mortality depends on the relative contribution of skill versus overconfidence: when skill is a larger share of the variance there will be a negative correlation between ICD use and conditional mortality (higher skill hospitals have higher ICD use and lower conditional mortality); when overconfidence is a larger share of the variance there will be a positive correlation between ICD use and conditional mortality (more overconfident hospitals will have higher ICD use and higher conditional mortality).

While these two moments identify the relative variance of skill and overconfidence, they are only identified up to a scale, analogous to a Probit. The scale term translates the utility-based parameters into observable mortality risk. Therefore, we use observable variation across hospitals in mortality risk to identify the scale parameter linking the variance in mortality with the variance in utility ($\sigma_v^2 + \sigma_w^2$). Along with the relative variance of utility values (σ_v^2/σ_w^2), we can identify the variances of patient utilities with and without treatment (σ_v^2, σ_w^2).

The two remaining parameters are the average level of overconfidence and the average value of ICDs in the population ($\bar{O}$ and $\bar{v}$). These are identified from the means of ICD use and conditional mortality. From equation 11 (the ICD use equation) the mean of ICD use depends on the mean of perceived skill ($\bar{O} + \bar{v}$). From equation 12 (the conditional mortality equation) the mean of conditional mortality depends on the both the mean of perceived skill ($\bar{O} + \bar{v}$) and the mean of actual skill ($\bar{v}$), which lets us separately identify the mean of overconfidence ($\bar{O}$). Intuitively, equation 12 has a Tobit-like structure, which identifies mean skill ($\bar{v}$) after controlling for the selection probability (through $\bar{O} + \bar{v}$).

The top panel of Table 4 reports the calibrated values for the aggregate parameters. There is large variation in perception bias across hospitals (variance 0.022, SD=.15) and a mean value for generalized overconfidence $\overline{O_\iota}$ equal to 0.134, implying 82 percent of hospitals have positive generalized overconfidence. The variance in provider bias (σ_O^2) is greater than the cross-hospital variance in true skill (σ_a^2), and the variance of patient-specific variables (σ_v^2 and σ_w^2) is much larger

than the variance of hospital-level variables (σ_a^2 and σ_O^2). The bottom panel of Table 4 reports the targeted moments in the data and in the model and shows that the model matches well each of the targeted moments.

The calibration of the variance and the mean of overconfidence are critical for the welfare implications from our analysis. Therefore, it is worth re-stating the data features that pin them down. That the variance of overconfidence is large relative the variance in skill follows from the positive correlation observed between conditional mortality and ICD utilization rates. Mean overconfidence depends on the fraction of eligible patients who receive an ICD (which comes from Al-Katib et al. (2012)). Naturally, a lower ICD utilization rate would lead to a lower average misperception bias $\bar{O}$, potentially even a negative value. But if physicians were on average underconfident, we might expect the average ICD use rate to rise over time between 2006-13 as doctors learn more about their true skill. Instead, we see a strong decline in ICD rates. Even though we do not use the evolution of the average ICD use rate to calibrate mean overconfidence explicitly, it is reassuring that the target level from the literature yields a calibrated value for mean overconfidence that is consistent with the aggregate evolution in ICD use.

Hospital level parameters. Once calibrated, we use hospital-year data on ICD use and conditional mortality to identify skill and the degree of misperception for each hospital and year. Specifically, equation (11) shows that, given the aggregate parameters, the ICD use rate is fully determined by the level of perceived skill. Furthermore, equation (12) shows that, given the aggregate parameters and perceived skill, conditional mortality depends only on true skill. These observations imply that we can use data on conditional mortality and ICD use to identify a_i and A_i^P by inverting equations (11) and (12). Appendix Figures C.1 and C.2 plot the histogram of the identified parameters. We note geographic differences in misperception; the regions with greater generalized optimism or overconfidence are in the South (in particular, Texas), the Southeast, and the Great Lakes region (Michigan, Indiana, and Ohio).

4.2 Determinants of ICD use and Mortality

To better understand model mechanics, we conduct comparative statics exercises where we document in Table 5 the impact of reducing the mean and variance of skill and overconfidence for utilization and health outcomes. Column 1 reports the moments from the

actual data.[26] We then report the moments holding true skill unchanged, but setting mean (column 2), variance (column 3) and both the mean and the variance (column 4) of overconfidence equal to zero. Finally, column (5) sets the variance in skill equal to zero.

Setting the mean of overconfidence equal to zero predicts a decline in ICD use of about one fifth, from 0.378 to 0.314 of eligible CHF patients (Column 2). Because on average physicians are estimated to be overconfident; this holds also when setting both mean and variance equal to zero (Column 4). In contrast, setting the variance of overconfidence (column 3) or the variance of skill (Column 5) equal to zero has little impact on *average* ICD utilization. When we set the variance of overconfidence equal to zero, the standard deviation of ICD use across hospitals declines from 0.091 to 0.036, a 60 percent reduction in regional variation and roughly double the reduction when the skill variance is set to zero (0.067). Setting both the mean and variance of overconfidence to zero (column 4) reduces the standard deviation of ICD use by roughly two thirds (to .033).

Mortality *conditional* on an ICD declines when we set average overconfidence to zero, as fewer inappropriate patients are chosen, resulting in better outcomes among those treated. The standard deviation of mortality is little affected by reducing the mean of overconfidence (or the variance of overconfidence or skill). As well, setting the variance in overconfidence to zero flips the correlation between ICD utilization and conditional mortality from 0.204 (Column 1) to -0.998 (Column 3) as now the only source of heterogeneity across hospitals is in true skill, and hospitals with greater skill conduct more ICDs but also have lower conditional mortality rates. Correspondingly, eliminating the variance in skill makes ICD use near-perfectly correlated with conditional mortality because now the only source of cross-hospital variation is overconfidence, and more overconfident hospitals conduct more ICDs and have higher mortality rates.

The key measure of welfare is the unconditional (overall) mortality across all eligible patients. While we don't observe unconditional mortality in the data, we can calculate it based on our model. When both the mean and variance of overconfidence is set to zero, 2-year mortality

[26] Note that, in general, moments do not coincide with those used as aggregate targets because those moments were for 2006 only and hospitals were weighted by volume. Additionally, the targeted correlation between ICD use and conditional mortality in the aggregate calibration is .0947, while the correlation between ICD use and our shrinkage estimate of conditional mortality in 2006 is notably higher (.2497). This is because the latter estimate is based on predicted (shrunken) hospital-specific random effects that exhibit a smaller variance than the true hospital-specific random effects.

declines from 0.175 to 0.170, or a reduction of 0.5 percentage points (Column 4). Relative to those who actually receive an ICD, the mortality effect is roughly 40% of the estimated 3 percentage point gains arising from ICD placement (Bardy et al., 2005).[27] Note that there are gains (of similar magnitude) both from reducing the average rate of overconfidence (Column 2) and reducing the variance of overconfidence (Column 3) since both result in inefficient allocation of ICDs.

To better understand the welfare effects of perception biases, we study how those vary across hospitals with different levels of overconfidence. Specifically, we sort hospitals into deciles based on their overconfidence, and focus just on the top and bottom decile hospitals. For the top decile, just over half (0.55) of eligible patients received an ICD; this declines to 0.35 when the mean and variance of overconfidence are set to zero.[28] Unconditional mortality on the other hand is predicted to decline by 0.021, from 0.180 to 0.159. What this means is that for these top-decile hospitals, setting overconfidence equal to zero leads to 20 fewer ICDs (per 100 overall patients) and 2.1 fewer deaths, or a reduction of more than 10 percent in mortality for this group (2.1/20). By contrast, in the lowest decile, hospitals are slightly underconfident, and when overconfidence is set to zero ICD rates rise from 0.25 to 0.28, with very little change in mortality (from 0.179 to 0.178). Therefore, the welfare gains from eliminating misperceptions are larger in hospitals with the greatest degree of overconfidence.

4.3 Dynamics of perceived skill

So far, we have used the static model developed in Section 3.1 to estimate the distribution of perceived skill and overconfidence from the cross-section of ICD use and mortality. In this section, we use the simple dynamic model developed in Section 3.2 to study how learning affects overconfidence and the evolution of perceived skill over time, and the implications this has for the evolution of the cross-hospital distribution in ICD use and conditional mortality.

Estimating the dynamic model is useful for three reasons. First, it provides evidence on the rate at which hospitals learn about their own over-confidence. This parameter is key to understanding how both the average ICD use rate and its variance across hospitals will decline over time due to learning. Second, as shown in section 3.2, our dynamic model allows us to

[27] That is, we divide the decline in overall mortality (-0.005) by the fraction receiving an ICD (0.378), or a 1.3 percentage point reduction in mortality among those receiving an ICD; this in turn is slightly more than 40 percent of the estimated gains of 3 percentage points in the RCT.

[28] Average ICD rates are slightly higher than the overall mean of 0.247 because these hospitals also exhibit above-average skill.

separately identify the proportion of generalized overconfidence that is due to private benefit. This component is not eliminated by learning and will lead to persistent inefficiencies due to over-use of ICDs. Finally, we can use estimates of the dynamic model from our first years of data to conduct out-of-sample forecasts of ICD use and conditional mortality, and compare these to actual values as a test of the validity of our model.

Full sample estimates. We use the identified hospital-level parameters from 2006-2013 to estimate specification (24), which is an econometric counterpart to the law of motion for perceived skill derived in the learning model (equation 23):

$$\Delta O^{p}_{it+1} = \beta_i - \delta_a O_{it} + \pi_t + \varepsilon_{it+1} \qquad (24)$$

In the model's simplest form, we set the intercept equal across hospitals ($\beta_i = \beta$) and assume there is no learning about the common bias so there are no time effects $(\pi_t = 0)$. This results in a simple regression of the change in perceived skill from t to t+1 on the level of generalized overconfidence in t. In a more general model, we allow for time effects and for β_i to be a function of hospital characteristics (adding dummies for hospital ownership and teaching status).

The first column of Table 6 provides estimates of the simplest model with a common constant term and no year effects. The slope coefficient on O_{it} (-0.215; s.e. 0.007) is an estimate of $-\delta_a$ in our learning model (e.g., $\delta_a = .215$). This implies rapid learning: On average, 21.5% of misperception about skill in year t is gone by year t+1, yielding an average half-life for hospital misperceptions of their skill of about 3 years. In addition, the intercept (0.0104, s.e. 0.0011) is an estimate of $\delta_a B$ in our learning model, where δ_a is the rate of learning (.215) and B is the average private benefit. Thus, dividing the intercept from this regression by our estimate of α provides an estimate of the average private benefit: B=.0104/.215=.048 (delta method s.e.=.004). Note that B is simply the steady-state point in our learning model; the generalized overconfidence of .048 is associated with no average change in perceived skill. This estimate implies that approximately 36% of the average generalized overconfidence (.134 from Table 4) is due to private benefit, while the remaining 64% is due to physician's being initially overconfident about their skill at performing ICDs. While the estimate of δ_a suggests that learning about skill will rapidly eliminate overconfidence about skill, our estimate of B suggests that there will remain a persistent bias toward over-use of treatment due to private benefit.

The private benefit that a hospital receives from doing ICDs may vary across hospitals due to how different hospitals weight the revenue from performing ICDs against their reputation for providing appropriate care. In column 2 of Table 6 we add dummies for hospital ownership (for-profit and government relative to not-for-profit) and if the hospital was a major teaching hospital. The coefficient on for-profit ownership is positive and significant. The estimated coefficients from column 2 imply that the private benefit at for-profit hospitals ((.0102+.0050)/.219 = .069) is about 50% larger than the private benefit at not-for-profit hospitals (.0102/.219=.047). This estimate is consistent with the view that for-profit hospitals place more weight on generating revenue than do not-for-profit hospitals. In contrast, the coefficient on being a major teaching hospital is a bit larger in magnitude but negative and significant, implying that teaching hospitals have roughly 60% lower private benefit than non-teaching hospitals. Again, this would be consistent with the view that teaching hospitals place more weight on providing appropriate care than non-teaching hospitals.

