Does Bad Medical News Reduce Preferences for Generic Drugs?*

Manuel Hermosilla†
Andrew T. Ching

Johns Hopkins University, Carey Business School

First version: November 2021
This version: January 2023

Abstract

Policy makers and insurers promote the use of generic drugs because they can deliver large savings without sacrificing quality. But these efforts meet resistance from the public, who perceive generic drugs as inferior substitutes for brand name counterparts. Building on literature showing that negative emotions reduce risk-taking, we hypothesize that "bad medical news" prompts patients to favor brand name drugs as means to safeguard their health. Our evidence exploits LDL cholesterol test results, where a discontinuity from clinical guidelines allows us to estimate the causal effect of bad medical news. Using data covering patients' prescription drug choices across drug classes, we find that patients receiving bad medical news become 8% more likely to choose the brand name alternative. Our findings are reinforced by a secondary analysis incorporating the similar context of Hemoglobin A1c (blood sugar) testing. We also find that bad medical news reduces preferences for generics most strongly among drugs of direct clinical relevance for each test, but the effect also manifests among non-clinically relevant drugs.

Keywords:
Healthcare, Marketing and Public Policy, Behavioral Marketing, Natural Experiments, Generic Drugs.

---

*This article was previously circulated under the title: "Bad Medical News and the Aversion of Generic Drugs." We thank Dan Goetz, Simha Mummalaneni, Haiyang Yang, and David Powell for providing detailed comments and feedback, as well as conference and seminar participants at Johns Hopkins University, University of Miami, the 2022 International Industrial Organization Conference, 2022 Marketing Science Conference, and the 2022 Annual Health Econometrics Workshop. We gratefully acknowledge the support of a Pilot Award Grant from the Hopkins Business of Health Initiative.

†Corresponding author. Email: mh@jhu.edu

INTRODUCTION

The large and increasing healthcare expenditures of developed countries has been a challenge for policy makers and the general public. In the U.S., the issue is best reflected by the sizable share of GDP spent on healthcare—now close to 20% (Hartman et al. 2022). Critics argue that much of these expenditures reflect inefficiencies (Nunn, Parsons, and Shambaugh 2020), suggesting that costs could be contained without sacrificing the quality of care received by patients.

Regarding the prescription drug component of these expenditures, policy makers view policies fostering the substitution of brand name with generic drugs as options of particular interest. In addition to using the same active ingredients and dosages, the FDA certifies that generic drugs have the same key pharmacological properties of brand name counterparts (within margin). Accordingly, many experts view generics as molecular replicas of brand name drugs and, thereby, as delivering the same objective therapeutic value. Given the much lower prices of generics, substituting brand name with generic consumption could therefore lower expenditures without sacrificing the quality of care received by patients. Estimates for the U.S. suggest that these savings could be large—about 10% of prescription drug expenditure (US$ 36 billion a year) if patients always chose the generic option when available.[1]

Prompted by these facts, public and private insurers have introduced a variety of incentives (e.g., coupons, free samples) aiming to encourage the use of generics. These efforts are nevertheless met with resistance from the public, who perceive generics as of inferior quality compared to brand name drugs (Dunne and Dunne 2015; Hassali et al. 2009). Prior research has rationalized such preferences based on informational gaps, i.e., the fact that patients lack information reassuring them of the therapeutic equivalency between the two types of drugs. For example, Bronnenberg et al. (2015) find that, compared with the general public, pharmacists—who know more about drugs' properties—are more likely to prefer generic

over brand name aspirin, while Carrera and Villas-Boas (2020) and Ching (2010a) provide evidence that bridging this informational gap increases generic choice.

We contribute to this literature by investigating how negative information shocks about the patients' own health—"bad medical news"—impact the relative preferences between brand name and generic drugs. We are motivated by a series of findings linking similar psychological stimuli with subconscious e↵ects on choice. For example, people make more risk-conservative gambling and job-selection decisions when experiencing anxiety (Raghunathan and Pham 1999). Similarly, decisions become biased towards the low-risk option in the presence of worry (Johnson and Tversky 1983), fear (Cohn et al. 2015; Lerner and Keltner 2000, 2001), trauma (Callen et al. 2014; Cameron and Shah 2015), and weather-induced bad mood (Bassi, Colacito, and Fulghieri 2013; Hirshleifer and Shumway 2003; Saunders 1993). By raising alarm about their own health condition, bad medical news may infuse some of these emotions, prompting patients to take action to safeguard their health, e.g., improving their diet or exercising more. Given that brand name drugs are perceived as of higher efficacy and safety than generics (thereby implying smaller health risks), such actions may also include favoring the brand name alternative whenever confronted with a drug choice. Accordingly, we hypothesize that bad medical news may increase patients' propensities to choose brand name drugs over generics.

Our interest in how bad medical news may impact drug choices is also premised on the idea that these e↵ects could operate at a large scale. For example, millions of women test each year for genetic markers of breast cancer, where a positive test outcome could act as a bad news event. Similarly, given the high prevalence of cardiovascular disease in the developed world, tens of millions of adults test regularly for low-density lipoproteins (LDL) cholesterol (a.k.a., "bad cholesterol"). For an individual who has routinely tested in "optimal" ranges, a "borderline high" result may also deliver bad medical news. In all, given the high prevalence of bad medical news shocks arising organically as patients interact with the medical system, their total e↵ects could be large.

Our main analysis focuses on the medical news implied by LDL testing results. Instead of inferring the presence of bad medical news from patients' testing histories (as in the above example), we rely on a comparison across patients. We compare patients who receive 129 and 130 mg/dL LDL results. We choose this narrow window because it marks the frontier between "near optimal" and "borderline high" ranges, as defined by clinical guidelines. Compared with individuals who test 129 mg/dL, those who test 130 mg/dL "cross" the frontier. We therefore posit that, compared with the former, the latter patients (130 mg/dL) receive a bad news treatment. Crucially, the di↵erences resulting from this comparison can be given a causal interpretation because LDL results include a measurement error. Since these errors are due to factors such as the extent of fasting prior to the test or the patient's posture while the blood is drawn, they can be deemed as plausibly exogenous. We therefore assume that patients are locally randomized between the 129 and 130 mg/dL measurements. This assumption is supported by the tight balancing between the two patient populations. We make use of the di↵erence-in-di↵erences (DID) approach to estimate the di↵erential e↵ect that receiving the LDL results has on the drug choices of treated (130 mg/dL) versus control (129 mg/dL) patients. Because the perception that generics are of inferior quality than brand name drugs is general (i.e., it applies to all drugs), the bad LDL news may a↵ect choice beyond drugs of direct clinical relevance to LDL results. Accordingly, our estimation utilizes comprehensive data including all prescription drug choices made by sample patients, which cover almost 500 drugs across six drug classes (anti-infectives, cardiovascular, gastrointestinal, etc.). Using a similar DID design, we also investigate whether the bad news shock changes the quantity and bundle of drugs used by patients. However, we do not find evidence that the bad news shock influences these decisions.

In our main analysis we estimate the impact of the bad news treatment on the probability that a patient chooses the generic over the brand name alternative. Consistent with our hypothesis, we find that, relative to control patients, those treated with bad medical news reduce their generic choice probability by about 0.01 after the test. The estimate represents

a 1.3% reduction in the average patient's propensity to choose the generic option. Given that brand name drugs have a smaller share of choices (14%) than generics (86%), this result can be equivalently expressed as an 8% increase in the propensity to choose the brand name option. Considering the average generic price discount relative to brand name drugs (80-85%), this e↵ect implies roughly a 3% increase in total prescription drug expenditures for the average patient.

Encouraged by this finding, we probe the e↵ect by investigating the consequences of varying treatment intensity. In a series of analyses, we find that the e↵ect on generic propensity changes in a way that is directionally consistent with the change in the intensity of the bad news treatment. For example, we find relatively larger e↵ects on generic propensity among patients whose LDL test happens around a medical office visit (for whom the test result may be more salient) and among healthier patients (who may be more surprised by the bad news). Among others, we find evidence of a markedly front-loaded e↵ect, i.e., the e↵ect concentrates in the immediate aftermath of the test (90 days). The main mechanism behind this result pertains to the adoption of new drugs, i.e., the bad news shock is particularly influential for patients who are purchasing a drug for the first time. To investigate the generalizability of the e↵ect, we next expand our analysis by incorporating the results of a di↵erent medical test, i.e., Hemoglobin A1c (measuring blood sugar levels), where we exploit the 7% threshold that diabetic patients use to manage the condition. The results of this analysis are broadly supportive of the idea that our main result may generalize outside the context of LDL tests.

As noted, the perceived quality di↵erences between brand name and generic drugs acts as a precondition for the bad news e↵ect on brand/generic choice. The e↵ect's size should therefore increase with the di↵erence in perceived qualities. We examine this implication through two additional analyses. We first exploit brand/generic price di↵erences. In a differentiated product market, price di↵erences of two products should be positively correlated with their perceived quality di↵erences. Consistent with this idea, we find that the bad news e↵ect primarily manifests when the branded option is sufficiently more expensive than the

generic option. We next leverage findings of prior literature highlighting that patients learn drugs' "true" properties through usage. In the presence of this form of learning, perceived quality di↵erences should progressively narrow as patients gain more consumption experience. Accordingly, we find that the e↵ect decays with the patient's cumulative consumption experience.

Our final analysis studies how the bad news e↵ect on generic choice varies between di↵erent types of drugs. Our main analysis compares the e↵ects unfolding on clinically relevant (CR) drugs (cholesterol drugs for LDL testers, diabetes drugs for A1c testers) against those on non-CR drugs (e.g., gastrointestinal drugs). While we expect the e↵ect to operate on both types of drugs due to generics being perceived as of inferior quality in general, the bad testing results can be particularly alarming for patients su↵ering from a related condition. Consistent with this intuition, we find that the bad news e↵ect is significantly stronger among CR than among non-CR drugs. We also provide evidence suggesting that accounting for consumption experience is important to correctly estimate the e↵ects of bad news on generic propensity for CR and non-CR drugs.

We organize the rest of the article as follows. We start by describing the institutional and literature background, and then introduce our dataset and research design. We proceed by investigating the potential impacts of bad LDL news on patients' choices on prescription drugs. We then investigate the robustness of our results using bad A1c news. Our last set of analyses probes the importance of quality perceptions as a key driver for the documented e↵ects. We conclude by discussing implications for practice and future research directions.

INSTITUTIONAL BACKGROUND AND RELATED LITERATURE

Generic Drugs

Pharmaceutical drugs combine active and inactive ingredients. Active ingredients deliver the intended pharmacological e↵ects while inactive ingredients enable auxiliary features, e.g., coloring and flavoring. Generics have the same active ingredients as their respective

branded incumbents. Accordingly, public health agencies (e.g., the FDA, Health Canada, etc.) view them as providing the same objective quality as brand name drugs. Moreover, they do so at much lower costs—on average, at an 80%-85% discount compared to brand name equivalents.

For most drugs, generic entry is limited by patents protecting the use of active ingredients. Since the Hatch-Waxman Act defined modern entry requirements in 1984, generic penetration has increased steadily, from 36% in 1994 to 88% of all prescriptions in 2014 (IMS Institute for Healthcare Informatics 2015), to 90% in 2019 (Woodcock 2019). This trend reflects the continued e↵orts of public and private payers to encourage generic substitution (Dunne and Dunne 2015). These e↵orts, which are primarily price-based, include tiered copayments (i.e., lower copayments for generic drugs), pay-for-performance schemes that target physicians and pharmacists, reference price schemes, as well as the use of promotional tools such as coupons and free samples. Some insurers have even o↵ered generic drugs at no cost to patients (Ching, Granlund, and Sundström 2022; O'Malley et al. 2006).

Despite these incentives, generic substitution faces some resistance from the public. This is mainly due to generics being perceived as of inferior quality compared with brand name drugs (Dunne and Dunne 2015; Hassali et al. 2009). These perceptions entail concerns about safety and efficacy, both of which associate generics with higher perceived health risks. The link between safety concerns and perceived health risks is straightforward: patients worry about the possibility of generics producing adverse side e↵ects, e.g., migraine and vomiting. In turn, efficacy concerns may fuel perceived health risks by hindering patients' ability to manage the drug's targeted condition. For example, less e↵ective cholesterol drugs may increase the likelihood of future cardiovascular events, while less e↵ective insulin products may increase the risk for diabetic seizures. Consistent with this view, Tootelian, Gaedeke, and Schlacter (1988) show that patients tend to disproportionally favor brand name drugs when they face larger health risks from the targeted condition. Similarly, Ganther and Kreling (2000) find that patients demand larger savings to purchase generic prescription

drugs of higher perceived risk.

While many of the studies on the preference bias against generics survey perceptions of generics in general, others focus on drugs of specific domains, e.g., antipsychotics (Roman 2009), asthma (Williams and Chrystyn 2007), cardiovascular drugs (Kesselheim et al. 2008). These findings converge on the idea that much of the difference in brand/generic quality perceptions applies to all generics, i.e., across drug classes. It has also been found that, beyond patients, healthcare professionals (providers and pharmacists) can be negatively predisposed towards generics (Chua et al. 2010; Hassali et al. 2010). It is possible that professionals adopt these predispositions to avoid frictions with patients who may exert pressure to receive the brand name option (Chua et al. 2010; Hassali et al. 2010; Williams and Chrystyn 2007).

Related Literature

Our main hypothesis builds on empirical findings which link negative emotional experiences (e.g., fear, worry, anxiety) to less risk-taking (e.g., Johnson and Tversky 1983; Kuhnen and Knutson 2011; Lerner and Keltner 2000, 2001). For example, in a lab experiment, Johnson and Tversky (1983) show that reading a newspaper homicide report designed to induce anxiety and worry leads one to have more pessimistic crime rate estimates. Interestingly, Johnson and Tversky (1983) find that such pessimistic bias also occurs for risks unrelated to the news article, such as the chance of dying in a fire or being a victim of leukemia. Similarly, experiencing anxiety has also been found to bias decisions in favor of lower-risk gambling and job-selection choices (Raghunathan and Pham 1999), while weather-induced bad mood reduces risk-taking in stock-market trading (Bassi, Colacito, and Fulghieri 2013; Hirshleifer and Shumway 2003; Saunders 1993). Like the effect of anxiety-provoking homicide reports on estimates for unrelated risks, weather-induced bad mood hinders financial risk-taking behavior, even though weather contains no information about the economy. Similarly, bad LDL news may affect drug choices beyond cardiovascular drugs (e.g., asthma drugs), even

though cholesterol levels may contain little information about non-cardiovascular health.

The choice between brand name and generic drugs has also been studied in the marketing and economics literatures (reviewed by Ching, Hermosilla, and Liu 2019). Ching (2010a) provides evidence of a preference bias against generics in the prescription drug market, while Bronnenberg et al. (2015) and Carrera and Villas-Boas (2020) do so for over-the-counter (OTC) drug choices. As noted in the introduction, these analyses focus on the role of information about generics, i.e., patients lacking information that reassures them of the equivalency between brand name and generic drugs. By contrast, we examine how the psychological stimulus delivered by bad medical news impacts brand/generic choice, even though such news contains no information about products.

Our work also contributes to a growing literature focusing on behavioral hazard in health-care decision making (Baicker, Mullainathan, and Schwartzstein 2015; Handel and Kolstad 2015; Mullainathan, Schwartzstein, and Congdon 2012). This literature highlights how behavioral biases may lead to treatment choices which do not improve consumer well-being. In our context, consumer well-being may be negatively a↵ected through lost monetary savings implied by a strengthened preference for brand name drugs. As highlighted by Shrank et al. (2006), patients may also experience impoverished health outcomes given that patients who use generics are significantly less likely to skip doses. Finally, it is important to di↵erentiate this paper from prior work investigating the impact of news media coverage on drug demand (Ching et al. 2016). Rather than product-specific journalistic news, we consider patient-specific medical news.

DATA

We utilize 2011 and 2012 MarketScan data, which track healthcare utilization (i.e., health-care expenditures) for individuals living in the United States. Data are compiled from de-identified administrative claims, from a large array of employers and health plans, including government and public organizations. By 2012, MarketScan included claims covering

almost 80 million employees (of up to 65 years of age) and their dependents.

A distinctive feature of the database is that it links individual records across different domains of healthcare in addition to drug utilization. Importantly, we have access to MarketScan's "Lab Files," which capture laboratory tests ordered in office-based practice settings. Our analysis leverages LDL cholesterol test results available from these files.

LDL Testing Results

LDL testing results are expressed in milligrams per deciliter (mg/dL). The distribution of all LDL test results observed in full the sample is shown in Figure A.5. The distribution has wide dispersion and is smooth around the 130 mg/dL threshold for "borderline high" levels, where our analysis takes place. We observe a total of 4,940 test results of 129 or 130 mg/dL. However, a majority of these results correspond to individuals who test more than once in the sample. Because multiple testing blurs our inference, for our analysis we primarily focus on the set of 2,282 individuals who have only one test in the data. (Multiple testers are analyzed separately.) Individuals in our main sample of single-testers are about evenly split between 129 and 130 mg/dL results (respectively, N=1,169 and 1,113). Their average age (at testing) is close to 49 years old; about 57% of them are female.

Although MarketScan data do not record the specific reasons why patients take the LDL test, some statistics suggest that a majority of these tests may occur in the context of routine checkups. Specifically, claims for outpatient services indicate that about 70% of the tests happen within a week of a primary care office visit. Of these visits, over 50% are conducted by family practitioners and 20% by internists. Furthermore, the types of medical problems informed by these tests (as revealed by diagnostic codes, when available) concentrate on problems typically addressed in the context of routine checkups, e.g., general prevention, cardiovascular health (see Table A.11).

Drug Choices

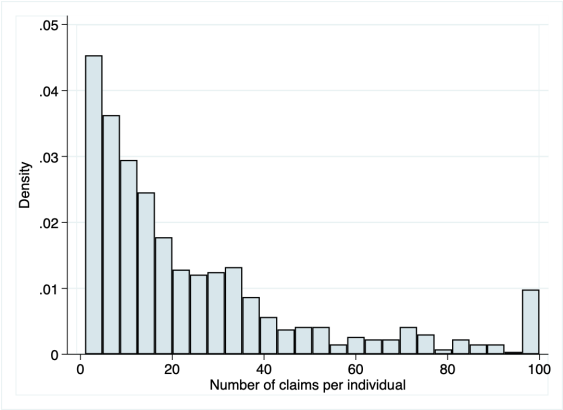
The second pillar for our dataset corresponds to claims for prescription drugs. To simplify our description of these data and results, we adopt a few terminology conventions. First, we use the terms "transaction," "purchase," and "prescription" interchangeably to refer to a single drug purchase. Second, we use the term "molecule" to refer to compounds or unique combinations of active ingredients, irrespective of branding. For example, we may say that the Alprazolam molecule has both brand name (Xanax, Niravam) and generic alternatives (generic Alprazolam). This nomenclature is helpful partly because we seek to control for unobserved variation at the molecule level, for example, from assortment sizes or brand/generic price gaps. Our regressions will include a fixed e↵ect for each molecule.

