Behaving Discretely
Heuristic Thinking in the Emergency Department

(Job Market Paper)

Stephen Coussens†

November 3, 2017

Link to most recent version: https://goo.gl/VyXCyj

Abstract

This paper explores the use of heuristics among highly-trained physicians diagnosing heart disease in the emergency department, a common task with life-or-death consequences. Using data from a large private-payer claims database, I find compelling evidence of heuristic thinking in this setting: patients arriving in the emergency department just after their 40th birthday are roughly 10% more likely to be tested for and 20% more likely to be diagnosed with ischemic heart disease (IHD) than patients arriving just before this date, despite the fact that the incidence of heart disease increases smoothly with age. Moreover, I show that this shock to diagnostic intensity has meaningful implications for patient health, as it reduces the number of missed IHD diagnoses among patients arriving in the emergency department just after their 40th birthday, thereby preventing future heart attacks. I then develop a model that ties this behavior to an existing literature on representativeness heuristics, and discuss the implications of this class of heuristics for diagnostic decision-making.

†Harvard Kennedy School, 79 JFK St, Cambridge, MA, 02138.
E-mail: coussens@fas.harvard.edu. Website: https://scholar.harvard.edu/coussens

I am grateful to my advisors Brigitte Madrian, David Cutler, and Ziad Obermeyer for their support and guidance. I thank Daniel Shoag, Nathan Hendren, Katie Coffman, and John Beshears for their helpful conversations. I also thank my classmates and all participants in the Harvard Economics Department Labor Workshop, Labor Lunch, and the Harvard Medical School Health Economics Seminar for their helpful comments and suggestions.
I Introduction

Most economic models make the simplifying assumption that the decision-maker is able to precisely optimize choice using all available information, both continuous and discrete in nature. Yet there also exists a well-established literature in both economics and psychology demonstrating the important role of heuristic decision-making (and the biases it can generate) in a variety of contexts (Gilovich et al., 2002; Kahneman et al., 1982). Evidence in both experimental (List, 2003, 2004; Alevy et al., 2015) and non-experimental (Lacetera et al., 2012) settings indicates that the effects of "nonstandard decision-making" (DellaVigna, 2009) on market transactions tend to be concentrated among naive agents with limited experience making the decision at hand. This suggests that anomalous behavior may not have a meaningful impact on settings in which the decision-makers are highly trained or experienced.

This paper provides evidence of an important environment in which this notion is violated. I explore the use of heuristics in a high-stakes setting in which the agents are highly trained professionals making familiar decisions. Specifically, I examine physician treatment patterns in the emergency department (ED) involving the diagnosis of ischemic heart disease (IHD), a common yet life-threatening condition. Given that age is a strong predictor of heart disease, the manner in which physicians incorporate this continuous attribute into their assessment of a patient’s risk for IHD is likely to meaningfully influence treatment decisions. So if physicians employ a cognitive shortcut that discretizes age into coarse categories, patients falling on either side of a category boundary may receive substantially different treatment, despite being otherwise similar.

Using private-payer health insurance claims relating to over 5 million ED visits in the U.S. between 2005 and 2013, I find evidence of emergency physicians using
a heuristic that classifies a subset of these patients as higher-risk for IHD if age 40 years or older, and lower-risk otherwise. This rule of thumb results in a sharp discontinuity in treatment patterns: patients entering the ED are roughly 10% more likely to be tested for a heart attack (a severe form of IHD) if they arrive shortly after their 40th birthday than if they had arrived just before this date. I argue that this is consistent with a representativeness-based heuristic (Kahneman and Tversky, 1972) whereby patients in their 30s are less representative of the prototypical heart attack patient than patients in their 40s. Further, I find that these discontinuities are most pronounced among ED patients possessing fewer of the characteristics most commonly associated with IHD, making their diagnosis less clear-cut.

What mechanisms appear to be driving this heuristic? There are a number of intertemporal factors that could influence physicians’ reliance on cognitive shortcuts like this one. Two of the most intuitive are fatigue and cognitive load. I examine the importance of these possibilities using ED records from a large Boston-area hospital. I find that the age-40 heuristic persists even when the emergency department patient volume is lower than usual (when cognitive load is below-average), as well as when patients arrive near the beginning of the physician’s shift (when the physician is less likely to be fatigued). This suggests that the observed behavior is not merely driven by unusually taxing environments, but is a relatively stable feature of physician decision-making in this setting.

As in Almond et al. (2010) and Almond and Doyle (2011), I also measure the effect of this discontinuous physician behavior on patients’ health outcomes. Specifically, I examine the rate of heart attacks subsequent to patients’ ED visits, which also serves as a proxy for IHD diagnoses that were likely missed during these visits. I find that the shock to diagnostic intensity generated by the heuristic substantially reduces the likelihood of missed diagnoses among a group of patients already known
by the medical community to be especially at risk for missed IHD diagnoses: women presenting without chest pain. This suggests that at least for this subpopulation, the marginal effort to detect IHD is of high value.

Despite this physiologically unjustified discontinuity in treatment intensity at age 40, it is important to note that in a general equilibrium sense, this heuristic is not necessarily suboptimal, as the cognitive effort it saves the physician may yield higher performance along other dimensions. Therefore I do not draw welfare conclusions about the use of heuristics more generally. Instead, I argue that undesirable features of heuristics represent a potential opportunity to leverage machine-learning prediction in decision-support systems. Heuristics generally provide an efficient means of arriving at near-optimal solutions, so there’s likely little reason to discourage their use in general. But if machines can identify narrow instances in which heuristics tend to lead physicians astray, " nudges" (Thaler and Sunstein, 2009) in these scenarios might provide a welfare improvement.

The remainder of this paper is organized as follows. Section II provides a discussion of the nature of decision-making in the ED with respect to patients who may be suffering from heart disease. Section III describes the data used in my analysis. Section IV discusses the empirical framework used to estimate the discontinuities generated by the age-40 heuristic. Section V presents the results. Section VI discusses the role of the representativeness heuristic in clinical decision-making, and presents a representativeness-driven model of diagnostic behavior. Section VII discusses the policy implications of heuristic thinking in medicine, and makes a case for an expanded role for electronic decision support systems in automatically identifying and attempting to correct problematic behavior. Section VIII concludes.
II The Clinical Decision

In many medical specialties, physicians are able to accrue knowledge of their patients’ health status over the course of several visits in which the physician observes, diagnoses, treats, and follows-up with their patients. However, physicians employed in a hospital’s emergency department are practicing in a uniquely low-information setting. Often the only evidence that the physician may use to assess a patient’s risk for a particular condition is the patient’s vital signs, demographic characteristics, medical history (as imperfectly recalled by the patient), and reported symptoms (Groopman and Prichard, 2007). Moreover, patient volume and limited hospital resources often require that this assessment be made in a matter of minutes.

Despite these limitations on the quantity and quality of the information readily available to emergency physicians, they are tasked with rapidly assessing the nature and severity of their patient’s condition, determining the extent of testing necessary, making a diagnosis, administering treatment, and deciding whether the patient should be admitted to the hospital or discharged. One of the most common diagnostic dilemmas facing ED physicians is the diagnosis of IHD. Given the severe health implications of missed IHD diagnoses, when assessing a patient presenting with symptoms that are consistent with this condition, a primary goal of the physician is to assess the likelihood of IHD as the underlying cause, despite the fact that the vast majority of such patients do not have a heart condition. The physician must therefore rule out an IHD diagnosis in a large number of low-risk patients in order to focus the hospital’s limited resources on those with the greatest risk.

When an ED physician considers the possibility that their patient has experienced a heart attack, she has the option to order first-line tests to aid in the diagnosis.
These tests are commonly used and inexpensive. While the tests are not monetarily costly, they require time to be processed – time during which the patient occupies an ED bed. EDs frequently operate at (or near) capacity, both in terms of bed and physician availability. Physicians in this setting are therefore unable to test patients indiscriminately. So among patients who appear to be relatively low-risk, how does the physician determine which patients will be tested, and which will be discharged without testing? As mentioned above, the patient history available to the emergency physician is limited, as is amount of time spent with the patient. Given the steep age gradient in incidence of heart attacks, age is almost certainly included as a component of this decision. Moreover, among younger patients, the incidence is at least twice as high for males as for females, making the sex of the patient a salient predictor as well (Mozaffarian et al., 2016). These facts, and their relationship to the representativeness heuristic, is discussed in greater detail in Section VI.