The third column of Table 6 adds year dummies. While the year dummies are jointly significant, adding them to the regression has little impact on the remaining coefficients. The coefficients on the year effects are estimates of $\delta_a B_i - (\delta_\mu - \delta_a)(1-\alpha)(\hat{\mu}_t - \bar{\mu})$ in the learning model, which capture both private benefit (constant over time) and common learning across all hospitals (which may vary over time). This means that the year dummy variables cannot separately identify private benefits from the average of common learning across years. Based on increasing clinical concerns over the effectiveness of ICDs (e.g., McMurray, 2016), we might expect the mean change in common learning to be negative over these years which would imply that, if anything, the intercept is an under-estimate of private benefits. The average intercept across years reported in column 3 is very similar to the intercept reported in column 2, and therefore implies similar estimates of private benefits under the conservative assumption that changes in common beliefs are mean zero.

There are two reasons that OLS estimates of the learning parameter δ_a could be biased. First, classical measurement error in the estimate of generalized overconfidence (O_{it}) would bias the OLS coefficient on O_{it} in the negative direction (since the dependent variable is $O_{it+1} - O_{it}$ leading to an over-estimate of learning. If measurement error is independent across years, then O_{it-1} is a valid instrument for O_{it}. Estimating the specification in Column 3 of Table 6 by 2SLS

using lagged overconfidence as an instrument yields an estimate of δ_a equal to 0.128 (s.e., 0.008), which is somewhat smaller than OLS but still suggests that learning eliminates overconfidence over time with a half-life of around 5 years.

Alternatively, variation across hospitals in private benefit (B_i) that are not captured by ownership and teaching status would bias the OLS coefficient on O_{it} in the positive direction (since both the unobserved intercept and O_{it} depend positively on B_i), leading to an under-estimate of learning. If the learning process about o_{it} follows an AR(1) and measurement error is independent across years, then differencing equation (24) to remove the hospital-specific intercept and instrumenting for ΔO_{it} using O_{it-2} would yield unbiased estimates of the learning parameter (Arellano and Bond, 1991). Estimating the specification in column 3 by this Arellano-Bond method yields an estimate of δ_a equal to 0.466 (s.e., 0.067), which is somewhat larger than OLS and suggests that learning eliminates overconfidence over time more rapidly with a half-life of around 1 year.

Both the measurement error correction and the Arellano-Bond correction may also suffer from bias, however, because both assume that measurement error in O_{it} is independent over time. This is unlikely to hold in our data: O_{it} is a function of both the ICD rate and the conditional mortality rate at each hospital, and the conditional mortality rate was estimated using Bayesian methods that smooth variation over time which may generate serial correlation in the measurement error. While all three approaches are consistent with a model of learning, we rely on OLS – the intermediate estimate -- as our preferred estimation approach.

In sum, the full-sample estimates are consistent with our dynamic learning model. They imply that hospital learning over time will reduce both average over-confidence and the variation across hospitals due to overconfidence, but that private benefits will lead to persistent over-use of ICDs.

Out-of-sample analysis. To study the role of learning for the evolution of (generalized) perceived skill and hospital level outcomes we conduct an out-of-sample prediction exercise. Specifically, we estimate the learning model using a single cross section of data from 2006 (regressing the change in generalized overconfidence from 2006 to 2007 on generalized overconfidence in 2006). We then use the estimated intercept and learning parameter to conduct

sequential out-of-sample forecasts of generalized overconfidence and perceived skill, holding hospital skill constant at its 2006 level. We then compare the resulting forecasts for 2013 to actual values as a test of the validity of our model and an illustration of the implications of our estimates.

In the final two columns of Table 6, we re-estimate the specifications from columns 1 and 2 using the cross section of hospitals in 2006 (there is no need for year dummies with a single year of data). The reversion parameter α is estimated to be very similar to that for the full sample (.213, s.e. .0227 in Column 4, for example), as is the constant term. Given that the coefficients on ownership and teaching are not estimated with any precision (Column 5), we use the simpler version from column 4 to perform the prediction models.

Figure 8A presents the scatter plot of the change in perceived skill induced by learning from 2006 to 2013 against the initial (generalized) overconfidence. There are large differences across hospitals in the changes in perceived skill induced by learning ranging from -0.6 to 0.4, with more overconfident doctors experiencing larger reductions in perceived skill. Overall, this negative correlation between initial level and the longer run predicted change in overconfidence suggests that learning leads to substantial reductions in the both the mean and variance of overconfidence.

Table 7 compares how the actual moments of ICD use and conditional mortality changed between 2006 and 2013 to the predictions based on our dynamic model. The model is fit to the 2006 parameters and so predicts them exactly, as shown in the first two rows of Table 7. The mean of ICD use fell from 43.8 percent in 2006 to 33.4 percent in 2013. Our learning model generates 66 percent of this decline in ICD use, as the average ICD use associated with the distribution of overconfidence levels predicted in 2013 is equal to 36.9 percent. Our model also predicts the decline between 2006 and 2013 in the standard deviation of ICD use across hospitals, although somewhat over-predicts the magnitude. Similarly, the model predicts the decline in the mean and cross-hospital standard deviation of conditional mortality but over-predicts their magnitudes.

Beyond these trends in aggregate moments, we can evaluate the importance of learning dynamics by studying the association across hospitals on actual and predicted changes in ICD use and conditional mortalities. Figures 8B and 8C plot the changes in ICD use (8B) and conditional mortality (8C) across hospitals in the model's out-of-sample forecasts and in the data. The correlations between the model's out-of-sample forecasts and the data are .68 for changes in ICD use and .43 for changes in conditional mortality. Thus, the learning model not only predict a

substantial portion of the large aggregate changes in ICD use observe in the data, but it correctly predicts the evolution for individual hospitals. We take this finding as evidence that learning is key for the evolution of ICD use rates and conditional mortality across hospitals.

Before concluding, it is pertinent to revisit the interpretation of o_i in our model. While we have interpreted this parameter as pure physician over- (or under-) confidence, there is the possibility that o_i reflects doctor-specific biases in the perceived value of ICDs (i.e. $\hat{\mu}_i - \hat{\mu}$), as opposed to the doctor' skill in implementing them. Under this view, the variation in generalized overconfidence occurs because some physicians read the medical journals, while others don't. In theory, this could be an important factor, but in practice it seems unlikely; to our knowledge there were no new landmark studies (between 2006 and 2013) shedding light on appropriateness for older ICD patients, and overall mortality among patients receiving ICDs was stable over time.[29] In sum, we believe the documented perception biases are more likely to be about the doctor's true skill than about the net value of ICDs in the larger population.

5. Conclusions

What determines the adoption and diffusion of new technologies? Research in economics has focused on factors primarily related to rates of return, input prices, differential factor productivity, or profitability. We find an additional factor in adoption and diffusion: Variation in the adopter's *perceptions* of their own skill and abilities in using the new technology. In the case of implantable cardioverter defibrillators (ICDs), we have estimated that variation in these perceptions – independent of true skill – accounts for roughly two-thirds of the cross-hospital variance in ICD use and reduced the average health benefits of the new technology by 40 percent. On average, physicians are overconfident, in the sense of having inflated views of their own skills leading to higher utilization rates but often with poor outcomes. We also found that physicians learn about these perception biases and reduce them over time. Learning dynamics account for two-thirds of the decline in ICD use over the period 2006-2013.

Misperceptions may be an important determinant of adoption dynamics in other settings where adopters are on average overconfident (as in our case) or underconfident. The average sign of the bias will affect the level of adoption and patterns of diffusion. In cases with average

[29] Additionally, in estimates not reported, we have observed that the learning rate of a hospital is decreasing in the doctor's true skill. The fact that learning rates are correlated with hospital-level variables/parameters is natural when learning is about hospital-specific parameters, such as skill.

under-confidence, misperceptions will cause (inefficiently) slow diffusion of technologies, a commonly observed phenomena in the diffusion literature (Rogers, 2004), while in cases with average overconfidence, diffusion is too rapid, and typically followed by a scaling back of investments as we observe in ICDs and for other innovations.[30] Understanding the determinants of the initial bias in the adopters' perceptions is beyond the scope of this paper, but certainly is an important question for future research.

A central focus of this paper is the presence of unobserved heterogeneity among adopters in the returns to using a given technology. We believe the methods developed in this paper can be applied more generally to distinguish between the roles of perception biases versus true fundamentals in technology diffusion. Indeed, an early example of this question arose in a debate between rural sociologists and economists regarding the seemingly slow adoption of hybrid corn among farmers during the 1930s and 1940s (Griliches 1957, 1960, 1962; Babcock 1962; Havens and Rogers 1961). While sociologists argued that it was due to heterogeneity in beliefs about the perceived benefits, Griliches argued that the variation was because of heterogeneity in the returns to using the technology (Skinner and Staiger, 2007). Our findings suggest a larger influence of beliefs, with a more modest role for the rate of return (e.g., true skill) in explaining such variations.

These findings are relevant for policies designed to improve the allocative efficiency of technology adoption even outside of health care. In an ideal world, strategies could be implemented to reduce the variance in skill (a_i) by encouraging physicians to learn from the most highly skilled physicians, much as pineapple farmers learned from their most productive neighbors (Conley and Udry, 2010). We know less about how existing physician networks contribute to physicians either learning about their skill or improving it (Agha and Molitor, 2018; Moen et al., 2018). But even in the absence of policies that directly improve skill levels, our findings suggest that efforts to reduce the degree of perception bias in technology adoption would improve allocative inefficiency and enhance consumer welfare.

[30] As Jupiter and Burke (2013) wrote, "Artelon® arthroplasty, thermal shrinkage, Vioxx®, metal-on-metal hip arthroplasty, and Infuse® bone grafting in the spine—all had come onto the "market" with enthusiastic reports only to fall from grace to unhappy outcomes, permanent disabilities, and malpractice litigation. (p. 249)."

References

Abaluck, J., L. Agha, C. Kabrhel, A. Raja, and A. Venkatesh, 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." *American Economic Review* 106, no. 12: 3730-64.

Agha, L., and D. Molitor, 2018. "The Local Influence of Pioneer Investigators on Technology Adoption: Evidence from New Cancer Drugs." *Review of Economics and Statistics* 100(1): 29-44.

Ahmed, A., W.S. Aronow, and J.L. Fleg, 2006. "Higher New York Heart Association Classes and Increased Mortality and Hospitalization in Heart Failure Patients with Preserved Left Ventricular Function." *American Heart Journal* 151(2): 444-450.

Al-Khatib, S.M., A.S. Hellkamp, A.F. Hernandez, G.C. Fonarow, K.L. Thomas, et al. 2012. "Trends in Use of Implantable Cardioverter-Defibrillator Therapy among Patients Hospitalized for Heart Failure: Have the Previously Observed Sex and Racial Disparities Changed over Time?" *Circulation,* 125(9): 1094-1101.

Arellano, M., and S. Bond, 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58(2): 277-297.

Babcock, J.M., 1962. "Adoption of Hybrid Corn: A Comment." *Rural Sociology* 27:332–38.

Barber, B.M. and T. Odean, 2001. "Boys will be Boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics,* 116(1): 261–292.

Bardy, G.H., K.L. Lee, D.B. Mark, J.E. Poole, D.L. Packer, et al., 2005. "Amiodarone or an Implantable Cardioverter–Defibrillator for Congestive Heart Failure." *New England J Medicine,* 352(3): 225-237.

