Risk Screening in Digital Insurance Distribution: Evidence and Explanations

Finbarr Murphy[1], Wei Xu[*,1], Xian Xu[2]

Abstract

Extant research holds that digital sales channels advances competition and service or goods availability, but rarely notes the relationship with information asymmetry. With a large and detailed dataset, this study examines the specific case where consumers have a choice between offline and digital channels in the context of insurance purchases. We find that digital channels screen in consumers with lower unobserved risk. For the term life, endowment and disease insurance products, the average risk of the policies purchased through digital channels were 75%, 21% and 31% respectively lower than those purchased offline. As a consequence, the lower unobserved risk leads to weaker information asymmetry and higher profitability of digital channels. We highlight three mechanisms of the risk screening effect: heterogeneous marginal influence of channel features on insurance demand, channel features directly related to risk control and the link between digital divide and risk.

Keywords: Digital Economy; Risk Screening; Information Asymmetry; Insurance; Mobile Application.

[1] Department of Accounting and Finance, Kemmy Business School, University of Limerick, Ireland

[*] Corresponding Author: Wei.Xu@ul.ie

[2] Department of Risk Management and Insurance, School of Economics, Fudan University, China

1 Introduction

The embedding of digital technologies in the global economy has attracted increasing attention from economists. A key insight into the economic mechanism of digital technology is that it lowers search and transaction costs (Brynjolfsson et al. 2011; Einav et al. 2014; Jolivet and Turon 2019; Goldfarb and Tucker 2019). This logic underpins a mass of microeconomics literature on how digital technology improves consumer surplus by advancing market competence and service or goods accessibility (Brown and Goolsbee 2002; Ellison and Ellison 2009; Jack and Suri 2014; Callen et al. 2019). However, few studies have addressed the heterogeneity in consumers' responses to digital technology adoption. That is, under equal access and acceptance of digital technology[1], consumers who choose to use digital technology inherently differentiate from those who do not and these differences, unsurprisingly, can affect the economic consequences of adopting digital technology for enterprise. In this article, we relate the sale of insurance policies with digital distribution[2] and show that digital distribution screens in consumers with lower unobserved risk on average.

Our analysis comes from studying the difference in average policy risk across channels for the same insurance product sold simultaneously through both digital distribution – Mobile applications (APPs) and traditional offline distribution – individual agent or bancassurance channels. Using unique data on policy purchases provided by a large Chinese life insurer operating nationwide, we document that after controlling for all observed policyholder and policy characteristics, enrollees who opt for digital channels have lower unobserved risk than those who opt for traditional offline channels. This risk screening effect of digital distribution mitigates information asymmetry because it is not adjusted into the unit premium by the insurer. The magnitude of the decrease in the accident probability of digital channel policies is

[1] This premises that the acceptance and ability to use digital technology-based services or products are indifferent across consumers.

[2] Throughout the paper, we regard digital distribution as the distribution channel that relies on 5G technology and mobile devices, such as the mobile App distribution channel.

substantial when compared with offline policies – accounting for 75%, 21% and 30% of the offline accident rate for the term life insurance, endowment insurance and disease insurance, respectively.

The risk screening effect makes digital distribution a potential insurer-side risk selection tool. Employing the fact that digital distribution was introduced in addition to existing traditional offline distribution during the sales period of the term life insurance, we empirically show that the insurer's introduction of digital distribution considerably lowers the average policy risk of the term life insurance by attracting more low-risk enrollees.

There are two important economic consequences of the risk screening effect for the insurance sector. The first important consequence is related to information asymmetry. Unobserved risk is a source of information asymmetry that impedes insurer's assessment on applicants' risks and results in the "*lemon market*" phenomenon. We present evidence that the risk-coverage correlation is weaker for digital channels, indicating that digital distribution channels have lower information asymmetry. The second important consequence is that the lower unobserved risk also increases the profitability of digital distribution. This is demonstrated by the evidence that the loss ratio is over 10% lower for digital channel policies than for offline policies.

A significant contribution of this paper is decomposing the drivers of the risk screening effect as *channel capability* and *channel preference*. The role of *channel capability* comes from the channel features (i.e., search cost) that correlate with both insurance demand and risks of attracted enrollees. Two mechanisms of *channel capability* are presented. The first mechanism is the heterogeneity in the marginal influence of channel features on the insurance demand of different risk groups. We study this mechanism based on an advantageous channel feature of digital distribution – reduced search cost. We theoretically and empirically show that consumers self-select digital channels because of a higher insurance demand sensitivity to the reduction of search costs afforded by digital distribution. The second mechanism of *channel capability* is the channel features directly related to risk control. We study this mechanism by showing that the poor underwriting service quality of offline agents leads to enrollment of high-risk consumers who should not have been eligible for insurance.

The role of *channel preference* into the risk screening effect relies on the relationship between policy risk and the ability or acceptance to use digital channels (i.e., digital divide). We examine this role by using advanced education as a typical risk characteristic not adjusted into the unit premium, and showing that advanced education positively correlates with the preference to use digital technology while negatively correlates with policy risk.

Finally, we explore consumer contributions to the risk screening effect by benchmarking the counter-fact. We find that the introduction of digital distribution mainly attracts both low-risk consumers switching from offline channels and new low-risk consumers. The former lead to average risk increases in offline policies, however, this only contributes to a small part of the risk screening effect. At least 74% of the risk screening effect sources from the average risk decreases in the policies of digital distribution, among which, ulteriorly, almost 77% can be attributed to the attracted new low-risk consumers. Therefore, the risk screening effect largely comes from the extensive margin (i.e., from new consumers).

This article makes three significant contributions to existing literature. First, it contributes to the digital economics literature by highlighting the efficacy digital technology adoption. Evidence of this has been largely anecdotal with the most relevant literature on the financial inclusion of mobile banking (Kochar 2018; Stein and Yannelis 2021). For example, Kochar (2018) emphasizes that savings of low-income households are more sensitive to the existence of mobile banking. Therefore, this article broadens the understanding of the heterogeneity of the economic impact of digital technology.

Second, this article is related to a large literature set that empirically studies screening in the insurance market. Two veins of this literature are most relevant to our article. The first vein is the growing studies showing the important engagement of choice frictions with risk selection. Consumers' lack of awareness of plan properties, choice complexity, choice overload, inertia and behavioral frictions can impact the plan choice of insurance consumers (Abaluck and Gruber 2011; Ketcham et al. 2012; Kling et al. 2012; Handel 2013; Handel and Kolstad 2015; Domurat et al. 2021). Such frictions can result in higher equilibrium pricing (Ericson

2014) and adverse selection welfare loss (Polyakova 2016; Handel et al. 2019). Our article enriches this literature by incorporating the reduction in search cost by digital technology as a new engagement factor. The second vein of related literature on the insurer-side risk selection. Risk adjustment systems, plan designs, hospital network and advertising have been empirically shown as insurers' risk selection tools (Brown et al. 2014; Carey 2017; Aizawa & Kim 2018; Shepard 2022). Our article's novelty is to show digital distribution as another tool for insurers to achieve advantageous risk selection.

Lastly, this article also adds to the literature on marketing in the insurance context. Different distribution channel strategies have different impacts on insurance business. While there are some empirical insurance studies investigating the effect of PC-Internet distribution on promoting insurance market competence and insurance demand (Brown and Goolsbee 2002; Butler 2021; Hu et al. 2022), they do not particularly address the effect on policy risk which is of utmost relevance to insurance institutions. They do not include digital distribution as a differentiating factor. The only exceptions are Venezia et al's (1999) and Hsieh et al.'s (2014) studies arguing that the higher claim service quality by independent agents compared to direct underwriters may lead to risk sorting. There are two main differences between their studies and our article. First, the offline channel in our setting consists of employed agents on behalf of the insurer's interest instead of independent agents, in theory with smaller claim service quality differences from direct underwriters. Second, our documented risk screening effect cannot be explained by their argument because the offline claims in our setting are actually more likely to be rejected, which motivates us to explore other explanations.

The remainder of this paper is structured as follows: the next section introduces a conceptual framework introducing the main ideas of this paper. Section 3 shows data description and baseline empirical specification. Section 4 presents the baseline results. In Section 5, two important consequences on insurance business induced by the risk screening effect are examined. Section 6 examines three mechanisms of the risk screening effect according to the conceptual framework. Section 7 decomposes the consumer contributions to the risk screening effect. The final Section 8 concludes this paper and discusses implications for the future research.

2 A Conceptual Framework

In this section, we provide a conceptual framework to theorize the meanings and mechanisms of the risk screening effect of digital insurance distribution. This framework motivates our empirical tests by (1) highlighting the engagement between information asymmetry and the adoption of digital distribution and (2) decomposing sources of the risk screening effect as *channel capability* and *channel preference*.

2.1 Information Asymmetry and Digital Insurance Distribution

Consider a simple market for a single insurance product offering coverage of a fixed insured amount L[3]. Based on our empirical settings, we focus on the insurer's decision to introduce the digital distribution channel in addition to the existing offline channels. Let $j \in \{0, 1\}$ denote the insurer's adoption of the offline distribution channel (taking value 0) and the digital distribution channel (taking value 1). Let D_{ij} indicate consumer i's demand of purchasing insurance through the distribution channel j, given the channel capability and channel preference. The insurer charges the unit premium based on the fixed loading factor λ and the expected risk of observed risk characteristics z_i (such as age and gender). Accordingly, both channels charge the same premium for the same consumer.