Because most health insurance plans do not cover over-the-counter drugs, our data has very little coverage among them (< 1%). We remove over-the-counter drugs altogether, and are thus left with a sample of prescription drug purchases only. Given our focus on brand/generic choice, we restrict our attention to molecules for which there is at least one generic and one brand name alternative in the market during the covered period (i.e., "multi-source" drugs).

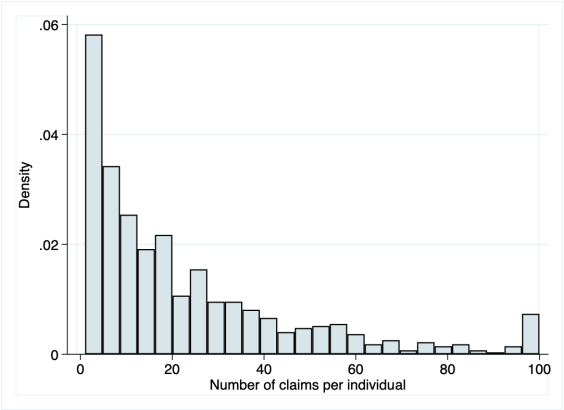
The resulting sample has claims for 35,080 prescription drug purchases, covering 484 di↵erent molecules. It is important to emphasize that this set of molecules covers the full set of drug needs of patients in our sample, not just those associated with high cholesterol. About 62% of individuals have at least one drug purchase. For these, the median number of purchases is 15 (IQR = [6,33]). Figures 1a and 1b show the distribution of total purchases per individual, respectively for individuals who test 130 and 129 mg/dL. For each type of individual, Figures 1c and 1d illustrate the distribution of purchases across the six drug classes (categories) listed in the data. Drugs targeting cardiovascular and central nervous system conditions command the chart. In line with independent data (IMS Institute for Healthcare Informatics 2015), most purchased drugs are generics (86% overall). However, note that there is significant generic share variation across drug classes. We will take advantage of this

Figure 1: Main descriptives of the drug purchases (claims) sample: treated (130 mg/dL) vs. control (129 mg/dL) individuals.

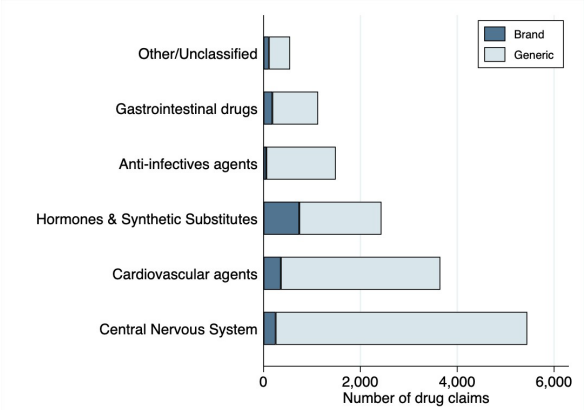
(a) Treated individuals (130 mg/dL): number of purchases per-individual.



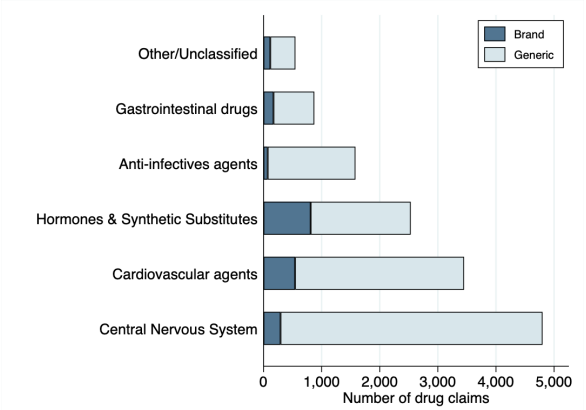
(b) Control individuals (129 mg/dL): number of purchases per-individual.



(c) Treated individuals (130 mg/dL): number of transactions across drug classes.



(d) Control individuals (129 mg/dL): number of transactions across drug classes.



Notes. Data described in these plots are from the 62% of individuals with at least one prescription drug claim in the analytic sample.

variation to investigate the generalizability of our results. As further stressed below, notice that the distributions do not significantly di↵er between the two types of individuals.

QUASI-EXPERIMENTAL FRAMEWORK

The Role of LDL in Lipid Management

Cholesterol is a waxy substance that circulates in the blood stream which assists a series of biologic processes. Cholesterol can also enter the body through food, particularly meat,

poultry, and dairy. There are two main kinds of cholesterol: high-density lipoprotein (HDL) or "good" cholesterol, and low-density lipoprotein (LDL) or "bad" cholesterol. The latter is problematic because it can narrow blood vessels and impede blood circulation. This problem can evolve into serious adverse health events, such as strokes. For this reason, LDL tends to be the main focus in clinical guidelines for lipid management.

Clinical guidelines for cholesterol management adopt a series of thresholds for LDL levels. In particular, 130 mg/dL marks the frontier between "near or above optimal" and "borderline high" level. For example, the formal NCEP ATP-III lipid management guideline prompts providers to consider a cholesterol-management treatment for patients who have up to two risks factors and receive an LDL test result of 130 mg/dL or higher (National Institutes of Health 2001). Similarly, the Cleveland and Mayo Clinics both state in their patient-oriented websites that LDL should be "less than 130 mg/dL." The popular medical website WebMD provides similar guidance.[2]

Local Randomization

We posit that, compared with an infra-marginal patient (129 mg/dL), a patient testing at 130 mg/dL is *treated* to a probabilistic shock of bad news about his/her health condition. For these "treated patients" (130 mg/dL), concepts such as "borderline high" and "abnormal" may be used with higher probability in connection to their health conditions, as compared with "control patients" (129 mg/dL). As a result, treated patients may become disproportionally likely to experience the kind of negative emotions which tilt decisions towards lower-risk alternatives.

To identify the e↵ect of such bad news on generic choice propensity, we rely on the assumption that on which side of the 129/130 mg/dL frontier an individual lands is a random event. The main argument supporting this assumption pertains to the relatively large LDL measurement error of LDL test results. The medical literature decomposes this error into "analytical" and "pre-analytical" variability (Marcovina, Gaur, and Albers 1994). Sources

of pre-analytical variability include factors such as the individual's posture during sampling, the duration of tourniquet application, how strictly and for how long the individual adhered to fasting prior to testing, among various others (Narayanan 1996). In turn, analytical variability is determined by how the blood sample is handled and analyzed in the lab, e.g., whether and for how long it was frozen prior to analysis.

Marcovina, Gaur, and Albers (1994) estimate that analytical variability corresponds to 1.3% of the mean LDL result while pre-analytical variability reaches 9.2%. In the context of our sample, these results would indicate that pre-analytical and analytical factors add a standard deviation of about 11.5 mg/dL around the "true" LDL level. Based on this large variability, we can assume local randomization between individuals testing 129 and 130 mg/dL.

Table 1 presents a series of pre-period (i.e., before testing) statistics that show treated and control individuals have similar characteristics, operate in similar contexts, have similarly generous insurance plans, and behave in similar ways with respect to healthcare utilization. Panel A focuses on demographics—age and sex. As shown by the small standardized di↵erences of Column 3, these variables balance tightly between groups. For example, the average age of treated and control patients di↵ers by less than 1% of a standard deviation. This difference is much smaller than the usual 25% and 10% standardized di↵erence thresholds used in the literature to declare covariate imbalance (Austin 2009). A similar result is observed for the female indicator and geographical location categories (regional indicators).

The next panel considers a series of variables that track the amount of utilization measured through the number of filed claims. These variables are: claims for drug purchases (total, generic, and cardiovascular-targeted), claims for in-patient admissions and outpatient services, and claims for medical tests (all measured as monthly averages). All variables balance tightly between groups. In Panel C we consider the number of (other) medical test claims filed within the same day as the LDL test under study. The average of around 12 suggests that patients take several tests at the same time, which would be consistent with

Table 1: Balancing in the quasi-experimental setup.

	(1)	(2)	(3)
	129mg/dL	130mg/dL	Std. Di↵.
(A) Demographics and location			
Age	48.78	48.83	0.00
	(10.92)	(11.04)	
Sex (fraction female)	0.59	0.54	-0.07
	(0.49)	(0.50)	
Regional indicators			0.06
(B) Utilization prior to LDL testing (monthly claims)			
All drugs	0.73	0.72	-0.00
	(2.02)	(1.32)	
Generic drugs	0.54	0.57	0.01
	(1.61)	(1.05)	
Cardiovascular drugs	0.18	0.15	-0.03
	(0.78)	(0.42)	
In-patient admissions	0.01	0.01	-0.04
	(0.08)	(0.05)	
Outpatient services	2.21	1.84	-0.05
	(7.13)	(3.48)	
All medical tests	0.50	0.39	-0.04
	(2.27)	(1.98)	
(C) Same-day medical testing			
Same-day medical tests	13.57	11.62	-0.07
	(20.61)	(18.75)	
(D) Insurance coverage (implicit)			
Coverage ratio	0.40	0.40	-0.01
	(0.28)	(0.28)	
N	1,169	1,113	

Notes. Parentheses show sample standard deviations.

routine yearly checkups. The small standardized di↵erence again suggests that treated and control individuals have a similar experience during the day the LDL test is taken.

As mentioned, healthcare insurance plan characteristics are not observed in the data. Nevertheless, we can assess the implicit generosity of the available insurance through a "coverage ratio." We compute this variable by dividing the patient's out-of-pocket payments (deductible+copay+coinsurance) on the total payments made by the insurer to the pharmacy. As shown in Panel D, this ratio averages about 40%, without much of a di↵erence between groups. Thus, unobserved insurance variation should not drive our results.

In Appendix A, we present a series of additional analyses that further document the ex-

periment's tight balancing. In particular, we show the absence of systematic treatment/control differences in the following respects: (i) medical diagnoses codes associated with the LDL test, (ii) kinds and amount of non-LDL testing carried out prior to the LDL test, (iii) quantitative results of non-LDL tests taken prior to the LDL test, and (iv) time when the tests (LDL and non-LDL) are administered. We interpret the summation of these results as strong evidence in favor of the validity or our research design.

IMPACTS ON DRUG CONSUMPTION BEHAVIOR

Although our main focus is on how bad medical news shocks may impact brand/generic choice, it is conceivable that these shocks may have broader impacts on drug consumption. For example, the bad news could lead to changes in the bundle of drugs used if patients are prompted to adopt preventive treatments or abandon those of relatively strong side effects. Similarly, the bad news could introduce changes in the quantity of drugs used, e.g., if patients react by improving treatment adherence (e.g., becoming less likely to "miss a pill"). In this section we investigate the extent to which these "bundle" and "quantity" effects are observed in practice. Results do not support the presence of these effects.

Quantity effects. To measure changes in the quantity of drugs that patients consume, we formulate an outcome variable that tracks the total number of prescription drug purchases by patient i during period t. Since most tests in the data (92%) are taken sometime during the month, we cannot define time periods as calendar months. (Testing dates are distributed relatively uniformly in time. See Figure A.1a.) Accordingly, we define time periods as 30-day windows relative to the testing date. For example, for an individual who took the test on September 18th of 2012 (9/18/12), this approach would give us t = –1 for 8/20/12- 9/17/12 (last period before the test), t = 0 for 9/18/12 - 10/17/12 (period starting the day the test is taken), t = 1 for 10/18/12 - 11/16/12, and so on, with t = 0 marking the beginning of the "post-testing" period. This person enters the final dataset through 23 observations, one for each of the fully covered 30-day periods, t = -20 (1/27/11 - 2/25/11), ..., 2 (11/17/12 -

12/16/12). The full dataset contains 31,567 observations derived from the 1,408 individuals associated with at least one drug purchase. On average, patients purchase 1.03 prescriptions each time period (SD = 1.71). About 57% of the observations in the panel are zero. Using the data formatted in this way, we estimate the following di↵erence-in-di↵erences specification:

$$\text{PURCHASES}_{it} = /3 \cdot \text{Treated}_i \rightarrow \text{Post}_t + >_i + 6_t + \blacksquare_{it}, \quad (1)$$

where PURCHASES_{it} represents the log of (one plus) the number of drug purchases (i.e., claims) within each (i, t) cell. The variable Treated_i is a (time-invariant) indicator for whether individual i was treated to the bad medical news, i.e., tested 130 mg/dL instead of 129 mg/dL. The variable Post_t identifies transactions that occur after the LDL test, i.e., Post_t = 1[t 2" 0]. The key parameter is /3. An estimate $\hat{\beta} > 0$ would indicate that the number of drug purchases exhibits a disproportionate increase for treated individuals compared to control individuals in the post period. The term ▪ is a disturbance; > and 6 are fixed e↵ects (described below). Standard errors are clustered at the individual level using the bootstrapping-based method (N=500) of Bertrand, Duflo, and Mullainathan (2004).

Several aspects of Equation 1 are important to highlight. First, since we cannot qualify whether patients directly observe the test result (probabilistic treatment), /3 estimates are formally described as intention to treat results (Angrist, Imbens, and Rubin 1996). Second, given that the sample only includes individuals testing 129 or 130 mg/dL, the bad news shock corresponds to the di↵erential information received by the latter compared to that received by the former individuals. That is, our framework does not allow potential treatment e↵ect asymmetries rooted on whether the news is positive or negative. In other words, similar to interpreting a 130 mg/dL result as bad news relative to a 129 mg/dL result, we could interpret the latter as good news relative to the former.

Also notice that Equation 1 includes narrowly defined fixed e↵ects. The individual-level fixed e↵ects > absorb the influence of all time-invariant characteristics of individuals,

e.g., socioeconomics and demographics, overall health condition and medical history, etc. Considering that most people rarely change insurance plans (Handel 2013), > e↵ects also help control for unobserved insurance coverage di↵erences. In turn, the time period fixed e↵ects 6 control for e↵ects operating in relation to the temporal proximity between the drug purchase and the medical test. For example, in anticipation of taking the test (e.g., during t = −1), patients could become less likely to "miss a pill," thereby increasing total consumption. Equation 1 would flexibly capture such e↵ects through a larger-than-average estimate for 6_{-1}. Similarly, the interaction with the medical system implied by taking the test could lead to a temporary increase in drug consumption, e.g., through refilled prescriptions. These e↵ects would be captured by 6 parameters for the immediate aftermath of the test, e.g., 6_0 and 6_1. Lastly, note that Equation 1 omits the variables Treated and Post in stand-alone form because > and 6 e↵ects make them redundant.

The estimated coefficient for /3 is presented in Column 1 of Table 2. While the coefficient's positive value associates the bad news shock with an increase in purchased quantities, the e↵ect is statistically non-significant at conventional levels. We take this result as evidence that bad medical news does not impact the quantity of drugs consumed.

Table 2: Main results.

	(1)	(2)	(3)	(4)	(5)	(6)
	Consumption behavior		Generic choice propensity			
Dependent variable	PURCHASES	BUNDLE	GENERIC	GENERIC	GENERIC	GENERIC
Treated⇥Post	0.0171	0.0172	−0.0109⇤	−0.0174⇤⇤		
	(0.0176)	(0.0159)	(0.0064)	(0.0076)		
Treated⇥Post⇥(1-HighSpender)					−0.0265⇤⇤	
					(0.0104)	
Treated⇥Post⇥HighSpender					−0.0085	
					(0.0067)	
Treated⇥Post⇥Periods0-2						−0.0169⇤⇤⇤
						(0.0065)
Treated⇥Post⇥Periods3-6						−0.0061
						(0.0078)
Treated⇥Post⇥Periods7+						−0.0079
						(0.0092)
Individual FEs	×	×	×	×	×	×
Time period FEs	×	×	×	×	×	×
Molecule FEs			×	×	×	×
Sample	Full	Full	Full	With office visit	Full	Full
Aggregation	Individual/month		Transaction			
N	31,567	31,567	34,935	20,512	34,935	34,935

Notes. Linear probability specifications for the probability of choosing the generic option (Equation 2). Parentheses show standard errors. For estimates in Columns 1-2, errors are clustered at the individual level; for estimates in Columns 3-6, they are clustered at the individual/molecule level. *Legend:* ⇤ $p < 0.1$,⇤⇤ $p < 0.05$,⇤⇤⇤ $p < 0.01$.

Bundle e↵ects. Recall that we conceptualize bundle e↵ects as changes in the set of drugs used by the patients induced by the bad LDL news. We use the same procedures described above to construct a variable that tracks the number of di↵erent molecules purchased by each patient during each time period. The resulting variable, $BUNDLE_{it}$, equals the log of (one plus) the number of di↵erent molecules purchased by patient i during time period t. On average, individuals purchase 0.96 di↵erent molecules each time period (SD = 1.55). Column 2 of Table 2 presents the /3 estimate that we obtain by estimating Equation 1 using BUNDLE as dependent variable. We again obtain a positive and statistically non-significant estimate, which fails to support the hypothesis that bad news prompts changes in the set of drugs consumed.