Research suggests that when interpreting a continuous characteristic, individuals may discretize it into coarse categories to which a decision rule can be more easily applied (Lacetera et al., 2012). Suppose that physicians were to exhibit such a tendency with respect to patient age while considering a diagnosis, and lumped patients into one of two groups: patients are "young" if under $\alpha$ years of age, and "old" otherwise. Were this the case, it would likely result in discrete treatment patterns – patients whose age falls just on either side of $\alpha$ may receive substantially different treatment, despite being otherwise comparable. Using the data described in Section III, I identify a pattern consistent with this hypothesis, which motivates the regression discontinuity (RD) design that I outline in Section IV.

\footnote{The two most common tests are the troponin blood test and the electrocardiogram (often abbreviated ECG or EKG). While the ECG is frequently used to detect a wide array of conditions, the troponin test is specifically used to detect heart attacks. Therefore my analyses regarding physician testing decisions will focus on the use of the troponin test.}
III Data

The empirical analysis in this paper makes use of two data sources: private-payer health insurance claims from the Truven Commercial Claims and Encounters database, and records from the emergency department of a large Boston-area hospital, hereafter referred to as the "Truven claims," and 'ED records', respectively. My analysis primarily focuses on the Truven claims, as it contains a far larger sample of emergency department visits, yielding superior statistical power. I supplement this approach with insights gleaned from the ED records, which offer a richer description of ED visits than the Truven claims.

Data Description - Truven Claims

The Truven claims provide data relating to private-payer health insurance claims (both inpatient and outpatient) for millions of individuals in the US. Truven obtains this information from a number of large employers, governments, and health plans, so individuals in this data are not a representative sample of the general population, although geographic coverage is fairly broad. All claims contain a unique individual-level identifier that allows for longitudinal tracking of all plan enrollees. However, I do not possess provider or geographic identifiers, so any heterogeneity that I explore in clinical decision-making will necessarily pertain only to the characteristics of the patient, and not those of the provider or region.

Throughout my analysis, I make use of the service dates, ICD-9 diagnosis codes and CPT procedure codes relating to each claim. These allow me to determine the diagnoses made and procedures performed during every patient interaction with a health care provider. Year of birth and sex (but not race or ethnicity) are available for each individual in the database. I am able to determine the month of birth for
roughly 95% of individuals in my analysis sample, which provides me with a finer measure of patient age during each visit than if I were to use the patient’s integer age. For the remaining 5% of the sample, I impute the month of birth.\footnote{My results are not sensitive to the exclusion of the 5% of individuals with an imputed month of birth. See Appendix I for details regarding the process by which month of birth was determined.}

**Analysis Sample - Truven Claims**

I impose minimal restrictions for inclusion in my analysis. The cohort of individuals included in the analysis sample is first restricted to those in the Truven claims who visit a hospital ED between 2005 and 2013.\footnote{Note that I was only able to identify ED heart attack testing claims for the period 2010 through 2013. Therefore, estimates of effects on testing rates are restricted to this period. Estimates for all other outcomes are generated using the full sample (2005 through 2013), which yields very similar point estimates, but provides much greater statistical power.} If a patient visits the ED more than once during this period, I include only the patient’s first visit in my sample, as physician behavior in subsequent ED visits may be endogenous to that of prior visits. All patients with a diagnosis of IHD prior to their first ED visit are also excluded. As the focus of this paper pertains to changes in physician treatment decisions around age 40, I further restrict the sample to only include individuals who were within 5 years of their 40th birthday at the time of their visit. Moreover, as my hypothesis hinges upon the physician’s knowledge of the patient’s integer age on the date of their ED visit, I also exclude patients whose visit takes place during the month of their 40th birthday, as I cannot determine these patients’ integer age on the day of their visit. After imposing these restrictions, my analysis sample consists of roughly 5.6 million patients, with more than one million patients arriving in the ED within one year of their 40th birthday.
Data Description - ED Records

While the Truven claims provide me with a large sample from which I can generate precise estimates, it lacks features that might allow me to test a wider range of hypotheses. To supplement my primary analysis of the Truven claims, I conducted an analysis of all ED visits to a large Boston-area hospital between January 2010 and May 2015. In addition to the types of data present in the Truven claims, these records also include timestamps for patient arrival and discharge, as well as attending physician shift schedules.

Analysis Sample - ED Records

The sample restrictions that I impose on the ED records are consistent with those of the Truven data: If a patient visits the ED more than once during this period, I keep only the patient’s first visit in my sample, and include only those who were within 5 years of their 40th birthday at the time of their visit. However, unlike in the Truven claims, since I observe the patients’ exact date of birth, there is no need for me to exclude patients visiting the ED during their month of birth. After imposing these restrictions, my analysis sample consists of approximately 20,000 patients.

IV Empirical Framework

Estimating Equation

Regression discontinuity designs are frequently employed in settings in which individuals’ ‘treatment’ status is at least partially determined by an arbitrary threshold in a continuous characteristic (‘running variable’) possessed by the population of interest. In this paper, my empirical strategy relies upon using the patients’ age as a running
variable. I posit that while patients arriving in the ED shortly before and after their 40th birthday are physiologically very similar, crossing this arbitrary threshold can change the way that they are perceived by the physician, which can in turn affect the manner in which the patient is treated. So while the impact of the age-40 threshold on a physician’s assessment of her patient’s IHD risk is unobserved, it can be inferred from the subsequent impact of this heuristic on the physician’s actions, such as testing and diagnosis.

In order to measure the discontinuity in physician behavior that is generated by the age-40 heuristic, I estimate a local linear regression using only patients visiting the ED near their 40th birthday, following standard methods such as those outlined in Lee and Lemieux (2010). Specifically, for patients whose age on the date of their ED visit falls within $h$ months of their 40th birthday, I estimate the following model:

$$Y_i = \beta_0 + \beta_1 \text{Turned}40 + \beta_2 (Age_i - 40) + \beta_3 \text{Turned}40 \times (Age_i - 40) + \epsilon_i$$  \hspace{1cm} (1)

where $Y_i$ represents a decision by a physician (such as tests, procedures, or diagnoses), or subsequent outcomes (such as future heart attacks) for patient $i$, and $\text{Turned}40_i = 1\{Age_i \geq 40\}$. The parameter of interest in this equation is $\beta_1$, which captures the discontinuous jump in the average value of $Y_i$ in patients arriving in the ED just after their 40th birthday relative to those arriving just before. This estimate of the discontinuity is unbiased if the underlying functional form of $Y$ is piecewise linear in age within $h$ months of the threshold. The methodology for determining the appropriate bandwidth $h$ is discussed in detail below.

Equation (1) is estimated using a triangular kernel centered at the age-40 threshold, so that the weight placed on each observation decays as the difference between the patient’s age and the threshold increases (Cheng et al., 1997). In the Truven
claims, the running variable is discrete since patient age is only available in month-age bins. To account for the implications of this discreteness for inference, I compute heteroscedasticity-robust standard errors that are clustered at the month-age level, following Lee and Card (2008). I select the bandwidth $h$ using the mean squared error-optimal procedure developed by Imbens and Kalyanaraman (2012) and implemented in Calonico et al. (2015). This method yields an optimal bandwidth of 1.24 years. However, since patient age is only observable in month-age bins in the Truven claims data, I round this number to the nearest month, resulting in a bandwidth of $h = 15$ months.

As alluded to in the introduction to this paper, my empirical strategy is similar to that of the instrumental variables framework applied in Almond et al. (2010): I estimate the effect of a shock on physician behavior, and then estimate the effect of the shock on subsequent health outcomes. While these estimates are intuitively akin to first stage and reduced form estimates, I do not interpret the ratio of the two as the local average treatment effect (LATE). I am able to demonstrate that the age-40 heuristic is a plausibly random shock to the treatment of patients near the threshold. However, there are likely many unobservable pertinent decisions that are also affected by the heuristic; these may impact subsequent health outcomes as well. So while the instrument is exogenous, the exclusion restriction does not hold with respect to any one observable action by the physician.