If the insurer has perfect information on the consumer's risk, the resulting unit premium would mitigate the adverse effects of information asymmetry (e.g., adverse selection or moral hazard). However, in general, there are always unobservables and it is impossible for the insurer to charge a perfect unit premium[4]. Let q_i denote consumer i's actual risk, the expected

---

[3] For simplicity, we keep the insured amount fixed. This is without loss of generality because a varied indemnity is similar in technical to a varied loss probability q_i and thus will generate similar conclusions.

[4] Especially for the pricing of life insurance in China, insurers usually comply with a uniform mortality table or disease table classifying risks simply by gender and age, which is regulated by the China Banking and Insurance Regulatory Commission. Under this pricing principle, the measurement of applicant risk does not take all

risk of observed risk characteristics is $\Phi_i = E(q_i|z_i)$. Then the total profits of only keeping the offline distribution π_0 and additionally introducing the digital distribution $\pi_{0,1}$ can be respectively written as

$$\pi_{0,1} = \sum_i (\lambda - q_i{}^{RA})\Phi_i L \cdot max\{D_{i1}, D_{i0}\} \tag{1}$$

$$\pi_0 = \sum_i (\lambda - q_i{}^{RA})\Phi_i L \cdot D_{i0} \tag{2}$$

Where $q_i{}^{RA} = q_i/\Phi_i$ is the unobserved risk after adjusting the insurer's expected risk. The outcome of interest is the change of the insurer's profits when introducing the digital distribution, which can be written as

$$\Delta\pi = \sum_i (\lambda - q_i{}^{RA})\Phi_i L \cdot max\{\Delta D_{ij}, 0\} \tag{3}$$

This equation clearly illustrates how the effect of information asymmetry on profits engages with the entry of digital distribution: there is advantageous/adverse risk screening if enrollees who select into the digital distribution channel ($\Delta D_{ij} > 0$) have lower/higher unobserved risk relative to λ. Therefore, the difference in unobserved risk between enrollees of the digital and offline distribution channels essentially reflects the impact of adopting digital distribution on information asymmetry, which motivates our following empirical tests.

2.2 Decomposing Sources of the Risk Screening Effect

What drives the advantageous risk screening of digital distribution? There are two potential sources: *channel capability* and *channel preference*. We call the channel features (i.e., search cost) that correlate with both insurance demand and risks of attracted enrollees as *channel capability*, and the difference in the ability and acceptance to use different channels (i.e., digital divide) as *channel preference*. To understand their roles, we decompose D_{ij} and ΔD_{ij} as

$$D_{ij} = p_{ij} \times CD_{ij} \tag{4}$$

$$\Delta D_{ij} = (CD_{i1} - CD_{i0}) \cdot p_{i0} + (p_{i1} - p_{i0}) \cdot CD_{i1} \tag{5}$$

characteristics into account.

Where p_{ij} indicates whether consumer i is willing and able to use the distribution channel j, and CD_{ij} indicates the insurance demand when consumer i is willing and able to use the distribution channel j. The logic of the decomposition in equation (4) is very natural. Only when consumers are willing and able to use a digital channel can they be affected by the introduction of digital distribution; when both channels can be accepted and used by consumers, the utility of channel choice will be affected by channel features. The roles of *channel capability* and *channel preference* are presented in equation (5) where term (i) represents the impact of channel capability (i.e., $CD_{i1} > CD_{i0}$) and term (ii) represents the impact of channel preference (i.e., $p_{i1} > p_{i0}$).

Three reasons suggest the relationships of *channel capability* and *channel preference* with unobserved risk. First, the marginal influence of channel features on the insurance demand of different risk groups is heterogeneous. For example, if there is some advantageous feature of digital distribution that increments its insurance demand compared with the offline channel, consumers with high unobserved risk would respond less to this feature in the marginal insurance demand than those with low unobserved risk, which results in a lower unobserved risk on average for the enrollees of digital distribution.

Second, the channel features related to risk control set up a direct link with unobserved risk. For example, if there are differences in the rigor of implementing underwriting rules between digital and offline distribution, it is natural to expect a difference in unobserved risk.

Third, *channel preference* may also directly negatively correlate with unobserved risk due to digital divide. Deursen and Dijk (2019) found that even if the Internet penetration has reached saturation in a country, differences in individuals' ability or acceptance to use Internet technology (e.g., mobile Apps) still lead to inequality in benefitting from the Internet, creating widening digital divide[5]. Generally, the old, poor-educated or low-socioeconomic-status are more likely to lack the ability or acceptance to use digital technology (Wei and Hindman 2011; Hall and Owens 2011). This is precisely the group that may have higher unobserved risk. In

[5] In literature (Büchi et al.; Scheerder et al., 2017), the digital divide due to differences in individuals' ability or acceptance to use Internet technology is also called second-level digital divide.

essence, the heterogeneous *channel preference* reflects different clientele being catered by two distribution channels.

We further empirically examine the impacts of *channel capability* (based on two channel features) and *channel preference* in detail in Section 6.

3 Data and Empirical Strategies

3.1 Settings of Digital Insurance Distribution

The adoption of digital distribution is particularly prevalent among Chinese life insurers, thanks to a developed communication infrastructure and widespread use of mobile devices. According to statistics from the Insurance Association of China (Securities Journal 2016), among the 55 life insurance companies that have disclosed their domestic Internet insurance business, 18 have launched mobile APPs to sell insurance products. The earliest ones have been running mobile APPs as a digital distribution channel for over 10 years. These mobile APPs usually provide the entire service journey from quotation to claim. Therefore, China's life insurance industry is a good window to observe the effects of digital distribution adoption.

We study the insurance products sold simultaneously through both digital and offline distribution channels from a life insurer operating nationwide in China. This is a good setting to study the risk screening implications of digital distribution for several reasons. First, the investigated life insurer is one of the first insurers to launch the mobile APP channel in China. Its mobile APP has reached maturity after many iterative updates, characteristic of the life insurance industry. Second, its mobile APP sells products of various insurance types, allowing our study to cover different insurance types and makes our research conclusions more generalized. Third, the investigated insurance products keep the same plan settings across digital and offline distribution channels, insulating our estimates from the potential bias of unobservable product differences.

The mobile APP can be downloaded for free from mobile application stores. For the insurance product on sale, plan information such as application qualifications (e.g., insurable ages), liability and exempted liability descriptions, insurance period, insurance terms and optional plan settings (e.g., optional insured amount, additional insurance) are displayed clearly. Consumers can click on the screen to choose their preferred plan settings and see premiums. In this process, consumers can communicate with a 24-hour human customer service operator for additional product information. After identifying the preferred plan, consumers complete the required health information to be automatically checked for compliance with underwriting rules. Once the underwriting is approved, consumers will check and confirm the application information and plan details before paying premiums online.

The insurer's mobile APP also provides after-sales services including purchased policy retrieval and policy claims. Since the mobile APP account is tied to the personal identity, consumers can see all their historical policies whether purchased offline or via the mobile APP. For each historical policy, consumers can file for claims by submitting required proofs (e.g, pictures or electronic files).

Compared to digital distribution channels, the insurance application process of tradition offline channels – offline agents and bancassurance – has several notable differences. The first difference is the obvious absence of convenience for offline purchasing. The second difference is the time limit on offline channel services. For example, the bancassurance channel has fixed operation times while APP consumers can purchase insurance at any time. Moreover, health information is usually checked for compliance with underwriting rules manually by offline agents instead of automatically. However, the after-sales process is similar between offline and digital distribution channels – because the offline agents are required by the insurer to let consumers download and register the mobile APP after selling the policy, so that consumers who purchase policies offline can still enjoy the after-sales services on the mobile APP.

3.2 The Data

We gather the detailed proprietary data on records of all purchased policies for the three investigated insurance products, term life, endowment and disease insurance from a large life insurer operating nationwide in China. The data on these three products is used because they were sold through both offline and digital distribution channels simultaneously and have accumulated large sales numbers. For each policy, policy characteristics include purchase time, purchase channel, insured amount, unit premium, insurance period, payment term, policy status, hesitation period, waiting period and other plan properties such as selected additional liabilities. These data allow us to distinguish between different purchase channel choices and control for other policy characteristics in empirical tests. Policyholder characteristics include gender, age, education, profession, address and the relationship with the insured person. These also serve as control variables. Claim characteristics include claim record, accident type, indemnity amount, claim rejection record and reasons for claim rejection. These allow us to construct policy risk measures and shed light on how digital distribution affects them.

Table III.1 reports the descriptive statistics of the main variables by insurance product. The term life insurance product was sold from 2017 to 2019 on the offline agent channel and introduced on the insurer's mobile APP channel on August 15th of 2018. It insures death or disability risk with a total of 97,495 policies, 86% of which were purchased via the APP. The endowment insurance was sold simultaneously on the insurer's APP channel, the offline agent channel and the PC-Internet channel from May to December in 2019, totaling 963,244 purchased policies. This endowment insurance product covers the liabilities of death or disability, disease medication, accident medication and severe illness. We see that the APP channel policies account for roughly 17%. In empirical tests, the PC-Internet channel policies (around 0.3% of all policies) are dropped because we focus on the digital and offline channels. The whole life disease insurance product was sold on the both the APP channel and offline bancassurance channel from October in 2016 to May in 2018. This product insures death or disability and severe diseases risks, totaling 53,296 purchased policies with 38% purchased via the digital channel.