IMPACTS ON GENERIC CHOICE PROPENSITY

Here we turn to our main objective, which is to estimate the impact of bad LDL news on individual's generic choice propensity, i.e., the probability that an individual chooses the generic over brand name option conditional on purchasing a drug. Accordingly, we use the data in disaggregated form, i.e., a dataset in which each observation corresponds to a drug purchase. Also recall that our dataset includes drug purchases covering patients' full set of drug needs, not just those associated with high cholesterol. Using this full dataset, we estimate the following di↵erence-in-di↵erences model which minorly adapts Equation 1:

$$GENERIC_{ijtm} = /3 \cdot Treated_i \rightarrow Post_t + >_i + 6_t + \mu_m + \blacksquare_{ijtm} \quad (2)$$

The dependent variable $GENERIC_{ijtm}$ is an indicator activated if the drug purchased in transaction j by individual i during period t was a generic (as opposed to brand name) product of molecule m. Given this definition, an estimate $\hat{\beta} < 0$ would indicate that the generic propensity of treated individuals disproportionally falls compared with that of control individuals after receiving the test result. We would interpret this result as evidence that

bad LDL news increases the preference for branded over generic drugs. Like Equation 1, Equation 2 includes > and 6 fixed e↵ects, respectively aimed at controlling for time-invariant unobservables at the individual level and consumption dynamics around the time of the test. In addition to these controls, Equation 2 includes molecule-level fixed e↵ects, µ. We include these to help control for variation arising from the number of generic options, average brand/generic price gaps, etc. We cluster errors at the individual/molecule level.[3]

Main E↵ect Estimate

The estimate for the coefficient /3 in Equation 2 is presented in Column 3 of Table 2. The estimate is negative and marginally significant (i.e., significant with 90% confidence), supporting that bad medical news increases the preference for brand name drugs. The estimate indicates that, compared with control individuals (129 mg/dL), treated ones (130 mg/dL) experience a reduced generic choice probability of an additional -0.0109 after the test. Considering the pre-testing generic propensity baseline, this result suggests that the bad news shock reduces the frequency of generic choice by about 1.3%. Equivalently, given the relatively modest share of brand name drugs (14% overall), this point estimate represents an 8% increase in the average patient's propensity to choose the brand name option. In addition to the plausibly random treatment assignment, recall that our model controls for unobserved individual- and molecule-level variation, as well as for potential anticipatory e↵ects. As such, it is difficult to attribute this result to reasons other than the LDL test outcome. We are further reassured by two additional sets of results. First, we implement two sets of placebo tests (falsified thresholds and testing dates), none of which falsifies the result. Second, a formal test rejects the presence of confounding pre-trends. These analyses and their respective results are presented in Appendix B.

Treatment Intensity

Encouraged by the previous finding, we now probe the e↵ect in relation to the intensity of the treatment experienced by di↵erent individuals. We consider a series of scenarios where the intensity of the bad news treatment is implicitly altered. Across these contexts, we find directionally consistent changes in the estimated e↵ects on generic propensity.

Co-occurrent medical appointments. We first consider a scenario where treatment intensity may vary based on the salience of test results to patients. In particular, we leverage the idea that test results may be more salient for those patients who also have a primary care medical appointment (office visit) around the time of the test. As noted before, information on the incidence of these visits is available from out-patient services claims, and about 70% of the sample register one such visit within a week of the test. The coefficient in Column 4 of Table 2 results from reproducing our estimation using only the drug claims data for this subset of patients. Consistent with increased treatment salience, the -0.0174 estimate for β_3 is larger in magnitude and more precisely estimated (statistically significant with 95% confidence) than its full-sample counterpart.

Health status. We now investigate how the patient's health status may moderate the bad news e↵ect. Compared with sick patients, healthy patients may be less used to receiving information that unveils a health deficiency. Bad LDL news may therefore imply a larger shock for healthy than for sick patients.

The primary empirical hurdle to investigating this hypothesis stems from the fact that MarketScan data do not contain variables describing the overall health condition of individuals. Given this limitation, we use total healthcare expenditures as a proxy for health status, under the assumption that higher healthcare expenditures reflect poorer health condition. In particular, we create the indicator variable $\text{HighSpender}_i = 1[u_i \geq 0]$, where u_i is the residual of an equation that regresses individual i's logged total healthcare expenditures prior to the LDL test (on drugs, in-patient and out-patient services) on a set of demographics and the number of in-sample days prior to the test. (Appendix A.4 presents estimation details and

the distribution of u.) We interpret HighSpender=1 (=0) as reflecting relatively poor (good) health status. In Appendix A.1, we present an additional analysis that lends support to this interpretation, despite the potential for reverse causality e↵ects.

We incorporate the HighSpender indicator into Equation 2, as illustrated in Column 5 of Table 2. Separate bad news e↵ects are estimated for patients in each group (High-Spender=0,1). The estimated bad news e↵ect parameter for individuals associated with HighSpender=0 (i.e., good health) is −0.0265, which is more than twice of that estimated from the full sample, and statistically significant with 95% confidence. For individuals associated with HighSpender=1 (i.e., poor health), the parameter is −0.0085, which is much smaller as well as statistically non-significant. These results suggest that healthier patients are more vulnerable to the bad news e↵ect on brand/generic choice.

Temporal e↵ects. Research from multiple disciplines converges on the finding that emotional reactions tend to be short-lived (e.g., Ekman 1999; Card and Dahl 2011; Depetris-Chauvin, Durante, and Campante 2020; Verduyn et al. 2009; Verduyn, Van Mechelen, and Tuerlinckx 2011). Based on this finding, we conjecture that the bad news treatment may have higher intensity in the immediate aftermath of the test, decaying afterwards (i.e., front-loading). To investigate this idea, we modify Equation 2 to allow for the estimation of separate bad news e↵ects over three post-testing time periods. We select the following cut-o↵s (which partition post-testing transactions into approximate terciles): (i) first 90 days after testing (periods t=0,1,2), (ii) days 91-210 after testing (periods t=3,..,6), and (iii) day 211 and after (periods t=7,..). Compared with the full sample estimate of Column 3 (−0.0109), the −0.0169 coefficient for the first of these periods (Table 2, Column 6) is larger and more precisely estimated (significant with 99% confidence). The signs of the estimates for the next two periods continue to be negative, but they are statistically non-significant. These results support the idea that the negative emotions infused by the bad news have front-loaded e↵ects.

The e↵ect's front-loading raises the question about the specific decisions through which

the e↵ect operates. There are two possible channels. First, the negative e↵ect on generic propensity could unfold via *switching* decisions, i.e., a combination of slowed-down brand-to-generic switching and accelerated generic-to-brand switching. However, with less than 2% of purchases representing switching between brand name and generic alternatives, this channel can at most play a minor role in explaining the e↵ect in our data. The second channel pertains to brand/generic choice when patients *adopt* a new molecule, i.e., when they purchase it for the first time. It is possible that, in the context of adopting a new molecule, bad medical news tilt adopters' preferences towards the brand name option. In Appendix C, we present evidence consistent with this hypothesis. We find that the bad news shock increases the propensity to choose the brand name option when patients adopt a new molecule. Consistent with the front-loading results above, our estimates show that the e↵ect on generic propensity for newly-adopted drugs is also short-lived.

Multiple testing. We conclude by focusing on the issue of multiple LDL testing. People who test for LDL more than once may not only be more driven to thoroughly analyze the results, but also be better acquainted with the measurement error. Accordingly, we conjecture that multiple testing may be associated with reduced treatment intensity.

To analyze the problem, we construct a dataset using the information of individuals (N=2,341) who record at least one LDL test in addition to that resulting in a frontier 129/130 mg/dL result. (Recall that these individuals were excluded from our main sample.) A series of empirical considerations is required to analyze these data (e.g., individuals vary in how many additional testing results they record), so we present our procedures and results in Appendix D. Consistent with the bad news e↵ect, we also obtain a negative /3 estimate from this sample. However, the estimate is about one third the magnitude of our main estimate in Column 1, as well as statistically non-significant. A second analysis on the same data suggests that the bad news e↵ect may concentrate on individuals who did not previously receive a result of 130 mg/dL or higher. This result coincides with the intuition that individuals who have not previously received bad LDL news may be more surprised to receive it. In parallel,

the 130 mg/dL result may represent good news to those individuals who have previously obtained a borderline high (2" 130 mg/dL) result. Given several limitations and lack of statistical significance, we interpret these results as merely suggestive for the idea that bad LDL news has a marginally decreasing impact on brand/generic choice.

ADDITIONAL EVIDENCE: BAD A1C NEWS

The analysis presented in this section assesses the generalizability of the bad news e↵ect on generic propensity. We do so by examining the impact of bad news generated by a di↵erent type of medical test, i.e., Hemoglobin A1c tests, which are also contained in MarketScan Lab files. These tests measure blood sugar levels and are used for diagnosing and managing diabetes.

Hemoglobin A1c Tests

Three shared features with LDL testing make A1c testing a good secondary candidate for our analysis. First, like LDL results, A1c results are also expressed on a continuous scale, i.e., as the percentage of red blood cells with sugar-coated hemoglobin (typically 5-8%). Second, A1c results also include a significant measurement error, around 0.5% (Phillipov and Phillips 2001). This error introduces the necessary local randomization around clinical thresholds. Third, like LDL tests, A1c tests are common, particularly among people who have been diagnosed with diabetes.

Despite these favorable features, two aspects of A1c testing introduce a measure of experimental noise. First, compared with LDL cholesterol, the interpretation guidelines for A1c tests are not as strict as those for LDL tests. For example, the American Diabetes Association writes "providers might reasonably suggest even lower A1C goals than the general goal of <7% [...] conversely, less-stringent A1C goals than the general goal of <7% may be appropriate for patients with a history of [...]" (American Diabetes Association 2010). The website WebMD, which is highly popular among patients, includes a similar emphasis. This

aspect suggests that the measurable impacts of bad A1c news on generic propensity may be mu▬ed compared with those of bad LDL news.

Second, A1c results inform two distinct clinical decisions. The 7% threshold referenced above is used by diagnosed patients to manage the condition. In addition, two other thresholds are used for diagnosing the condition, 5.7% and 6.5% (entry to pre-diabetic and diabetic ranges, respectively). Whereas the 7% for diabetes management has been consistently adopted by official diabetes management guidelines (e.g., American Diabetes Association 2010, 2021), there has been an on-going debate about whether the latter two should be relied upon for diagnosis.[4] Consistent with this scenario, we only detect bad news e↵ect on generic choice propensity around the 7% frontier. Estimates obtained using the 5.7% and 6.5% cut-o↵s are presented in Table A.9. In line with our results for the LDL sample, we fail to detect statistically significant impacts on the quantity or bundle of consumed drugs (see results in Table A.8).

Analysis

With these caveats in mind, we incorporate into our analysis the 143,165 drug claims associated with the 3,725 patients in the 6.9/7% A1c frontier, with virtually no patient overlap with the LDL sample.[5] In our first analysis, we estimate Equation 2 using these data only. Consistent with our results from LDL testers, we also obtain a negative estimate, $\hat{\beta} = -0.0049$ (SE=0.0033), which is about half the magnitude of its analog from the LDL analysis, and statistically non-significant. This result suggests that the relatively more flexible A1c guidelines may mu▬e the bad news e↵ect on generic propensity. Another possibility stems from the fact that, as revealed by total expenditures, A1c testers may be in worse health condition than LDL testers in general (see Appendix A.1). Coupled with our results in relation to the moderating role of health status, the smaller e↵ect for A1c testers may reflect that these patients are more used to receiving bad medical news than LDL testers.[6]

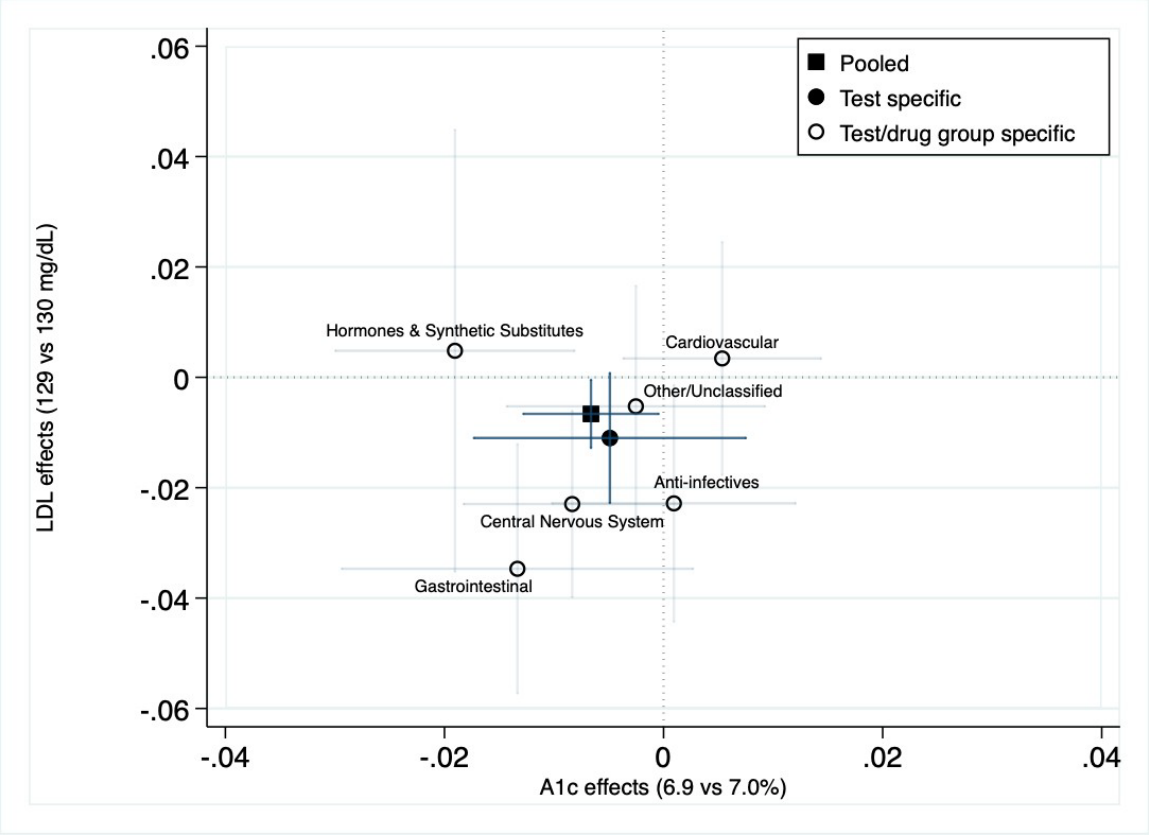
We next re-estimate Equation 2 on a dataset that combines the drug purchases of frontier

LDL (129 vs. 130 mg/dL) and frontier A1c (6.9 vs. 7%) individuals (N=5,131). In this "pooled" regression, baseline differences between the two testing contexts are absorbed by the individual-level fixed effects included in the model. The resulting estimate, $\hat{\beta} = -0.0066$ (SE=0.0029), is statistically significant with 95% confidence, although about half the size the estimated baseline effect from the LDL sample.

To further compare the effects of bad LDL and A1c news on brand/generic choice, we analyze the heterogeneity of the bad news effect across molecule classes. In particular, we use a slight modification of Equation 2 to estimate bad news effects specific to each of the six drug classes listed in Figure 1. Results are summarized by a set of estimates $\{\hat{\beta}^{k,test}\}$, where test 2 {LDL, A1c}, and k indexes drug classes (anti-infectives, gastrointestinal, etc.). (Models are separately estimated using the data of each test. Further estimation details are presented in Appendix F.2.) These estimates, which have the same interpretation of prior analyses, are presented in Figure 2 along their 95% CIs. To illustrate this graphical presentation, consider Gastrointestinal molecules. For this class, we estimate $\hat{\beta}^{\mathrm{Gast,LDL}} = -0.034$ (SE=0.011) using data from LDL testers, and then estimate $\hat{\beta}^{\mathrm{Gast,A1c}} = -0.013$ (SE=0.008) using data from A1c testers. Figure 2 represents this result through the hollow circled marker in coordinates $(\hat{\beta}^{\mathrm{Gast,A1c}}, \hat{\beta}^{\mathrm{Gast,LDL}}) = (-0.013, -0.034)$.

Two findings from Figure 2 support the generalizability of the bad news effect. First, most estimates are negative. That is, in both the LDL and A1c samples, estimates are directionally consistent with the bad news effect across the six drug classes covered by the data. Second, there is a positive correlation of 0.13 across the vector of six class-specific point estimates obtained from each test; the correlation increases to 0.82 when we omit the outlier drug class (Hormones & Synthetic Substitutes). We interpret this result as evidence that the bad news effect may generalize across different medical tests.

Figure 2: E↵ects of bad LDL and bad A1c news across drug classes.



Notes. The pooled estimate (squared marker) is obtained by estimating a single regression (Equation 2) on the full dataset composed of LDL (129/130 mg/dL) and A1c (6.9/7%) testers. (Note that there is a single pooled estimate—its value is reproduced in both axes.) Test-specific estimates (full circled marker) are obtained by estimating Equation 2 separately on each of the two samples, with LDL testers on the one hand and A1c testers on the other. Test/drug group specific estimates (hollow circled markers) are obtained through separate estimations on the samples of LDL and A1c testers, as described in Appendix F.2. In all cases, bars represent 95% Confidence Intervals from standard errors clustered at the patient/molecule level.

THE ROLE OF PERCEIVED QUALITY DIFFERENCES

As we have noted, a precondition for the bad news e↵ect is the presence of perceived quality di↵erences between brand name and generic drugs. Here we present two analyses that probe this idea. Our first analysis leverages the observation that, in a di↵erentiated product market, perceived quality di↵erences between brand name and generic options should be positively correlated with their corresponding price di↵erences. Accordingly, we find that the bad news e↵ect focuses on molecules where the brand name option is sufficiently more expensive than the generic counterpart. Our second analysis leverages a prediction from research on consumer learning, namely, that the accumulation of consumption experience helps patients grasp a drug's "true" therapeutic properties. Consistent with this prediction, we find that the bad news e↵ect decays with experience, arising only for relatively inexperienced patients.[7] We use these results to inform our final analysis, which characterizes the bad news e↵ect's heterogeneity with respect to molecules that are clinically relevant to each test and those which are not. We find that the bad news e↵ect operates on both types of molecules. However, the e↵ect is considerably larger on clinically relevant molecules.

Evidence from Pricing Di↵erentials

Given that generics are molecular replicas of their brand name counterparts, the bias against generics tends to be rationalized based on their perceived quality di↵erences. Accordingly, we should expect a stronger bad news e↵ect when the perceived quality di↵erences are larger. To test this implication, the main empirical hurdle is that we do not directly observe perceived quality di↵erences. Here we circumvent this challenge by leveraging pricing di↵erentials.

Our analysis builds on two strands of literature, both of which associate larger brand/generic price di↵erences with larger di↵erences in perceived quality. First, in structural models of drug choice, consumption utility is increasing in perceived quality and decreasing in price (e.g., Ching 2010a; Crawford and Shum 2005; Narayanan and Manchanda 2009). In this

framework, profit-maximizing firms would charge higher prices for drugs of higher perceived quality (Anderson, De Palma, and Thisse 1992; Ching 2010b). Second, considering that drugs can be described as experience goods (Berndt 2002), patients may use prices to make inferences about quality (e.g., Erdem, Keane, and Sun 2008; Milgrom and Roberts 1986; Wathieu and Bertini 2007). Based on these rationales, we posit that if the bad news e↵ect stems from generics being perceived as of inferior quality than brand name drugs, then the e↵ect should be stronger when the brand name option is relatively more expensive than the generic option.