In other cases, such as Almond et al. (2010), this problem might be reasonably side-stepped by using hospital charges as a summary measure of treatment intensity; 

---

4 However, my results are insensitive to whether or not the standard errors are clustered.

5 This optimal bandwidth was estimated using the full analysis sample, with IHD diagnosis as the outcome. Re-estimating the optimal bandwidth for each outcome of interest yields very similar results. Therefore, for ease of interpretation, I use the same 15-month bandwidth for estimation of all discontinuities in the Truven claims. See Appendix II for the robustness of my estimates to alternative bandwidths.
this would not be feasible in my setting. In Almond et al. (2010), crossing the "very low birthweight" threshold triggered an increase in treatment intensity and hospital costs for newborns. In my paper, when the age-40 heuristic makes a physician more likely to expend resources pursuing a heart attack diagnosis, it might also make her less likely to expend resources considering an alternative diagnosis. This violates the monotonicity assumption that is required for a valid LATE estimate.

V Results

Impact on Physician Decisions

Figure 1 presents the percentage of all ED patients that are administered heart attack testing. This, like all other regression discontinuity plots that follow it, is presented as a function of patient age on the ED visit date. Raw means are plotted in quarter-age bins, with a local-linear smoother fitted separately to the underlying data on either side of the age-40 threshold. The figure shows an approximately linear relationship between testing and age on both sides of the threshold, with a sharp discontinuity at age 40. The slopes of these lines imply that the average increase in the probability of testing for each additional quarter of age is approximately 0.1 percentage points, while the increase in the probability of testing from the quarter before to the quarter after the patients’ 40th birthday is nearly an order of magnitude greater than this. Local linear estimation (as described in the previous section) yields an estimated 0.89 percentage point discontinuity in testing rates, which corresponds to a relative increase of 9.5% (Table 1).

Figure 2 shows a similar result with respect to the percentage of ED patients who are diagnosed with IHD: a fairly steep linear upward trend in the rate of diagnosis on either side of the threshold, with a sharp discontinuity at age 40. The local lin-
ear regression estimate of this discontinuity is approximately 0.13 percentage points, representing an even larger relative increase of 19.3% (Table 2). Likewise, Figure 3 and Table 3 show that the same pattern exists with respect to the proportion of ED patients that are admitted to the hospital with an IHD diagnosis.

Notable in Tables 1 through 3 is the disparity in the impact of the heuristic by patient sex. The baseline rate of testing and diagnosis for IHD in males is substantially higher than for females, which is to be expected, given the higher prevalence of IHD among males in this age group. However, the relative effects of crossing the age-40 threshold is substantially higher for women in each of the three measures: testing rates, IHD diagnosis rates, and IHD admission rates. Interestingly, a similar pattern is present with respect to whether the patient presented with chest pain, the most common symptom of IHD: the relative impact of the heuristic is nearly three times as large among patients presenting without chest pain as among patients presenting with it (Table 4). This suggests that less typical IHD patients are relatively more likely to have their treatment affected by the heuristic.

**Addressing Alternative Hypotheses**

Throughout this paper I argue that the differential treatment patterns around the age 40 threshold are driven by physician reliance upon an informal heuristic. But one could reasonably posit that this pattern is driven by patient choice instead – that is, suppose that it is the patient and not the physician for whom the age 40 threshold is salient. If there exists some subset of the population that, upon experiencing distressing symptoms, is significantly more likely to visit an ED as a result of turning 40, a difference in mean testing rates might emerge without any difference in physician

---

6 See Appendix II to examine the robustness of these local-linear estimates to alternative band-widths.
behavior. If this were the case, it would likely result in an observable shift in patient composition at age 40, generating discontinuities in the baseline prevalence of risk factors for IHD at this age. However, Table 5 demonstrates that patients’ relevant risk factors are well-balanced at the threshold, casting doubt on this hypothesis.

Additionally, if an influx of patients at age 40 were the primary driver of this pattern, one would expect to see a discontinuous increase in the number of patients arriving in the ED at age 40. Visual inspection of the distribution of ED visits by patient age (Figure 5) shows no such pattern. More formally, using the method developed in Frandsen (2017) to test for the presence of manipulation around a regression discontinuity cutoff when the running variable is discrete, no statistically significant difference across the threshold can be detected.\footnote{Even under the most stringent parameterization (i.e. $k = 0$) of the Frandsen (2017) test, $p > 0.5$.}

On the other hand, one might wonder whether the discontinuity in testing rates is driven by formal decision rules grounded in medical literature or hospital policy. However, risk assessment algorithms that ED physicians might use to aid their clinical decisions such as HEART (Six et al., 2008) or TIMI (Antman et al., 2000) do not include any guidance for patients specifically around age 40, and such algorithms are generally intended to be referenced \textit{after} initial tests are conducted. These diagnostic aides are therefore unlikely to be driving the observed behavior. Given the inexpensive nature of first-line tests to detect heart attacks, it also seems unlikely that hospitals would adopt a formal policy to actively avoid testing patients under 40 due to their low heart disease incidence. This is particularly improbable since younger patients are known to be at a higher than average risk for missed IHD diagnoses.\footnote{I thank Brigham Frandsen for making the code for this test readily available.} Moreover, the number of heart attacks among younger patients is non-trivial: as many as 10% of

\footnote{A piece of anecdotal evidence that is consistent with my argument: no such policy exists at the Boston-area hospital from which I obtained records, yet I nonetheless find the effect of the age-40 heuristic on physician behavior to be strong there.}
Behaving Discretely

Stephen Coussens

Heart attacks are experienced by individuals under the age of 45 (Hals and LoVecchio, 2009).

Moreover, as discussed above, I detect a discontinuous jump in the likelihood that patients are diagnosed with IHD as they cross the age-40 threshold (Figure 2), a pattern that holds true even within patients that are tested for IHD. This is likely to be inconsistent with any alternative hypothesis that predicts a jump in testing that is independent of physicians' perception of patient risk for IHD. Were any such hypothesis the primary driver of the observed discontinuity in testing, the average patient being tested in the ED just after their 40th birthday would have lower risk of IHD than those arriving just beforehand. This would generate a discontinuous drop in diagnosis rates within the tested group, rather than the observed jump.

Impact on Subsequent Health Outcomes

While the number of observations in the data allow for precise estimation of the heuristic’s impact on physician behavior, the number of patients affected by it — the 'compliers' (Angrist et al., 1996) — relative to the number of all ED visits is small. Although this group makes up approximately 10% of 40-year-old patients tested for a heart attack, they represent only about 1% of all ED visits. This poses issues for precise estimation of casual effects on many broad subsequent outcomes of interest, such as hospital readmission or charges. One way to address this problem is to exclude irrelevant portions of the population from the analysis, thereby reducing statistical noise in the outcomes. However, in my analysis, identifying irrelevant portions of the population is particularly difficult: the patients who present in the ED without the most common characteristics of IHD are the ones who are most likely to be affected by the heuristic (as discussed earlier in this section). For instance, restricting the sample to patients presenting with chest pain (the most common symptom of IHD)
would exclude the majority of the population of interest.

So, the relevant population cannot be readily identified ex-ante, and impacts of the heuristic on broad measures of health outcomes are likely to be swamped by statistical noise generated by the remaining 99% of the sample. I therefore focus instead on a subsequent health outcome that is specifically responsive to an ED physician’s decision to pursue an IHD diagnosis: subsequent heart attacks. Among all patients not diagnosed with IHD during their ED visit in my sample, the rate of heart attack diagnoses within 90 days following their visit is low – less than 1 in 1,000. However, conditional on being diagnosed with a heart attack in this time frame, it is very likely that the patient’s ED visit pertained to undiagnosed IHD. The frequency of subsequent heart attack diagnoses among patients not diagnosed with IHD can therefore serve as a proxy for missed IHD diagnoses in the ED.

There are a few caveats to bear in mind with respect to this proxy. On one hand, it will inevitably include some subsequent heart attacks that were unrelated to the patient’s ED visit, thus overstating the number of likely missed diagnoses. This bias increases in the length of the period observed following the visit. On the other hand, the shorter this period is, the less likely it is to capture further contacts with health care providers, mechanically restricting the likelihood of observing a subsequent heart attack diagnosis. The net effect of these two forces is ambiguous, but given that at least one-third of heart attacks go undiagnosed altogether (Hals and LoVecchio 2009), it seems likely that this proxy is a conservative measure of missed diagnoses.