We also compare the plan settings between the policies of digital and offline distribution channels, as shown in Table A1, Appendix A. One can note that the ranges of plan settings[6]

---

[6] Also, each investigated insurance product shares the same premium rate table across different distribution channels.

such as waiting period and insurance period are effectively the same across distribution channels, indicating no differences between the plan settings offered by digital and offline distribution. It is normal that insurers are usually very reluctant to vary plan properties for the same product across different channels, because variations tend to intensify channel benefit conflicts (Geyskens et al. 2002). The consistency between digital and offline distribution channels is a crucial advantage of our data, as it avoids potential bias from unobservable product differences that are difficult to resolve in prior literature (Brown and Goolsbee 2002).

Table III.1: Descriptive Statistics of the Data

	Term life Insurance		Endowment Insurance		Disease Insurance	
Variables	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Age	36.586	7.421	38.922	8.757	37.697	7.711
Female	1.869	0.991	0.488	0.500	0.723	0.448
Financial Profession	0.386	0.487	0.136	0.343	0.182	0.386
Advanced Education	0.251	0.434	0.137	0.344	0.285	0.451
Policy Cancellation	0.005	0.068	0.010	0.098	0.001	0.035
Insurance Period	21.087	3.112	37.164	9.100	Whole life	Whole life
Insured Amount	32.858	18.778	9.428	15.735	48.475	24.370
Unit Premium	0.547	0.227	0.138	0.154	0.010	0.008
Additional Liability	0	0	0.381	0.457	0.858	0.349
Payment Term	17.766	4.165	15	0	19.182	2.741
Self-insured Relation	0.960	0.205	0.725	0.461	0.438	0.496
Digital Channel Choice	0.861	0.346	0.166	0.372	0.382	0.486
Policy Claim	0.001	0.036	0.035	0.185	0.008	0.089
Observations	97,495		963,244		53,296	

Note: Insured amount is in ten thousand Yuan and unit premium is in Yuan. Financial profession indicates whether the policyholder is employed in the financial industry. Advanced education indicates whether the policyholder has an undergraduate degree or above. Insurance period is in years. Policy cancellation indicates whether the policy has been cancelled. Additional liability indicates whether the policy includes additional insurance. Payment term is the period of premium payment in years. Self-insured relation indicates whether the policyholder is the insured. Digital channel choice is a dummy indicating whether the policy is purchased through the digital channel.

To examine evidence on the *channel capability* mechanism of the risk screening effect, we construct measures of offline insurance search costs with a variety of datasets on geography, population and weather. The first dataset is the API service offering the longitude and latitude of insurer branches throughout China in 2019, provided by BAIDU Map[7], one of the largest private digital map providers in China. The second one is the grid-cell data of population density at the accuracy level of one square kilometer throughout China in 2019, which is publicly available on the website of WorldPop[8]. Each insurer branch and population grid-cell are

[7] Seen on https://lbsyun.baidu.com/

[8] Available on https://www.worldpop.org/

matched to the corresponding prefecture according to their coordinates. The third dataset is the daily rainfall for each prefecture of China in 2019, available on the website of the China Meteorological Data Science Center[9]. We also offer description of related variables in Table A2, Appendix A.

3.3 Empirical Specifications

We now estimate the risk screening effect of digital distribution on policy risk. Our first estimate is performed at the individual policy level. For each insurance product, we relate policy risk with purchase channel choice using the following baseline specification at the individual policy level:

$$\boldsymbol{AC}_{i,r,t} = \alpha + \beta \boldsymbol{digital}_{i,r,t} + \theta \boldsymbol{D}_t + \boldsymbol{X'}_{i,r,t}\boldsymbol{\Gamma} + \boldsymbol{X'}_{r,t}\boldsymbol{\Phi} + \varepsilon_{i,r,t} \tag{6}$$

Where i, r and t index policyholder, prefecture and purchase date, respectively. $\boldsymbol{AC}$ is a dummy indicating whether the policy has claimed for an accident. $\boldsymbol{digital}$ is the independent variable of interest that represents the choice of the digital distribution channel. $\boldsymbol{D}_t$ is a trend term capturing the natural time influence on policy risk. $\boldsymbol{X'}_{i,r,t}$ represents a set of controls for policyholder and policy characteristics: gender, age, age squared, a dummy of advanced education, a dummy of whether employed in the financial industry, dummies of relationships with the insured person, logarithmic insured amount, unit premium, dummies of insurance period, dummies of payment term, a dummy of policy status (whether the policy was cancelled) and a dummy of having additional liabilities, as listed in Table III.1. $\boldsymbol{X'}_{r,t}$ is a vector of prefecture-year interactive fixed effects (also including separate prefecture and year fixed effects), month, day-in-month and day-in-week fixed effects. Robust standard errors are clustered at the prefecture level. This OLS estimate directly captures the difference in unobserved policy risk between digital and offline distribution channels. The regression samples consist of the policies purchased via digital and offline distribution channels for the endowment insurance and disease insurance. While for the term life insurance, the regression sample

[9] Available on http://www.nmic.cn/data/cdcindex/cid/0b9164954813c573.html

consists of the policies purchased after introducing digital distribution (from August 15th in 2018 to December 31st in 2019), keeping the investigated sales periods of both digital and offline channels consistent.

Our second estimate pursues the evidence on the effect of the introduction of digital distribution and is performed at the prefecture aggregation level. We construct a difference-in-difference (DID) counterfactual framework by employing the introduction of digital distribution on August 15th, 2018 for the investigated term life insurance product. This DID framework takes the investigated term life insurance product as the treatment product and another term life insurance product sold purely through offline agents[10] in 2018 as the control product. Both products have simple and similar plans[11]: they only insure death or disability risk without any additional liabilities. We aggregate policies by product, prefecture and date. The DID specification can be written as

$$\boldsymbol{LnAR}_{p,r,t} = \alpha + \pi \boldsymbol{Treat}_p \times \boldsymbol{\mathcal{L}}_t + \boldsymbol{X''}_{p,r,t}\boldsymbol{\Gamma} + \boldsymbol{X''}_t\boldsymbol{\Phi} + \boldsymbol{X'}_{r,p}\boldsymbol{\Omega} + \varepsilon_{p,r,t} \qquad (7)$$

where for product p, prefecture r and date t, $\boldsymbol{LnAR}_{p,r,t}$ denotes natural log of 1 plus the accident rate[12] of sold policies, $\boldsymbol{Treat}$ is defined as 1 for treatment product and 0 for control product, $\boldsymbol{\mathcal{L}}_t$ is defined as 1 after August 15th and 0 otherwise. $\boldsymbol{X''}_{p,r,t}$ is a vector including the averages of policyholder and policy characteristics. For example, we calculate the average of the unit premiums of all the policies purchased in prefecture r on date t for product p; in a similar way, dummy controls (such as relationships between the policyholder and the insured) are averaged into percentages. $\boldsymbol{X''}_t$ is a vector including date fixed effects and $\boldsymbol{X'}_{r,p}$, a vector including product and prefecture fixed effects. π captures the decrease in the average risk of

---

[10] Our criteria for selecting products suitable as the control group are: (1) covering the same liabilities, (2) being sold during the same period and (3) being sold only in the same offline channel as the investigated product. Amongst all three investigated insurance products, we only find out such a product that meets all three criteria for the term life insurance product. This is presumably because relative to other products, term life insurance products are simpler and tend to be homogeneous.

[11] The mere difference in plan setting is that the control product offers an optional insurance period of 10 or 15 years, while the treatment product offers a longer optional insurance period of 20 or 30 years.

[12] For product p, the accident rate is calculated as the share of the claimed policies among all the policies sold in prefecture r on date t.

policies during the period with digital distribution relative to the counter-fact without digital distribution. The econometric intuition is that the estimated effect should reflect the influence of introducing digital distribution on the average policy risk across prefectures. Here, one should note a key difference in results interpretation from above OLS estimates that, the identified effect reflects the difference in average policy risk from the offline channel in the counter-fact instead of in reality.

4 Results

4.1 OLS Estimates

Table Ⅳ.1 presents the results of estimating the risk screening effect of the digital distribution on policy risk with the baseline OLS specification. Panel A, B and C present the results for the term life insurance, endowment insurance and disease insurance product respectively. Column 1 employs the full sample, while Columns 2, 3 and 4 exclude cancelled, non-self-insured and claim-rejected policies respectively. We find that the results are statistically significant and negative for all three panels, regardless of policy cancellations, claim rejections and relationships with the insured. Specifically, Column 1 implies that on average, the accident probability of the policies purchased through digital distribution channels are 0.22, 0.36 and 0.34 percent points lower than that of the policies purchased offline for the term life insurance, endowment insurance and disease insurance, respectively.

The magnitude of the risk screening effect of digital distribution on policy risk, documented in Table Ⅳ.1, is substantial. The accident rates of offline policies are presented in the bottom row of each panel as the basis to understand the magnitude of the risk screening effect. The reduction in risk probability of the digital distribution channel policies takes up at least 75% (=0.0022/0.0029), 21% (=0.0029/0.0137) and 31% (=0.0018/0.0059) of the corresponding offline accident rate for the term life, endowment and disease insurance product, respectively.

Since our dependent and independent variables are both dummy, we complement Logit regressions using the same empirical specification for robustness. We also include Cloglog regressions given the rarity of claims. Their estimate results, presented in Table A3 of Appendix A, are qualitatively and quantitatively consistent with those of OLS estimates.

An insurance product usually covers a package of liabilities insuring different risks, generating a possibility that the risk screening effect of digital distribution would lose universality if it only occurs to specific risks. To address this concern, for the endowment and disease insurance, we group the policy claims by accident types including medication for diseases, medication for accidents, severe diseases and death or disability, forming four corresponding subsamples along with unclaimed policies. Columns 1 to 4 in Table Ⅳ.2 present the estimate results pertinent to each subsample. We find that the results are statistically significant and negative for all four accident types, with only a lower significance of 6% for the death or disability risk of the endowment insurance. Notably, medication for disease has the largest reduction of the accident probability among all risk types of the endowment insurance. Therefore, the risk screening effect is unconditional on most risk types insured by the investigated products.