To implement this test, we take advantage of the widespread brand/generic price di↵erences observed in the market. For a molecule j, we operationalize the price di↵erential as $\delta_j = (p_j^{Brand} - p_j^{Gen})/p_j^{Gen}$. Following the literature cited above, a positive δ_j would imply that there is a perceived quality gap favoring the branded option. Hence, a near-zero δ_j would suggest that perceived qualities are similar. We should expect the bad news e↵ect to be stronger for molecules where δ values are larger.

We construct δ di↵erentials leveraging data from Average Wholesale Prices (AWPs), which are the equivalent to sticker/list prices in traditional retailing (Alpert, Duggan, and Hellerstein 2013; Gencarelli 2002). Details on how we construct p_j^{Brand} and p_j^{Gen} using AWP data are provided in Appendix G. The resulting distribution showcases wide variation (see Figure A.4a), with a median of 55% price di↵erence in favor of brand name drugs.

In our econometric model, we account for the variation of δ through a median split, as shown in Table 3. Whereas the two-way interaction Treated⇥Post captures a baseline bad news e↵ect that applies to all molecules, the triple interaction Treated⇥Post⇥AboveMedianδ captures an additional e↵ect which would apply only to molecules for which the price differential is large enough in favor of the brand name option. (The specification is otherwise identical to Equation 2.) Columns 1 and 2 of Table 3 show the estimates obtained from the LDL and A1c samples, respectively; Column 3 shows the estimates from the pooled sample. (Estimation samples are somewhat smaller than in our previous analyses due to

Table 3: Price as a signal of quality.

	(1) LDL	(2) A1c	(3) Pooled
Treated→⇥Post	−0.0081	0.0084⇠	0.0057
	(0.0097)	(0.0043)	(0.0040)
Treated→⇥Post→⇥AboveMedian$\hat{\theta}$	−0.0055	−0.0281⇠⇠⇠	−0.0254⇠⇠⇠
	(0.0100)	(0.0045)	(0.0040)
N	33,840	138,245	172,126

Notes. Linear probability specifications for the probability of choosing the generic option (Equation 2). The pooled sample includes individuals from both the LDL and A1c samples. All models include fixed e↵ects for molecules, individuals, and time periods. Parentheses show standard errors clustered at the level of patient/molecule pairs. *Legend:* ⇠ $p < 0.1$,⇠⇠ $p < 0.05$,⇠⇠⇠ $p < 0.01$..

missing price information.) Consistent with our prediction, the bad news e↵ect strengthens for molecules associated with above-median $\hat{\theta}$ values. This strengthening is clearer (and statistically significant) in the A1c and pooled samples. Moreover, for these two samples, it is possible to conclude that the bad news e↵ect arises primarily for molecules associated with above-median $\hat{\theta}$ values.

In Appendix G we present an analog analysis that is based on approximated Out-of-Pocket prices instead of AWPs. We obtain consistent results that the bad news e↵ect is primarily observed where the brand name alternative is sufficiently more expensive than the generic one.

Evidence from Consumption Experience

Several studies document how the accumulation of consumption experience leads to patients learning about drugs' "true" therapeutic properties (e.g., Ching 2010a,b; Crawford and Shum 2005). Since generic drugs are molecular replicas of brand name counterparts (and hence have the same "true" properties), this form of learning should progressively level the perceived qualities of brand name and generic drugs. Building on this observation, we hypothesize that the bad news e↵ect decays with consumption experience. The analysis presented in this section finds support for this hypothesis. The crucial input needed to implement the test

Table 4: Consumption inexperience scores.

	Mean	Std. Dev.
Cardiovascular	0.45	0.39
Hormones & Synthetic Substitutes	0.48	0.40
Central Nervous System	0.53	0.41
Gastrointestinal	0.54	0.41
Unclassified	0.61	0.40
Anti-infectives	0.83	0.30
Total	0.53	0.41

Notes. Inexperience scores are computed at the patient/molecule level, as per Equation 3. The scores summarize the amount of consumption experience that the patient has with respect to a given molecule, with lower scores reflecting more experience. Scores tabulated here are for the sample used to estimate the models of Table 5, which is composed of 2012 purchases by patients who tested for LDL or A1c during 2012.

is a measure of consumption experience. Since we cannot measure consumption experience for patients purchasing drugs early in the sample, we reserve the early portion of our data to assess patients' experience levels. Accordingly, we use drug purchase data for 2011 (first year of our sample) to compute the experience measure. We then incorporate this measure into the sample of 2012 purchases, which we use to estimate the models. To counteract the sample size reduction, we rely on the sample that pools the data of LDL and A1c testers for estimation.[8]

To facilitate the interpretation of our econometric estimates, we formulate a metric of inverse experience, or "inexperience." This metric is defined as:

$$\text{Inexperience}_{im} = \frac{1}{1 + \#\ \text{purchases of molecule m by patient i during 2011}} \in (0, 1]. \quad (3)$$

According to this formulation, if patient i did not purchase molecule m during 2011, s/he is given the maximum inexperience score of one (i.e., minimum experience). In turn, the larger the number of 2011 purchases is, the smaller the inexperience score becomes. The score approaches zero as the number of purchases goes to infinity.[9]

Table 4 describes the variation of the Inexperience score, presented separately for the six

molecule classes codified in the data. The Cardiovascular class is associated with the most experienced patients, with inexperience scores that average 0.45. As for other classes, these scores exhibit a significant amount of within-class variability (SD=0.39). In the other extreme, the Anti-infectives class has the least experienced patients, with scores averaging 0.83 (SD=0.30). This comparison between the Cardiovascular and Anti-infectives class is intuitive in that, given that most cardiovascular conditions are chronic, the scope for experience accumulation is much larger. By contrast, most conditions treated with anti-infectives are acute (i.e., short-lived), thereby providing fewer opportunities for experience accumulation. This relationship generalizes to the full sample, where patients purchasing drugs for chronic conditions are associated with 0.2 lower inexperience scores than the average other patient ($p < 0.01$).

Column 4 of Table 5, Panel A, presents estimation results for a model that incorporates the inexperience score as a moderator for the bad news e↵ect. In addition to the Treated⇥Post interaction, the model includes the triple interaction Treated⇥Post⇥Inexperience, and is otherwise identical to Equation 2. As in our previous analyses, the coefficient for Treated⇥Post captures a baseline bad news e↵ect operating regardless of consumption experience. In turn, the coefficient for Treated⇥Post⇥Inexperience captures an additional bad news e↵ect, which operates in direct proportion to inexperience. The coefficient estimate for the baseline bad news e↵ect (Treated⇥Post) is positive, although quite small as well as statistically non-significant. By contrast, the estimate for the triple interaction parameter is negative, marginally significant (90% confidence), and large in magnitude. In Panel A, we have also included the bad news e↵ect estimates from previous analyses to highlight that the bad news e↵ect becomes contingent on sufficiently high levels of inexperience. Evaluated at maximum inexperience, the bad news e↵ect amounts to a 0.0105 reduction of the generic choice probability. In addition, given the formulation of the inexperience score (Equation 3), Column 4 estimates imply that the bad news e↵ect disappears after 2.7 purchases. All in all, we interpret these results as broadly supportive of the idea that, by reducing brand/generic

di↵erences in perceived quality, consumer learning reduces the scope of operation for the bad news e↵ect.

Do Bad News Matter More for Clinically Relevant Drugs?

Recall that we have defined clinically relevant molecules (CRMs) as those targeting the medical condition that is managed based on the results of the medical tests considered for our analyses. As such, CRMs correspond to cholesterol drugs for LDL testers and to diabetes drugs for A1c testers. Non-CRMs correspond to all other molecules covered by the data. Here we study how the bad news e↵ect varies between CRMs and non-CRMs.

In the background section we highlighted two points derived from prior literature which rationalize the bad news e↵ect operating on all drugs, i.e., including non-CRMs. The first is the observation that generics are perceived as inferior to brand name drugs in general, i.e., across the spectrum of all drugs used by patients. Second, emotional stimuli like bad medical news appear to function by altering how individuals weigh alternatives rather than what they know about them, meaning that the e↵ect does not require choice-relevant information (e.g., weather impacts on stock returns). Nevertheless, we may still expect a stronger e↵ect on CRMs, e.g., if the bad testing results are particularly alarming for patients su↵ering a related condition.

To implement our analysis, we begin by formally classifying CRMs. We do so by parsing through products' approved usages, as described by their FDA labels. This codification reveals that the Cardiovascular and Hormones & Synthetic Substitutes (HSS) classes contain many molecules in addition to CRMs. For LDL testers, CRMs (primarily statins) account for only about a quarter of purchases in the Cardiovascular class. In turn, for A1c testers, CRMs (mainly metformin and related products) represent about three quarters of purchases in the HSS class.

Another important consideration is that, as discussed in the previous subsection, high cholesterol and diabetes are chronic conditions and thus generate persistent drug needs. In

Table 5: Inexperience and class spillover e↵ects.

	(1) LDL	(2) A1c	(3) Pooled	(4)
(A) All molecules				
Treated⇥Post	−0.0109←	−0.0049	−0.0066←←	0.0039
	(0.0064)	(0.0033)	(0.0029)	(0.0067)
Treated⇥Post⇥Inexperience				−0.0144←
				(0.0086)
N	34,935	143,165	178,162	42,012
(B) Clinically relevant molecules (CRMs)				
Treated⇥Post	0.0563	−0.0073	−0.0023	0.0092
	(0.039)	(0.0057)	(0.0059)	(0.0231)
Treated⇥Post⇥Inexperience				−0.0597←←
				(0.0277)
N	1,831	24,142	25,973	5,557
(C) Non-CRMs				
Treated⇥Post	−0.0148←←	−0.0043	−0.0076←←	0.0066
	(0.0060)	(0.0039)	(0.0033)	(0.0069)
Treated⇥Post⇥Inexperience				−0.0156←
				(0.0092)
N	33,045	118,741	151,847	36,304
Sample	Full	Full	Full	2012 purchases by 2012 testers

Notes. Linear probability specifications for the probability of choosing the generic option. Columns 1-3 show results for Equation 2; Column 4, for an specification that enriches with Equation 2 with the triple interaction displayed above. The Inexperience score used in this triple interaction is computed at the patient/molecule level (Equation 3). This summarizes the amount of consumption experience that the patient has with respect to a given molecule, with lower scores reflecting more experience. Clinically relevant molecules (CRMs) are molecules targeting cholesterol in the case of LDL testers and molecules targeting diabetes in the cases of A1c testers. We estimate the models of Column 4 on the sample of 2012 purchases by 2012 testers, as described in the text. All models include fixed e↵ects for molecules, individuals, and time periods. Parentheses show standard errors clustered at the level of patient/molecule pairs. *Legend:* ← $p < 0.1$,←← $p < 0.05$,←←← $p < 0.01$..

other words, patients use these drugs over long periods of time, acquiring high levels of consumption experience and knowledge about them compared to drugs used sporadically. This element is evidenced in Table 4, where the Cardiovascular and HSS classes are associated with the least inexperienced patients across all six drug classes. These statistics suggest that accounting for consumption experience may be important to correctly estimate how the bad news e↵ect varies between CRMs and non-CRMs.

In Panel B of Table 5 we present a series of estimates for the bad news e↵ect obtained from the sample of CRMs. Estimates in Columns 1-3 show results for our main specification (Equation 2), separately estimated on the LDL, A1c, and pooled samples. The estimate obtained from the LDL sample (Column 1) is positive, although estimated with significant error and ultimately statistically non-significant. This lack of precision may be attributable to the small size of the sample available to estimate the e↵ect. On the contrary, from the A1c (Column 2) and pooled (Column 3) samples we obtain negative estimates, which align with the presence of the bad news e↵ect among CRMs. However, both of these estimates are statistically non-significant. Column 4 presents the estimates for the specification that incorporates the experience moderator. The obtained estimates describe a statistically significant bad news e↵ect, one that is driven by consumption (in)experience. The estimate indicates that, for fully inexperienced patients (no prior purchases), the bad news shock reduces the probability of generic choice by about 0.06 (7%).

We now turn to estimating the bad news e↵ect that operates on non-CRMs. Results are presented in Panel C of Table 5. The presence of these e↵ects is supported by the negative estimates obtained from the LDL, A1c, and pooled samples (Columns 1-3, respectively), of which the former (LDL) and latter (pooled) are statistically significant (with 95% confidence). From the specification that incorporates experience e↵ects (Column 4), estimates again suggest that the bad news e↵ect is driven by the lack of consumption experience. The most important aspect of these parameter estimates pertains to their magnitude. Holding consumption experience constant, the bad news e↵ect (Column 4) for non-CRMs is about

one fourth the size of its counterpart for CRMs. This indicates that even though the bad news e↵ect seems to operate over non-CRMs, it is considerably smaller than that operating on CRMs.

To conclude, we would like to draw the attention back to Figure 2. Recall that this figure provides class-specific estimates for the bad news e↵ect, for each of the six drug classes in the data. A key aspect of these estimates is that they do not control for consumption experience. As a result of this omission, we estimate relatively bad news e↵ects of relatively large magnitude for some non-CRM classes (e.g., Gastrointestinal, Central Nervous System). Combined with the results of this section, these estimates illustrate that controlling for consumption experience may be important to correctly estimate the bad news e↵ect.

IMPLICATIONS FOR PRACTICE

Our findings have implications for several key stakeholders in the healthcare industry. First, health policy makers, generic drug manufacturers, and insurers (public and private) all share the common goal of encouraging patients to choose generics over brand name drugs. To design policies aimed at achieving this goal, insurers currently rely on two primary toolkits. The first corresponds to a set of demographic and socioeconomic predictors of generic-averse attitudes, which are leveraged for the targeting of interventions. The second toolkit corresponds to possible intervention tools, which in practice boils down to a set of price-based promotional activities (e.g., discounts, coupons, free samples, etc.). Our analysis delivers important new insights to the application of these frameworks.

Concerning the targeting decision, our findings suggest that relying solely on demographic and socioeconomic predictors may lead to neglect of an important observable—the arrival of bad medical news. Accordingly, enriching the targeting framework with variables for recency with respect to these events could improve the targeting campaigns' allocative efficiency.

Our analysis of brand/generic price di↵erentials also raises a potential concern about the common use of price-based incentives to encourage generic use. In line with prior literature

(Dunne et al. 2014; Lambert et al. 1980; Verger et al. 2003), our results are consistent with the idea that patients may use prices to draw inferences about brand/generic qualities. If this behavior is pervasive in the field (which is not confirmed by our analysis), it would introduce a previously unrecognized trade-o↵. Namely, a price incentive that increases the generic option's share-of-wallet appeal (e.g., through a discount) may also deteriorate its perceived quality. In such scenario, the optimal design of price-based campaigns would need to strike a delicate balance between the direct, short-term share-of-wallet e↵ects and the indirect, more slowly-unfolding potential impacts operating through quality inferences.

For providers and administrators, it may be helpful to consider strategies to neutralize the impacts of bad medical news on brand/generic preferences. A first step in this direction consists of generating awareness that even routine medical tests can trigger behavioral responses such as the one we have documented. Although the medical profession places marked emphasis on adequately breaking bad news to patients (Baile et al. 2000; Buckman 1992; Faulkner 1998), the traditional focus has been on cases related to severe outcomes (e.g., death, cancer diagnoses). Expanding this focus to include the much more subtle type of medical news that we consider could have a positive impact on patients' wellbeing as well as on the system's efficiency. A simpler approach would be to remind patients of the equivalency of generic drugs following bad medical news, e.g., via text messages after test results are shared with the patient (Pop-Eleches et al. 2011).

Finally, regulators should take note of the implications on brand name direct-to-consumer advertising, which routinely encourages patients to get tested for a variety of symptoms. Our results suggest that such advice is also consistent with the goal of hindering generic adoption. If drug manufacturers understand these mechanics, they may promote medical testing above and beyond medically justifiable levels. Such distortion would imply excess healthcare spending in terms of both additional testing and forgone savings from generic use.

CONCLUSION AND DIRECTIONS FOR FUTURE RESEARCH

To reduce inefficient healthcare spending, the substitution of brand name with generic drugs is one of the policies that attracts close attention from both policy makers and insurers. Given that generic drugs are molecular replicas of their brand name counterparts, these policies could deliver large savings without sacrificing patient health. Nevertheless, these policies are met with resistance from the public, who exhibit a preference bias against generics. In this article, we contribute to the literature by uncovering a new source of this bias—bad medical news. Our evidence supports the idea that patients may become more reluctant to use generics when they receive news that highlight a deficiency in their health. Since receiving bad medical news is an inherent component of patients' interaction with the health system, the identified e↵ects might operate at a large scale and be responsible for a large amount of over-spending.

One question that remains open from our analysis pertains to attribution: Does the bad news e↵ect reflect the patient's decision or doctor's decision? When considering this question, we first note that our results related to the importance of patients' consumption experience suggests that patients play a role in the bad news e↵ect. However, since several prior works find that doctors mediate the brand/generic decision (Hellerstein 1998; Iizuka 2012), our analysis cannot rule them out, and future research should aim to characterize their role. For example, to what extent do doctors acquiesce to patients' pressure in favor of the brand name option?

Second, although our analysis demonstrates the existence and basic properties of the bad news e↵ect, it does not disentangle the specific psychological mechanisms at play. Related literature highlights two possible mechanisms: patients could shy away from generics because they become more pessimistic about their health status or because they become more averse to health risks.[10] It is also possible that, rather than become motivated by avoiding adverse health events, patients may start to aspire to to improve their overall health

condition. Eliciting the specific channel(s) at play could be helpful to inform the design of remedy interventions and guide future literature. Highlighting the difficulties of addressing the question based on observational data, Bassi, Colacito, and Fulghieri (2013) propose an experimental approach which could also be deployed in the context of the bad news effect.