Berner, E. S., and M.L. Graber, 2008. "Overconfidence as a Cause of Diagnostic Error in Medicine." *The American Journal of Medicine,* 121(5), S2-S23.

Birkmeyer, J.D., J.F. Finks, A. O'Reilly, M. Oerline, A.M. Carlin, A.R. Nunn et al., 2013. "Surgical Skill and Complication Rates after Bariatric Surgery," *NEJM* 369: 1434-1442.

Camerer, C. and D. Lovallo, 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review,* 89(1): 306-318.

Caselli, F. and W. J. Coleman, 2006. "The World Technology Frontier." *American Economic Review,* 96(3):499–522.

Chan, D. C., M. Gentzkow, and C. Yu, "Selection with Variation in Diagnostic Skill: Evidence from Radiologists," *Quarterly Journal of Economics* (2022, forthcoming).

Chandra, A. and D. O. Staiger, 2007. "Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks." *The Journal of Political Economy,* 115(1): 103-140.

Chandra, A. and D.O. Staiger, 2010. "Identifying Provider Prejudice in Healthcare," (No. w16382). National Bureau of Economic Research WP 16382.

Chandra, A. and D.O. Staiger, 2020. "Identifying Sources of Inefficiency in Healthcare." *The Quarterly Journal of Economics,* 135(2): 785-843.

Comin, D., and B. Hobijn, 2004. "Cross-country Technological Adoption: Making the Theories Face the Facts." *Journal of Monetary Economics* 51(1): 39-83.

Comin D., and B. Hobijn, 2007. "Implementing Technology." NBER Working Paper 12886.

Comin, D., and B. Hobijn, 2010. "An Exploration of Technology Diffusion." *American Economic Review* 100(5): 2031-59.

Comin, D., and M. Mestieri, 2018. "If Technology Has Arrived Everywhere, Why Has Income Diverged?" *American Economic Journal: Macroeconomics*, 10(3): 137-78.

Conley, T.G. and C. R. Udry, 2010. "Learning about a New Technology: Pineapple in Ghana." *American Economic Review*, 100(1): 35-69.

Creswell, J., 2015. "A Small Indiana Town Scarred by a Trusted Physician." *The New York Times*, October 17.

Currie, J., and W. B. MacLeod, 2017. "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians." *Journal of Labor Economics* 35(1): 1-43.

Currie, J., W. B. MacLeod, and J. Van Parys, 2016. "Physician Practice Style and Patient Health Outcomes: The case of Heart Attacks." *Journal of Health Economics*, 47:64-80.

Cutler, D., J. Skinner, A.D. Stern, and D. Wennberg, 2019. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," *AEJ: Economic Policy*, 11(1): 192-221.

David, G., and B. Wu, 2022. "Information, Relative Skill, and Technology Abandonment," *Journal of Health Economics*, Vol. 83 102596.

DeGroot, Morris, 1971. *Probability and Statistics*, Addison-Wesley.

Duflo, E., M. Kremer, and J. Robinson, 2011. "Nudging Farmers to use Fertilizer: Theory and Experimental Evidence from Kenya. *The American Economic Review*, 101(6), 2350-2390.

Foster, A.D., and M. R. Rosenzweig, 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy*, 1176–1209.

Freeman, J.V., Y. Wang, J.P. Curtis, P.A. Heidenreich, and M.A. Hlatky, 2010. "The Relation between Hospital Procedure Volume and Complications of Cardioverter-Defibrillator Implantation from the Implantable Cardioverter-Defibrillator Registry." *Journal of the American College of Cardiology*, 56(14): 1133-1139.

Friedrich, E.B., and M. Böhm, 2007. "Management of End Stage Heart Failure." *Heart*, 93(5): 626–631.

Galasso, A., and Simcoe, T.S., 2011. "CEO Overconfidence and Innovation." *Management Science*, 57(8): 1469-1484.

Glaser, M., and M. Weber, 2007. "Overconfidence and Trading Volume." *The Geneva Risk and Insurance Review* 32(1): 1-36.

Gong, Q. "Physician Learning and Treatment Choices: Evidence from Brain Aneurysms," working paper, University of North Carolina, 2018.

Griliches, Z., 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, 25(4): 501–522.

Grilliches, Z., 1960. "Congruence versus Profitability: A False Dichotomy." *Rural Sociology* 25:354–56.

Grilliches, Z., 1962. “Profitability versus Interaction: Another False Dichotomy.” *Rural Sociology* 27:325–30.

Havens, A.E. and E. Rogers. 1961. “Adoption of hybrid corn: Profitability and the Interaction Effect.” *Rural Sociology* 26:409–14.

Jovanovic, B., and Y. Nyarko, 1995. “A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves.” *Brookings Papers on Economic Activity. Microeconomics,* 247–305.

Jovanovic, B., and Y. Nyarko, 1996. “Learning by Doing and the Choice of Technology.” *Econometrica,* 64(6): 1299–1310.

Jupiter, J., and D. Burke, 2013. "Scott's Parabola and the Rise of the Medical-Industrial Complex." *Hand*, 8(3): 249-252.

Kolata, G., 2017. “For Patients with Heart Failure, Little Guidance as Death Nears,” *New York Times,* November 7.

Malmendier, U., and G. Tate, 2015. “Behavioral CEOs: The Role of Managerial Overconfidence,” *Journal of Economic Perspectives* 29(4) (Fall): 37–60.

McMurray, J. J., 2016. “The ICD in Heart Failure - Time for a Rethink?” *New England J of Medicine,* 375(13): 1283-1284.

Moen, E. L., Bynum, J. P., Austin, A. M., Skinner, J., Chakraborti, G., and O’Malley, A. J., 2018. “Assessing Variation in Implantable Cardioverter Defibrillator Therapy Guideline Adherence with Physician and Hospital Patient-Sharing Networks.” *Medical Care* 56(4): 350-357.

Moore, D. A., and Healy, P. J., 2008. “The Trouble with Overconfidence.” *Psychological Review,* 115(2): 502-517.

Mullianathan, S., and Z. Obermeyer, 2022. “Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care,” Quarterly Journal of Economics 137(2): 679-727.

Olley, G. S. and A. Pakes, 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64(6): 1263-97.

O’Neill, H.M., R. W. Pouder, and A. K. Buchholtz, 1998. “Patterns in the Diffusion of Strategies across Organizations: Insights from the Innovation Diffusion Literature.” *Academy of Management Review,* 23(1):98–114.

Rogers E.M., 2004. *Diffusion of Innovations*. Simon and Schuster.

Roth, G.A., J.E. Poole, R. Zaha, W. Zhou, J. Skinner, and N.E. Morden, 2016. “Use of Guideline-Directed Medications for Heart Failure before Cardioverter- Defibrillator Implantation.” *Journal of the American College of Cardiology*, 67(9): 1062–1069.

Skinner J., and D. Staiger, 2007. “Technological Diffusion from Hybrid Corn to Beta Blockers.” *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches.* University of Chicago Press and NBER.

Skinner, J., and D. Staiger, 2015. “Technology Diffusion and Productivity Growth in Health Care.” *Review of Economics and Statistics*, 97(5): 951-964.

Suri, T., 2011. “Selection and Comparative Advantage in Technology Adoption.” *Econometrica,* 79(1):159–209.

Virani S.S., A. Alonso, E.J. Benjamin, M.S. Bittencourt, C.W. Callaway, A.P. Carson, et al., 2020.

"Heart Disease and Stroke Statistics—2020 Update: A Report from the American Heart Association." *Circulation*. 141(9): e139-596.

Wennberg, D.E., S.M. Sharp, G. Bevan, et al., 2014. "A Population Health Approach to Reducing Observational Intensity Bias in Health Risk Adjustment: Cross Sectional Analysis of Insurance Claims." *BMJ* 348: g2392.

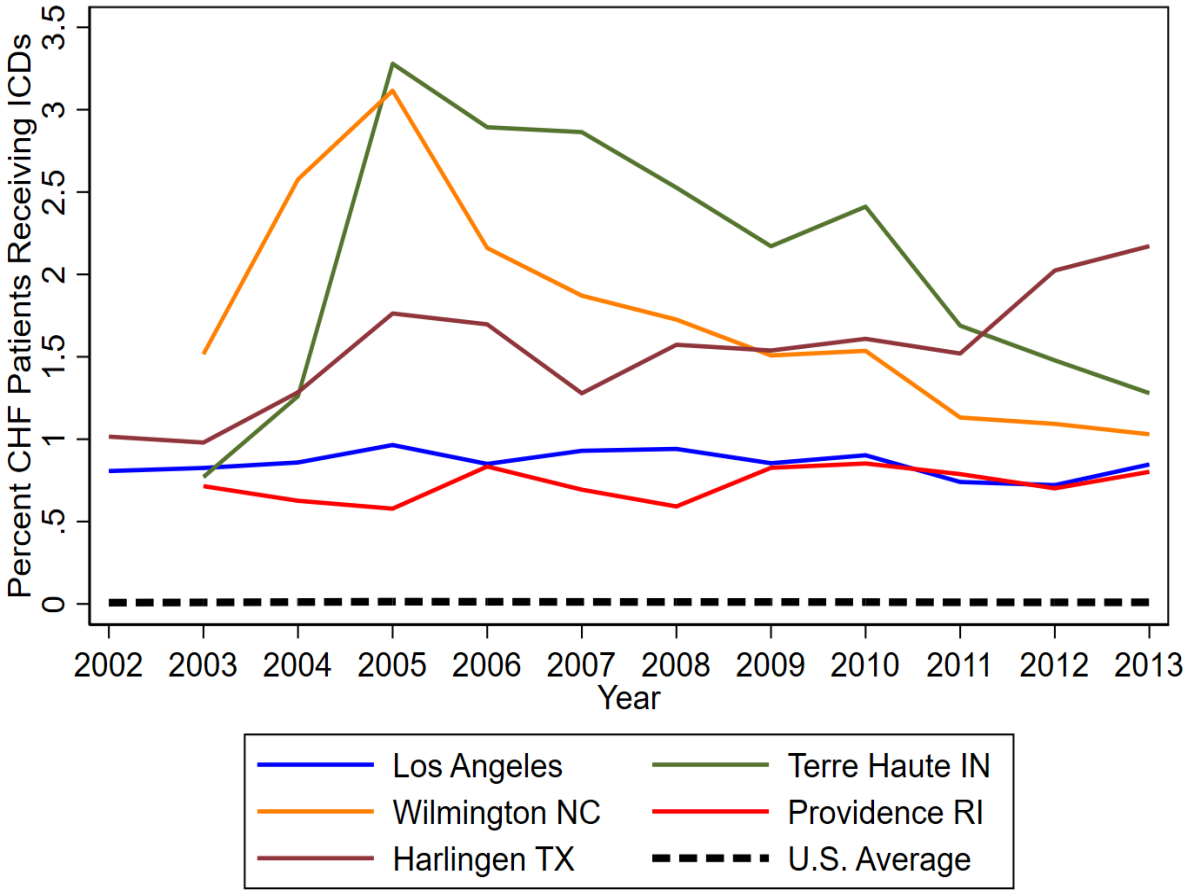

Figure 1: Implantable Cardioverter Defibrillator (ICD) Use for Patients with Diagnosed Congestive Heart Failure (CHF) for Selected Hospital Referral Regions: 2002-13