As the prevalence of IHD increases with age, so does the number of missed diagnoses, as measured by the number of heart attacks diagnosed within 90 days of the

\[10\text{This assertion is supported by the data: roughly half of these patients had chest pain symptoms recorded during their ED visit, suggesting that their visit was indeed related to their subsequent heart attack diagnosis.}\]
patients’ ED visit. But while the difference in the cross-threshold missed diagnosis rate is not statistically significant for men, I find that women visiting the ED just after their 40th birthday experience significantly fewer missed diagnoses than those arriving just before (Figure 4, Table 6). Moreover, I find that this effect appears to be concentrated among women presenting without chest pain (Figure 5). Using the local linear estimation framework described in the previous section, for this group of women I estimate a drop in the missed IHD diagnosis rate of 21.3 per 100,000 ED visits upon crossing the threshold, implying a decrease of 53.9% (Table 7).

However, I qualify this point estimate with a caveat. Recall that patient age in this data is observed only at the month-age level. As discussed in Section III above, patients arriving in the ED in the month of their 40th birthday are therefore excluded from the analysis sample. As a result, my local linear regression estimates represent a ‘donut’ RD design, extrapolating from the linear functions’ support to the threshold. When the underlying functional form of interest is linear, this practice yields unbiased estimates of the discontinuity. In my estimates of the effect of the age-40 heuristic on physician testing and diagnosis patterns, this seems to be a reasonable assumption, as the global trends (at least between age 35 and 45) in these measures are nearly linear and very precisely estimated. Visual inspection of Figure 5 suggests that the same cannot be said for the relationship between age and missed IHD diagnoses to the right of the threshold. As a result of the steep slope of the regression line at 40 years and one month, linearly extrapolating to the age-40 threshold substantially increases the estimated magnitude of the discontinuity. A more conservative estimate

---

11For robustness, I also conducted my analysis using periods of 30 and 180 days – these estimates are consistent with those presented here (Appendix II).

12However, I have insufficient power to rule out a similar effect among women with chest pain.

13To demonstrate that this discontinuity is unlikely to be spurious, I conducted a falsification test in which I re-estimated the discontinuity at a series of placebo thresholds consisting of all feasible month-age bins in the analysis sample ($p < 0.05$).
of the discontinuity that does not rely upon functional form assumptions can be obtained by comparing mean misdiagnosis rates above and below age 40 within a very narrow neighborhood around the threshold. Given that IHD risk increases in age, the larger this neighborhood is, the more likely this estimate is to understate the true discontinuity. Using a neighborhood of 6 months, this yields an estimated drop in the missed IHD diagnosis rate among women without chest pain of 16.8 per 100,000 ED visits, or 42.5%\textsuperscript{14}. This estimate is statistically significant ($p < 0.01$), under both asymptotic and randomization inference procedures described in Cattaneo et al. (2015).

The reason that I observe this effect among female and not male patients is likely twofold. First, as discussed earlier in this section, the heuristic itself appears to generate a larger discontinuity in physicians’ behavior when treating female patients. Second, it is well-documented in the medical literature that women with IHD are under-diagnosed and under-treated relative to men with the same condition (Pope et al., 2000; Kim et al., 2016). Both of these facts suggest that women would be relatively more likely then men to benefit from the increase in diagnostic intensity that the heuristic generates at age-40. These results also imply that the marginal returns to additional diagnostic effort in screening women of this age for IHD are likely to be positive, an implication which I discuss in greater detail below.

Are women under-tested for IHD?

Given the estimated effect of the age-40 heuristic on subsequent heart attacks among women, it is relatively straightforward to conduct a back-of-the-envelope cost-benefit analysis demonstrating that on the margin, women under 40 do in fact appear to be

\textsuperscript{14}Note that this estimate is not very sensitive to the choice of neighborhood: considering all candidate neighborhoods 1 through 6 months from the age-40 threshold yields point estimates ranging from -11.8 to -17.0, with a median of -15.2.
under-tested for IHD. Doing so first necessitates assigning a monetary value to the
cost of testing for IHD, as well as to the benefits of preventing a heart attack. In
Appendix III I estimate the average cost of testing this population for IHD to be
approximately $250 per patient tested, or $250,000 per 100,000 ED visits.\footnote{Recall that in Table I I estimate that an additional 1\% of women visiting the ED are tested for IHD upon turning 40. Therefore, for 100,000 ED visits, only 1,000 additional patients are tested, at a total cost of $250,000 per 100,000 ED visits.}

On the other hand, if a diagnosis of IHD is made during the ED visit, the patient
will likely be placed on an inexpensive prophylactic treatment regimen, such as aspirin
or statin drugs, which greatly reduces the risk of a future heart attack. Roughly 15\% of
heart attacks are fatal\cite{Mozaffarian2016}, while\cite{Mahoney2002} estimate
that among patients with IHD, a non-fatal heart attack reduces life expectancy by
about one-eighth. If patients are assumed to have a relatively normal lifespan (e.g.
roughly 80 years) if their IHD is diagnosed and treated, the average heart attack
for someone around 40 years of age results in a loss of approximately 11 life-years.\footnote{15\% \times (80 - 40) + \frac{1}{8} \times (80 - 40) = 11 \text{ life-years.}}

Using the suggested benchmark valuation of $100,000 per life-year\cite{Neumann2014}, this implies that the value of preventing a heart attack for a patient of this age
to be roughly $1.1 million.\footnote{This likely to be a conservative valuation, as it implies that that a fatal heart attack at age 40 is valued at $4 million, which is less than half of the value of a statistical life used in US Department of Health and Human Services regulatory impact analyses\cite{HHS2016}.} Based on my more conservative estimate of the effect
of the age-40 heuristic on subsequent heart attacks among women, this implies a
benefit of $18.48 million per 100,000 female ED patients.\footnote{16.8 \times $1.1m = 18.48m}$ This is almost two orders
of magnitude greater than the estimated costs of testing this marginal population
for IHD. Thus it appears that under any reasonable set of alternative assumptions,
women at this age are under-tested for IHD.
Considering Role of Physician Cognitive Load and Fatigue

Most leading hypotheses attempting to address when individuals employ heuristics invoke the notion of bounded rationality ([Shah and Oppenheimer 2008](#)). In many settings, decision-makers have insufficient time or cognitive resources to make choices via exhaustive search or formal probabilistic reasoning, forcing them to rely upon heuristics instead. [Kahneman (2011)](#) discusses the tendency of individuals to switch from deliberative reasoning ("System 2") to heuristic thinking ("System 1") when available cognitive resources are low – that is, during or following particularly cognitively taxing tasks. This begs the question: is the age-40 heuristic only relevant to decision-making during especially demanding periods in the ED?

To explore this hypothesis, I analyze detailed records from the ED of a large Boston-area hospital, as described in Section III. Using the ED arrival timestamps recorded for each patient, I am able to generate proxies for physician cognitive load (as measured by ED patient volume) and physician fatigue (as measured by the order in which patients arrive during the physician’s shift). Since more physicians are scheduled to work during the typically busy times of the day, it may not be the case that the burden on any individual physician is actually higher during periods of peak patient flow. Therefore, I define an hour in which an above-median (for that hour of the day) number of patients arrive as a "high volume" period, and "low volume" otherwise. This measure of patient volume thus reflects deviations from typical patient flows for any given time of the day. Then, after merging the attending physicians’ shift schedules to these patient-level records, I can also determine the order in which patients were assigned to attending physicians within a given shift.[19](#)

Using these records, I am able to replicate the pattern found in the Truven claims

---

[19] However, the shift schedule that I obtained does not cover the entire sample; roughly 75% of patient visits could be linked directly to scheduled shifts.
data. Employing the same regression discontinuity methodology presented above, I find a large and statistically significant jump in the heart attack testing rate at age 40 (Figure 7). The point estimate of this discontinuity is substantially larger than in the claims data (although less precisely estimated), implying an 51.7% increase in the testing rate at the threshold (Table 8 column 1). In order to examine whether use of the age-40 heuristic is primarily driven by periods in which cognitive resources are depleted, I re-estimate the discontinuity separately among patients arriving in the ED during high vs. low-volume hours (Figure 8, Table 8 columns 2 & 3), and again for patients arriving during the second vs. first half of their physician’s shift (Figure 9 Table 8 columns 4 & 5). In each case, a large discontinuity in testing rates is present at age 40.