It is important to note that, despite the signs of generalizability provided by the consistency of bad LDL and A1c news' effects, the treatment effect of bad medical news could vary outside these contexts. This is because the LDL and A1c settings have two key commonalities that may not be shared by other types of medical news: (i) the patient is not directly confronted with outcomes of utmost severity (e.g., death, losing a limb), and (ii) one therapeutic alternative (brand name) strictly dominates the other (generic) in terms of perceived efficacy and safety. The effects of bad medical news could be qualitatively different compared to our evidence when these conditions are not met. For example, Harmon (2010) describes the story of two patients who, after being diagnosed with end-stage skin cancer, made a dramatic plea to be treated with an experimental drug of formally unverified properties. That is, confronting a highly probable death outcome, and lacking therapeutic alternatives, bad medical news could lead to patients choosing options of high associated risk. Exploring these treatment effect differences across decision contexts would be an important avenue to improve our understanding about the bad medical news effect.

We conclude by highlighting two limitations of our analysis. First, as noted before, available data do not contain variables designed to measure individuals' overall health status. As a result, we have relied on a health status proxy constructed based on total healthcare expenditures (see Appendix A.1). Since this proxy could include a reverse causality bias, readers should be cautious when interpreting our results. Second, our dataset lacks information about the individuals' income level or health insurance plan. Although the absence of this information does not introduce a bias into our estimates (because of the randomized treatment), it prevents us from examining whether the response to the bad news shock could vary with income or insurance coverage.

REFERENCES

Alpert, Abby, Mark Duggan, and Judith K. Hellerstein (2013), "Perverse Reverse Price Competition: Average Wholesale Prices and Medicaid Pharmaceutical Spending," Journal of Public Economics, 108, 44–62.

American Diabetes Association (2010), "Glycemic Targets: Standards of Medical Care in Diabetes—2010," Diabetes Care, 33 (1, Suppl.), S11–S61.

American Diabetes Association (2021), "Glycemic Targets: Standards of Medical Care in Diabetes—2021," Diabetes Care, 44 (1, Suppl.), S73–S84.

Anderson, Simon P., Andre De Palma, and Jacques-Francois Thisse (1992), Discrete Choice Theory of Product Di↵erentiation. MIT Press.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin (1996), "Identification of causal e↵ects using instrumental variables," Journal of the American statistical Association, 91 (434), 444–455.

Austin, Peter C. (2009), "Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples," Statistics in Medicine, 28 (25), 3083–3107.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein (2015), "Behavioral Hazard in Health Insurance," The Quarterly Journal of Economics, 130 (4), 1623–1667.

Baile, Walter F., Robert Buckman, Renato Lenzi, Gary Glober, Estela A. Beale, and Andrzej P. Kudelka (2000), "SPIKES—A Six-Step Protocol for Delivering Bad News: Application to the Patient with Cancer," The Oncologist, 5 (4), 302–311.

Bassi, Anna, Riccardo Colacito, and Paolo Fulghieri (2013), "'O Sole Mio: An Experimental Analysis of Weather and Risk Attitudes in Financial Decisions," The Review of Financial Studies, 26 (7), 1824–1852.

Berndt, Ernst R (2002), "Pharmaceuticals in US health care: determinants of quantity and price," Journal of Economic perspectives, 16 (4), 45–66.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How Much Should We Trust Di↵erences-In-Di↵erences Estimates?," The Quarterly Journal of Economics, 119 (1), 249–275.

Bloomgarden, Zachary T. (2009), "A1C: Recommendations, Debates, and Questions," Diabetes Care, 32 (12), e141–e147.

Bronnenberg, Bart J., Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro (2015), "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium," The Quarterly Journal of Economics, 130 (4), 1669–1726.

Buckman, Robert (1992), How to Break Bad News: A Guide for Health Care Professionals. University of Toronto Press.

Callen, Michael, Mohammad Isaqzadeh, James D. Long, and Charles Sprenger (2014), "Violence and Risk Preference: Experimental Evidence from Afghanistan," American Economic Review, 104 (1), 123–48.

Cameron, Lisa and Manisha Shah (2015), "Risk-Taking Behavior in the Wake of Natural Disasters," Journal of Human Resources, 50 (2), 484–515.

Card, David and Gordon B. Dahl (2011), "Family Violence and Football: The E↵ect of Unexpected Emotional Cues on Violent Behavior," The Quarterly Journal of Economics, 126 (1), 103–143.

Carrera, Mariana and Sofia B. Villas-Boas (2020), "Generic Aversion and Observational Learning in the Over-The-Counter Drug Market," CUDARE Working Paper, eScholarship Publishing (December 16), https://escholarship.org/uc/item/7ks7s9jf.

Ching, Andrew T. (2010a), "Consumer Learning and Heterogeneity: Dynamics of Demand for Prescription Drugs after Patent Expiration," International Journal of Industrial Organization, 28 (6), 619–638.

Ching, Andrew T. (2010b), "A Dynamic Oligopoly Structural Model for the Prescription Drug Market after Patent Expiration," International Economic Review, 51 (4), 1175–1207.

Ching, Andrew T., Robert Clark, Ignatius Horstmann, and Hyunwoo Lim (2016), "The E↵ects of Publicity on Demand: The Case of Anti-cholesterol Drugs," Marketing Science, 35 (1), 158–181.

Ching, Andrew T., David Granlund, and David Sundström (2022), "Quantifying the Zero-Price E↵ect in the Field: Evidence from Swedish Prescription Drug Choices," Journal of the Association for Consumer Research, 7 (2), 175–185.

Ching, Andrew T., Manuel Hermosilla, and Qiang Liu (2019), "Structural Models of the Prescription Drug Market," Foundations and Trends® in Marketing, 13 (1), 1–76.

Ching, Andrew T and Hyunwoo Lim (2020), "A structural model of correlated learning and late-mover advantages: The case of statins," Management Science, 66 (3), 1095–1123.

Chua, Gin Nie, Mohamed Azmi Hassali, Asrul Akmal Shafie, and Ahmed Awaisu (2010), "A Survey Exploring Knowledge and Perceptions of General Practitioners Towards the Use of Generic Medicines in the Northern State of Malaysia," Health Policy, 95 (2–3), 229–235.

Cohn, Alain, Jan Engelmann, Ernst Fehr, and Michel André Maréchal (2015), "Evidence for Countercyclical Risk Aversion: An Experiment with Financial Professionals," American Economic Review, 105 (2), 860–85.

Crawford, Gregory S. and Matthew Shum (2005), "Uncertainty and Learning in Pharmaceutical Demand," Econometrica, 73 (4), 1137–1173.

Depetris-Chauvin, Emilio, Ruben Durante, and Filipe Campante (2020), "Building Nations through Shared Experiences: Evidence from African Football," American Economic Review, 110 (5), 1572–1602.

Dunne, Suzanne, Bill Shannon, Colum Dunne, and Walter Cullen (2014), "Patient Perceptions of Generic Medicines: A Mixed-Methods Study," The Patient: Patient-Centered Outcomes Research, 7 (2), 177–185.

Dunne, Suzanne S. and Colum P. Dunne (2015), "What Do People Really Think of Generic Medicines? A Systematic Review and Critical Appraisal of Literature on Stakeholder Perceptions of Generic Drugs," BMC Medicine, 13 (173), https://doi.org/10.1186/s12916-015-0415-3.

Ekman, Paul (1999), "Basic Emotions," in Handbook of Cognition and Emotion, Tim Dalgleish and Mick J. Power, eds. Wiley Online Library, 45–60.

Erdem, Tulin, Michael P. Keane, and Baohong Sun (2008), "A Dynamic Model of Brand Choice When Price and Advertising Signal Product Quality," Marketing Science, 27 (6), 1111–1125.

Faulkner, Ann (1998), When the News Is Bad: A Guide for Health Professionals on Breaking Bad News. Nelson Thornes.

Ganther, Julie M. and David H. Kreling (2000), "Consumer Perceptions of Risk and Required Cost Savings for Generic Prescription Drugs," Journal of the American Pharmaceutical Association (1996), 40 (3), 378–383.

Gencarelli, Dawn M. (2002), "Average Wholesale Price for Prescription Drugs: Is There a More Appropriate Pricing Mechanism?," Issue Brief No. 775, National Health Policy Forum (June 7).

Haas, Jennifer S., Kathryn A. Phillips, Eric P. Gerstenberger, and Andrew C. Seger (2005), "Potential Savings from Substituting Generic Drugs for Brand-Name Drugs: Medical Expenditure Panel Survey, 1997–2000," Annals of Internal Medicine, 142 (11), 891–897.

Handel, Benjamin R (2013), "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," American Economic Review, 103 (7), 2643–82.

Handel, Benjamin R. and Jonathan T. Kolstad (2015), "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare," American Economic Review, 105 (8), 2449–2500.

Harmon, Amy (2010), "New Drugs Stir Debate on Rules of Clinical Trials," The New York Times (September 18), https://www.nytimes.com/2010/09/19/health/research/19trial.html.

Hartman, Micah, Anne B Martin, Benjamin Washington, Aaron Catlin, National Health Expenditure Accounts Team et al. (2022), "National Health Care Spending in 2020: Growth Driven by Federal Spending in Response to the COVID-19 Pandemic: National Health Expenditures Study Examines US Health Care Spending in 2020," Health A↵airs, 41 (1), 13–25.

Hassali, Mohamed A. A., Asrul A. Shafie, Shazia Jamshed, Mohamed I. M. Ibrahim, and Ahmed Awaisu (2009), "Consumers' Views on Generic Medicines: A Review of the Literature," International Journal of Pharmacy Practice, 17 (2), 79–88.

Hassali, Mohamed Azmi, Asrul Akmal Shafie, Ahmed Awaisu, Mohamed Izham Ibrahim, Chong Chee Ping, and Shazia Jamshed (2010), "Physicians' Views on Generic Medicines: A Narrative Review," Journal of Generic Medicines, 7 (1), 30–39.

Hellerstein, Judith K. (1998), "The Importance of the Physician in the Generic versus Trade-Name Prescription Decision," The RAND Journal of Economics, 29 (1), 108–136.

Hirshleifer, David and Tyler Shumway (2003), "Good Day Sunshine: Stock Returns and the Weather," The Journal of Finance, 58 (3), 1009–1032.

Iizuka, Toshiaki (2012), "Physician Agency and Adoption of Generic Pharmaceuticals," American Economic Review, 102 (6), 2826–58.

IMS Institute for Healthcare Informatics (2015), "Medicines Use and Spending Shifts: A Review of the Use of Medicines in the US in 2014," https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/medicines-use-and-spending-shifts-in-the-us-in-2014.pdf.

Johansen, Michael E. and Caroline Richardson (2016), "Estimation of Potential Savings through Therapeutic Substitution," JAMA Internal Medicine, 176 (6), 769–775.

Johnson, Eric J. and Amos Tversky (1983), "A↵ect, Generalization, and the Perception of Risk," Journal of Personality and Social Psychology, 45 (1), 20–31.

Kesselheim, Aaron S., Alexander S. Misono, Joy L. Lee, Margaret R. Stedman, M. Alan Brookhart, Niteesh K. Choudhry, and William H. Shrank (2008), "Clinical Equivalence of Generic and Brand-Name Drugs Used in Cardiovascular Disease: A Systematic Review and Meta-Analysis," JAMA, 300 (21), 2514–2526.

Kuhnen, Camelia M. and Brian Knutson (2011), "The Influence of A↵ect on Beliefs, Preferences, and Financial Decisions," Journal of Financial and Quantitative Analysis, 46 (3), 605–626.

Lambert, Zarrel V., Paul L. Doering, Eric Goldstein, and William C. McCormick (1980), "Predispositions toward Generic Drug Acceptance," Journal of Consumer Research, 7 (1), 14–23.

Lerner, Jennifer S. and Dacher Keltner (2000), "Beyond Valence: Toward a Model of Emotion-Specific Influences on Judgement and Choice," Cognition and Emotion, 14 (4), 473–493.

Lerner, Jennifer S. and Dacher Keltner (2001), "Fear, Anger, and Risk," Journal of Personality and Social Psychology, 81 (1), 146–159.

Marcovina, Santica M., Vinod P. Gaur, and John J. Albers (1994), "Biological Variability of Cholesterol, Triglyceride, Low-Density And High-Density Lipoprotein Cholesterol, Lipoprotein (A), and Apolipoproteins A-I and B," Clinical Chemistry, 40 (4), 574–578.

Milgrom, Paul and John Roberts (1986), "Price and Advertising Signals of Product Quality," Journal of Political Economy, 94 (4), 796–821.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon (2012), "A Reduced-Form Approach to Behavioral Public Finance," Annual Review of Economics, 4 (1), 511–540.

Narayanan, S. (1996), "Pre and Post Analytical Errors in Lipid Determination," Indian Journal of Clinical Biochemistry, 11 (1), 12–16.

Narayanan, Sridhar and Puneet Manchanda (2009), "Heterogeneous Learning and the Targeting of Marketing Communication for New Products," Marketing Science, 28 (3), 424–441.

National Institutes of Health (2001), "ATP III Guidelines At-A-Glance Quick Desk Reference," NIH Publication No. 01-3305.

Nunn, Ryan, Jana Parsons, and Jay Shambaugh (2020), "A Dozen Facts about the Economics of the US Health-Care System," research report, The Hamilton Project (March 10).

O'Malley, A. James, Richard G. Frank, Atheer Kaddis, Barbara M. Rothenberg, and Barbara J. McNeil (2006), "Impact of Alternative Interventions on Changes in Generic Dispensing Rates," Health Services Research, 41 (5), 1876–1894.

Pechlivanoglou, Petros, Willem Jan van der Veen, Jens H. Bos, and Maarten J. Postma (2011), "Analyzing Generic and Branded Substitution Patterns in the Netherlands Using Prescription Data," BMC Health Services Research, 11 (89), https://doi.org/10.1186/1472-6963-11-89.

Phillipov, George and Patrick J. Phillips (2001), "Components of Total Measurement Error for Hemoglobin A1C Determination," Clinical Chemistry, 47 (10), 1851–1853.

Pop-Eleches, Cristian, Harsha Thirumurthy, James P. Habyarimana, Joshua G. Zivin, Markus P. Goldstein, Damien De Walque, Leslie Mackeen, Jessica Haberer, Sylvester Kimaiyo, John Sidle, Duncanm Ngare, and David R. Bangsberg (2011), "Mobile Phone Technologies Improve Adherence to Antiretroviral Treatment in a Resource-Limited Setting: A Randomized Controlled Trial of Text Message Reminders," AIDS, 25 (6), 825–834.

Raghunathan, Rajagopal and Michel Tuan Pham (1999), "All Negative Moods Are Not Equal: Motivational Influences of Anxiety and Sadness on Decision Making," Organizational Behavior and Human Decision Processes, 79 (1), 56–77.

Roman, Beatriz (2009), "Patients' Attitudes Towards Generic Substitution of Oral Atypical Antipsychotics: A Questionnaire-Based Survey in a Hypothetical Pharmacy Setting," CNS Drugs, 23 (8), 693–701.

Saunders, Edward M. (1993), "Stock Prices and Wall Street Weather," The American Economic Review, 83 (5), 1337–1345.

Shrank, William H., Tuyen Hoang, Susan L. Ettner, Peter A. Glassman, Kavita Nair, Dee DeLapp, June Dirstine, Jerry Avorn, and Steven M. Asch (2006), "The Implications of Choice: Prescribing Generic or Preferred Pharmaceuticals Improves Medication Adherence for Chronic Conditions," Archives of Internal Medicine, 166 (3), 332–337.

Tootelian, Dennis H., Ralph M. Gaedeke, and John Schlacter (1988), "Branded versus Generic Prescription Drugs: Perceptions of Risk, Efficacy, Safety, and Value," Journal of Health Care Marketing, 8 (3), 26–29.

Verduyn, Philippe, Ellen Delvaux, Hermina Van Coillie, Francis Tuerlinckx, and Iven Van Mechelen (2009), "Predicting the Duration of Emotional Experience: Two Experience Sampling Studies," Emotion, 9 (1), 83–91.

Verduyn, Philippe, Iven Van Mechelen, and Francis Tuerlinckx (2011), "The Relation between Event Processing and the Duration of Emotional Experience," Emotion, 11 (1), 20–28.

Verger, P., J. C. Gourheux, P. Villani, B. Desquins, G. Bouvenot, and A. Paraponaris (2003), "Willingness to Prescribe by International Non Proprietary Names: General Practitioners Attitudes in South East France in 2002," La Presse Médicale, 32 (22), 1022–1025.

Wathieu, Luc and Marco Bertini (2007), "Price as a Stimulus to Think: The Case for Willful Overpricing," Marketing Science, 26 (1), 118–129.

Williams, Angela E. and Henry Chrystyn (2007), "Survey of Pharmacists' Attitudes Towards Interchangeable Use of Dry Powder Inhalers," International Journal of Clinical Pharmacy, 29 (3), 221–227.

Woodcock, Janet (2019), "Safety, Efficacy, and Quality Remain Top Priorities as We Continue Our Work to Expand Access to Cost-Saving Generic Drugs for the American Public," US Food and Drug Administration, Center for Drug Evaluation and Research, (accessed April 10, 2022), https://www.fda.gov/news-events/fda-voices/safety-efficacy-and-quality-remain-top-priorities-we-continue-our-work-expand-access-cost-saving.

NOTES

[1]This figure follows the findings of Haas et al. (2005) and Johansen and Richardson (2016), who estimate that prescription drug expenditures in the U.S. could fall by 10% in the (partial equilibrium) scenario of full generic substitution.

[2]For the formal NIH recommendations, see https://www.nhlbi.nih.gov/files/docs/public/heart/chol_tlc.pdf. For the Cleveland clinic, Mayo clinic, and WebMD recommendations, respectively see https://my.clevelandclinic.org/health/articles/16866-cholesterol-guidelines--heart-health, https://www.mayoclinic.org/diseases-conditions/heart-disease/in-depth/heart-disease/art-20049357, and https://www.webmd.com/heart-disease/ldl-cholesterol-the-bad-cholesterol.

[3]We cluster at this level because individual preferences for generics may vary by molecule, e.g., because the patient's insurance is more generous for some drug classes over others. Also notice that the fixed e↵ects A remove individual-level variation.

[4]Bloomgarden (2009), who summarizes the debate, suggests that the root cause of the problem is the relationship between long-term glycemia and blood sugar, which can systematically di↵er across individuals.

[5]Of these 3,725 patients in the A1c frontier, only two are also in the LDL frontier examined in our main analyses. Consistent with local randomization, treated and control groups in the A1c sample are well balanced (see Table A.10).

[6]Based on the evidence of Appendix A.1 (population-level health condition is between that in the LDL and A1c samples), we expect the population-level response to bad news to also fall between the responses documented for the LDL and A1c samples.