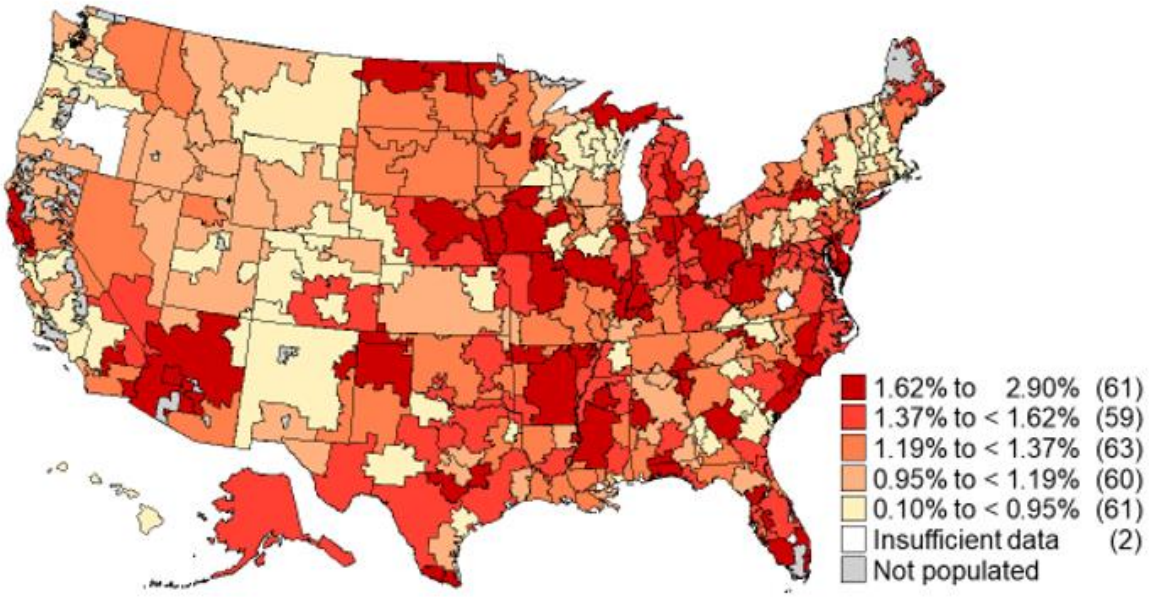

Figure 2. Percent of Congestive Heart Failure (CHF) Patients Receiving an Implantable Cardioverter Defibrillator (ICD), by Hospital Referral Region: 2006

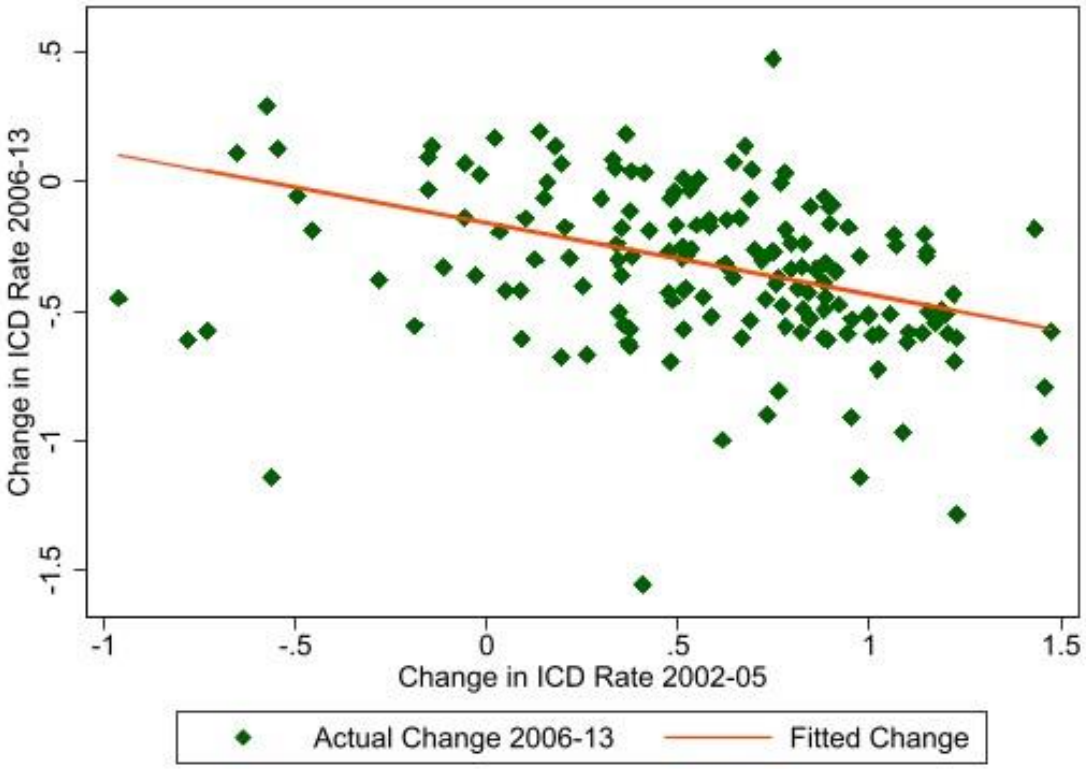

Figure 3: Correlation between 2002-05 and 2006-13 Changes in ICD Implantation Rates for Heart Failure Patients, by Hospital Referral Regions

Note: This sample is restricted to Hospital Referral Regions with at least 11 ICDs.

Figure 4a: Assigning regional ICD use rates to hospitals

Figure 4b: Assigning hospital-level mortality to HRRs

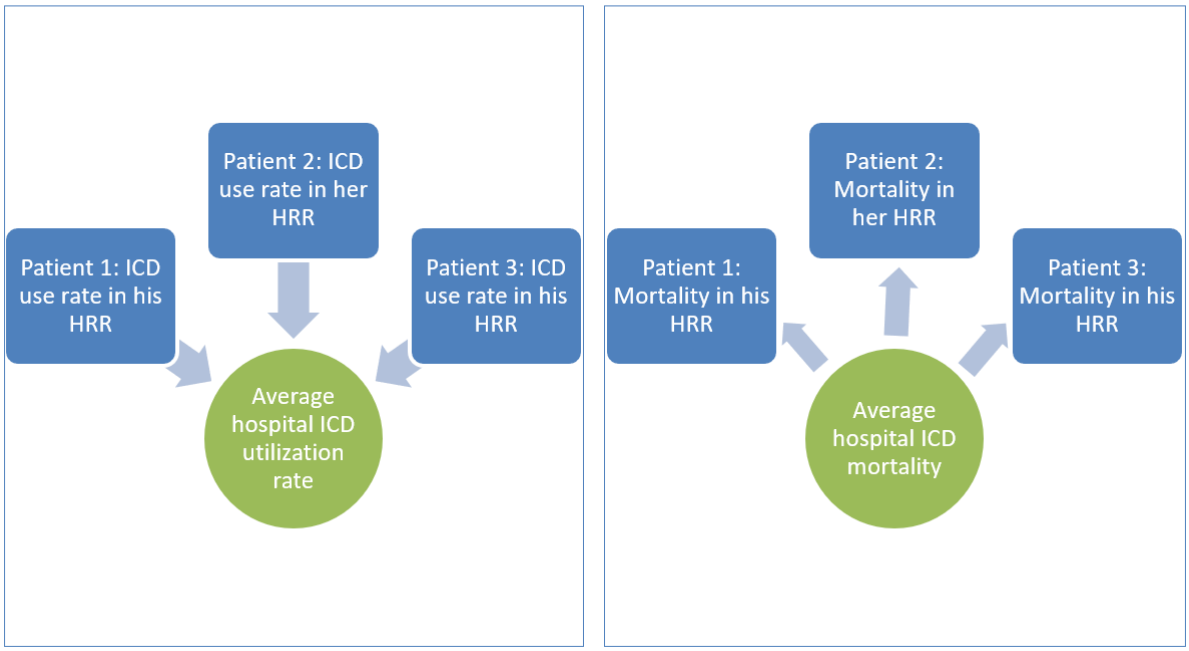

Figure 4: Schematic to Show How ICD Utilization Rates are Assigned to Hospitals, and How Hospital Mortality Rates are Assigned to Hospital Referral Regions (HRRs).

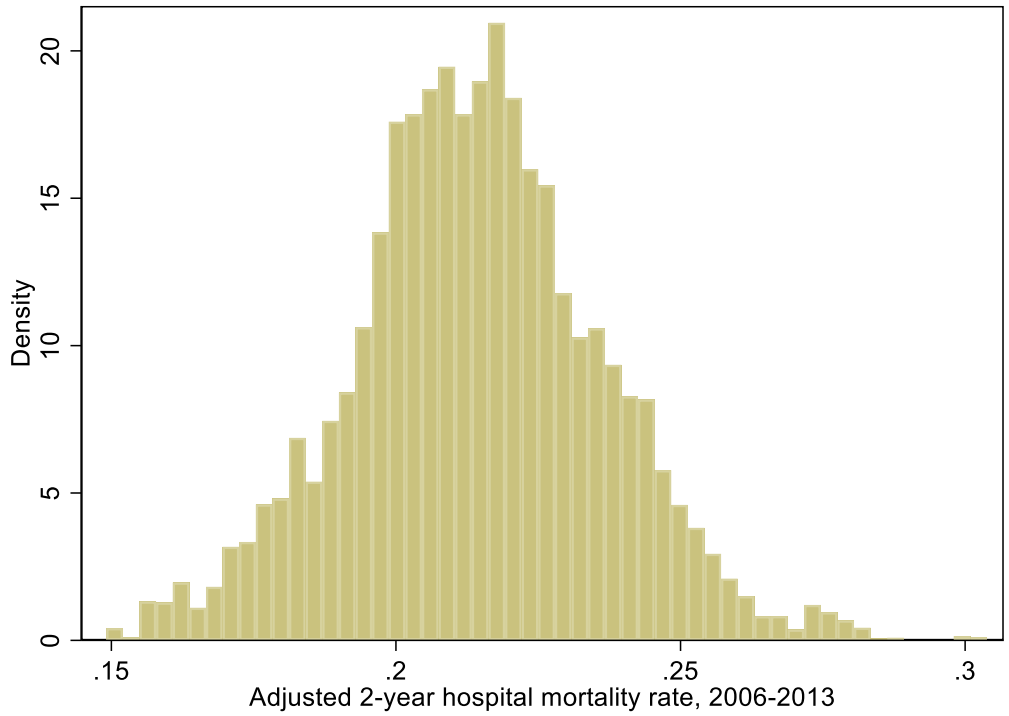

Figure 5: Distribution of Risk-adjusted Random-Effects 2-Year Mortality by Hospital: 2006-13