Table Ⅳ.1: Baseline Results of OLS Estimates

Variables	(1) AC	(2) AC	(3) AC	(4) AC
Panel A. Term Life Insurance				
$digital$	-0.0022***	-0.0024***	-0.0019**	-0.0018**
	(-2.63)	(-2.62)	(-2.32)	(-2.51)
Observations	97,495	97,068	93,623	97,458
Adj. R-squared	0.032	0.032	0.018	0.029
Offline accident rate	0.0029	0.0029	0.0020	0.0018
Panel B. Endowment Insurance				
$digital$	-0.0036***	-0.0036***	-0.0029***	-0.0032***
	(-9.53)	(-9.51)	(-7.82)	(-8.50)
Observations	928,291	919,368	672,562	927,584
Adj. R-squared	0.027	0.027	0.029	0.025
Offline accident rate	0.0139	0.0140	0.0137	0.0132
Panel C. Disease Insurance				
$digital$	-0.0034***	-0.0035***	-0.0071***	-0.0018**
	(-3.43)	(-3.48)	(-3.95)	(-2.14)
Observations	53,296	53,232	23,343	53,186
Adj. R-squared	0.052	0.052	0.150	0.037
Offline accident rate	0.0079	0.0079	0.0109	0.0059
Controls	Y	Y	Y	Y
Fixed Effects				
Prefecture-Year	Y	Y	Y	Y
Month	Y	Y	Y	Y

Day-in-Month	Y	Y	Y	Y
Day-in-Week	Y	Y	Y	Y

Note: Columns 1 employs the full sample, while Columns 2, 3 and 4 excludes cancelled, non-self-insured and claim-rejected policies respectively. For each regression sample, the accident ratio calculates the percentage of claims among offline policies. Throughout the paper, robust t-statistics in parentheses with standard errors clustered at the prefecture level and *** p<0.01, ** p<0.05, * p<0.1, which is not repeated in the table notes hereafter.

Table Ⅳ.2: Estimated Effects by Accident Type for Endowment and Disease Insurance

Variables	(1) **Death or Total Disability**	(2) **Medication for Accidents**	(3) **Medication for Disease**	(4) **Severe Disease**
Panel B. Endowment Insurance				
digital	-0.0002*	-0.0006**	-0.0030***	-0.0001***
	(-1.73)	(-2.03)	(-15.06)	(-6.06)
Observations	916,105	924,065	918,750	915,434
Adj. R-squared	0.005	0.020	0.014	0.001
Panel C. Disease Insurance				
digital	-0.0006**			-0.0029***
	(-2.03)			(-3.05)
Observations	52,911			53,259
Adj. R-squared	0.027			0.047
Controls	Y	Y	Y	Y
Fixed Effects				
Prefecture-Year	Y	Y	Y	Y
Month	Y	Y	Y	Y
Day-in-Month	Y	Y	Y	Y
Day-in-Week	Y	Y	Y	Y

Note: for each column, the policies claiming the specific accident type and unclaimed policies consist the sample. Since the term life insurance only insures one accident type - the death or disability risk, it is not presented here.

4.2 Difference-in-Difference Estimates

Table Ⅳ.3 reports the DID estimate result in Column 1, confirming the risk screening effect of digital distribution. It shows that the average accident probability of total policies of the term life insurance reduces by 0.24 percent points after the introduction of digital distribution. This result is very close to the corresponding OLS estimate result (in Panel A, Column 1 of Table Ⅳ.1).

Table Ⅳ.3: DID Estimates of the Risk screening effect

Variable	DID estimate	Falsification Tests	
	(1)	(2)	(3)
$\boldsymbol{Treat_p} \times \boldsymbol{\mathcal{L}_t}$	-0.0024**	-0.0005	0.0029
	(-2.27)	(-0.87)	(1.02)

Observations	41,295	27,321	26,462
Adj. R-squared	0.037	0.016	0.015
Controls	Y	Y	Y
Fixed Effects			
Prefecture	Y	Y	Y
Product	Y	Y	Y
Date	Y	Y	Y

Note: column 1 uses the policies of treatment and control products in 2018, column 2 uses the policies of treatment product in 2017 and the policies of control product in 2018, column 3 uses the policies of treatment product in 2019 and the policies of control product in 2018.

In the remainder of this subsection, we provide several robustness checks. A potential concern of our DID specification is pre-trends. If the average policy risk of the treatment product presented a decreasing trend relative to the control product before the digital distribution introduction, the identified effect would be endogenous. To address this concern, we use a similar DID specification analyzing monthly leads and lags as follows.

$$\boldsymbol{LnAR}_{p,r,t} = \alpha + \sum_{\substack{m-j=8 \\ m=1,2,3\ldots7}} \tau_{-j}\boldsymbol{Treat}_p \times \boldsymbol{mpre}_{-j} + \sum_{\substack{m-k=8 \\ m=8,9,10,11,12}} \tau_k \boldsymbol{Treat}_p \times \boldsymbol{mpost}_k + \boldsymbol{X''}_{p,r,t}\boldsymbol{\Gamma} + \boldsymbol{X''}_t\boldsymbol{\Phi} + \boldsymbol{X'}_{p,r}\boldsymbol{\Omega} + \varepsilon_{r,t} \quad (8)$$

Where $\boldsymbol{mpre}_{-j}$ and $\boldsymbol{mpost}_k$ are dummies defined as 1 for j months before and k months post August respectively. Fixed effects and controls are identical with Equation (7). The point estimates along with their confidence intervals for the coefficients are illustrated in Figure A1, Appendix A. Consistent with the parallel pre-trends assumption, we find that none of the monthly leads has a significant effect on the average policy risk; Rather, the coefficients of all four monthly lags fall sharply to be negative.

We also preform two falsification tests by replacing the sample with the policies of both products sold in 2017 and 2019 - the year prior and next to the introduction year. Under the same DID specification, if there was a natural decrease in average policy risk of the treatment product on August 15th in each year, the results would still be significant in these falsification tests. The insignificant results presented in Columns (2) and (3) of Table Ⅳ.3 falsify this suspicion.

5 Consequences of the Risk Screening Effect

This section examines the economic consequences born by the difference in unobserved risk between digital and offline channels.

5.1 Information Asymmetry

As indicated in our conceptual framework, a direct consequence of the lower unobserved risk for policies purchased via the digital distribution channel is the weaker effect of information asymmetry. That is, compared to digital channel consumers, offline consumers have stronger motivation to employ the private risk information advantage to behave in a way detrimental to the insurer.

A standard test for the effect of information asymmetry is the risk-coverage relation (Cohen and Siegelman 2010). The principle is intuitive that information asymmetry incentivizes high-risk consumers to buy more insurance (adverse selection), or higher insurance coverage results in less cautious behavior (moral hazard), both forging a positive risk-coverage relation. In this spirit, we add an interaction between $\boldsymbol{digital}$ and logarithmic policy insured amount $\boldsymbol{cov}$ into Equation (6) to capture the difference in risk-coverage relation between digital and offline channels, keeping the same set of controls and fixed effects as in the Main Results. The results, reported in Column 1 of Table Ⅴ.1, are all negative and only insignificant for the term life insurance. This demonstrates that overall, digital distribution has a weaker positive risk-coverage relation and thus weaker information asymmetry than offline distribution.

We also provide a piece of descriptive evidence in the bottom two rows of each panel. Since the waiting period[13] is widely adopted in practice to discourage insurance applications with known diseases or imminent risks, the claims in the waiting period are typical adverse selection behaviors under information asymmetry. As shown in Table Ⅴ.1, the percentage of the claims during the waiting period is higher for offline distribution than for digital distribution, which holds for all three products and indicates lower adverse selection of digital

[13] The waiting period, also known as the observation period, is a short period (usually 30 to 180 days) after the policy takes effect, during which the insurer is not liable and only refunds the premium paid in the event of an insured accident. Only after the waiting period does the insurer assume the agreed liability for the occurrence of an insured accident. This can effectively reduce adverse selection such as applications with imminent risks.

distribution. Therefore, this descriptive evidence also confirms the weaker information asymmetry for digital distribution.

Table Ⅴ.1: Examining Consequences of the Difference in Unobserved Risk

Variables	(1) AC	(2) $LnInd$	(3) $LnInd$	(4) $LnCostRatio$
Panel A. Term Life Insurance				
$digital \times cov$	-0.0003			
	(-0.74)			
$digital$	0.0012	-0.0209***		-0.0010
	(0.815)	(-2.65)		(-1.49)
Observations		97,495		95,094
Adj. R-squared		0.028		0.012
Digital Channel Claims in Waiting Period	0.0000			
Offline Channel Claims in Waiting Period	0.0132			
Panel B. Endowment Insurance				
$digital \times cov$	-0.0011***			
	(-10.09)			
$digital$	-0.0001	-0.0257***	-0.1473***	-0.0018***
	(-0.45)	(-9.73)	(-3.31)	(-7.52)
Observations	928,291	928,291	12,203	928,291
Adj. R-squared	0.026	0.024	0.398	0.005
Digital Channel Claims in Waiting Period	0.0000			
Offline Channel Claims in Waiting Period	0.0015			
Panel C. Disease Insurance				
$digital \times cov$	-0.0103***			
	(-5.05)			
$digital$	0.1310***	-0.0202**	-0.2371*	-0.0066**
	(4.94)	(-1.98)	(-1.80)	(-2.00)
Observations	53,296	53,296	275	53,296
Adj. R-squared	0.054	0.035	0.902	0.039
Digital Channel Claims in Waiting Period	0.0417			
Offline Channel Claims in Waiting Period	0.0556			
Controls	Y	Y	Y	Y
Fixed Effects				
Prefecture-Year/Province-Year	Y	Y	Y	Y
Month	Y	Y	Y	Y
Day-in-Month	Y	Y	Y	Y
Day-in-Week	Y	Y	Y	Y
Accident Type	N	N	Y	N

Note: for each product, columns 1, 2 and 4 use the full sample, column 3 uses only indemnified policies. In column 3, due to limited sample size, we control the Province-Year fixed effects instead of Prefecture-Year fixed effects. While other columns control Prefecture-Year fixed effects.