[7]Notice the di↵erence between this and our previous analysis on multiple testers (i.e., more than one LDL test). Rather than experience derived from taking multiple medical tests, the analysis herein considers experience derived from continued drug consumption.

[8]For individuals who test during 2011, 2012 purchase data do not include a pre-testing period. Accordingly, we exclude these individuals from the estimation sample.

[9]Measuring consumption experience at the molecule level contrasts with the prior work of Ching (2010a,b) and Crawford and Shum (2005), where learning occurs separately for each product. In Appendix E we present arguments that reconcile the two approaches.

[10]Literature findings are mixed in this regard. For example, while Johnson and Tversky (1983) and Hirshleifer and Shumway (2003) favor the probability assessment channel, Bassi, Colacito, and Fulghieri (2013) and Kuhnen and Knutson (2011) favor the risk aversion channel.

Web Appendix

"Does Bad Medical News Reduce Preferences for Generic Drugs?"

Manuel Hermosilla (mh@jhu.edu)
Andrew T. Ching (andrew.ching@jhu.edu)

Contents

A Additional Sample Descriptives And Experimental Balancing Checks 2

A.1 The Health Condition of Sample Patients Through the Lens of Total Healthcare Expenditures 2

A.2 Amount, nature, and results of non-LDL testing 6

A.3 Timing of testing 6

A.4 The HighSpender indicator 8

B Specification checks and supplemental results 9

B.1 Falsification 9

B.2 Analysis of pre-trends 9

C Generic Propensity in the Adoption of New Molecules 11

D The Bad News E↵ect on Multiple Testers 14

E Consumption Experience and Generic Propensity 16

F Complementary results for the A1C analysis 18

F.1 Pre-trends and falsification 18

F.2 Treatment E↵ect Heterogeneity 18

G Perceived Qualities and Pricing: Analytical Details and Additional Results 19

G.1 Additional Details 19

G.2 Price Di↵erentials Based on Out-of-Pocket Prices 19

H Supplemental Figures and Tables 23

A Additional Sample Descriptives And Experimental Balancing Checks

A.1 The Health Condition of Sample Patients Through the Lens of Total Healthcare Expenditures

As noted in the main text, MarketScan data do not contain variables designed to describe the health condition of individuals. We will therefore investigate how the health status of sample patients compares to that of the broader population using healthcare expenditures as a yardstick.

Expenditure di↵erences in the raw data. Column 1 of Table A.1 describes the average dollar expenditure of individuals in our main sample, i.e., the sample based on frontier LDL results used for our main generic propensity analyses. Expenditures are measured as the value of services, i.e., the total dollar amount (US$) paid to the service provider (out-of-pocket expenses plus amount covered by insurance). Expenditures cover the full 2-year sample (2011 and 2012) and are separated by type, i.e., expenditures on drugs, outpatient services, and inpatient services (expenditures on medical tests are omitted because they are not reported in the data available to us). On average, individuals in our main sample spent $1,844 on drugs, $5,363 on out-patient services (e.g., medical visits), and $1,942 on in-patient services (i.e., surgeries). Total average expenditure amounted to $9,149 over the full 2-year period.

Column 2 of Table A.1 reproduces these statistics for individuals in the broad MarketScan database, as reflected by a random sample of 5,000 individuals. Compared with individuals in our main sample, individuals in this random MarketScan sample spent about $400 more on drugs, $2,000 more on out-patient services, and $200 more on in-patient services. In turn, Column 3 focuses on individuals of the A1c sample (6.9 vs 7%). Compared with individuals in our main (LDL) sample (Column 1), individuals in the A1c sample spent about $300 less on drugs, but significantly more on out-patient and in-patient services (about $3,000 and $3,500, respectively).

Demographics. Table A.2 describes the demographic composition of each of the samples. Compared with individuals in the LDL sample (57% female, born in 1962 on average), those in the random MarketScan sample (Column 2) are about equally likely to be female but are significantly younger (by 9 years on average). Individuals in the A1c sample are generally older (compared with the LDL sample, by 14 years) and less likely to be female (48% are female). These statistics suggest that the demographics are significantly di↵erent across samples. The econometric results that follow incorporate these di↵erences as potential determinants of expenditures.

Econometric results. In Table A.3 we investigate the utilization di↵erences across samples using linear regression. As a dependent variable, we consider each individual's log total expenditures (sum of expenditures on drugs, out-patient, and in-patient services). As independent variables, we include indicators for the LDL and A1c samples, i.e., MarketScan individuals are used as the baseline. Column 1 presents the most basic set of results. Given the above definitions, the -0.1180 estimate for the LDL indicator suggests that individuals in our main sample spent about 12% less than individuals in the random MarketScan sample. Analogously, the 0.2702 estimate for the A1c indicator suggest that individuals in the A1c sample spent about 27% more than those in the random MarketScan sample.

Column 2 of Table A.3 reproduces the analysis but also controls for available demographic characteristics, i.e., sex and year of birth. After controlling for these demographic characteristics,

the expenditure di↵erences between patients in the LDL and random MarketScan sample increase to about 37%. By contrast, the di↵erences between patients in the A1c and random MarketScan sample vanish. Overall, these results suggest that (i) the LDL sample contains individuals that, after accounting for sex and age, are healthier compared with those in the random MarketScan sample, and (ii) the A1c sample contains individuals that, after accounting for sex and age, have about the same expenditure levels as those in the random MarketScan sample.

Interpretation: does less healthcare expenditure reflect better health? A leading interpretation for the first of these results would be that the lower expenditures of individuals in the LDL sample reflects their less need for healthcare (compared with individuals in the random MarketScan sample), and thereby their better health condition. We will adopt this rationale to interpret the above results. That is, we will interpret these results as reflective of the fact that individuals in our main sample (LDL) exhibit on average better health condition than those in the random MarketScan sample.

Nevertheless, it is important to highlight that this interpretation (i.e., less healthcare consumption signals better health) is subject to an important reverse causality caveat given that some healthcare consumption may reflect e↵orts to preserve health. In other words, it is possible that some of the additional expenditure by individuals in the random MarketScan sample (compared with those in the LDL sample) reflects their preventive e↵orts, and thereby their better health condition.

Having noted on this possibility, we think that the reverse causality e↵ect that it alludes to is not of first order importance in our data. Our argument is based on the following observations. If present, we expect preventive healthcare expenditure to primarily occur within the context of out-patient services (e.g., preventive consultations). Patients are often able to book these services on their own, without medical justification. By contrast, in-patient admissions usually occur in the context of surgery or emergency room visits, for which the medical need is often well justified. Based on these observations, we may test for the importance of preventive healthcare utilization by studying the substitutability of these two types of expenditures. In particular, if preventive healthcare utilization drove healthcare expenditures in our data, we would expect a negative correlation between the expenditures on out-patient and in-patient services. That is, we would expect to see that individuals who utilize more (prevention-oriented) out-patient services have lesser need for and therefore lower utilization of in-patient services. Contrary to this implication, we find that the two types of expenditures have a positive and statistically significant correlation ($\dashrightarrow = 0.3316, p < 0.01$). This result is ratified by a patient-level regression of logged out-patient expenditures on logged in-patient expenditures (plus demographic controls), which delivers a 0.3064 ($p < 0.013$) elasticity of out-patient expenditures to in-patient expenditures. We conclude that, although preventive healthcare utilization is likely to be present in our data, it is unlikely to drive overall spending.

How representative of the broad U.S. population is the MarketScan sample? Thus far, our analysis has not addressed how representative the random MarketScan sample is of the broader United States (U.S.) population. We now turn to examining this question.

Addressing this question is possible because MarketScan data include population weights, which indicate how representative of the overall U.S. population each individual in the sample is. Specifically, each individual i in the sample is associated with a number $w_i \geq 0$, which corresponds to the estimated number of U.S. residents that are represented by individual i in the data. Using these values, we first computed a weighted average of expenditures in the random MarketScan

sample, $\sum_i \frac{w_i}{\sum_{i'} w_{i'}} \cdot \text{TotalExpenditures}_i$. This calculation yielded \$10,644, which differs by less than 2% compared with its non-weighted counterpart of \$10,855 (see Table A.1). We next investigated whether there exists a correlation between an individual's (logged) total expenditures and his/her representative weight. We obtained a small correlation between the two variables, $\rho = -0.0101$ ($p = 0.4737$), which suggests that expenditures and national representativeness are largely statistically independent. Combined, these results would suggest that our conclusions would not change by a lot if instead of considering a random sample from the MarketScan population we used a random sample from the broad U.S. population.

Table A.1: Healthcare utilization in each sample (US\$).

	(1)	(2)	(3)
Type of	Sample		
utilization	LDL	MarketScan	A1c
Drug	1,844	2,232	1,569
Outpatient	5,363	6,410	7,462
Inpatient	1,942	2,109	5,402
Total	9,149	10,855	14,838

Notes. Figures correspond to the average total expenditures across individuals included in each sample. Expenditures are measured as the value of services (total amount paid to the provider). Column 2 corresponds to a random sample of 5,000 individuals from the broad MarketScan data.

Table A.2: Demographics in each sample.

	(1)	(2)	(3)
Type of	Sample		
claim	LDL	MarketScan	A1c
Share female	0.57	0.55	0.48
Birth year	1962	1971	1958

Notes. Column 2 corresponds to a random sample of 5,000 individuals from the broad MarketScan data.

Table A.3: Utilization di↵erences across samples: LDL and A1C samples compared to the random MarketScan sample (regression baselines).

	(1)	(2)
LDL	−0.1180⋆⋆⋆	−0.3673⋆⋆⋆
	(0.0427)	(0.0437)
A1c	0.2702⋆⋆	−0.0193
	(0.0347)	(0.0366)
Fixed e↵ects		
Female		X
Year of birth		X
N	16,167	16,167

Notes. Parentheses show standard errors. *Legend:* ⋆p < 0.1,⋆⋆ p < 0.05,⋆⋆⋆ p < 0.01..

A.2 Amount, nature, and results of non-LDL testing

Panel B of Table 1 showed that treated and control individuals balance tightly in terms of healthcare utilization, as measured by di↵erent types of average monthly claims. In this section we complement those results through a fine-grained analysis of the amount, nature, and results of non-LDL testing carried out prior to the studied LDL results.

We begin our analysis focusing on the amount of testing. We construct the variable $\text{NumberNonLDLTests}_{ij}$ as the count of claims for non-LDL test j filed by individual i prior to his/her LDL test. The total count of such tests is 40,815. The index j for test type corresponds to the di↵erent entries in the "Logical Observation Identifiers Names and Codes" system (LOINC, see `https://loinc.org/`), which has about 700 di↵erent entries in our data. To examine whether there is a systematic treatment/control di↵erence in the amount of testing, we estimate:

$$\text{NumberNonLDLTests}_{ij} = \text{Poisson}(\cdot + (3 \cdot \text{Treated}_i + \cdot_j),$$

where $\cdot_j$ is a fixed e↵ect for each type of test. This fixed e↵ect helps to account for the fact that some tests (such as LDL) are more common than others. The key coefficient is (3. We obtain an estimate of $\hat{\beta} = 0.034$, although with a large standard error (0.371) which renders the estimate non-significant at all standard levels of statistical confidence (p=0.926). Thus, there does not appear to exist a significant di↵erence in the amount and nature of testing carried out by treated and control individuals.

Next, we evaluate whether there exists a systematic di↵erence in the results obtained by each group of individuals. Here it is important to note that the results of di↵erent medical tests are expressed in di↵erent units. For example, whereas LDL results are expressed in mg/dL, the results for "Sodium [Moles/volume] in Serum or Plasma" tests (LOINC=2951-2) are expressed in Millimols per liter (MMOL/L). To account for this feature of the data, our specification also includes fixed e↵ects for each test type. The specification is:

$$\text{LogNonLDLTestResult}_{ijt} = \cdot + (3 \cdot \text{Treated}_i + \cdot_j + \bullet_{ijt},$$

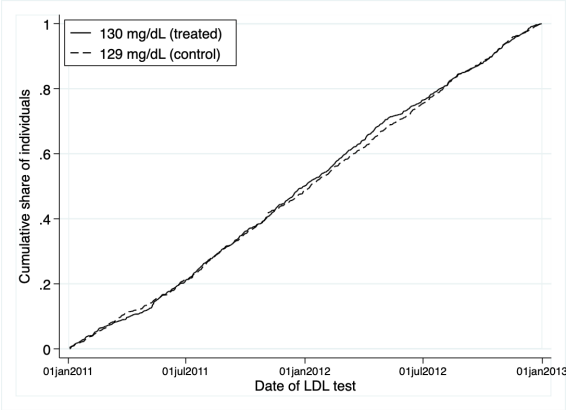
where LogNonLDLTestResult is the log of one plus the quantitative result obtained by individual i from test type j taken at date t. As before, > represents the test specific fixed e↵ects that control for the varying scales (units). The term • is an error. Our estimate for the key parameter of (3 is $\hat{\beta} = 0.0002$, with SE=0.008 and p=0.976. That is, this result supports the conclusion that treated and control patients obtained the same medical test results prior to the LDL test. We interpret this finding as strong evidence that, prior to the LDL result, there were no systematic di↵erences in terms of the health condition of treated and control groups.

A.3 Timing of testing

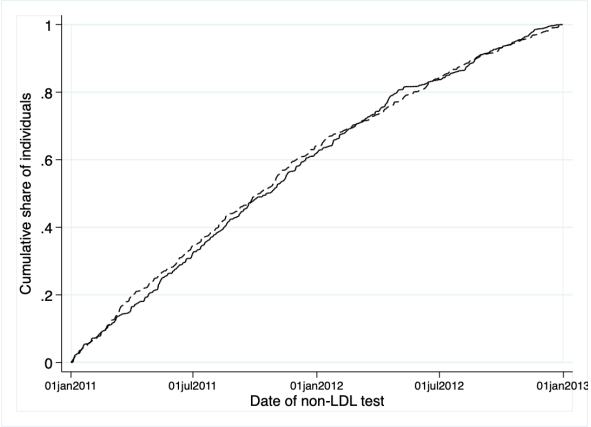
Figure A.1A describes the timing of LDL tests. Curves plot the cumulative density of individuals of each group who had taken the LDL test by each date. For example, both curves indicate that by July 2011, about 20% of individuals in each group had been tested. One year later, by July 2012, close to 80% had been tested. The overall similarity of the two curves indicates that there is no systematic di↵erence in terms of when individuals of each group took the test. Figure A.1B presents the same analysis for non-LDL tests taken by each individual before their respective LDL test. Again in this case, the curves for treated and control group individuals trace each

Figure A.1: Timing of medical tests (when they are taken).

(A) LDL tests.



(B) Other tests.



Notes. Curves show the cumulative distributions of individuals of each group (treated, control) who have taken the LDL test (Panel A) and non-LDL test (Panel B) by each date. That is, for example, both curves of Panel A indicate that by early July of 2011, about 20% of individuals in each group had taken the LDL test. In both cases (LDL and non-LDL), Kolmogorov-Smirnov tests fail to reject the null hypothesis of distribution equality at all conventional statistical significance levels.

other tightly. In both cases (Figure A.1A and Figure A.1B), Kolmogorov-Smirnov tests fail to reject the null hypothesis of distribution equality at all conventional statistical significance levels.

A.4 The HighSpender indicator

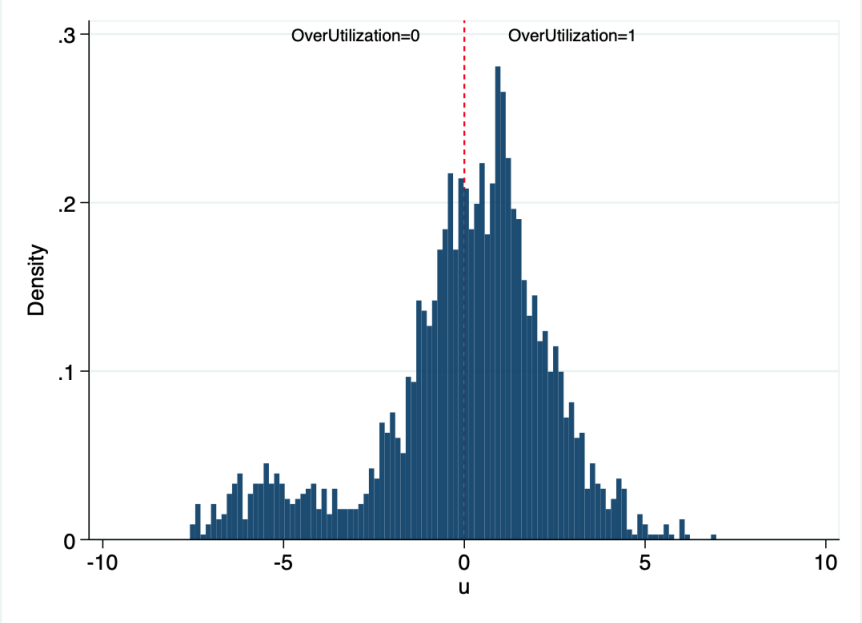
To construct the HighSpender indicator used in the analysis of Column 5 of Table 2, we first estimate the following equation using the full sample of individuals:

$$\log(1 + \text{TotalExpenditure}_i) = \checkmark_0 + \checkmark_1 \cdot \text{Female}_i + \checkmark_2 \cdot \log(\text{Days}_i) + \text{A}_{yob(i)} + \text{u}_i \qquad \text{(A.1)}$$

where TotalExpenditure is the sum of individual i's expenditures on drugs, in-patient, and out-patient services made prior to taking the LDL test. Female is an indicator for female individuals, Days is individual i's number of in-sample days prior to the LDL test (e.g., given that the earliest day in the sample is 1/1/2011, Days = 10 for someone who took the LDL test on 1/10/2011), and A are year-of-birth (yob) fixed e↵ects.

Based on the residuals obtained from this regression, we construct $\text{HighSpender}_i = 1[\text{u}_i \; 2: \; 0]$, as shown in Figure A.2.

Figure A.2: The HighSpender indicator.



Notes. Plots the errors u obtained from estimating Equation A.1.