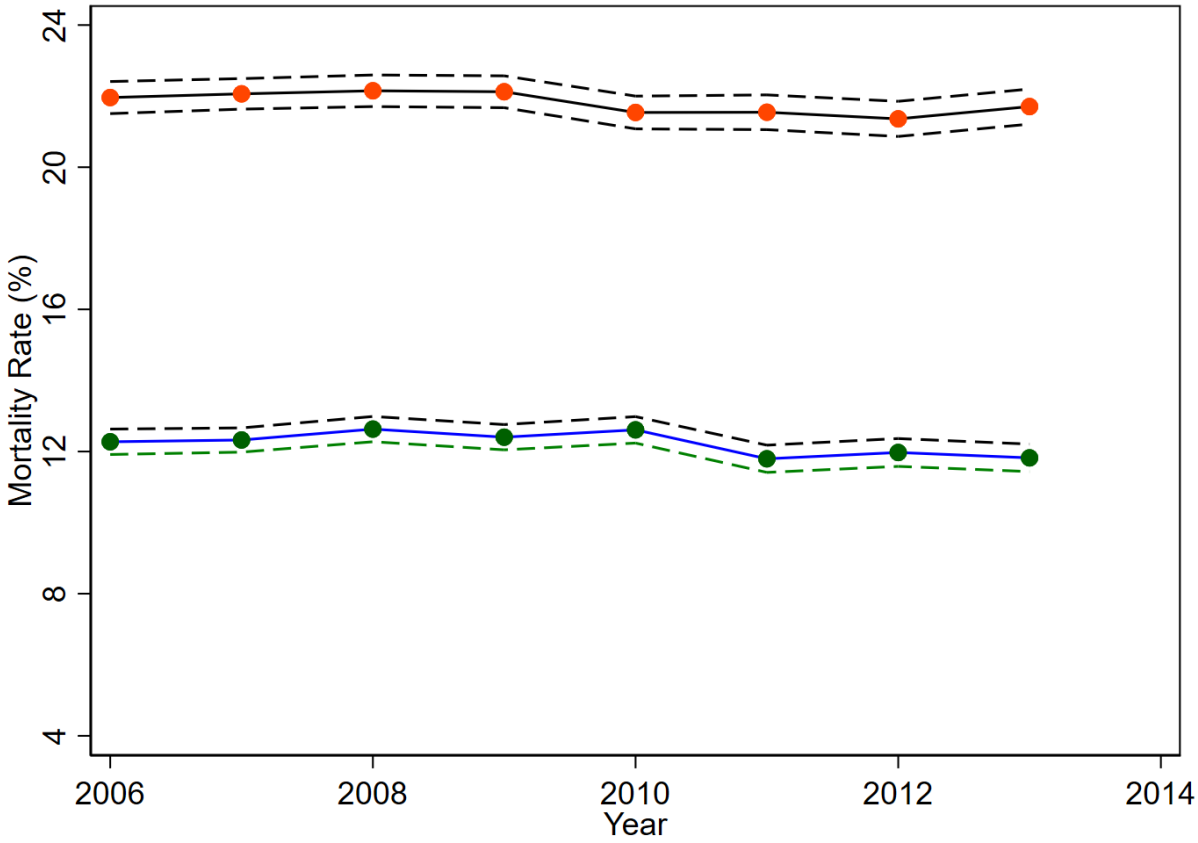

Figure 6: Evolution of One-Year (Bottom) and Two-Year (Top) Mortality Following ICD Procedure: 2006-13

Notes: 95% Confidence Intervals shown by dashed lines. Estimates based on patients receiving an ICD in the Registry with follow-up in the Medicare claims data. Aggregated data not risk-adjusted.

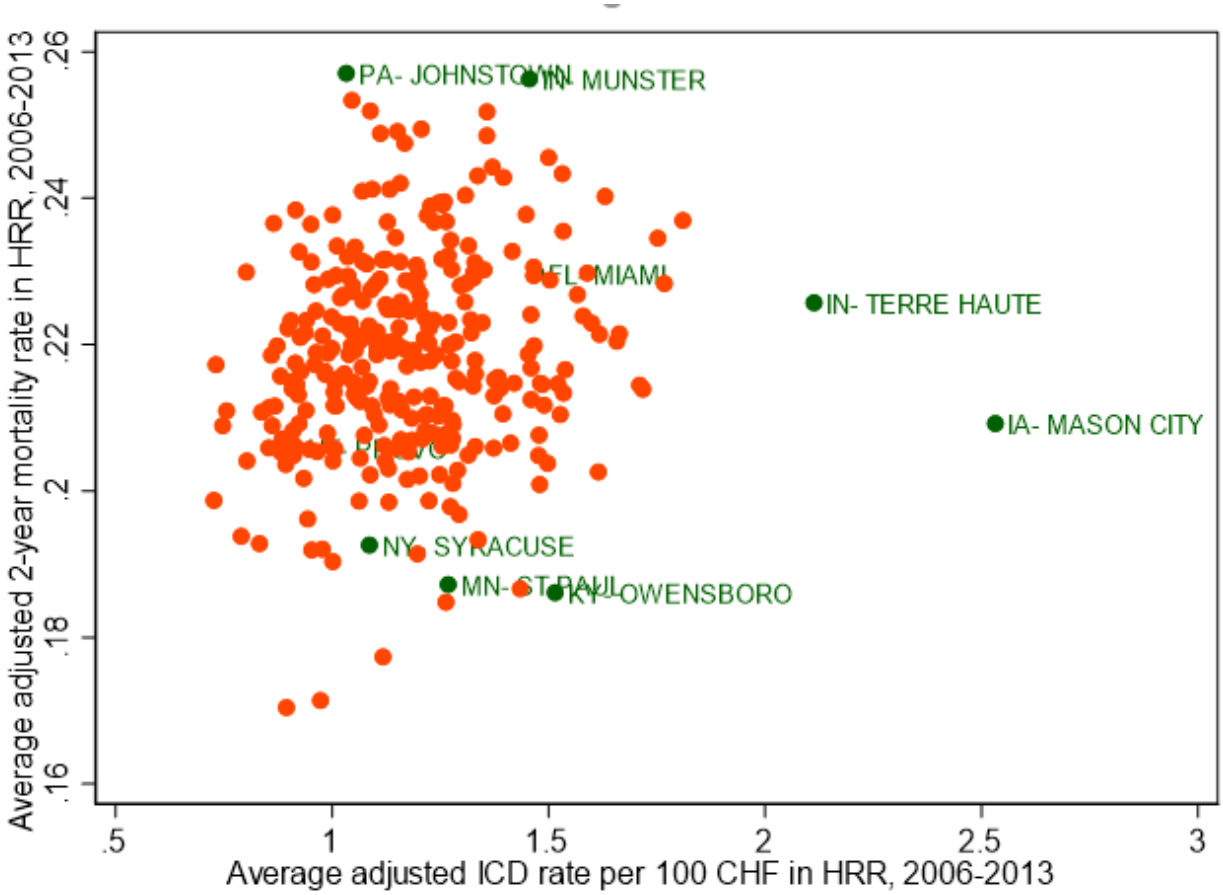

Figure 7: Correlation Between Average Implantable Cardioverter Defibrillator (ICD) Utilization (2006-13) and 2-Year Risk-adjusted Mortality

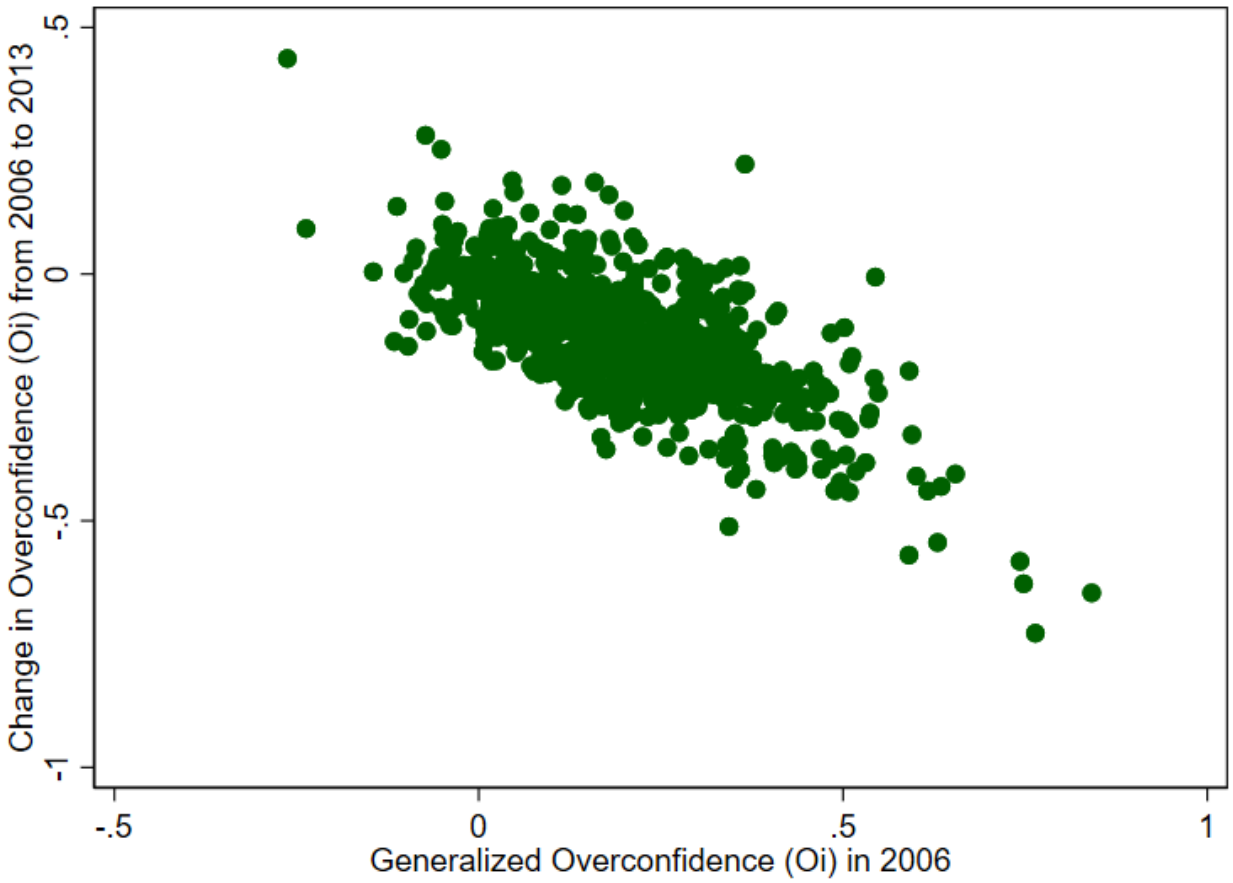

Figure 8A: Change in Overconfidence (2006-2013) from out-of-sample forecasts based on learning model vs. generalized overconfidence in 2006, by hospital.

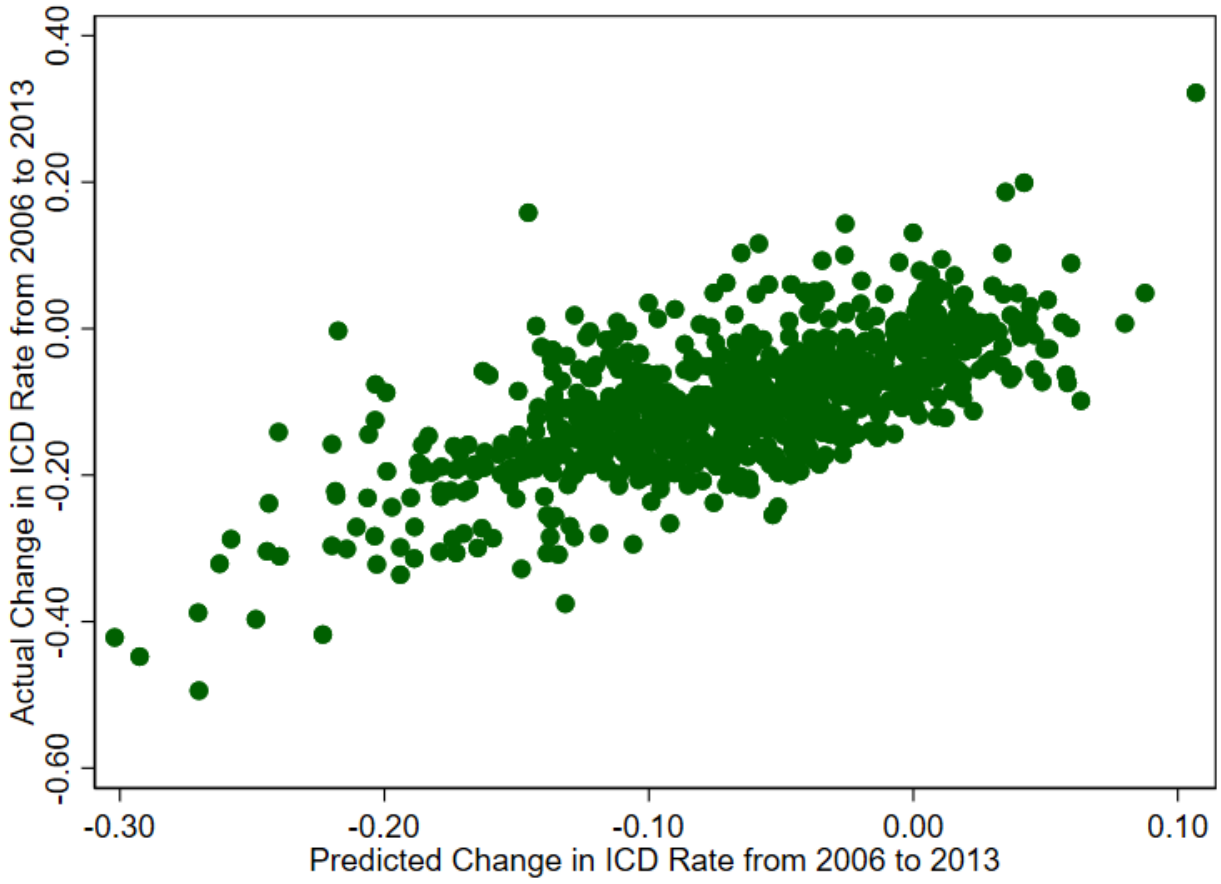

Figure 8B: Actual Change in ICD Utilization Rates versus Predicted Out-of-Sample Change in ICD Utilization Rates, 2007-13, by Hospital