5.2 Indemnity and Profitability

The other direct consequence elicited by the difference in unobserved risk is the difference in indemnity and profitability across channels. We first present evidence on indemnity by replacing the outcome variable with $\boldsymbol{LnInd}$, which measures indemnity in natural logarithm as follows for each policy:

$$\boldsymbol{LnInd} = \ln(1 + Indemnity)$$

In this form, the outcome variable equals zero for uncompensated policies. By redoing the same OLS analyses as in Main Results, the results are shown in Column 2 of Table Ⅴ.1. As expected, they are significant and negative for all three panels, implying an average decrease in indemnity deriving from digital distribution by 2.1% (=$e^{-0.0209}$-1) for the term life insurance, by 2.6% (=$e^{-0.0257}$-1) for the endowment insurance and by 2.0% (=$e^{-0.0202}$ -1) for the disease insurance.

In Column 3 of Table Ⅴ.1, we further limit the samples only to compensated policies and redo the same analyses with additional dummy controls of accident types for the endowment and disease insurance[14]. They show that even among compensated policies, the average indemnity for policies of digital distribution is still lower than offline policies by 13.7% (=$e^{-0.1473}$-1) for the endowment insurance and by 21.1% (=$e^{-0.2371}$-1) for the disease insurance. These estimates strengthen the evidence supporting the risk screening effect.

To test the difference in profitability between digital and offline channels, we use the logarithmic loss ratio $\boldsymbol{LnLossRatio}$ as the outcome variable, which is calculated for each policy by:

$$\boldsymbol{LnLossRatio} = \ln(1 + \frac{Indemnity}{Premiums\ Paid})$$

where $Premiums\ Paid$ refers to the accumulated premiums received by the insurer. The higher loss ratio, the lower profitability. The results of the OLS estimates, reported in Column 4 of Table Ⅴ.1, are significantly negative for the endowment and disease insurance, except

[14] Here the data of the term life insurance is not used for regression due to very few compensated policies (only 121 observations).

the term life insurance has a negative coefficient with only 14% significance. Overall, our evidence shows lower indemnity and higher profitability of digital distribution as a result of the risk screening effect.

6 Understanding What Drives the Risk Screening Effect

In this section, we examine the decomposed sources suggested by the conceptual framework. Our first two examinations focus on *channel capability* by employing two channel features - the reduced search cost of digital distribution and biased services of offline agents. The final examination focuses on *channel preference* by studying how advanced education correlates with digital divide and policy risk.

6.1 Channel Capability: Search Cost

Search cost occurs everywhere in an economy and sometimes causes very substantial welfare loss of consumers (Jolivet and Turon 2019). As implied in the settings, a remarkable advantage of digital insurance distribution is the lower search cost compared with offline insurance distribution. Taking this channel feature into account, we first propose a simple model of insurance purchase decisions to formalize the search cost mechanism and then carry out empirical examinations.

6.1.1 A Search-Based Model on Risk Screening

The model is based on a simple two-period intertemporal consumption framework and captures two key frictions in the insurance market: search cost and premium uncertainty. Suppose that the loading factor λ follows a distribution Ψ across all products which is known to consumers before searching. To perceive the actual loading factor of an insurance product, a consumer has to spend a search cost s in the form of disutility to access the insurance product information at period 0. Given the pricing principle of expected risk based on observed risk characteristics, a Φ-type consumer would pay the premium $\lambda\Phi L$ at period 0 for the fixed coverage L at period 1[15] with a random unobserved risk q^{RA}. We assume that the unobserved risks of the Φ-type consumers follow a common distribution.

Since the loss rate is not the focus in this paper, it is assumed equal to 1 for simplicity[16]. For a representative consumer of the Φ-type, the utility gain of insurance search U is made up of the difference in consumption utility $u(\cdot)$ between the insured and uninsured cases, given by

$$\mathrm{U} = u(y - \lambda\Phi L) - u(y) + \alpha\Phi q^{RA}u(L) \tag{9}$$

Where $\alpha \in (0,1)$ is a constant subjective discount factor, y is the constant income in each period and consumers are risk averse: $u' > 0$ and $u'' < 0$. Then, the expectation of search utility gain is given by

$$E(\mathrm{U}) = \int_{\lambda} max\{\mathrm{U}, 0\}\, d\Psi(\lambda) \tag{10}$$

Prior to receiving product information, the consumer does not know the actual loading factor (that is, innocent of unit premium) and has to weigh the expected search utility gain against the disutility of offline search cost to decide whether to search. Thus, the search rule can be written as[17]

$$E(\mathrm{U}) > s \tag{11}$$

Similarly, the purchase rule after receiving product information and knowing the actual loading factor λ^r is

$$\mathrm{U}(q^{RA} \mid \lambda = \lambda^r) > 0 \tag{12}$$

A purchase decision is made only when both rules are met. Therefore, the insurance demand (the probability of purchasing insurance) is given by

$$\mathrm{Prob}(\varphi(s, q^{RA}) > 0) \tag{13}$$

where $\varphi(s, q^{RA}) = \min\{E(\mathrm{U}) - s, \mathrm{U}(q^{RA} \mid \lambda = \lambda^r)\}$.

---

[15] The effective insurance period usually lags behind the time of the first premium payment, because most insurance products set a waiting period during which the insured cannot be compensated but can only be refunded premiums for an accident.

[16] This is without loss of generality as the following proof does not rely on variations in the loss rate.

[17] Equation (11) can also be interpreted in another way: people will search only if the expected maximum utility between buying insurance and not buying insurance after searching is higher than the utility of not searching.

PROPOSITION: *If digital distribution has a search cost of insurance smaller than offline distribution,*

(i) *the average unobserved risk of digital channel policies should be lower than that of offline channel policies;*

(ii) *the introduction of digital distribution should lower the average risk of total purchased policies.*

PROOF: By partially differentiating with respect to s and q, it is easy to see that φ decreases with s and U increases with q. Then we can derive that $\varphi_q^{-1}(s)$, the inverse function of $\varphi(s, q^{RA})$ with respect to q, should increase with s. By rewriting Equation (13), the expected policy risk of this income group is

$$E\left(q^{RA} \mid q^{RA} > \varphi_q^{-1}(s)\right) \tag{14}$$

Obviously, Equation (14) increases with s and this shows effect *(i)*. Similarly, when introducing digital distribution, the search costs of the consumers who are able and willing to use both digital and offline channels also reduce and thus the average unobserved risk of the enrollees from them decreases, which shows effect *(ii)*.

PROPOSITION describes the effect of the reduced search cost of digital distribution on the average policy risk of enrollees. From the proof, it can be seen that high-risk consumers tend to be more motivated to purchase insurance, regardless of search costs. In contrast, low-risk consumers are relatively less motivated to purchase insurance and thus more sensitive to the search cost reduction. Taken together, the insurance demand growth due to the reduced search costs of digital distribution is higher for low-risk consumers than for high-risk consumers, leading to the risk screening effect.

6.1.2 Testing for the Search Cost Mechanism

To test the above prediction that the reduced search costs of digital distribution compared with offline distribution leads to the screening of policy risk, we examine whether the risk screening effect diminishes with lower offline search costs. Offline search costs are measured in three ways: a) the population weighted average distance to the nearest insurer branch, b) the correlation between the population distribution and insurer branch distribution, and c) daily rainfall.

We start with an intuitive measure based on distance to the insurer branch. For each prefecture, the offline search cost of insurance products is indexed as the distance to the nearest branch of the investigated insurer per capita, as shown in Equation (15). Specifically, for each grid cell in each prefecture, we weight the distance from the grid cell center to its nearest insurer branch by the grid cell population.

$$\boldsymbol{SC}_r = \frac{\sum_{k\in r} \boldsymbol{dist}_{r,k} \times \boldsymbol{pop}_{r,k}}{\sum_{k\in r} \boldsymbol{pop}_{r,k}} \tag{15}$$

Where for prefecture $r$and grid cell $k \in r$, $\boldsymbol{SC}$ denotes the offline search cost per capita, $\boldsymbol{dist}$ denotes the distance from the grid cell center to the nearest branch and $\boldsymbol{pop}$, the grid cell population. Given that there may be branches of other life insurers in the vicinity and consumers tend to search for multiple insurers' products and shop around before making the final purchase decision, we further recalculate $\boldsymbol{SC}$ with the distance to the nearest branch of local life insurers and the average distance to the top three nearest branches of local life insurers. Hence, we construct three measures of $\boldsymbol{SC}$ with different distances to branches of insurance companies.