Counter-intuitively, these point estimates turn out to be larger when patient volume is lower and when patients are seen during the first half of the physician’s shift; these estimates are statistically significant ($p < .05$)\textsuperscript{20}. Though noisy, these results demonstrate that the discontinuity in testing rates at age 40 appears to persist even in settings in which available cognitive resources are presumably higher than average. Thus it seems that reliance upon this heuristic is not merely limited to unusually taxing conditions, but is instead a more integral part of physicians’ diagnostic process in the ED.

Why age 40?

It seems intuitive that a physician might informally categorize patients into coarse age bins to simplify diagnostic decisions involving patient age. However, it is not obvious why 40 years of age might be a particularly salient number for establishing a cutoff

\textsuperscript{20}However, the limited number of observations in these ED records preclude me from finding a statistically significant difference between these two sets of point estimates.
between "young" and "old" patients for the purpose of diagnosing IHD. Literature in both psychology and economics demonstrates that individuals seem to have a preference for round numbers (Lynn et al., 2013), and in some cases appear to anchor their decisions to them (Lacetera et al., 2012; Pope and Simonsohn, 2011), so perhaps it is not surprising that the cutoff age is round.

Among integers ending in a '0,' a few objective criteria seem to suggest that age 40 might be a reasonable choice. First, 40 years is approximately the midway point of the average lifespan in the US, making it a relatively natural (if naive) point at which to bifurcate age. Alternatively, suppose that the physician’s choice of threshold is driven by the relationship between age and IHD risk. Figure 10 shows the US annual per-capita heart attack mortality rate as a smooth, continuous function of age. If the physician is unable to fully internalize such a non-linear relationship, it could be simply approximated as a piece-wise linear function. Using the same data, the MSE-minimizing split point for such a function is approximately 40 years (Figure 11). Moreover, age 40 is roughly the inflection point in the heart attack testing rate curve (Figure 12), further supporting the notion of age 40 being a pivotal point with respect to testing decisions.

VI Theoretical Framework

The Role of Representativeness in Clinical decision-making

The process by which a physician learns to practice medicine is through an on-the-job training environment. Medical students and residents observe more senior physicians in action, and begin making diagnoses and performing procedures under their supervision. It is through this process that they learn what it looks like for a patient to be suffering from a heart attack, angina, acid reflux, or pneumonia. In economics,
a common way to model the diagnostic process would be to assume that the physician is a Bayesian agent. This implies that the physician arrives at a diagnosis by computing the posterior probability that a patient suffers from a medical condition given a set of observed characteristics. In reality, despite having completed a long and arduous education both in and outside of the classroom, physicians (like most people) struggle mightily with the estimation of conditional probabilities even in situations that are familiar to them (Casscells et al. 1978; Eddy 1982; Manrai et al. 2014). Yet on the whole, they are reasonably effective diagnosticians; how do we reconcile these two facts? I posit that the problem they are solving is not governed directly by objective probability, but by a form of reasoning that is nonetheless correlated with it: representativeness.

Kahneman and Tversky (1972) define representativeness as a commonly used heuristic in which a person judges the subjective probability of an event by the extent to which it is "(i) similar in essential properties to its parent population; and (ii) reflects the salient features of the process by which it is generated." I argue that this form of reasoning seems particularly relevant in the context of emergency department diagnoses, where algorithmic approaches to medicine are commonplace, and often amount to a matching exercise: How closely does a patient’s characteristics resemble the prototypical presentation of a particular condition? If "very close," make this diagnosis. If "not close at all," rule out this diagnosis. If "close enough," collect additional information and then re-evaluate.

To formalize this process, I propose a representativeness-based model of diagnostic decision-making which bears similarity to the "drift diffusion" model of Krajbich et al. (2014). Patient \( i \) with vector of characteristics \( X_i \) arrives in the emergency department, and is considered for a diagnosis of medical condition \( c \). \( X^c \) represents the prototypical presentation of characteristics consistent with condition \( c \). The sub-
objective risk of patient $i$ for condition $c$ is inversely related to the distance $d$ between $X_i$ and $X^c$: 

$$R^c_i \equiv 1 - \frac{d_c(X_i, X^c)}{d^{MAX}_c}; \quad R^c_i \in [0, 1]$$  

(2)

Distance $d$ is normalized by $d^{MAX}_c$ so that a patient perfectly matching the prototypical presentation of condition $c$ has a risk of 1, while a patient bearing no resemblance to the prototypical presentation has a risk of 0. The physician pursues a diagnosis of condition $c$ if the following are true:

1. $\underline{R^c} < R^c_i < \overline{R^c}$, and

2. $R^c_i > \frac{K_t}{B^c_i}$

where $K_t$ is the cost of test $t$, and $B^c_i$ is the benefit of running the test on a patient with $c$. That is, diagnostic tests are conducted (the results of which are incorporated into patient characteristics $X_i$) until subjective risk $R^c_i$ falls below $\underline{R^c}$, at which point the diagnosis is ruled out, or rises above $\overline{R^c}$, at which point a diagnosis of condition $c$ is made, or until the cost of any remaining test is greater than its expected benefits.

Note that in equation (2) I make no restrictions on the functional form of distance function $d$, thereby allowing for the possibility that physicians may misweight the relevance of characteristics in determining the patient’s risk. So upon what information might the physician (mis)weight the relative importance of patient characteristics when making a diagnosis? As alluded to earlier, numerous studies have demonstrated the proclivity of physicians to make rather serious mistakes when estimating conditional probabilities in a clinical setting. In [Casscells et al. (1978)], physicians at Harvard Medical School teaching hospitals are asked the following question:

"If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5 per cent, what is the chance that a person found to have a positive
result actually has the disease, assuming that you know nothing about the
person’s symptoms or signs?*

A straightforward application of Bayes’ theorem shows that this conditional prob-
ability is less than 2%, yet the modal response from those surveyed was 95%[21], a
finding that has been replicated by others, including recently in [Manrai et al.] (2014).
As [Eddy (1982)] points out, this pattern is consistent with physicians inadvertently
substituting the attribute of interest \( P(\text{disease}|\text{positive}) \) with the more easily recalled
\( P(\text{positive}|\text{disease}) \)[22]. This suggests that when physicians are presented with a char-
acteristic \( X_j \) from which to make a diagnosis, they tend to draw upon the likelihood
that someone with the condition possesses that characteristic \( (P(X_j|c)) \) rather than
the posterior probability that the patient has condition \( c \) given the characteristic
\( (P(c|X_j)) \). Note that this tendency is consistent with the [Kahneman and Tversky
(1972)] definition of representativeness quoted above.

This proclivity also fits naturally with the research on stereotypes found in [Bordalo
et al. (2016a,b)], which formally model representativeness in terms of these likelihoods.
Following their model and adapting it to the specific question at hand, I define the
representativeness of a binary characteristic \( X_j \) for a patient with condition \( c \) relative
to a patient without \( c \) to be:

\[
\pi^c_j = \frac{P(X_j = 1|c)}{P(X_j = 1| \bar{c})}. \tag{3}
\]

I call \( X_j \) a "representative characteristic" for condition \( c \) if \( \pi^c_j > 1 \) (that is, when \( X_j \) is
more prevalent among those with \( c \) than among those without \( c \)). It is a set of these
characteristics that comprise \( X^c \), the prototypical presentation of condition \( c \).