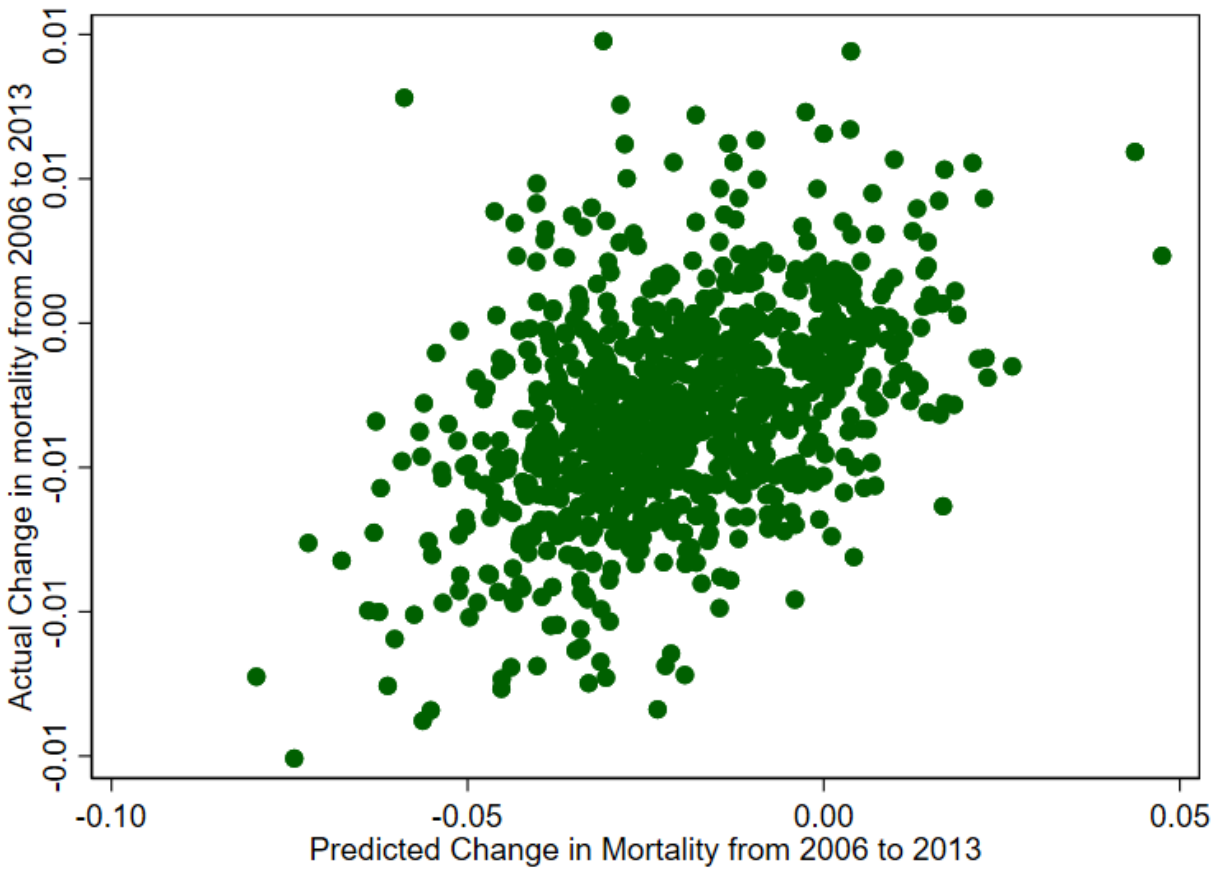

Figure 8C: Actual Change in Conditional Mortality Rates versus Predicted Out-of-Sample Change in Conditional Mortality Rates, 2007-13, by Hospital

Table 1: Summary Statistics for ICD Registry Data

Variable	Mean	Standard Deviation
2-Year Mortality: 2006-13	0.218	0.413
2-Year Mortality: 2006	0.219	0.414
2-Year Mortality: 2013	0.216	0.411
1-Year Mortality: 2006-13	0.123	0.328
1-Year Mortality: 2006	0.122	0.328
1-Year Mortality: 2012	0.118	0.323
Fraction Inappropriate	0.098	0.297
Ejection Fraction (Percentage)	25.76	7.319
Fraction with EF > 35%	0.034	0.182
Fraction Class I	0.029	0.169
Fraction Class IV	0.043	0.202
Age	74.90	6.248
Previous cardiac arrest	0.020	0.142
Family history: Sudden death	0.030	0.171
Ventricular tacchycardia	0.225	0.418
Non-isch, dilated cardiomyopathy	0.320	0.467
Ischemic heart disease	0.696	0.460
Previous myocardial infarction	0.548	0.498
Previous CABG	0.395	0.489
Previous PCI	0.345	0.475
Electrophysiology study	0.083	0.276
VT indication (ES study)	0.021	0.143
Female	0.282	0.450
Black	0.101	0.301
Hispanic (Medicare)	0.052	0.222
Other race	0.025	0.157
Hispanic ethnicity (Registry)	0.051	0.219

Notes: N=238,059. Sample includes all patients age 65 and over with CHF receiving a primary ICD between January 1, 2006 and December 31, 2013.

Table 2: Selected Characteristics of Patients Treated in Hospitals in the Top and Bottom Quartile of risk-adjusted ICD Rate

Variable	Bottom Quartile (low rate)	Top Quartile (high rate)
Risk-adjusted ICD rate at hospital	0.88%	1.59%
Inappropriate for ICD	9.02%	10.6%
>20% 1-year mortality risk	9.45%	11.71%
>20% 2-year mortality risk	55.2%	60.5%
Class IV CHF	3.69%	4.90%

Notes: N=59,107 in bottom quartile and 60,029 in top quartile. Patients were sorted into quartiles based on the risk-adjusted ICD rate for the hospital in which they received their ICD. Mortality risk based on prediction from mortality probit using clinical risk factors and demographics described in the text. All differences reported in the table between top and bottom quartile are highly statistically significant ($p<.001$).

Table 3: Regression Coefficients for OLS, Random, and Fixed Effects Models: Two-Year Mortality

VARIABLES	(1) OLS	(2) OLS	(3) Random Effect	(4) Random Effect	(5) Fixed Effect	(6) Fixed Effect
Hospital-level ICD Rate	1.049	1.427	0.952	1.288	0.238	0.590
	(0.441)	(0.432)	(0.434)	(0.429)	(0.683)	(0.694)
Ln(hospital volume)		-0.0126		-0.0126		-0.00951
		(0.00139)		(0.00132)		(0.00289)
Hospital-level Rx Rate		-0.121		-0.126		-0.120
		(0.0193)		(0.0187)		(0.104)
Observations	238,059	237,466	238,059	237,466	238,059	237,466
R-squared	0.046	0.047			0.058	0.058
Number of Groups			1,608	1,617		

Notes: Dependent variable is 2-year mortality. All regressions control for year effects and full set of clinical and demographic variables as reported in appendix table B2(a). Standard errors reported in parentheses, and are clustered at the hospital level in the OLS and FE models. Random effect model includes hospital and hospital-year random effects. Fixed effect models include hospital fixed effects.

Table 4: Calibration of Aggregate Parameters and Target Moments in the Model and Data

Means and Variances of Model Parameters		
	Mean	**Variance**
Potential value of the ICD: treatment (v_j)	-0.324	0.475
Potential value without an ICD: control (w_j)	0*	0.534
Skill (a_i)	0*	0.010
Overconfidence (o_i)	0.134	0.022
Moments in the Data and the Model		
	Data	**Model**
Average ICD rate for appropriate patients	.399	0.399
Standard Deviation of ICD rates across providers	.095	0.093
Average mortality conditional on ICD	.2187	0.2175
Standard deviation of mortality conditional on ICD	0.03	0.0385
Correlation of ICD rate and conditional mortality	0.092	0.092

Notes: Normalized means denoted by an asterisk. The variance of ε is assumed to be 1.0.

Table 5: Comparative Statics: Simulated ICD Utilization and Mortality Effects of Changes in Parameter Values

	(1) Data Baseline	(2) Mean of $o_i = 0$	(3) Variance $Var(o_i) = 0$	(4) Mean $o_i = 0$ and Variance $Var(o_i) = 0$	(5) Skill Variance $Var(a_i) = 0$
ICD Utilization (Fraction)	0.378	0.314	0.376	0.309	0.377
Standard Deviation of ICD Utilization Across Hospitals	0.091	0.086	0.036	0.033	0.067
Overall Mortality	0.175	0.172	0.172	0.170	0.174
Mortality Conditional on ICD	0.222	0.200	0.222	0.200	0.222
Standard Deviation of Conditional Mortality Across Hospitals	0.020	0.019	0.020	0.019	0.022
Correlation between ICD Utilization and Conditional Mortality	0.204	0.263	-0.998	-0.994	0.999

Table 6: Regressions Explaining the Hospital-Level Change in Generalized Overconfidence

Years:	(1) All	(2) All	(3) All	(4) 2006-7	(5) 2006-7
Generalized Overconfidence in Year t	-0.215	-0.219	-0.221	-0.213	-0.212
	(0.00711)	(0.00725)	(0.00769)	(0.0227)	(0.0233)
For-Profit Hospital		0.00497	0.00525		-0.00671
		(0.00184)	(0.00185)		(0.00547)
Government Hospital		0.00436	0.00447		-0.00491
		(0.00221)	(0.00221)		(0.00671)
Major Teaching Hospital		-0.00633	-0.00690		-0.00306
		(0.00184)	(0.00187)		(0.00571)
Constant	0.0104	0.0102	0.0105	0.0107	0.0124
	(0.00111)	(0.00121)	(0.00127)	(0.00432)	(0.00445)
Year Effects?	No	No	Yes	No	No
Observations	8,032	8,032	8,032	1,109	1,109
R-squared	0.145	0.147	0.221	0.156	0.157

Notes: The dependent variable is the change in generalized overconfidence from year t to t+1. The constant term reported in (3) is the average of the year effects.

Table 7: Calibration of Aggregate Parameters and Target Moments in the Model and Data

	ICD Use		**Conditional 2 Year Mortality**	
	Mean	Standard Deviation	Mean	Standard Deviation
2006: Data	0.438	0.102	0.224	0.021
2006: Model	0.438	0.102	0.224	0.021
2013: Data	0.334	0.071	0.221	0.021
2013: Model	0.369	0.047	0.204	0.018

Notes: Model initially calibrated to match 2006 actual data. Correlation by hospital between change in model and change in data is 0.68 (ICD use) and 0.43 (conditional 2-year mortality).