We confirm the search cost mechanism by adding into equation (6) an interaction between $\boldsymbol{digital}$ and $\boldsymbol{SC}$. The results are reported in Columns 1 to 3 of Table Ⅵ.1. An average population weighted distance to the nearest insurer's branch, the nearest and top three nearest branch of local life insurers enlarges the risk screening effect of digital distribution – the risk probability reduction by 0.12 (=-0.0011×1.0714×100), 0.06 (=-0.0019×0.3393×100) and 0.06 (=-0.0014×0.4394×100) per cent points for the term life insurance, while by 0.09 (=-0.0008×1.0714×100), 0.13 (=-0.0039×0.3393×100) and 0.15 (=-0.0034×0.4394×100) per cent points for the endowment insurance. To understand the magnitude of the influence of offline search cost, we take the risk screening effect presented in Column 1 of Table Ⅳ.1 as a basis. Simple calculations show that the search cost mechanism can explain 27% to 55% of the risk screening effect for the term life insurance and 15% to 42% of the risk screening effect for the endowment insurance. In this way we show that the explanatory power of search cost is considerable.

Using distance to measure offline search cost could incur endogeneity with the insurer's preference on branch locations. For instance, a rational insurer is probably more inclined to set up more branches in higher-income residential districts. To alleviate this concern, by following Roca and Puga (2017), our second measure of offline search cost adopts the correlation between the population distribution and insurer branch distribution. This measure exploits the covariance of $\boldsymbol{dist}_{r,k}$ and $\boldsymbol{pop}_{r,k}$, as shown in Equation (16).

$$\boldsymbol{SC}_r = \frac{Covariance(\boldsymbol{dist}_{r,k}, \boldsymbol{pop}_{r,k})}{\boldsymbol{AD}_r \times \boldsymbol{AP}_r} \tag{16}$$

Where the numerator is a covariance, $\boldsymbol{AD}_r$ denotes the average of $\boldsymbol{dist}_{r,k}$ and $\boldsymbol{AP}_r$ denotes the average of $\boldsymbol{pop}_{r,k}$ across grid-cells of the prefecture. This measure, illustrates how the distance to the insurer branch changes with the population density across grid-cells inside a prefecture, capturing the degree to which the insurer branches are located in more populated areas. Specifically, if a prefecture presents such a landscape across grid-cells that the higher population, the more insurer branches, then the distance to the insurer branch should strongly negatively correlate with the population density in that prefecture; on the contrary, the less negative correlation, the higher distribution bias between the population and insurer branch, indicating higher offline search cost. Therefore, the smaller $\boldsymbol{SC}$ in Equation (16), the lower offline search cost. Of note is that this measure standardizes $\boldsymbol{dist}_{r,k}$ and $\boldsymbol{pop}_{r,k}$ by dividing their averages, excluding the influence of endogenous factors associated with prefectures.

The interaction between $\boldsymbol{digital}$ and $\boldsymbol{SC}$ is of our interest. The results, reported in Columns 4 to 6 in Table Ⅵ.1, are significant and negative for both products. They show that the less negative correlation between the population distribution and insurer branch distribution (the higher offline search cost) leads to a larger risk screening effect.

Our third more exogenous measure is daily rainfall. There are two reasons for indexing offline search cost with daily rainfall. Firstly, empirical evidence shows that in rainy days, catching a taxi becomes more difficult and taxi fees increase due to surged demand (Faber 2015; Brodeur and Nield 2018). Commercial taxis are an important transportation means in

China where the per capita car ownership is low (around one third of UK and one fourth of USA[18]). Secondly, travel is usually riskier and slower in rainy days (Faber 2015). Specifically, we use the median daily rainfall per month as a more exogeneous measure of offline search cost, which is added into the regression along with its interaction with $\boldsymbol{digital}$. Column 7 in Table Ⅵ.1 presents the results. It shows that for one millimeter increase in the median daily rainfall per month, the magnitude of the risk probability reduction of digital distribution increments by 0.08 percent points (28% of the offline accident rate in Column 1, Table Ⅳ.1) for the term life insurance and by 0.17 percent points (12% of the offline accident rate in Column 1, Table Ⅳ.1) for the endowment insurance.

So far, regressions with different measures of offline search cost all verify that search cost mediates the relationship between digital distribution and policy risk, serving as a channel for the risk screening effect.

Table Ⅵ.1: Examining the Search Cost Mechanism

	Distance to Insurer Branch			Correlation between Distributions of Population and Insurer Branch			Rainfall
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	$\boldsymbol{AC}$	$\boldsymbol{AC}$	$\boldsymbol{AC}$	$\boldsymbol{AC}$	$\boldsymbol{AC}$	$\boldsymbol{AC}$	$\boldsymbol{AC}$
Panel A. Term Life Insurance							
$\boldsymbol{digital} \times \boldsymbol{SC}$	-0.0011*	-0.0019**	-0.0014**	-0.0080*	-0.0092**	-0.0087**	
	(-1.75)	(-2.06)	(-2.17)	(-1.77)	(-1.98)	(-2.02)	
$\boldsymbol{digital} \times \boldsymbol{rain}$							-0.0008*
							(-1.74)
Observations	97,495	97,495	97,495	97,495	97,495	97,495	97,495
Adj. R-squared	0.030	0.030	0.030	0.018	0.018	0.018	0.040
Panel B. Endowment Insurance							
$\boldsymbol{digital} \times \boldsymbol{SC}$	-0.0008*	-0.0039**	-0.0034**	-0.0080***	-0.0114***	-0.0113***	
	(-1.91)	(-2.25)	(-2.36)	(-4.31)	(-5.19)	(-5.33)	
$\boldsymbol{digital} \times \boldsymbol{rain}$							-0.0017***
							(-4.04)
Observations	928,291	928,291	928,291	928,291	928,291	928,291	928,291
Adj. R-squared	0.025	0.025	0.025	0.025	0.025	0.025	0.025
Controls	Y	Y	Y	Y	Y	Y	Y
Fixed Effects							
Prefecture-Year	Y	Y	Y	Y	Y	Y	Y
Month	Y	Y	Y	Y	Y	Y	Y
Day-in-Month	Y	Y	Y	Y	Y	Y	Y

[18] According to the data from WIKIPEDIA (https://en.wikipedia.org/wiki/List_of_countries_by_vehicles_per_capita), China has 219, USA has 868 and UK has 575 road motor vehicles per 1,000 inhabitants in 2022.

Day-in-Week	Y	Y	Y	Y	Y	Y	Y

Note: The search cost is indexed by grid-cell population weighting the distance to the nearest insurer branch for Columns 1 and 4, the distance to the nearest branch of local life insurers for Columns 2 and 5, the average distance to the top three nearest branches of local life insurers for Columns 3 and 6, respectively. Each Column uses an interaction between the search cost and the independent variable to confirm the mechanism based on the full sample.

6.2 Channel Capability: Offline Underwriting Service Quality

Another channel feature directly linked to the risk screening of digital distribution is the underwriting service quality of offline agents. Most offline agents in China life insurance industry are employed by insurers and are not independent. It is probable that they do not strictly comply with underwriting rules in pursuit of more sales and commission, leading to underwrite more high-risk consumers on average. This biased underwriting service is particularly strong for the straight commission institution without base pay (Cummins and Doherty 2006; Hilliard et al. 2013), such as in the life insurance industry of China. Obviously, If this is true, offline policyholders' claims would be more likely to be rejected than those purchasing via the digital distribution channel. Based on this tenet, we investigate the effect of digital distribution on claim rejection.

We use the sample of only claimed policies and replace the dependent variable with the dummy of whether the claim was rejected in the baseline OLS specification. Accident types are also fixed. The results are presented in Column 1 of Table Ⅵ.2, showing that the rejection probability of the offline policies claiming indemnity is 1.42 percent points higher than that of digital channel policies for the endowment insurance. For disease insurance, the coefficient of interest is also negative but insignificant possibly due to the limited subsample size.

We also exclude the claims rejected for the reasons unrelated to ineligibility and redo the same analysis. The results are presented in Columns 2 and 4 of Table Ⅵ.2. We find that they are qualitatively consistent but have a reduction in the magnitude.

We also provide a piece of direct descriptive evidence in the bottom two rows of each panel, by calculating the percentage of rejections due to ineligibility among all claim rejections. As shown, this percentage is lower for the digital distribution channel than for the offline distribution channel, which holds for both products. Overall, these results suggest low-quality underwriting services provided by offline agents, resulting in underwriting more ineligible consumers with high unobserved risk.

The above results seem to contradict Venezia et al.'s (1999) argument that independent agents provide a higher quality service by helping claim for compensation on behalf of policyholders compared with direct underwriters such as digital distribution. There are two explanations to this contradiction. First, relative to independent agents, employed agents act more on behalf of the insurer's interest and less on behalf of policyholders' interests. Second, information asymmetry and inconsistent interests between consumers and offline agents may lead to low-quality services (Eckardt and Räthke 2010; Focht et al. 2013), such as misleading sales.

Table Ⅵ.2: The Difference in Claim Rejection Between Digital and Offline Distribution Channels

	Endowment Insurance		Disease Insurance	
	(1)	(2)	(3)	(4)
Variables	***Rej***	***Rej***	***Rej***	***Rej***
digital	-0.0142***	-0.0066***	-0.1110	-0.0056
	(-2.79)	(-2.68)	(-1.63)	(-0.11)
Observations	13,128	12,659	410	266
Adj. R-squared	0.054	0.049	0.130	0.002
Rej. for Ineligibility in Digital Channels	0.2143		0.1200	
Rej. for Ineligibility in Offline Channels	0.3640		0.2222	
Controls	Y	Y	Y	Y
Fixed Effects				
Province-Year /Prefecture-Year	Y	Y	Y	Y
Month	Y	Y	Y	Y
Day-in-Month	Y	Y	Y	Y
Day-in-Week	Y	Y	Y	Y
Accident Type	Y	Y	Y	Y

Note: for the corresponding product, columns 1 and 3 both use all claimed policies, column 2 and 4 both use claimed policies the reasons related to ineligibility. Here, term life insurance is not used due to its too tiny size of claimed policies. For the same reason, Columns 3 and 4 only control Province-Year fixed effects. While other columns control Prefecture-Year fixed effects.