[21] The mean estimate given by the respondents was 55.9%.
[22] \( P(\text{positive}|\text{disease}) = 0.95 \) assuming that the test’s false negative rate is also 5%.
Implications of Representativeness for Missed Diagnoses

A fundamental attribute of representativeness is that the decision weights placed on representative characteristics are exaggerated relative to their actual predictive power (Bordalo et al., 2016a,b). This feature of representative thinking is plausibly responsible for the heterogeneity in the empirical results discussed in Section V. Recall that the discontinuous drop in missed IHD diagnoses upon crossing the age-40 threshold appears to be concentrated among females presenting in the ED without chest pain, indicating that this group is under-diagnosed on the margin. The attributes that define this group (female and without chest pain) are simply the absence of two representative characteristics for IHD, which would lead physicians to underestimate this group’s risk. This decreases the intensity with which they are screened for IHD, thereby increasing the likelihood that those with IHD will be misdiagnosed.

It is also important to keep in mind that representative characteristics are not universal, but context-dependent – what is representative for one condition is not necessarily representative for another. For example, consider the results of Abaluck et al. (2016), which demonstrate that substantial welfare losses are generated by physicians’ misweighting of patient characteristics when making testing decisions for another potentially fatal acute condition, pulmonary embolism (PE). Yet patient sex is not one of these misweighted characteristics: the authors find that testing for PE appears to be reasonably balanced across males and females according to their relative risks. But this is in fact consistent with my model, as unlike the male-skewed incidence of heart attacks around age 40, the prevalence of PE is relatively similar among males and females (Stein et al., 1999), implying that being male is not a representative characteristic for PE.
VII Policy Remedies

Given the importance of clinical decision-making in health care, there are potentially large welfare gains to be made by implementing policies to address the types of systematic biases discussed in this paper. When policymakers are confronted with the errors in physicians’ probabilistic judgment as illustrated in Casscells et al (1978) (as described in Section VI above), a natural response is to call for additional training in statistics during medical school. Certainly, greater statistical sophistication is a worthy goal, as it might not only improve clinical decisions directly, but also physicians’ ability to interpret important findings in the medical literature. However, it is unreasonable to assume that enhancing medical education in this manner is a panacea, as this solution fails to address with three key issues.

First, it is not clear that further education should substantially reduce the types of biases described in this paper. Indeed, Manrai et al. (2014) suggests that physician behavior has changed little in this respect in the decades since Casscells et al. (1978), despite changes made to medical education curriculum in the interim. This should not come as a surprise – these biases are a result of heuristics that are endemic to human judgment under uncertainty more generally, not because individuals do not possess the ability to compute probabilities. And to be clear, such tendencies are not somehow unique to physicians – after all, some of the empirical evidence underlying Tversky and Kahneman (1971) was gathered from questionnaires that the authors administered to colleagues in the field of mathematical psychology, individuals for whom knowledge of statistics is likely to be deeply ingrained.

Second, it is conceivable that an attempt to compel physicians to behave in a more Bayesian manner could actually result in a deterioration, rather than an improvement, in the quality of care. Heuristics are employed to quickly solve problems while con-
serving cognitive effort, and in most cases, they result in near-optimal choices. By forcing physicians to expend more effort estimating probabilities, less effort may be expended on other aspects of their jobs, a trade-off that may very well result in a welfare loss.

Third, even if improvements in this area can be made through training, any effort to improve physician behavior through medical education will likely take decades to come to fruition, as after physicians complete their medical training, many fail to stay abreast of recent developments. In the words of a prominent cardiologist, "it’s a job, and they’re trying to earn money, and they don’t necessarily keep up. So really major changes have to be generational" [Epstein, 2017]. So, a new generation of physicians must be trained before they can join the profession and eventually influence common practices.

Implicit in arguments for more training is the desire to somehow make physicians more like computers. But why expend great effort training individuals to think in a way that is unnatural to them when machines can perform these tasks at exceptionally low cost? It is likely more productive to encourage physicians to focus on improving high-value aspects of their jobs that machines cannot readily handle, such as eliciting and interpreting the patient’s symptoms and patient history, conducting careful physical exams, noticing subtle details about the patient that are not easily captured in a patient chart, or improving bedside manner.

To the extent that research can identify scenarios in which physicians’ heuristics tend to lead them astray, electronic decision support systems (DSS) can be developed to 'nudge' [Thaler and Sunstein, 2009] them away from undesirable actions. Clearly, the benefits to implementing these systems will depend heavily on successful design and implementation. Such efforts should be guided by a principle of minimal intrusion in the diagnostic process – the decision-maker should only be nudged in the limited
settings in which biased behavior is prevalent and meaningful, thereby minimizing DSS "alert fatigue" (Kesselheim et al., 2011) and allowing them benefit from the efficiency that heuristic thinking typically offers. Just as Kleinberg et al. (2017) demonstrates the potential to improve decisions in the judicial system by harnessing machine learning prediction techniques, a similar approach could be taken in the field of medicine.

VIII Conclusion

Most of the evidence regarding heuristics and biases in human decisions has been demonstrated among relatively unsophisticated or inexperienced agents operating in low-stakes experimental settings. In this paper, I provide evidence that such behavior is in fact present even among extensively trained professionals facing common dilemmas in a high-stakes environment. This finding suggests that if heuristic behavior is prevalent in this setting, it’s likely to have meaningful impacts in many domains, making this an important area of research going forward.

Empirically, I find that physicians in the ED are roughly 10% more likely to test and 20% more likely to diagnose patients with IHD just after their 40th birthday than just beforehand, despite the fact that the underlying incidence of IHD increases smoothly with age. This shock to diagnostic intensity appears to substantially reduce the likelihood of missed diagnoses among women, a population that is known to be at elevated risk of misdiagnosis. This suggests that the marginal returns to screening for IHD in female ED patients at this age are relatively large. I also demonstrate a more general pattern in the misdiagnosis of IHD that is also consistent with heuristic thinking: patients presenting without characteristics that are representative of IHD appear more likely to be misdiagnosed than can be justified by their objective risk.
It is clear that the quality of diagnostic decisions in the ED can have large impacts on patient outcomes. And from a policy perspective, the role of EDs in medical spending has expanded greatly in recent years, as the share of hospital admissions that originate in the ED has risen sharply in recent decades ([Schuur and Venkatesh 2012](#), [Kocher et al. 2013](#)), making emergency physicians increasingly important gatekeepers for expensive inpatient care. However, the solution to the issues described in this paper is not to somehow purge decision-makers of heuristic behavior. Decision-makers have good reason to think heuristically – in general, it serves them well, efficiently generating near-optimal conclusions with minimal cognitive effort. Instead, effort should be made to accommodate heuristic thinking, and only implement safeguards in settings where it is likely to lead decision-makers astray. As technological advances in data quality and computation continue, machine-learning algorithms provide the opportunity to systematically identify these settings.
Figure 1: Proportion of ED patients tested for heart attack

Source: Truven Health MarketScan Database, 2010 - 2013
Figure 2: Proportion of ED patients diagnosed with IHD

Source: Truven Health MarketScan Database, 2005 - 2013

Note: IHD diagnosis defined as ICD-9 Diagnosis Codes 410-414
**Figure 3:** Proportion of ED patients admitted to hospital with IHD diagnosis

*Source:* Truven Health MarketScan Database, 2005 - 2013

*Note:* IHD diagnosis defined as ICD-9 Diagnosis Codes 410-414
Figure 4: Missed IHD diagnoses (90 Days), by patient sex

(a) Female

(b) Male

Source: Truven Health MarketScan Database, 2005 - 2013
**Figure 5:** Missed IHD diagnoses (90 Days), females with and without chest pain

(a) With chest pain

(b) Without chest pain

*Source:* Truven Health MarketScan Database, 2005 - 2013
Figure 6: Number of ED visits, by month-age bin

Source: Truven Health MarketScan Database, 2005 - 2013
**Figure 7:** Proportion of patients tested for heart attack in Boston-area ED

*Source:* Emergency department records from a Boston-area hospital, January 2010 - May 2015
**Figure 8:** Proportion of patients tested for heart attack in Boston-area ED, above-median (solid blue) vs. below-median (dashed red) patient volume

*Source:* Emergency department records from a Boston-area hospital, January 2010 - May 2015

*Note:* The solid blue line (dashed red line) is a local linear fit among patients arriving during an hour in which there were an above-median (below-median) number of patients arriving in the ED, conditional on the hour of day. The blue o’s (red x’s) represent the mean testing rate among patients within each three-month age bin in the above-median (below-median) group.
Figure 9: Proportion of patients tested for heart attack in Boston-area ED, first half (dashed red) vs. second half (solid blue) of attending physician’s shift.