B Specification checks and supplemental results

B.1 Falsification

We probe the causal interpretation of our main 129 or 130 mg/dL estimate (Column 1, Panel A, Table 2) through two separate placebo tests. First, we reproduce our analysis using falsified clinical thresholds—120, 125, 135, and 140 mg/dL. Results for each of these scenarios are presented in Columns 1-4 of Table A.4A. Estimates are all positive, small (about one order of magnitude smaller than our main estimate), and statistically non-significant at conventional levels. Also notice that, like the actual 130 mg/dL threshold, the first and last of these falsified thresholds (120 and 140) are multiples of 10. Next, we falsify testing dates. After dropping all post-testing data, we estimate Equation 2 while assuming that each individual's LDL testing date occurred T months before the actual date. Columns 1-4 of Table A.4B respectively show the obtained results for T =3, 6, 9, and 12 months. Again, in addition to the lack of statistical significance, these estimates exhibit a much smaller magnitude than the estimate obtained from the un-falsified setting.

Even though the similarity of treated and control individuals leaves little room for confounding pre-trends, we also conducted a formal assessment of the issue. Concerns that the estimated impact of bad LDL news on generic choice propensity is driven by the influence pre-trends are not supported.

Table A.4: Falsification tests.

	(1)	(2)	(3)	(4)
	(A) Falsified clinical thresholds			
	119 vs 120 mg/dL	124 vs 125 mg/dL	134 vs 135 mg/dL	139 vs 140 mg/dL
Treated⇥Post	0.004	0.006	0.001	0.001
	(0.005)	(0.006)	(0.007)	(0.007)
N	44,200	35,847	30,598	25,361
	(B) Falsified test date			
	(only uses pre-test data; below are falsified test dates in months prior to actual test)			
	3	6	9	12
Treated⇥Post	0.001	0.001	0.003	-0.001
	(0.007)	(0.008)	(0.009)	(0.010)
N	16,467	16,467	16,467	16,467

Notes. Linear probability specifications for the probability of choosing the generic option (Equation 2). All models include fixed e↵ects for molecules, individuals, and time periods. Parentheses show standard errors clustered at the level of patient/molecule pairs. *Legend:* $^{\cdot}p < 0.1$, $^{\cdot\cdot}p < 0.05$, $^{\cdot\cdot\cdot}p < 0.01$..

B.2 Analysis of pre-trends

We specify a model that addresses the question of whether a generic-choice-reducing trend existed among treated individuals prior to the LDL test. Using pre-testing data only, we estimate:

$$\text{GENERIC}_{ijtm} = \leftarrow_0 \cdot \tilde{t} + \leftarrow_1 \cdot \text{Treated}_i \rightarrow \tilde{t} + >_i + b_t + \mu_m + \bullet_{ijtm},$$

This is the same specification as in Equation 2, except that $\beta \cdot \text{Treated}_i \times \text{Post}_t$ is replaced by $\gamma_0 \cdot \tilde{t} + \gamma_1 \cdot \text{Treated} \times \tilde{t}$. The equation's main variable is $\tilde{t}$, which corresponds to a measure of chronological time. The confounding pre-trend concern would be supported by a $\hat{\gamma}_1 < 0$ estimate. In this case, treated individuals would have already been experiencing (i.e., prior to LDL testing) a negative trend of generic choice propensity, above and beyond that experienced by control individuals. The estimate of Column 1 in Table 2 could be partly attributable to this trend.

We estimate the equation for two measures of chronological time, calendar months and time periods t (as used in our main analysis). When using calendar months, the suspected trend operates relative to the beginning of the sample (January 2011); when using time periods, the trend operates with respect to the time of testing. (Recall that time periods were defined based on the difference between the day of LDL testing and that of the drug claim. Since the equation is estimated on pre-testing data only, all time periods entering the estimation have negative values (since they are before the test), approaching zero as time passes. Also note that, since the equation includes fixed effects for time periods, when we use this variable to operationalize $\tilde{t}$ we cannot estimate γ_0.) In the first case, we estimate $\hat{\gamma}_1$ = -0.00005 (SE=0.0009); in the second, $\hat{\gamma}_1$ = -0.00003 (SE=0.0009). Because the estimates are statistically insignificant in both cases, we interpret these results as inconsistent with the presence of confounding pre-trends.

C Generic Propensity in the Adoption of New Molecules

In this section, we explore a possible mechanism behind the bad news effect on generic propensity, namely, an increased preference for brand-name when patients adopt new molecules (i.e., when patients first start to consume a new molecule). Our analysis has three components:

- First, we define a sample where we can identify new molecule adoption with reasonable confidence.
- Second, we use Equation A.2 to test whether there exist any treatment/control differences in the propensity to adopt new molecules. Since we fail to detect these differences, we can treat the adoption of new molecules as exogenous events. (This result is consistent with the lack of bad news effects on drug bundles, which we presented in the main text.)
- Third, we use Equation A.4 to investigate treatment/control differences in brand/generic choice for patients who adopt a new molecule. A marginally significant effect (90% confidence) provides support for our main hypothesis. Consistent our front-loading results, this effect concentrates in the first 90 days following the LDL test.

Estimation sample and variable definition. We define the variable $Adoption_{ijtm}$ as an indicator for whether transaction j (occurring during a 30-day period t) represents the first time that patient i purchases molecule m. Particularly for transactions made early in the sample, a natural concern is that some patients may have adopted the molecule before the start of the sample. In this case, $Adoption_{ijtm}$ would over-represent adoption. To deal with this problem, our approach is to: (i) use 2012 data (second half of our sample) to estimate regressions, and (ii) use 2011 data (first half of our sample) to verify prior purchases. Hence, if $Adoption_{ijtm}=1$, we can state that transaction j is the first time patient i purchases molecule m within at least one year. Also note that the 2012 transactions used for estimations do not include a "pre" period (i.e., period before the LDL test is taken) for individuals who took the medical test during 2011. Accordingly, we exclude these individuals from the sample (i.e., those who tested during 2011). In the resulting sample, $Adoption_{ijtm}=1$ for about a quarter of all observations.

Exogenous adoption. The null effects on consumption bundle (see Section "Impacts on Drug Consumption Behavior") suggest that we can view adoption rates as exogenous. For reassurance, we investigate whether bad LDL news may impact the extent of molecule adoption based on the estimation sample defined above. We estimate Equation A.2, where the dependent variable is Adoption. The specification includes interactions of the treatment indicator with indicators for each of the three time periods used for our analysis of temporal effects. For example, an estimate $\beta_1 > 0$ would indicate that the bad news shock leads to increased adoption rates within the first 90 days after the bad news are received. Like the equations used in the main text (Equation 1 and Equation 2), Equation A.2 includes period- and molecule-specific fixed effects (respectively θ and μ). However, due to the reduced sample size available to estimate the model, we exclude the individual-level fixed effects. Estimation results are shown in Column 1 of Table A.5. The estimates of interest are those for coefficients β_1, β_2, and β_3. These estimates are all statistically non-significant. Thus, consistent with the absence of bundle effects documented in the main text, our analysis in this section suggests that the bad news shock does not induce changes in

molecule adoption rates.

$$\text{Adoption}_{ijtm} = \alpha_0 + \alpha_1 \cdot \text{Treated}_i + \delta_t + \mu_m + \beta_1 \cdot 1[t = 0, 1, 2] \times \text{Treated}_i + \beta_2 \cdot 1[t = 3, .., 6] \times \text{Treated}_i + \beta_3 \cdot 1[t = 7, ..] \times \text{Treated}_i + \varepsilon_{ijtm}. \quad \text{(A.2)}$$

Differential generic propensity effects when molecules are being adopted. We formulate a model to investigate whether the effects of bad news on generic propensity are particularly pronounced when a molecule is being adopted (first time a patient purchases the molecule, i.e., $\text{Adoption}_{ijtm}=1$) compared to when it has been previously adopted ($\text{Adoption}_{ijtm}=0$). In the following regression (which adapts our main generic propensity specification, i.e., Equation 2), this effect would be picked up by the coefficient β:

$$\text{GENERIC}_{ijtm} = \alpha \cdot \text{Adoption}_{ijtm} + \gamma \cdot \text{Treated}_i \times \text{Post}_t + \phi \cdot \text{Adoption}_{ijtm} \times \text{Post}_t + \beta \cdot \text{Treated}_i \times \text{Post}_t \times \text{Adoption}_{ijtm} + \lambda_i + \delta_t + \mu_m + \varepsilon_{ijtm}. \quad \text{(A.3)}$$

In this specification, whereas γ picks up the general bad news effect on generic propensity that is common for newly- and previously-adopted molecules (as in Equation 2), β captures additional effects for patients who are just adopting the molecule. Also, α picks up any potential differences in generic propensity for newly- and previously-adopted molecules (but which are unrelated to receiving the bad news), while ϕ captures any additional differences unfolding after the test.

Since we are interested in unpacking the temporal variation of the bad news effect (i.e., front-loading), we estimate a model which allows the effects of interest to vary over time. We employ the following specification:

$$\begin{aligned}\text{GENERIC}_{ijtm} = {} & \alpha_0 + \alpha_1 \cdot \text{Treated}_i + \alpha_2 \cdot \text{Adoption}_{ijtm} + \\ & 1[t = 0, 1, 2] \cdot \gamma_1 \cdot \text{Treated}_i + \phi_1 \cdot \text{Adoption}_{ijtm} + \beta_1 \cdot \text{Treated}_i \times \text{Adoption}_{ijtm} + \\ & 1[t = 3, .., 6] \cdot \gamma_2 \cdot \text{Treated}_i + \phi_2 \cdot \text{Adoption}_{ijtm} + \beta_2 \cdot \text{Treated}_i \times \text{Adoption}_{ijtm} + \\ & 1[t = 7, ..] \cdot \gamma_3 \cdot \text{Treated}_i + \phi_3 \cdot \text{Adoption}_{ijtm} + \beta_3 \cdot \text{Treated}_i \times \text{Adoption}_{ijtm} + \\ & + \mu_m + \delta_t + \varepsilon_{ijtm}.\end{aligned} \quad \text{(A.4)}$$

This specification has the same basic structure of Equation A.3, with two differences. First, due to the reduced sample size available to estimate the model, we replace individual-level fixed effects with a Treatment indicator. The second and most important difference is that Equation A.4 allows for the estimation of time-differentiated effects. To illustrate this, consider β_1, which is the coefficient of main interest. Whereas β in Equation A.3 captures such potential effect over the full post-testing period, β_1 in Equation A.4 captures the effect as it may unfold over the first 90 days after the LDL test (i.e., periods t = 0, 1, 2). An estimate $\hat{\beta}_1 < 0$ would indicate that, during the 90 days following the receipt of the bad news, the decline in generic propensity provoked by the bad news is particularly strong among purchases representing the adoption of a molecule. An estimate $\hat{\beta}_2 < 0$ would point to the same effect but unfolding over purchases made during periods t = 3, .., 6. Similarly, $\hat{\beta}_3 < 0$ would point to the same effect over purchases made during periods t 2 7. Estimation results are presented in Column 2 of Table A.5. However, of the three parameter estimates $\{\hat{\beta}_k\}_{k=1,2,3}$, $\hat{\beta}_1$ has the largest magnitude and is the only one to achieve some statistical significance (significant with 90% confidence). The results suggest that the front-loading of bad news effect is due to its temporary impact on patients who newly adopt a molecule.

Table A.5: Adoption analysis.

Dependent variable	(1) Adoption Equation A.2	(2) Generic Equation A.4
β_1	0.0004	-0.0285*
	(.0208)	(0.0165)
β_2	-0.0148	-0.0039
	(0.0247)	(0.0224)
β_3	0.0262	-0.0015
	(0.0291)	(0.0335)
N	9,553	9,553

Notes. Parentheses show standard errors clustered at the level of patient/molecule pairs. *Legend:* * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$..

D The Bad News E↵ect on Multiple Testers

In this section we consider the impact of bad LDL news on individuals who have more than one LDL test in the data ("known multiple testers"). We are particularly interested in studying whether having received bad LDL news in the past plays a moderating role for our main e↵ect.

There are several important points we need to consider in order to implement this analysis. First, the majority of multiple LDL testers in the data (N=60,527) do not have a frontier (129 or 130 mg/dL) result. Given our research design, we retain data for those who have at least one such frontier result (N=2,341). Among this population, about 53% of individuals have 2 LDL tests, 24% have 3, 12% have 4, and the remainder have 5 or more. The main observation about this distribution regards the idea that the larger the number of tests an individual has, the more likely the individual finds him/herself in a state of high medicalization. This possibility opens the door for a variety of confounds. For example, more medicalized patients may be more accustomed to receiving and critically interpreting test results and thereby less likely to react to bad LDL news as we have defined it. In parallel, more medicalized patients may also be sicker and thereby more sensitive to bad medical news, making them more likely to react to the bad news. Given that the sign of the combined bias is ambiguous, we retain individuals with exactly 2 sample LDL tests (N=1,241) as means to minimize the influence of said confounds. Lastly, because we are interested in the moderating role of having received bad news in the previous test, we keep the subset of these individuals for whom the frontier result was received in the second test (N=556). The resulting sample contains 9,289 drug claims. The share of generic is very similar to that as in our main sample (84% in this sample versus 86% in the main sample).

In a first analysis, we use this sample to replicate our main generic propensity analysis (based on Equation 2). We estimate $\hat{\beta} = -0.003$ (SE= 0.013). That is, although the estimate is negative (bad news strengthen the preference for brand name drugs), the e↵ect is about one third as large as that in the main sample single-testers. In addition, the parameter is statistically non-significant. To investigate the role of having received bad news in the past, we next estimate:

$$\begin{aligned} \text{GENERIC}_{ijtm} = & \; /3_1 \cdot \text{Treated}_i \rightarrow \text{Post}_t \rightarrow (1 - \text{PriorBadNews}_i) + \\ & /3_2 \cdot \text{Treated}_i \rightarrow \text{Post}_t \rightarrow \text{PriorBadNews}_i + \\ & >_i + 6_t + \mu_m + \bullet_{ijtm} \end{aligned} \tag{A.5}$$

This specification has the same basic structure as our main specification of Equation 2. As in the previous analysis, Post is based on the timing of the frontier test (which, for all considered individuals, is the second one). The variable PriorBadNews is defined as PriorBadNews = 1[Result of first LDL test is borderline high (i.e., 2: 130)]. This variable is activated for 43% of sample individuals (accounting for 42% of transactions). This variable is used as the moderator for the main e↵ect. In particular, in the above specification, parameter $/3_1$ will capture the e↵ect of the (second test's) bad LDL news for those individuals who did not receive a borderline high in their first sample test. In turn, parameter $/3_2$ captures the same e↵ect for patients who did receive such borderline high result in their first sample LDL test.

We estimate the following parameters: $\hat{\beta}_1 = -0.013$ (SE=0.017) and $\hat{\beta}_2 = 0.007$ (SE=0.013). That is, the point estimate for patients who did not previously receive bad news is very similar to that as in our main analysis of Table 2. However, the large standard error renders the parameter statistically non-significant at conventional levels. In turn, the parameter estimate for patients who did previously receive bad LDL news, $\hat{\beta}_2$, has the opposite sign and is much smaller in

magnitude (about half) than for patients that did not receive them. The parameter's standard error is disproportionately large. This result is also consistent with the idea that the 130 mg/dL result may represent good news to the patients who had previously received bad news.

We conclude by interpreting this evidence as a mild suggestion that bad LDL news has a marginally decreasing impact on generic choice propensity. Given the strong marked sample selection and lack of statistical significance, we promote a cautious interpretation of this conclusion.

E Consumption Experience and Generic Propensity

Pechlivanoglou et al. (2011) finds evidence that consumption experience gradually increases the probability of generic adoption. In this section, we investigate the extent to which these patterns are observed in our data. Results provide strong evidence in favor of the e↵ect's presence.

The analysis is based on the findings of Pechlivanoglou et al. (2011), who show that consumption experience gradually increases the probability of generic adoption. Consistent with this finding, we estimate robust positive correlations between the number of in-sample purchases and the probability of generic choice.

For our analysis, rather than raw generic propensities (probability of generic purchase), we consider residual generic propensities. These are computed as the residual from a linear regression of the transaction level indicator for generic choice on patient and molecule fixed e↵ects. That is, this residual variable measures abnormal generic propensity choice relative to the patient's and molecule's associated averages. We prefer this residual measure over the raw measure because it focuses on within patient/molecule variation that is unfolding over time (as experience is accumulated).

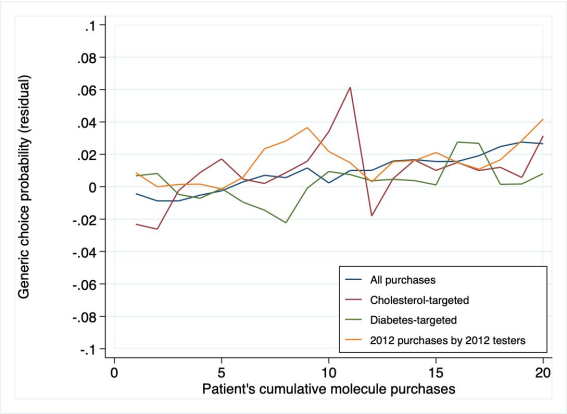
The second element of our analysis is the patient's number of cumulative molecule purchases. If, when making a purchase j for molecule $m(j)$, patient i has previously purchased molecule $m(j)$ a total of C times, the patient's number of cumulative molecule purchases associated with transaction j becomes C.

Results are presented graphically by averaging the residual generic propensity at each number of cumulative molecule purchases. Figure A.3A focuses on the sample of LDL testers. The upward-sloping blue curve (all transactions) associates the accumulation of consumption experience with higher abnormal generic propensity. The e↵ect is subtle: an additional purchase is associated with 0.02 higher abnormal generic propensity. The red and green curves repeat the analysis for cholesterol- and diabetes-targeted drugs, respectively, which are the focus of our analyses in the main text. The green curve repeats the analysis for the sample of 2012 testers, which also play an important role for the analysis in the main text. The positive slope is observed in all of these three cases.

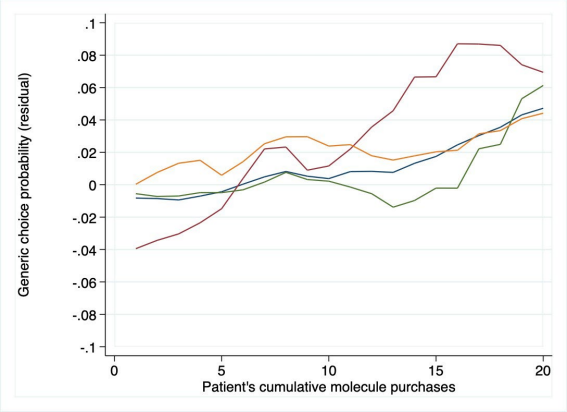
Figure A.3B and Figure A.3C repeat the analyses of Figure A.3A using the A1c sample and pooled samples instead, respectively. Again here we observe that generic propensity positively correlates with the cumulative number of molecule purchases. The e↵ects are more pronounced in the LDL sample, particularly for cholesterol-targeted drugs (red curves).