6.3 Channel Preference

Our third and last examination on the mechanism is direct at *channel preference*. Although there are many factors of the ability and acceptance to use digital technology, we choose education level as the investigated factor of digital divide for two reasons. First, education level is a typical risk characteristic not adjusted into the unit premium by the insurer. Second, there has been much literature supporting the positive relationship between internet use and advanced education (Hargittai 2002; Wei and Hindman, 2011; Cruz et al., 2016).

The logic of our test is to examine whether advanced education positively correlates with the choice of the digital distribution channel while negatively correlates with unobserved risk. Empirical specifications are presented below

$$\boldsymbol{digital}_{i,r,t} = \alpha + \beta \boldsymbol{edu}_{i,r,t} + \theta \boldsymbol{D}_t + \boldsymbol{X'}_{i,r,t}\boldsymbol{\Gamma} + \boldsymbol{X'}_{r,t}\boldsymbol{\Omega} + \varepsilon_{i,r,t} \tag{17}$$

$$\boldsymbol{AC}_{i,r,t} = \alpha + \beta \boldsymbol{edu}_{i,r,t} + \theta \boldsymbol{D}_t + \boldsymbol{X'}_{i,r,t}\boldsymbol{\Gamma} + \boldsymbol{X'}_{r,t}\boldsymbol{\Omega} + \varepsilon_{i,r,t} \tag{18}$$

Where $\boldsymbol{edu}_{i,r,t}$ is the dummy of advanced education. The controls (except education) and fixed effects are the same as the specification of equation (6). We limit the sample to only self-insured policies to ensure that $\boldsymbol{edu}_{i,r,t}$ measures the advanced education of the insured.

Prior to the examinations, we start with regressing unit premium on $\boldsymbol{edu}_{i,r,t}$ to show that education levels are not adjusted into the unit premium, keeping other controls (except education and unit premium) and fixed effects the same as in equation (6). The results, reported in Column 1 of Table Ⅵ.3, are insignificant but have very large R-squared for all three products, indicating that unit premiums do not take into account education levels and are largely explained by gender and age.

The results of equations (17) and (18) are presented in Columns (2) and (3), Table Ⅵ.3, showing significant relationships between advanced education and channel choice as well as policy risk for all three columns. Specifically, the advanced education increases the probability of choosing the digital distribution channel by 1.00, 25.38 and 4.86 percent points while decreases the risk probability by 0.07, 0.43 and 0.13 percent points for the term life, endowment and disease insurance respectively. Taken together, we show that consumers with advanced education prefer to use the digital distribution channel while have lower policy risk than those with poorer education levels.

Table Ⅵ.3: Relationships between Unit Premium, Channel Choice, Risk and Advanced Education

	(1)	(2)	(3)

Variables	$unit\ premium$	$digital$	AC
Panel A. Term Life Insurance			
edu	0.0007	0.0100***	-0.0007***
	(0.277)	(3.19)	(-2.71)
Observations	93,623	93,623	93,623
Adj. R-squared	0.876	0.641	0.018
Panel B. Endowment Insurance			
edu	0.0090	0.2538***	-0.0043***
	(1.35)	(10.04)	(-13.07)
Observations	672,562	672,562	672,562
Adj. R-squared	0.654	0.163	0.028
Panel C. Disease Insurance			
edu	-0.00002	0.0486***	-0.0013*
	(-0.87)	(5.70)	(-1.70)
Observations	23,343	23,343	23,343
Adj. R-squared	0.940	0.280	0.052
Controls	Y	Y	Y
Fixed Effects			
Prefecture-Year	Y	Y	Y
Month	Y	Y	Y
Day-in-Month	Y	Y	Y
Day-in-Week	Y	Y	Y

Note: for each product, all columns use only self-insured policies. Column 1 uses unit premium as the outcome variable, keeping other controls (except education and unit premium) and fixed effects the same as in equation (6). Column 2 and 3 respectively use the specification of equations (17) and (18).

7 Who Contributes to the Risk screening effect?

Customers of the digital distribution channel are (i) new consumers who would not have purchased insurance and (ii) consumers who switched from offline. Correspondingly, the risk screening effect should originate from the risk differences of two consumer sources: the average risk decrease for consumers in the digital distribution channels and the average risk increase for retained offline consumers. To make it more explicit, let $\boldsymbol{R}_0$ denote the average risk of offline consumers in the counter-fact where there exists no digital distribution, $\boldsymbol{R}_d$ denote the average policy risk of digital distribution channel consumers and $\boldsymbol{R}_f$, the average policy risk of offline consumers in reality. The risk screening effect $\boldsymbol{RD}$ reflecting the policy risk difference between the offline and digital distribution channel consumers can be decomposed as

$$\boldsymbol{RD} = \boldsymbol{R}_d - \boldsymbol{R}_f = (\boldsymbol{R}_d - \boldsymbol{R}_0) + (\boldsymbol{R}_0 - \boldsymbol{R}_f) \tag{19}$$

On the RHS, the left expression is the average risk decrease for digital channel consumers and the right expression, the average risk increase for offline consumers.

Next, we further decompose the average risk decrease of the digital distribution channel $(\boldsymbol{R}_d - \boldsymbol{R}_0)$ by new consumers and the consumers who switched from offline. We notate that relative to the counter-fact, the actual number of total policies increases by $(\boldsymbol{K} - 1)$ while the actual number of offline policies decreases by $(1 - \boldsymbol{K}_f)$. Through calculations (details seen in Appendix B), the contribution in ratio to $(\boldsymbol{R}_d - \boldsymbol{R}_0)$ from the consumers who switched from offline to the digital distribution channel is

$$\frac{\boldsymbol{K}_f}{\boldsymbol{K} - \boldsymbol{K}_f} \cdot \frac{\boldsymbol{R}_0 - \boldsymbol{R}_f}{\boldsymbol{R}_d - \boldsymbol{R}_0} \tag{20}$$

Equation (20) shows that the offline crowded-out consumers contribute to $(\boldsymbol{R}_d - \boldsymbol{R}_0)$ through both the proportion of offline consumers and the risk increase of offline policies relative to the counter fact.

So far, we have decomposed the risk screening effect into a risk increase of offline consumers and risk decreases of new consumers and the consumers who switched from offline. Next, to quantify the decomposition, we further exploit the above DID framework based on the term life insurance.

Estimating $(\boldsymbol{R}_d - \boldsymbol{R}_0)$: we retain the control group, but the treatment group consists of two components: the offline policies before the introduction date and the digital distribution channel policies after the introduction date. In this way, we can capture the difference in average risk between the offline policies and the digital distribution channel policies. Using the same specification as in Equation (7), π captures the average policy risk difference between digital distribution and the counter-fact, which corresponds to $(\boldsymbol{R}_d - \boldsymbol{R}_0)$. The result, presented in Column (1) of Table Ⅶ.1, shows that the introduction of the digital distribution channel leads to an average decrease of 0.26 percent points (92% of the offline accident ratio in Column 1, Table Ⅳ.1) in the accident probability.

Estimating $\boldsymbol{RD}$: we further estimate the risk screening effect using the policies of the treatment product purchased in 2018 but after the introduction date of digital distribution, keeping the same controls and fixed effects as in Column 1. The result, presented in Column 2 of Table Ⅶ.1, shows that the risk screening effect on policy risk is as substantial as 0.31 percent points for the current sample. We find that most of the risk screening effect derives from the digital distribution channel consumers. As shown in the bottom row of Column 2 in Table Ⅵ.1, the risk decrease of the digital distribution channel consumers contributes to over 74% of the risk screening effect of the disease insurance.

Estimating $\boldsymbol{K_f}$ and $\boldsymbol{K}$: we regress Equation (7) but replace the dependent variable with the logarithmic number of purchased policies in each date and each prefecture. We use the same sample as in Column 1 to estimate $\boldsymbol{K_f}$ while using the full sample of the policies of both control and treatment products to estimate $\boldsymbol{K}$. Columns 3 and 4 in Table Ⅶ.1 report the results for estimating $\boldsymbol{K_f}$ and $\boldsymbol{K}$ respectively. The introduction of digital distribution contributed to 140% (=$e^{0.8735}$-1) growth in the total number of policies while a slight drop in the number of offline policies by 4% (=$e^{-0.0391}$-1). From Equation (20), calculations show that nearly 77% (=1-23.22%) of the average risk decrease in the policies of digital distribution are attributed to the new consumers. This is reasonable because as analyzed above, the magnitude of the average risk increase of offline policies is small compared to the average risk decrease of the policies of digital distribution. One can conclude that most of the risk screening effect derives from the attracted new consumers with lower risk.