Source: Emergency department records from a Boston-area hospital, January 2010 - May 2015

Note: The solid blue line (dashed red line) is a local linear fit among patients treated in attending physician’s second (first) half of shift. The blue o’s (red x’s) represent the mean testing rate among patients within each three-month age bin in second half (first half) group.
Figure 10: Annual Heart Attack Deaths per Capita

Figure 11: Piecewise Linear Fit - Annual Heart Attack Deaths per Capita

Figure 12: Proportion of ED patients tested for heart attack

Source: Truven Health MarketScan Database, 2010 - 2013
**Table 1: Heart Attack Testing Rates**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turned40</td>
<td>0.887***</td>
<td>0.998***</td>
<td>0.752**</td>
</tr>
<tr>
<td></td>
<td>(0.150)</td>
<td>(0.219)</td>
<td>(0.323)</td>
</tr>
<tr>
<td>Intercept</td>
<td>9.328***</td>
<td>8.404***</td>
<td>10.463***</td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.164)</td>
<td>(0.228)</td>
</tr>
</tbody>
</table>

Relative Change

- Turned40: 9.5%
- Intercept: 11.9%
- All: 7.2%

Observations

- Turned40: 622,783
- Intercept: 343,577
- All: 279,206

Note:

- \*p<0.1; \**p<0.05; \***p<0.01
- Cluster-robust standard errors
- Bandwidth = 15 Months

**Table 2: IHD Diagnosis Rates**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turned40</td>
<td>0.131***</td>
<td>0.087***</td>
<td>0.191***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.021)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.679***</td>
<td>0.370***</td>
<td>1.061***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.013)</td>
<td>(0.033)</td>
</tr>
</tbody>
</table>

Relative Change

- Turned40: 19.3%
- Intercept: 23.4%
- All: 18%

Observations

- Turned40: 1,325,771
- Intercept: 735,228
- All: 590,543

Note:

- \*p<0.1; \**p<0.05; \***p<0.01
- Cluster-robust standard errors
- Bandwidth = 15 Months
**Table 3: IHD Admission Rates**

<table>
<thead>
<tr>
<th>Dependent variable: Admitted with IHD x100</th>
<th>All (1)</th>
<th>Female (2)</th>
<th>Male (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turned40</td>
<td>0.068***</td>
<td>0.052**</td>
<td>0.091**</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.021)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.382***</td>
<td>0.188***</td>
<td>0.621***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.013)</td>
<td>(0.024)</td>
</tr>
</tbody>
</table>

Relative Change: 17.9%  27.9%  14.6%

Observations: 1,325,771  735,228  590,543

*Note:*  *p < 0.1; **p < 0.05; ***p < 0.01
Cluster-robust standard errors
Bandwidth = 15 Months

**Table 4: Heart Attack Testing Rates, w.r.t. Chest Pain (CP)**

<table>
<thead>
<tr>
<th>Dependent variable: Tested for Heart Attack x100</th>
<th>With CP (1)</th>
<th>No CP (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turned40</td>
<td>1.996**</td>
<td>0.414***</td>
</tr>
<tr>
<td></td>
<td>(0.817)</td>
<td>(0.097)</td>
</tr>
<tr>
<td>Intercept</td>
<td>49.089***</td>
<td>3.542***</td>
</tr>
<tr>
<td></td>
<td>(0.776)</td>
<td>(0.064)</td>
</tr>
</tbody>
</table>

Relative Change: 4.1%  11.7%

Observations: 79,597  543,186

*Note:*  *p < 0.1; **p < 0.05; ***p < 0.01
Cluster-robust standard errors
Bandwidth = 15 Months
**Table 5:** Prevalence of Relevant Characteristics Above and Below the Age-40 Threshold

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Below</th>
<th>Above</th>
<th>Diff</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.552</td>
<td>0.555</td>
<td>0.003</td>
<td>0.110</td>
</tr>
<tr>
<td>Prior Diabetes DX</td>
<td>0.048</td>
<td>0.047</td>
<td>-0.000</td>
<td>0.809</td>
</tr>
<tr>
<td>Prior High Blood Pressure DX</td>
<td>0.134</td>
<td>0.133</td>
<td>-0.001</td>
<td>0.341</td>
</tr>
<tr>
<td>Prior High Cholesterol DX</td>
<td>0.044</td>
<td>0.045</td>
<td>0.001</td>
<td>0.343</td>
</tr>
<tr>
<td>Prior Obesity DX</td>
<td>0.045</td>
<td>0.044</td>
<td>-0.001</td>
<td>0.121</td>
</tr>
<tr>
<td>Days Enrolled Before Visit</td>
<td>535.652</td>
<td>533.891</td>
<td>-1.761</td>
<td>0.511</td>
</tr>
</tbody>
</table>

*Note:* All differences and p-values in this table were computed using the same model, observations and bandwidth (15 months) used to compute the discontinuities presented in Tables 2 and 3.

**Table 6:** Likely Missed IHD Diagnosis Rates

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: Missed IHD, 90 Days (per 100,000)</th>
<th>All</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Turned40</td>
<td>−0.552</td>
<td>(9.206)</td>
<td>−24.537***</td>
<td>29.977</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.199)</td>
<td>(22.908)</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>82.345***</td>
<td>(3.776)</td>
<td>57.020***</td>
<td>113.809***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6.224)</td>
<td>(10.733)</td>
<td></td>
</tr>
<tr>
<td>Relative Change</td>
<td>-0.7%</td>
<td></td>
<td>-43%</td>
<td>26.3%</td>
</tr>
<tr>
<td>Observations</td>
<td>1,315,741</td>
<td>732,085</td>
<td>583,656</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* *p<0.1; **p<0.05; ***p<0.01
Cluster-robust standard errors
Bandwidth = 15 Months
### Table 7: Likely Missed IHD Diagnoses Rates, Female Only

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Missed IHD, 90 Days (per 100,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With CP</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Turned 40</td>
<td>-55.509</td>
</tr>
<tr>
<td></td>
<td>(64.529)</td>
</tr>
<tr>
<td>Intercept</td>
<td>200.837***</td>
</tr>
<tr>
<td></td>
<td>(40.541)</td>
</tr>
<tr>
<td>Relative Change</td>
<td>-27.6%</td>
</tr>
<tr>
<td>Observations</td>
<td>79,264</td>
</tr>
</tbody>
</table>

*Note:* *p*<0.1; **p*<0.05; ***p*<0.01
Cluster-robust standard errors
Bandwidth = 15 Months

### Table 8: Heart Attack Testing Rates, Boston-Area ED

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Tested for Heart Attack x100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Turned 40</td>
<td>4.033***</td>
</tr>
<tr>
<td></td>
<td>(1.607)</td>
</tr>
<tr>
<td>Intercept</td>
<td>7.807***</td>
</tr>
<tr>
<td></td>
<td>(1.021)</td>
</tr>
<tr>
<td>Relative Change</td>
<td>51.7%</td>
</tr>
<tr>
<td>Observations</td>
<td>6,271</td>
</tr>
</tbody>
</table>

*Note:* *p*<0.1; **p*<0.05; ***p*<0.01
Robust standard errors
Bandwidth = 1.52 Years
References


Appendices

Appendix I Data

Determining Month-Age of Patients

The Truven claims Enrollment Detail Table records individual-level information for each month that the individual is enrolled in a health plan. This information includes the individual’s year of birth and [integer] age as of the beginning of each calendar month. Given these two facts, it is straightforward to determine the individual’s month of birth so long as during the observed period (2005 - 2013), she is enrolled in her month of birth and in the month following it. This is true of roughly 95% of all individuals in the database. However, the remaining 5% of individuals are not enrolled long enough for this to be observed. For each individual whose month of birth cannot be determined with certainty, I randomly assign a month of birth drawn from the individual’s potential months of birth (i.e. the calendar months in which the individual is never enrolled). This imputation could lead to classical measurement error of the discontinuities at age 40, thereby attenuating my estimates. However, as noted in Section III, my results are robust to the inclusion/exclusion of individuals with imputed months of birth.