Figure A.3: Consumption experience and generic propensity.

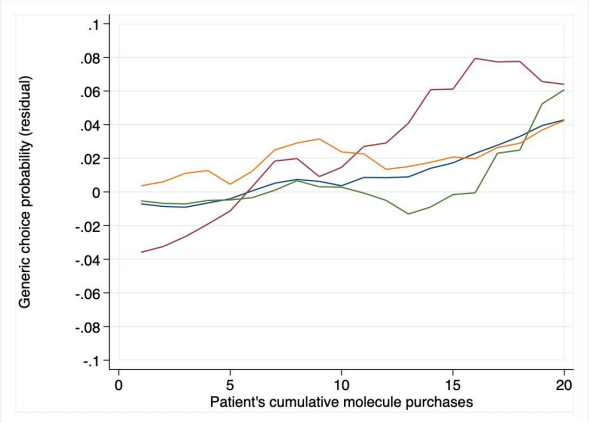
(A) LDL sample.



(B) A1c sample.



(C) Pooled sample (LDL+A1c).



Notes. Curves show the residual propensity of choosing a generic drug, averaged at each consumption instance. Residual generic propensities are computed by linearly regressing the generic indicator on patient and molecule fixed e↵ects.

F Complementary results for the A1C analysis

F.1 Pre-trends and falsification

Pre-trend and falsification analyses are presented in relation to the estimate of bad A1c news on generic propensity, which is represented by the full circle marker in the horizontal axis of Figure 2. We first repeat the pre-trends analysis described in subsection B.2. When calendar months are used to implement the time trends, our key estimate $\hat{\alpha}_1 = 0.00003$ (SE= 0.000452). When the time trend is implemented via differential months, $\hat{\alpha}_1 = -0.000044$ (SE= 0.0003). Both of these estimates are inconsistent with the presence of confounding pre-trends. Table A.6 presents placebo tests. They have the same structure as those presented in Table A.4, Panel B.

Table A.6: A1C Falsification tests (falsified test dates).

	(1)	(2)	(3)	(4)
	Falsified test date (# months prior to actual date)			
	3	6	9	12
Treated×Post	0.0038	0.0015	−0.0004	−0.0017
	(0.0039)	(0.0039)	(0.0049)	(0.0056)
N	64,706	64,706	64,706	64,706

Notes. Linear probability specifications for the probability of choosing the generic option (Equation 2). All models are estimated on pre-testing data only, and include fixed effects for molecules, individuals, and time periods. Parentheses show standard errors clustered at the level of patient/molecule pairs. *Legend:* $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$..

F.2 Treatment Effect Heterogeneity

Here we describe the approach employed to estimate the class-specific differences-in-differences estimates presented by the hollow circle markers of Figure 2. For each drug group k listed in Figure 1, we estimate the following specification:

$$\begin{aligned} \text{GENERIC}_{ijtm} = {} & (\beta_1^k \cdot \text{Treated}_i \times \text{Post}_t \times 1[\text{Group}_j = k] + \\ & (\beta_2^k \cdot \text{Treated}_i \times \text{Post}_t \times (1 - 1[\text{Group}_j = k]) + \\ & \gamma_i + \delta_t + \mu_m + \varepsilon_{ijtm} \end{aligned} \tag{A.6}$$

The model is otherwise the same as in Equation 2. This approach delivers one estimate $\hat{\beta}_1^k$ for each group k, which reflects the class-specific impacts of bad news on brand/generic preferences. The hollow circled markers plotted in Figure 2 correspond to the set of $\{\hat{\beta}_1^k\}$ estimates delivered by this procedure. The equation is estimated separately on data associated with LDL testing on the one hand and A1c testing on the other.

G Perceived Qualities and Pricing: Analytical Details and Additional Results

G.1 Additional Details

In this section we describe the construction of the brand/generic price differentials Δ used for the analysis in the main text. For a molecule m, the differential is defined as:

$$\Delta_m = \frac{p_m^{\text{Brand}} - p_m^{\text{Gen}}}{p_m^{\text{Gen}}}. \tag{A.7}$$

The main step involved in computing the set $\{\Delta_m\}_{m \in \mathcal{M}}$ (where M is the set of molecules covered by the data) corresponds to computing the pairs $\{(p_m^{\text{Brand}}, p_m^{\text{Gen}})\}_{m \in \mathcal{M}} \equiv \{(p_m^0, p_m^1)\}_{m \in \mathcal{M}}$.

We describe the involved procedures focusing first on the case of Average Wholesale Prices (AWPs). To do this, it is important to first clarify the structure of the data. Although AWPs have some temporal variation, they primarily vary across products. A product can be either brand name ($g = 0$) or generic ($g = 1$), and each product is associated with a unique molecule m.

In this context, let K_{mg} represent the set of molecule m products of generic type g. A set K_{mg} may (and often does) contain more than one product if there is more than one firm producing the molecule, or if the molecule is available in more than one format (due to different dosages and/or delivery routes). For $g = 0$ (Brand) and 1 (Generic), the price p_m^g entering Equation A.7 is computed as:

$$p_m^g = \sum_{k \in K_{mg}} \omega_k \tilde{p}_k, \tag{A.8}$$

where $\tilde{p}_k$ is the median AWP of product k within the full set of drug claims contained in the 2011 and 2012 MarketScan datasets (about 700 million claims, averaging about 12,000 transactions per product). In turn, ω_k represents a frequency weight for product k computed based on the number of associated transactions, with $\sum_{k \in K_{mg}} \omega_k = 1$. The resulting distribution of Δ values is shown in Figure A.4A (one value per molecule, truncated at 10).

G.2 Price Differentials Based on Out-of-Pocket Prices

Out-of-Pocket (OOP) prices refer to the share of a drug's total cost that is recorded by the patient. To compute the OOP version of Δ_m, a preliminary step is needed. For each individual i we compute the AWP-coverage ratio, γ_{ig}. This ratio is computed as:

$$\gamma_i^g = \frac{\sum_{j \in C_{ig}} (\text{Deductible}_j + \text{Copay}_j + \text{Coinsurance}_j)}{\sum_{j \in C_{ig}} \text{AWP}_j}. \tag{A.9}$$

In Equation A.9, the set C_{ig} corresponds to all drug claims filed (prior to the LDL/A1c test) by individual i for drugs of generic type g. The numerator sums all the OOP components associated to a claim j, across all claims in C_{ig}. The denominator, in turn, sums all the AWPs for the same set of transactions. As such, γ_i^0 represents the average OOP payment (as a share of AWPs) that individual i records when s/he purchases brand name products. (Instead of the average of

ratios, we compute the ratio of sums to avoid instabilities.) Analogously, $\checkmark_i^1$ represents the AWP fraction a↵orded OOP when the individual purchases generics. Figure A.4B showcases the wide coverage variability manifesting via $\checkmark^0$ and $\checkmark^1$ parameters (rare cases outside the $[0, 1]^2$ square are omitted from the figure). With these values in hand, we compute the OOP version of $\blacklozenge$ as:

$$\blacklozenge_{mi} = \frac{p^0_m \cdot \checkmark^0_i - p^1_m \cdot \checkmark^1_i}{p^1_m \cdot \checkmark^1_i}. \tag{A.10}$$

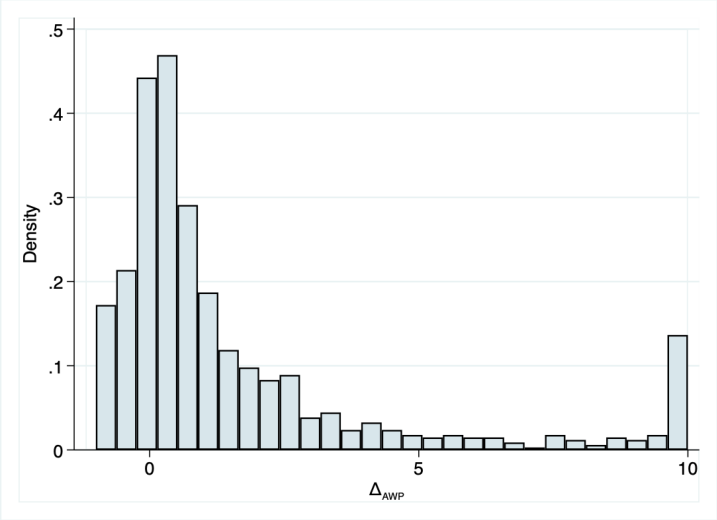
It is important to note that:

- We have aggregated coverage at the individual/generic type level because the number of individuals who consume brand name and generic versions of the same molecule is relatively small.
- Many individuals consume only brand name or only generic drugs. For these people, one of the two $\checkmark$ values is missing and we cannot compute $\blacklozenge$. For this reason, the OOP analysis is subject to a significant sample size loss compared with the AWP analysis.

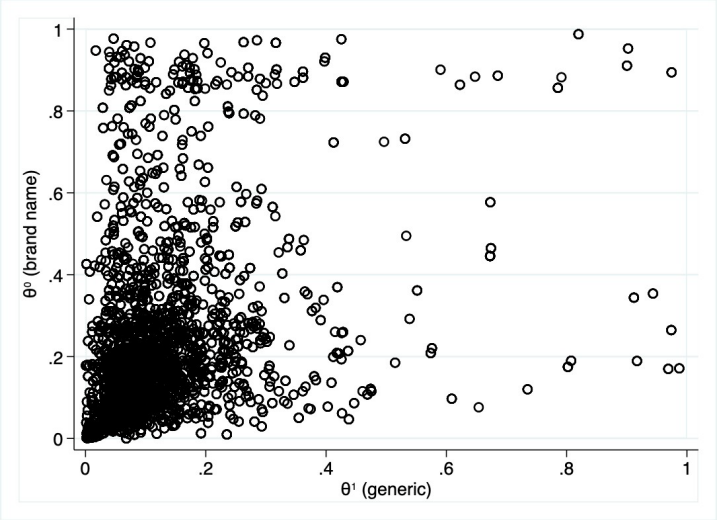
Estimation results using the partition induced by OOP prices are presented in Table A.7. These results o↵er the same qualitative conclusion as its counterpart results based on AWP prices, as presented in Table 3 of the main text.

Figure A.4: Pricing di↵erences between branded and generic drugs.

(A) Brand/generic price di↵erentials Δ based on AWPs.



(B) Scatterplot (θ^1, θ^0).



Notes. Procedures used to construct these values are described in Appendix G.

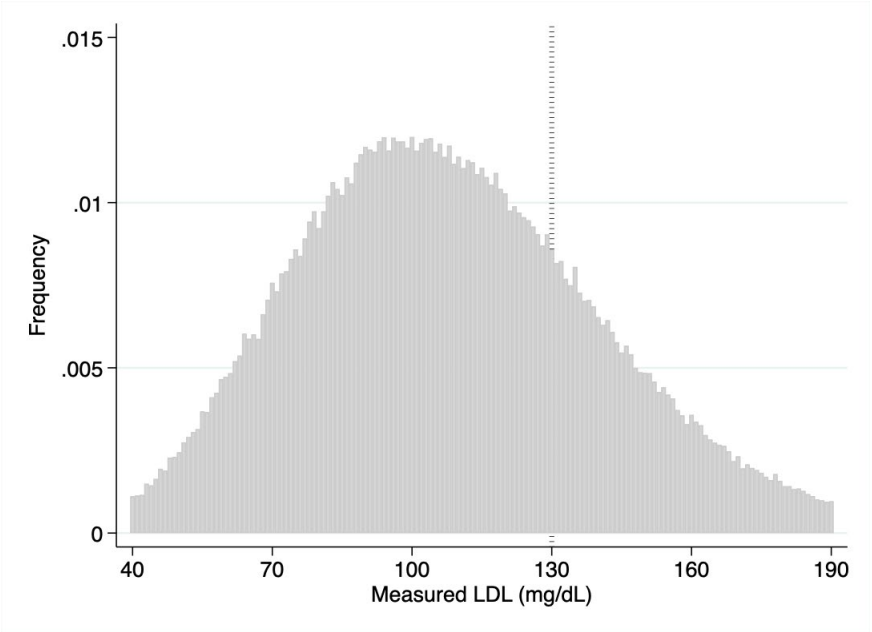
Table A.7: Price as a signal of quality (Out-of-Pocket Prices).

	(1) LDL	(2) A1c	(3) Pooled
Treated→Post	0.0024	0.0108*	0.0091
	(0.0015)	(0.0064)	(0.0057)
Treated→Post→AboveMedian6	−0.0436**	−0.0223***	−0.0240***
	(0.0188)	(0.0067)	(0.0062)
N	13,061	72,810	85,903

Notes. Linear probability specifications for the probability of choosing the generic option (Equation 2). The pooled sample includes individuals from both the LDL and A1c sample. All models include fixed effects for molecules, individuals, and the month difference between the test date and drug claim. Parentheses show standard errors clustered at the level of patient/molecule pairs. *Legend:* * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$..

H Supplemental Figures and Tables

Figure A.5: Distribution of LDL results contained in the 2011-2012 MarketScan data.



Notes. The vertical dotted line marks the frontier between "acceptable" and "borderline high" LDL measurements. Results are expressed in mg/dL, corresponding to milligrams per deciliter.

Table A.8: Quantity and Bundle e↵ects in the A1c sample.

	(1)	(2)
Dependent variable	CLAIMS	BUNDLE
Treated⇥Post	−0.0011	−0.0003
	(0.0159)	(0.0190)
Individual FEs	X	X
Time period FEs	X	X
Sample	Full	Full
Aggregation	Individual/month	
N	83,448	83,448

Notes. Log linear estimates based on Equation 1 using the sample of A1c testers. Standard errors (shown in parentheses) are clustered at the individual level. *Legend:* ⇤ $p < 0.1$,⇤⇤ $p < 0.05$,⇤⇤⇤ $p < 0.01$..

Table A.9: Bad A1c News: All Clinical Thresholds.

	(1)	(2)	(3)
	Diagnosing Diabetes		Managing Diabetes
	5.6 vs 5.7%	6.4 vs 6.5%	6.9 vs. 7.0%
Treated⇥Post	−0.0004	0.0017	−0.0049
	(0.0013)	(0.0025)	(0.0063)
N	1,186,264	269,153	143,282

Notes. Linear probability specifications for the probability of choosing the generic option (Equation 2). All models include fixed e↵ects for molecules, individuals, and the month di↵erence between the test date and drug claim. Parentheses show standard errors clustered at the level of patient/molecule pairs. *Legend:* ⇤ $p < 0.1$,⇤⇤ $p < 0.05$,⇤⇤⇤ $p < 0.01$..

Table A.10: Bad A1c News: Balancing (6.9 vs 7.0%).

	(1)	(2)	(3)
	6.9%	7.0%	Std. Di↵.
(A) Demographics and location			
Age	53.03	52.56	-0.04
	(9.36)	(9.31)	
Sex (fraction female)	0.48	0.47	-0.01
	(0.50)	(0.50)	
(B) Utilization prior to LDL testing (monthly claims)			
All drugs	1.54	1.66	0.03
	(2.02)	(3.13)	
Generic drugs	1.26	1.36	0.03
	(1.70)	(2.55)	
Cardiovascular drugs	0.53	0.59	0.03
	(0.95)	(1.70)	
In-patient admissions	0.01	0.01	-0.03
	(0.11)	(0.05)	
Outpatient services	3.31	3.07	-0.03
	(7.19)	(5.29)	
All medical tests	1.21	1.19	-0.00
	(6.24)	(6.62)	
(C) Same-day medical testing			
Same-day medical tests	28.75	29.03	0.01
	(21.29)	(21.25)	
(D) Insurance coverage (implicit)			
Coverage ratio	0.36	0.37	0.04
	(0.25)	(0.26)	
N	2,382	2,122	

Notes. Parentheses show sample standard deviations.

Table A.11: Major diagnostic categories associated with studied LDL tests.

	(1)	(2)	(3)	(4)
	Freq.	Pct.	Freq.	Pct.
	129mg/dL		130mg/dL	
Diseases and Disorders of the Nervous System	11	0.94%	10	0.90%
Diseases and Disorders of the Eye	1	0.09%	1	0.09%
Diseases and Disorders of the Ear, Nose, Mouth and Throat	2	0.17%	9	0.81%
Diseases and Disorders of the Respiratory System	4	0.34%	6	0.54%
Diseases and Disorders of the Circulatory System	116	9.92%	111	9.97%
Diseases and Disorders of the Digestive System	22	1.88%	16	1.44%
Diseases and Disorders of the Hepatobiliary System and Pancreas	1	0.09%	2	0.18%
Diseases and Disorders of the Musculoskeletal System and Connective	21	1.80%	16	1.44%
Tissue Diseases and Disorders of the Skin, Subcutaneous Tissue and Breast				
Endocrine, Nutritional and Metabolic Diseases and Disorders	3	0.26%	8	0.72%
Diseases and Disorders of the Kidney and Urinary Tract	200	17.11%	166	14.91%
Diseases and Disorders of the Male Reproductive System	12	1.03%	13	1.17%
Diseases and Disorders of the Female Reproductive System	3	0.26%	7	0.63%
Pregnancy, Childbirth and the Puerperium	3	0.26%	3	0.27%
Newborns and Other Neonates with Conditions Originating in Perinatal Period	13	1.11%	8	0.72%
Diseases and Disorders of the Blood, Blood Forming Organs, Immunological				
Disorders Myeloproliferative Diseases and Disorders, Poorly Di↵erentiated Neoplasm				
Infectious and Parasitic Diseases, Systemic or Unspecified Sites	2	0.17%	0	0.00%
Mental Diseases and Disorders	4	0.34%	6	0.54%
Alcohol/Drug Use and Alcohol/Drug Induced Organic Mental Disorders	0	0.00%	0	0.00%
Injuries, Poisonings and Toxic E↵ects of Drugs	0	0.00%	1	0.09%
Burns	0	0.00%	0	0.00%
Factors Influencing Health Status and Other Contacts with Health Services	243	20.79%	216	19.41%
Multiple Significant Trauma	0	0.00%	0	0.00%
Human Immunodeficiency Virus Infections	0	0.00%	0	0.00%
Missing/Unspecified	508	43.46%	514	46.18%
Total	1169	100%	1113	100%