Table Ⅶ.1: Decomposed Consumer Contributions to the Risk screening effect

	Estimating $(\boldsymbol{R_d}-\boldsymbol{R_0})$	*Estimating* $\boldsymbol{RD}$	*Estimating* $\boldsymbol{K_f}$	*Estimating* $\boldsymbol{K}$
	(1)	(2)	(3)	(4)
$\boldsymbol{Treat} \times \boldsymbol{\mathcal{L}_t}$	-0.0026**		-0.0391*	0.8735***
	(-2.23)		(-1.89)	(14.92)
$\boldsymbol{digital}$		-0.0035**		
		(-2.00)		
Observations	41,295	50,962	27,041	41,295
Adj. R-squared	0.039	0.010	0.478	0.423
$(\boldsymbol{R_d}-\boldsymbol{R_0})/\boldsymbol{RD}$		74.29%		
$\dfrac{\boldsymbol{K_f}}{\boldsymbol{K}-\boldsymbol{K_f}} \cdot \dfrac{\boldsymbol{R_0}-\boldsymbol{R_f}}{\boldsymbol{R_d}-\boldsymbol{R_0}}$				23.22%
Controls	Y	Y	Y	Y
Fixed Effects				
Prefecture-Year	Y	Y	Y	Y
Product	Y	N	Y	Y
Date	Y	N	Y	Y
Month	Y	Y	Y	Y
Day-in-Month	Y	Y	Y	Y
Day-in-Week	Y	Y	Y	Y

Note: Column 1 uses the sample of the control product policies, offline policies of the treatment product before the introduction date and digital channel policies of the treatment product after the introduction date in 2018; Column 2, the OLS estimate, uses the policies of the treatment product purchased post the introduction date in 2018; column 3 uses the sample of the control product policies and offline policies of the treatment product in 2018; column 5 uses all policies of the treatment product and control product in 2018. The second row from the bottom of column

2 is calculated by 0.0026/0.0035; The last row of column 4 is calculated by $[e^{-0.0391}/(e^{0.8735} - e^{-0.0391})] \times [(0.0035 - 0.0026)/0.0026]$.

8 Conclusion

Using a unique data set on the purchased policies of the term life, endowment and disease insurance products sold on both digital and offline distribution channels, we show that digital distribution attracts more low-risk applicants than traditional offline distribution, leading to an advantageous screening of unobserved policy risk. This risk screening effect generates important economic consequences by reducing information asymmetry, lowering the average indemnity and increasing the profitability of digital distribution.

We theoretically and empirically show three mechanisms of the risk screening effect from the roles of *channel capability* and *channel preference*. First, the advantageous channel feature of digital distribution, such as reduced search costs, has higher marginal incentives to the insurance demand of low-risk consumers than high-risk consumers. Second, channel features that relate to risk control, such as the offline manual underwriting service, may directly link to the risk of enrollees. Third, the ability or acceptance to use digital channels may positively correlate with unobserved policy risk via the risk characteristics not adjusted into unit premium, such as advanced education.

In addition, we decompose the consumer contributions to the risk screening effect. We find that 74% of the risk screening effect comes from the risk decrease in digital distribution channel policies, with only a small part attributed to the risk increases in offline policies due to crowded-out, low-risk consumers. Furthermore, new low-risk consumers through the digital distribution contribute most to the risk screening effect.

There are two significant implications as a result of this article. First, the consumer selection process that this article highlights from the adoption of digital technology may also occur in other industries undergoing digital transformation. For example, mobile APPs for investment and finance have also proliferated in recent years, meaning we can likely apply similar principles to the associated effects on the stock market. Second, this article indicates lowering search cost as a new way to mitigate adverse selection for insurance and other industries suffering from adverse selection. Our finding suggests that the measures that reduce search cost, such as digital distribution, improves the risk profile by creating higher incentives to low-risk consumers. However, one caveat of this article deserves noting: Since the data used in this article sources from a large Chinese insurer, the persistence of the estimated effects for insurers in other countries remains an open question.

References

Abaluck, J., & Gruber, J. (2011). Choice inconsistencies among the elderly: evidence from plan choice in the Medicare Part D program. *American Economic Review*, *101*(4), 1180-1210.

Aizawa, N., & Kim, Y. S. (2018). Advertising and risk selection in health insurance markets. *American Economic Review*, *108*(3), 828-67.

Brodeur, A., & Nield, K. (2018). An empirical analysis of taxi, Lyft and Uber rides: Evidence from weather shocks in NYC. *Journal of Economic Behavior & Organization*, *152*, 1-16.

Brynjolfsson, E., Hu, Y., & Simester, D. (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*, *57*(8), 1373-1386.

Brown, J. R., & Goolsbee, A. (2002). Does the Internet make markets more competitive? Evidence from the life insurance industry. *Journal of Political Economy, 110*(3), 481-507.

Brown, J., Duggan, M., Kuziemko, I., & Woolston, W. (2014). How does risk selection respond to risk adjustment? New evidence from the Medicare Advantage Program. *American Economic Review*, *104*(10), 3335-64.

Butler, R. J. (2021). Information access and homeowners insurance purchases. *The Geneva Papers on Risk and Insurance-Issues and Practice, 46*(4), 649-663.

Callen, M., De Mel, S., McIntosh, C., & Woodruff, C. (2019). What are the headwaters of formal savings? Experimental evidence from Sri Lanka. *The Review of Economic Studies, 86*(6), 2491-2529.

Carey, C. (2017). Technological change and risk adjustment: Benefit design incentives in Medicare Part D. *American Economic Journal: Economic Policy*, *9*(1), 38-73.

Cohen, A., & Siegelman, P. (2010). Testing for adverse selection in insurance markets. *Journal of Risk and Insurance, 77*(1), 39-84.

Cruz-Jesus, F., Vicente, M. R., Bacao, F., & Oliveira, T. (2016). The education-related digital divide: An analysis for the EU-28. *Computers in Human Behavior*, *56*, 72-82.

Cummins, J. D., & Doherty, N. A. (2006). The economics of insurance intermediaries. *Journal of Risk and Insurance, 73*(3), 359-396.

Domurat, R., Menashe, I., & Yin, W. (2021). The role of behavioral frictions in health insurance marketplace enrollment and risk: evidence from a field experiment. *American Economic Review, 111*(5), 1549-74.

Eckardt, M., & Räthke‐Döppner, S. (2010). The quality of insurance intermediary services—

empirical evidence for Germany. *Journal of Risk and Insurance, 77*(3), 667-701.

Einav, L., Levin, J., Popov, I., & Sundaresan, N. (2014). Growth, adoption, and use of mobile E-commerce. *American Economic Review, 104*(5), 489-494.

Ellison, G., & Ellison, S. F. (2009). Search, obfuscation, and price elasticities on the Internet. *Econometrica, 77*(2), 427-452.

Farber, H. S. (2015). Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *The Quarterly Journal of Economics*, *130*(4), 1975-2026.

Focht, U., Richter, A., & Schiller, J. (2013). Intermediation and (Mis‐) Matching in Insurance Markets—Who Should Pay the Insurance Broker? *Journal of Risk and Insurance, 80*(2), 329-350.

Geyskens, I., Gielens, K., & Dekimpe, M. G. (2002). The market valuation of Internet channel additions. *Journal of Marketing, 66*(2), 102-119.

Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of economic literature, 57*(1), 3-43.

Hall, T. E., & Owens, J. (2011, September). The digital divide and e-government services. In *Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance* (pp. 37-44).

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, *103*(7), 2643-82.

Handel, B., Hendel, I., & Whinston, M. D. (2015). Equilibria in health exchanges: Adverse selection versus reclassification risk. *Econometrica*, *83*(4), 1261-1313.

Handel, B. R., Kolstad, J. T., & Spinnewijn, J. (2019). Information frictions and adverse selection: Policy interventions in health insurance markets. *Review of Economics and Statistics*, *101*(2), 326-340.

Hargittai, E. (2002). Second-level digital divide.

Hilliard, J. I., Regan, L., & Tennyson, S. (2013). Insurance distribution. In *Handbook of Insurance* (pp. 689-727): Springer.

Hsieh, S. H., Liu, C. T., & Tzeng, L. Y. (2014). Insurance marketing channel as a screening mechanism: Empirical evidences from Taiwan automobile insurance market. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *39*(1), 90-103.

Hu, X., Wang, Z., & Liu, J. (2022). The impact of digital finance on household insurance purchases: evidence from micro data in China. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 1-31.

Jack, W., & Suri, T. (2014). Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution. *American Economic Review, 104*(1), 183-223.

Jolivet, G., & Turon, H. (2019). Consumer search costs and preferences on the Internet. *The Review of Economic Studies, 86*(3), 1258-1300.

Ketcham, J. D., Lucarelli, C., Miravete, E. J., & Roebuck, M. C. (2012). Sinking, swimming, or learning to swim in Medicare Part D. *American Economic Review*, *102*(6), 2639-73.

Kling, J. R., Mullainathan, S., Shafir, E., Vermeulen, L. C., & Wrobel, M. V. (2012). Comparison friction: Experimental evidence from Medicare drug plans. *The quarterly journal of economics, 127*(1), 199-235.

Kochar, A. (2018). Branchless banking: Evaluating the doorstep delivery of financial services in rural India. *Journal of Development Economics, 135*, 160-175.

Roca, J. D. L., & Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, *84*(1), 106-142.

Polyakova, M. (2016). Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D. *American Economic Journal: Applied Economics*, *8*(3), 165-95.

Shepard, M. (2022). Hospital network competition and adverse selection: evidence from the Massachusetts health insurance exchange. *American Economic Review*, *112*(2), 578-615.

Stein, L. C., & Yannelis, C. (2020). Financial inclusion, human capital, and wealth accumulation: Evidence from the freedman's savings bank. *The Review of Financial Studies*, *33*(11), 5333-5377.

Van Deursen, A. J., & Van Dijk, J. A. (2019). The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *New media & society*, *21*(2), 354-375.

Venezia, I., Galai, D., & Shapira, Z. (1999). Exclusive vs. independent agents: A separating equilibrium approach. *Journal of Economic Behavior & Organization*, *40*(4), 443-456.