Appendix

A: Description of risk-adjustment to create regional measures of ICD use, 2002-13

The goal was to create HRR-level measures of ICD utilization in a cohort limited to congestive heart failure (CHF) patients. The cohort was created in each year by including Medicare enrollees aged 65+ (as of January 1), living in the U.S., and with full enrollment for the calendar year in fee-for-service Parts A and B, who fulfill the CHF definition from the Carrier Files with the following DX codes in any position; 40201, 40211, 40291, 425x-4259, 428x-4289, 4293, 40401, 40403, 40411, 40413, 40491, 40493. An ICD was determined by the Part B CPT code 33249. We used a 20 percent sample for 2002 (thus some of the regions do not report ICD rates because the total number of ICDs is less than 11), a 40 percent sample for 2003-05, and 100 percent for 2006 going forward.

We further seek to risk adjust ICD rates by considering additional health and socioeconomic factors that could affect the likelihood that the individual CHF patient receives an ICD. These include ZIP-code income (based on American Community Survey data), an individual measure for dual eligibility in Medicaid (which is either a marker for poor health, low income, or both), and county-level estimates of smoking, diabetes, obesity, and binge drinking. (Because the CDC ceased reporting some of the county level data in 2012, the later year-specific county health measures use the last year reported).

The regression estimates were estimated separately by year using a Probit model with HRR fixed effects; the HRR fixed effects, evaluated for a patient with average ICD risk, were used to create the risk-adjusted measures of ICD utilization rates. As can be seen from the results in Table B.2 below, once one conditions on CHF, the only strong predictors of ICD are age (with younger Medicare enrollees far more likely to receive the ICD) and dual-eligibility, which is negatively associated with ICD use. This negative association could be the consequence of poorer access to health care; more likely is that for the very sickest CHF patients, those with Class IV (well-advanced) CHF, ICDs are contraindicated.

Table B.1: Risk Adjustment Probit Regression Coefficients for ICD Implantation in Selected Years: 2002, 2006, and 2013

	2002		2006		2013	
Variable	Coefficient Estimate	Standard Error	Coefficient Estimate	Standard Error	Coefficient Estimate	Standard Error
Age 65-69	0.6989	*0.0224*	0.6332	*0.00687*	0.6096	*0.00761*
Age 70-74	0.6531	*0.0219*	0.6123	*0.00672*	0.5841	*0.00753*
Age 75-79	0.5738	*0.0218*	0.5611	*0.00663*	0.5361	*0.00759*
Age 80-84	0.4279	*0.0228*	0.4118	*0.00689*	0.4205	*0.00783*
Female	-0.5229	*0.0169*	-0.455	*0.00503*	-0.3801	*0.00599*
Black	0.0604	*0.0382*	0.0236	*0.012*	0.0146	*0.0127*
White	0.1177	*0.0320*	0.0457	*0.010*	0.00635	*0.0103*
Log ZIP Median Income	0.0168	*0.0301*	0.0122	*0.0103*	0.0183	*0.012*
% ZIP in Poverty	-0.2123	*0.1415*	-0.0247	*0.0468*	0.0531	*0.0541*
Dual-Eligible	-0.1981	*0.0169*	-0.1594	*0.00533*	-0.1334	*0.00606*
County Smoking Rate	-0.00548	*0.00226*	-0.00148	*0.00071*	0.00008	*0.00089*
County Diabetes Rate	-0.00773	*0.00685*	-0.00832	*0.00205*	-0.00824	*0.00208*
County Obesity Rate	0.00052	*0.00244*	0.00235	*0.00076*	0.00301	*0.00082*
County Binge Drinking	0.00027	*0.00348*	0.00281	*0.00127*	0.0021	*0.00136*
Sample % of Medicare	20%		100%		100%	
Sample Size	891,352		4,494,897		3,840,332	
Mean ICD Rate	0.0066		0.0127		0.0099	

Sample comprises people who have been diagnosed with CHF in that year (HCC85 = 1).

Excluded variables: Age = 85+, Race/ethnicity = Other (Includes Hispanic)

Table B.2(a): Random-Effects (RE) Regression Estimates (Full Set of Regression Coefficients)

	(1) Death1yr	(2) Death1yr	(3) Death1yr	(4) Death1yr	(5) Death2yr	(6) Death2yr	(7) Death2yr	(8) Death2yr
HRR-level ICD Rate	1.076***	1.464***	1.287***	0.910***		0.952**	1.555***	1.288***
	(0.340)	(0.340)	(0.338)	(0.339)		(0.434)	(0.431)	(0.429)
Ln(volume)		-0.00795***	-0.00794***				-0.0125***	-0.0126***
		(0.00104)	(0.00104)				(0.00133)	(0.00132)
HRR-level Rx Rate			-0.0895***	-0.0908***				-0.126***
			(0.0145)	(0.0148)				(0.0187)
Ejection Fraction (EF) <20%	-0.00344***	-0.00345***	-0.00345***	-0.00344***	-0.00467***	-0.00468***	-0.00469***	-0.00468***
	(0.000353)	(0.000353)	(0.000354)	(0.000354)	(0.000442)	(0.000442)	(0.000443)	(0.000443)
EF 20-25%	-0.00463***	-0.00463***	-0.00463***	-0.00463***	-0.00556***	-0.00556***	-0.00554***	-0.00554***
	(0.000407)	(0.000407)	(0.000408)	(0.000408)	(0.000510)	(0.000510)	(0.000510)	(0.000510)
EF 25-30%	-0.00212***	-0.00213***	-0.00213***	-0.00212***	-0.00361***	-0.00363***	-0.00364***	-0.00362***
	(0.000396)	(0.000396)	(0.000397)	(0.000397)	(0.000496)	(0.000496)	(0.000497)	(0.000497)
EF 30-35%	-0.000750	-0.000741	-0.000738	-0.000747	-0.00113*	-0.00112*	-0.00110*	-0.00111*
	(0.000481)	(0.000480)	(0.000481)	(0.000481)	(0.000601)	(0.000601)	(0.000602)	(0.000602)
EF > 35%	0.00161***	0.00161***	0.00160***	0.00160***	0.00213***	0.00213***	0.00210***	0.00210***
	(0.000304)	(0.000304)	(0.000305)	(0.000305)	(0.000381)	(0.000381)	(0.000382)	(0.000382)
EF Missing	0.0191***	0.0185***	0.0189***	0.0196***	0.0236***	0.0227**	0.0218**	0.0228**
	(0.00712)	(0.00712)	(0.00715)	(0.00715)	(0.00891)	(0.00891)	(0.00895)	(0.00895)
NY Heart Assoc. Class II	0.00221	0.00217	0.00171	0.00178	0.00378	0.00373	0.00324	0.00332
	(0.00407)	(0.00407)	(0.00408)	(0.00408)	(0.00510)	(0.00510)	(0.00510)	(0.00511)
NY Heart Assoc. Class III	0.0476***	0.0477***	0.0473***	0.0472***	0.0690***	0.0691***	0.0686***	0.0685***
	(0.00400)	(0.00400)	(0.00401)	(0.00401)	(0.00501)	(0.00501)	(0.00502)	(0.00502)
NY Heart Assoc. Class IV	0.155***	0.154***	0.154***	0.154***	0.189***	0.188***	0.188***	0.188***
	(0.00508)	(0.00508)	(0.00508)	(0.00509)	(0.00636)	(0.00636)	(0.00636)	(0.00637)
NY Heart Assoc. Class missing	0.0497***	0.0484***	0.0486***	0.0499***	0.0783***	0.0763***	0.0768***	0.0789***
	(0.0121)	(0.0121)	(0.0121)	(0.0122)	(0.0152)	(0.0152)	(0.0152)	(0.0152)
Age 70-74	0.0154***	0.0154***	0.0156***	0.0155***	0.0279***	0.0280***	0.0279***	0.0278***
	(0.00188)	(0.00188)	(0.00189)	(0.00189)	(0.00236)	(0.00236)	(0.00236)	(0.00236)
Age 75-79	0.0363***	0.0364***	0.0365***	0.0364***	0.0647***	0.0649***	0.0649***	0.0647***
	(0.00190)	(0.00190)	(0.00190)	(0.00190)	(0.00238)	(0.00238)	(0.00238)	(0.00238)
Age 80-84	0.0624***	0.0625***	0.0626***	0.0625***	0.109***	0.110***	0.110***	0.109***
	(0.00208)	(0.00208)	(0.00208)	(0.00208)	(0.00260)	(0.00260)	(0.00260)	(0.00260)
Age 85-89	0.0988***	0.0989***	0.0987***	0.0986***	0.171***	0.172***	0.171***	0.171***

	(0.00297)	(0.00297)	(0.00297)	(0.00297)	(0.00371)	(0.00371)	(0.00372)	(0.00372)
Age 90+	0.179***	0.179***	0.179***	0.179***	0.271***	0.272***	0.272***	0.272***
	(0.00801)	(0.00801)	(0.00801)	(0.00802)	(0.0100)	(0.0100)	(0.0100)	(0.0100)
Previous cardiac arrest	0.0551***	0.0546***	0.0546***	0.0551***	0.0544***	0.0537***	0.0537***	0.0544***
	(0.00470)	(0.00470)	(0.00470)	(0.00470)	(0.00588)	(0.00588)	(0.00588)	(0.00589)
Family history sudden arrest	-0.0121***	-0.0121***	-0.0121***	-0.0121***	-0.0182***	-0.0183***	-0.0183***	-0.0183***
	(0.00390)	(0.00389)	(0.00390)	(0.00390)	(0.00488)	(0.00488)	(0.00488)	(0.00488)
Ventricular tachycardia	0.0444***	0.0444***	0.0443***	0.0443***	0.0566***	0.0566***	0.0564***	0.0564***
	(0.00166)	(0.00166)	(0.00166)	(0.00166)	(0.00208)	(0.00207)	(0.00208)	(0.00208)
Non-ischemic dilated cardiomyopathy	-0.0199***	-0.0195***	-0.0195***	-0.0198***	-0.0298***	-0.0293***	-0.0291***	-0.0296***
	(0.00239)	(0.00238)	(0.00239)	(0.00239)	(0.00299)	(0.00298)	(0.00299)	(0.00299)
Ischemic heart disease	0.0191***	0.0195***	0.0195***	0.0191***	0.0276***	0.0281***	0.0282***	0.0276***
	(0.00254)	(0.00254)	(0.00255)	(0.00255)	(0.00318)	(0.00318)	(0.00319)	(0.00319)
Previous MI	0.00900***	0.00907***	0.00908***	0.00901***	0.0137***	0.0138***	0.0138***	0.0137***
	(0.00170)	(0.00170)	(0.00171)	(0.00171)	(0.00213)	(0.00213)	(0.00214)	(0.00214)
Previous CABG	0.00810***	0.00813***	0.00799***	0.00797***	0.0196***	0.0196***	0.0196***	0.0195***
	(0.00161)	(0.00161)	(0.00161)	(0.00161)	(0.00202)	(0.00202)	(0.00202)	(0.00202)
Previous PCI	-0.0106***	-0.0106***	-0.0107***	-0.0107***	-0.0124***	-0.0124***	-0.0125***	-0.0125***
	(0.00158)	(0.00158)	(0.00158)	(0.00158)	(0.00198)	(0.00198)	(0.00198)	(0.00198)
Electrophysiology study	-0.0181***	-0.0172***	-0.0175***	-0.0184***	-0.0259***	-0.0246***	-0.0248***	-0.0261***
	(0.00287)	(0.00287)	(0.00287)	(0.00287)	(0.00360)	(0.00359)	(0.00360)	(0.00360)
VT indication (ES study)	-0.00426	-0.00392	-0.00392	-0.00427	-0.00352	-0.00295	-0.00299	-0.00356
	(0.00539)	(0.00538)	(0.00539)	(0.00539)	(0.00674)	(0.00674)	(0.00674)	(0.00675)
Female	-0.00443**	-0.00438**	-0.00456**	-0.00459**	-0.0129***	-0.0129***	-0.0131***	-0.0132***
	(0.00197)	(0.00197)	(0.00197)	(0.00197)	(0.00247)	(0.00247)	(0.00247)	(0.00247)
Black	0.0196***	0.0200***	0.0197***	0.0193***	0.0340***	0.0347***	0.0343***	0.0337***
	(0.00282)	(0.00281)	(0.00281)	(0.00282)	(0.00354)	(0.00353)	(0.00353)	(0.00354)
Hispanic	0.00123	0.000668	0.000814	0.00140	0.00216	0.00132	0.00121	0.00209
	(0.00333)	(0.00333)	(0.00333)	(0.00334)	(0.00418)	(0.00418)	(0.00418)	(0.00419)
Other race	0.00772*	0.00724*	0.00719*	0.00770*	0.0125**	0.0118**	0.0114**	0.0121**
	(0.00433)	(0.00433)	(0.00433)	(0.00433)	(0.00542)	(0.00542)	(0.00542)	(0.00543)
County Smoking Rate	-0.000106	-9.76e-05	-0.000118	-0.000123	0.000245	0.000253	0.000204	0.000201
	(0.000238)	(0.000237)	(0.000236)	(0.000237)	(0.000301)	(0.000298)	(0.000297)	(0.000300)
County Obesity Rate	-0.00121***	-0.00120***	-0.00116***	-0.00117***	-0.00118***	-0.00117***	-0.00113***	-0.00114***
	(0.000246)	(0.000245)	(0.000244)	(0.000245)	(0.000311)	(0.000309)	(0.000308)	(0.000310)