Imputing IHD Testing in Truven Inpatient Claims

Due to the manner in which inpatient and outpatient claims are filed with insurers, the troponin test, which I use as a proxy for IHD testing, can only be readily identified in outpatient ED claims. But at minimum, the troponin test should be ordered for all patients in the ED who are diagnosed with IHD. I therefore make the assumption that
all ED patients who are admitted to the hospital with an IHD diagnosis received this test, and assume that all other admitted patients did not. To the extent that this fails to account for ED patients who are admitted to the hospital without a diagnosis of IHD but nonetheless received a troponin test, this will understate the overall testing rate. To the extent that there exists a discontinuous jump in admissions at the age-40 threshold, this assumption might attenuate my estimated discontinuity in testing. However, empirically I find that this assumption has no meaningful impact on the magnitude of the discontinuity in testing, which is unsurprising given that admitted patients diagnosed with IHD only comprise ~0.5% of my analysis sample.
Appendix II  Robustness Checks

RD Bandwidth

Tables A1 through A3 present local linear estimates of the discontinuities in physician behavior using the optimal bandwidth (15 months) described in Section IV, as well as using bandwidths half and twice this size, as suggested by [Lee and Lemieux (2010)]. The estimated discontinuity in the heart attack testing rate is very similar in each specification, as would be expected given the highly linear relationship between age and testing rates shown in Figure 1. On the other hand, the magnitude of the discontinuities in the IHD diagnosis and admission rate estimates are diminishing in the size of the bandwidth (although still statistically significant). This is consistent with the patterns displayed in Figures 2 and 3, which show that the slope of these rates grows flatter near the threshold.

Table A1: Heart Attack Test Rates, By Bandwidth

<table>
<thead>
<tr>
<th></th>
<th>Tested for Heart Attack x100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8 Months</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Turned40</td>
<td>0.712***</td>
</tr>
<tr>
<td></td>
<td>(0.191)</td>
</tr>
<tr>
<td>Intercept</td>
<td>9.448***</td>
</tr>
<tr>
<td></td>
<td>(0.150)</td>
</tr>
<tr>
<td>Relative Change</td>
<td>7.5%</td>
</tr>
<tr>
<td>Observations</td>
<td>312,452</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01
Cluster-robust standard errors
### Table A2: IHD Diagnosis Rates, By Bandwidth

<table>
<thead>
<tr>
<th></th>
<th>8 Months</th>
<th>15 Months</th>
<th>30 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turned40</td>
<td>0.175***</td>
<td>0.131***</td>
<td>0.092***</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.024)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.634***</td>
<td>0.679***</td>
<td>0.701***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.015)</td>
<td>(0.017)</td>
</tr>
</tbody>
</table>

- Relative Change: 27.7% 19.3% 13.1%
- Observations: 663,437 1,325,771 2,742,363

*Note:* *p<0.1; **p<0.05; ***p<0.01
Cluster-robust standard errors

### Table A3: IHD Admission Rates, By Bandwidth

<table>
<thead>
<tr>
<th></th>
<th>8 Months</th>
<th>15 Months</th>
<th>30 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turned40</td>
<td>0.091***</td>
<td>0.068***</td>
<td>0.048**</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.023)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.349***</td>
<td>0.382***</td>
<td>0.394***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.015)</td>
<td>(0.013)</td>
</tr>
</tbody>
</table>

- Relative Change: 26.1% 17.9% 12.3%
- Observations: 663,437 1,325,771 2,742,363

*Note:* *p<0.1; **p<0.05; ***p<0.01
Cluster-robust standard errors
Duration of Heart Attack Follow-up Period

In the Tables A4 (a) and (b) below, I demonstrate the robustness of the discontinuity in female missed diagnoses found in Table 6. Using follow-up periods of 30 and 180 days, the estimates are consistent with my primary specification of 90 days, and as expected, grow in the number of days observed post-visit.

Table A4: Likely Missed IHD Diagnosis Rates

(a) 30 Days

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turned40</td>
<td>-5.985</td>
<td>-22.822**</td>
<td>15.481</td>
</tr>
<tr>
<td>Intercept</td>
<td>65.913***</td>
<td>45.151***</td>
<td>91.713***</td>
</tr>
<tr>
<td>Relative Change</td>
<td>-9.1%</td>
<td>-50.5%</td>
<td>16.9%</td>
</tr>
<tr>
<td>Observations</td>
<td>1,315,741</td>
<td>732,085</td>
<td>583,656</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01

(b) 180 Days

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turned40</td>
<td>-1.434</td>
<td>-27.002**</td>
<td>31.223</td>
</tr>
<tr>
<td>Intercept</td>
<td>117.748***</td>
<td>86.661***</td>
<td>156.342***</td>
</tr>
<tr>
<td>Relative Change</td>
<td>-1.2%</td>
<td>-31.2%</td>
<td>20%</td>
</tr>
<tr>
<td>Observations</td>
<td>1,315,741</td>
<td>732,085</td>
<td>583,656</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01
Cluster-robust standard errors
Bandwidth = 15 Months
Appendix III  Cost of Testing for IHD

Throughout this paper, I use the troponin blood test as a proxy for IHD testing. I do so not only because is it the first-line test for ruling out a heart attack, but also because it is used almost exclusively for this purpose, making it a uniquely precise proxy for IHD testing. However, it is not the only procedure that would be conducted when exploring an IHD diagnosis, as the troponin test will generally be negative for patients whose IHD has not risen to the level of heart attack. To quantify the effect of the age-40 heuristic on relevant IHD testing costs, I must consider its effects on other tests as well.

Ideally, I would do so by measuring the discontinuity in testing at the threshold for each procedure separately, but the nature of these procedures make it difficult to do so. First, by far the two most common are the ECG and chest x-ray, but both are also frequently used in pursuing diagnoses unrelated to IHD. Second, the remaining procedures, which include stress testing and imaging, are more costly and thus much less frequently used than the troponin, ECG, or x-ray tests. As a result, it difficult to precisely measure the effect of the age-40 heuristic on these procedures individually. However, since they are typically only used in diagnosing IHD in addition to a troponin test, the rate at which these other procedures are performed within the population that is administered the troponin test serves as a reasonable approximation of the frequency with which these tests are used to diagnose IHD.

Table A5 below presents how I arrive at the IHD testing cost estimate used in my back-of-the-envelope cost-benefit analysis performed in Section V. Among the 40-year-old female outpatients in my analysis sample who are given the troponin test, column (a) states the proportion that are also given additional screening procedures. Column (b) displays the 2017 national CMS fee-for-service reimbursement rates (physician
plus outpatient hospital service payments, or lab payments in the case of the troponin test), which I use as an estimate of the cost of performing these procedures. Column (c) is the product of columns (a) and (b). Thus the sum of column (c) is my estimate of the average aggregate cost of IHD testing per woman administered a troponin test.

Table A5: Average Aggregate IHD Testing Cost

<table>
<thead>
<tr>
<th>Procedure Name</th>
<th>Testing Rate (a)</th>
<th>CMS Payment (b)</th>
<th>Weighted Cost (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Troponin</td>
<td>1.000</td>
<td>13.50</td>
<td>13.50</td>
</tr>
<tr>
<td>Chest X-Ray</td>
<td>0.792</td>
<td>79.48</td>
<td>62.95</td>
</tr>
<tr>
<td>ECG</td>
<td>0.925</td>
<td>65.64</td>
<td>60.72</td>
</tr>
<tr>
<td>Stress ECG</td>
<td>0.048</td>
<td>269.35</td>
<td>12.93</td>
</tr>
<tr>
<td>Echo</td>
<td>0.030</td>
<td>505.88</td>
<td>15.18</td>
</tr>
<tr>
<td>Stress Echo</td>
<td>0.016</td>
<td>513.42</td>
<td>8.21</td>
</tr>
<tr>
<td>Chest CT</td>
<td>0.133</td>
<td>349.57</td>
<td>46.49</td>
</tr>
<tr>
<td>MPI</td>
<td>0.025</td>
<td>1223.10</td>
<td>30.58</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td><strong>$250.56</strong></td>
</tr>
</tbody>
</table>

To the extent to which the marginal patient affected by the age-40 heuristic is less likely to receive these additional procedures than the average patient receiving a troponin test, this estimate likely overstates the aggregate cost of IHD testing for marginal patients.