County Diabetes Rate	0.00353***	0.00344***	0.00327***	0.00336***	0.00435***	0.00424***	0.00399***	0.00411***
	(0.000575)	(0.000572)	(0.000570)	(0.000573)	(0.000730)	(0.000723)	(0.000720)	(0.000728)
County Binge Drinking Rate	1.72e-05	-4.64e-05	0.000145	0.000210	1.60e-05	-7.21e-05	0.000196	0.000286
	(0.000281)	(0.000278)	(0.000277)	(0.000281)	(0.000360)	(0.000354)	(0.000353)	(0.000359)
ZIP Code median income	-2.21e-08	-1.70e-08	-2.87e-08	-3.34e-08	-1.12e-07	-1.04e-07	-1.25e-07	-1.33e-07
	(6.68e-08)	(6.67e-08)	(6.66e-08)	(6.68e-08)	(8.40e-08)	(8.37e-08)	(8.36e-08)	(8.39e-08)
ZIP code poverty rate	0.0626***	0.0598***	0.0607***	0.0637***	0.0732***	0.0689***	0.0702***	0.0746***
	(0.0139)	(0.0139)	(0.0139)	(0.0139)	(0.0174)	(0.0174)	(0.0174)	(0.0174)
2007.year	0.00368	0.00524**	0.00499*	0.00346	0.00432	0.00675**	0.00620*	0.00380
	(0.00266)	(0.00267)	(0.00267)	(0.00267)	(0.00330)	(0.00330)	(0.00331)	(0.00330)
2008.year	0.00856***	0.0104***	0.0101***	0.00835***	0.00748**	0.0103***	0.00975***	0.00695**
	(0.00278)	(0.00279)	(0.00279)	(0.00278)	(0.00345)	(0.00345)	(0.00346)	(0.00345)
2009.year	0.00713**	0.00919***	0.00872***	0.00670**	0.00848**	0.0117***	0.0109***	0.00775**
	(0.00284)	(0.00285)	(0.00285)	(0.00284)	(0.00352)	(0.00353)	(0.00353)	(0.00353)
2010.year	0.0103***	0.0120***	0.0115***	0.00982***	0.00381	0.00657*	0.00558	0.00289
	(0.00300)	(0.00300)	(0.00301)	(0.00300)	(0.00373)	(0.00373)	(0.00374)	(0.00374)
2011.year	0.00657**	0.00783**	0.00703**	0.00583*	0.00850**	0.0104**	0.00901**	0.00716*
	(0.00333)	(0.00333)	(0.00333)	(0.00334)	(0.00416)	(0.00415)	(0.00415)	(0.00417)
2012.year	0.00976***	0.0103***	0.00966***	0.00916***	0.00869**	0.00952**	0.00826*	0.00753*
	(0.00341)	(0.00340)	(0.00340)	(0.00341)	(0.00426)	(0.00424)	(0.00424)	(0.00426)
2013.year	0.0106***	0.0112***	0.0102***	0.00974***	0.0157***	0.0166***	0.0153***	0.0145***
	(0.00339)	(0.00338)	(0.00338)	(0.00339)	(0.00423)	(0.00421)	(0.00422)	(0.00424)
ln(SD of Hospital RE)	-3.833***	-3.879***	-3.929***	-3.880***	-3.513***	-3.579***	-3.625***	-3.555***
	(0.0534)	(0.0558)	(0.0591)	(0.0563)	(0.0474)	(0.0503)	(0.0528)	(0.0495)
ln(SD of Hosp x Year RE)	-4.418***	-4.419***	-4.416***	-4.416***	-4.400***	-4.418***	-4.413***	-4.395***
	(0.181)	(0.181)	(0.181)	(0.180)	(0.262)	(0.271)	(0.268)	(0.260)
ln(SD of Individual Error)	-1.135***	-1.135***	-1.135***	-1.135***	-0.910***	-0.910***	-0.910***	-0.910***
	(0.00147)	(0.00147)	(0.00147)	(0.00147)	(0.00147)	(0.00147)	(0.00147)	(0.00147)
Constant	0.0901***	0.118***	0.175***	0.147***	0.152***	0.195***	0.277***	0.234***
	(0.0147)	(0.0150)	(0.0176)	(0.0174)	(0.0185)	(0.0190)	(0.0224)	(0.0222)
Observations	238,059	238,059	237,466	237,466	238,059	238,059	237,466	237,466
Number of Hospitals	1,617	1,617	1,608	1,608	1,617	1,617	1,608	1,608

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. For purpose of comparison, the log standard error of hospital, hospital x year, and individual random effects in the absence of the first three variables in the table is -3.821 (0.0526), -4.436 (0.187), and -1.135 (0.00147) for one-year mortality and -3.508 (0.0472), -4.416 (0.270), and -0.910 (0.00147) for two-year mortality. VT denotes ventricular tachycardia, MI myocardial infarction, CABG coronary artery bypass graft, PCI percutaneous coronary intervention,

Table B.2(b): Least-Squares Regression Estimates (Partial Set of Regression Coefficients)

	(1) Death1yr.	(2) Death1yr	(3) Death1yr	(4) Death1yr	(5) Death2yr	(6) Death2yr	(7) Death2yr	(8) Death2yr
HRR-level ICD Rate	1.180***	1.604***	1.386***	0.971***	1.049**	1.746***	1.427***	0.740*
	(0.330)	(0.335)	(0.327)	(0.322)	(0.441)	(0.447)	(0.432)	(0.426)
Ln(volume)		-0.00761***	-0.00760***			-0.0125***	-0.0126***	
		(0.00109)	(0.00108)			(0.00140)	(0.00139)	
HRR-level Rx Rate			-0.0880***	-0.0885***			-0.121***	-0.121***
			(0.0143)	(0.0149)			(0.0193)	(0.0203)
Observations	238,059	238,059	237,466	237,466	238,059	238,059	237,466	237,466
R-squared	0.035	0.035	0.036	0.035	0.046	0.047	0.047	0.047

Notes: All risk adjusters (as in Table B.2(a)) included in the regression analysis but not reported. Standard errors clustered at hospital level in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table B.2(c): Fixed-Effects Regression Estimates (Partial Set of Regression Coefficients)

	(1) death1yr	(2) death1yr	(3) death1yr	(4) death1yr	(5) death2yr	(6) death2yr	(7) death2yr	(8) death2yr
HRR-level ICD Rate	0.551	0.757	0.749	0.540	0.238	0.609	0.590	0.209
	(0.536)	(0.545)	(0.547)	(0.538)	(0.683)	(0.689)	(0.694)	(0.688)
Ln(volume)		-0.00512**	-0.00523**			-0.00925***	-0.00951***	
		(0.00228)	(0.00230)			(0.00286)	(0.00289)	
HRR-level Rx Rate			-0.0795	-0.0814			-0.120	-0.123
			(0.0810)	(0.0808)			(0.104)	(0.103)
Observations	238,059	238,059	237,466	237,466	238,059	238,059	237,466	237,466
R-squared	0.046	0.046	0.046	0.046	0.058	0.058	0.058	0.058

Notes: All risk adjusters (as in Table B.2(a)) included in the regression analysis but not reported. All models also include hospital fixed effects. Standard errors clustered at the hospital level in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

C: Hospital Level Measures of ICD Use and Mortality

Figure C.1 and C.2 plot the histograms of a_i and O_i identified from the data, as described in section 4.1.

Figure C.1: Histogram of Hospital-Level True Skill Estimates (a_i)

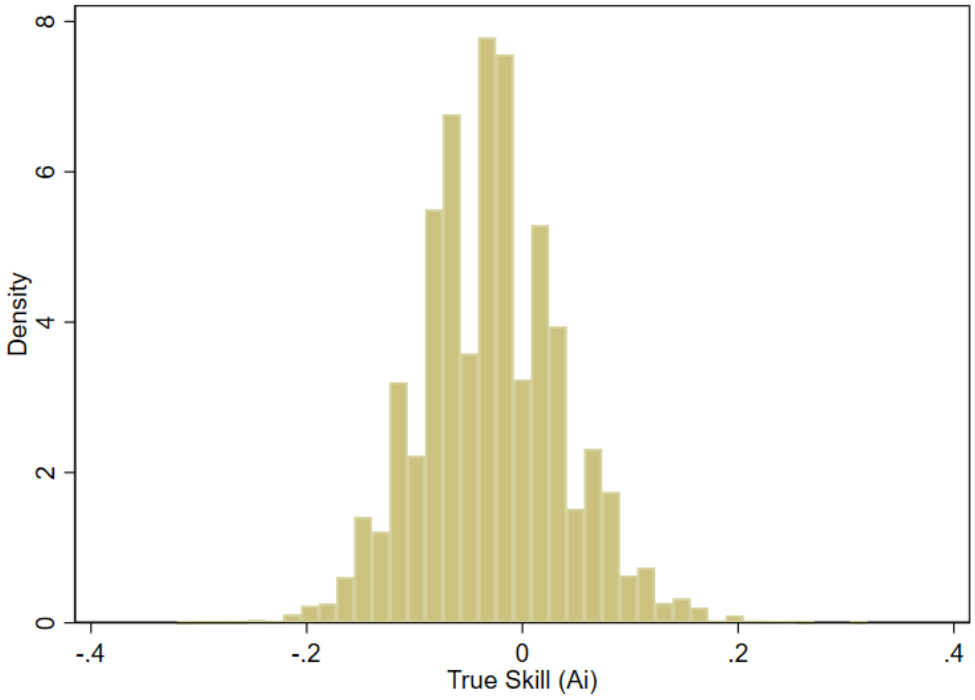

Figure C.2 Histogram of Hospital-Level Generalized Overconfidence Estimates (O_i).