Risk Screening in Digital Insurance Distribution: Evidence and Explanation

1 Introduction

The embedding of digital technology in the global economy has attracted increasing attention from economists. A key insight into the economic mechanisms 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). In turn, this logic underpins an abundance of microeconomics literature on how digital technology improves consumer surplus by increasing market efficiency 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 enterprises. 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

[1] On the premise 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 channels that rely on mobile devices, such as mobile APPs.

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 substantial compared with offline policies – accounting for 21%, 54% and a comparable size of the offline accident rate for the endowment insurance, disease insurance and term life insurance products, respectively.

The risk screening effect makes digital distribution an effective insurer-side risk selection tool. Given that digital distribution was introduced in addition to existing 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. This risk screening effect persists through the sales period and reflects more adverse selection than moral hazard due to the limited policy maturity.

We find that the risk screening effect of digital distribution largely comes from the extensive margin (i.e., from *new consumers*) and thus improves market efficiency. The introduction of digital distribution attracts both low-risk, *switchers* who switched from offline channels and low-risk, *new consumers* who were previously not covered. The former lead to average risk increases in retained offline policies, however, this contributes to only a small part of the risk screening effect. At least 81% of the risk screening effect is sourced from the new low-risk consumers attracted.

There are two important economic consequences of the risk screening effect for the insurance sector. First, unobserved risk is a source of information asymmetry that impedes insurers' assessments of 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 has lower information asymmetry. Second, lower unobserved risk also increases the profitability of digital distribution. This is demonstrated by the average indemnity amount and loss ratio being lower for digital channel policies than for offline policies.

This paper distinguishes *channel capability* and *channel preference* as the drivers of the risk screening effect. This distinction unravels the prevalence of risk screening independent of the difference in clienteles catered to by digital and offline distribution. The role of *channel preference* relates to the relationship between policy risk and the ability or acceptance to use digital channels (i.e., digital divide). For example, we show that advanced education, as a typical risk characteristic not adjusted into pricing, positively correlates with the preference to purchase via digital channels while negatively correlates with policy risk.

The role of *channel capability* comes from the channel features (i.e., search cost) that correlate with both insurance demand and the 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 low-risk consumers self-select into digital channels because of the higher insurance demand sensitivity to the search cost reduction afforded by digital distribution. This mechanism explains the source of market efficiency improvement. The second mechanism of *channel capability* is the channel features directly related to risk control. We study this mechanism by showing that offline agents' less rigorous implementation of underwriting rules leads to the enrollment of higher-risk consumers who should not have been eligible for insurance. Finally, our evidence suggests that *channel capability* plays a dominantly larger role than *channel preference*.

This article makes threefold significant contributions to the literature. First, it contributes to digital economics literature by highlighting consumer selection resulting from 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 the savings of low-income households are more sensitive to the introduction of mobile banking than those of high-income households. Therefore, this article broadens the understanding of the heterogeneity of the economic impacts of digital technology.

Second, this article relates to a large literature set that empirically studies screening in the insurance market. Two strands of this literature are most relevant to our article. The first strand is the growing corpus on the engagement between choice frictions and 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 as a new engagement factor. The second strand of related literature is on insurer-side risk selection. Risk adjustment systems, plan designs, hospital networks and advertising have been empirically shown to be insurers' risk selection tools (Brown et al. 2014; Carey 2017; Aizawa and Kim 2018; Shepard 2022). Our article's novelty is to show digital distribution as another tool for insurers to achieve advantageous risk screening.

Last, this article also adds to the literature on marketing in the insurance context. Different distribution channel strategies have different impacts on the 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 neither particularly address the effect on policy risk which is of utmost relevance to insurance institutions, nor 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 of independent agents than that of direct underwriters may lead to risk sorting. There are two main differences between their studies and ours. First, the offline channel in our setting consists of agents employed 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 fully explained by their argument, because in our setting, the claims of offline policies are more likely to be rejected than the claims of digital channel policies, 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 the data description and baseline empirical specification. Section 4 presents the baseline results. In Section 5, two important consequences on the insurance business induced by the risk screening effect are examined. Section 6 decomposes the consumer contributions to the risk screening effect. Section 7 examines three mechanisms of the risk screening effect according to the conceptual framework. The final Section 8 discusses implications and concludes this paper.

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) distinguishing *channel capability* and *channel preference* as the drivers of the risk screening effect.

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 $j \in \{0, 1\}$

[3] For simplicity, we keep the insured amount fixed. This is without loss of generality because a varied indemnity relates varied loss probability q_i.

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 for purchasing insurance via distribution channel j, given *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 policy.

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 unobservable variables and it is impossible for the insurer to charge a perfect unit premium[4]. Let q_i denote consumer i's actual risk, and the expected risk of the 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. Here, we assume the implementation cost of digital distribution is sunk cost of this insurer and the marginal cost of digital platforms is negligible (Bakos and Brynjolfsson, 2000; Aguiar and Waldfogel, 2018). The outcome of interest is the change of the insurer's profits when introducing the digital distribution, which can be written as

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

[4] For 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 characteristics into account. This group-specific pricing practice has been shown in Section 3.1.2.

This equation clearly illustrates how information asymmetry affects profits by engaging 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 channels essentially reflects the impact of adopting digital distribution on information asymmetry, which motivates our following empirical tests.

2.2 Mechanisms of the Risk Screening Effect

What drives the engagement between digital distribution and information asymmetry? There are two drivers: *channel capability* and *channel preference*. We call the channel features (i.e., search cost) that correlate with both insurance demand and risks of enrollees attracted as *channel capability*, and the difference in the people's ability or acceptance to use digital channels (i.e., digital divide) as *channel preference*. A distinction between these two drivers is necessary for three reasons. First, this distinction can tell whether the risk screening effect persists as people's ability to use digital technology continues to advance until saturation is reached. Second, this distinction can tell whether the risk screening effect is widespread regardless of the insurer's marketing strategies targeting different clienteles[5]. Third, more importantly, *channel capability* and *preference* respectively explain the extensive margin and the intensive margin. Highlighting this explanation difference is crucial as the extensive margin, i.e., *new consumers*, will improve the risk profile and increase market efficiency; while the intensive margin, i.e., *switchers* who switched from the offline channel, will only lead to cross-platform risk selection and not affect the risk profile[6].

[5] For example, income, a characteristic not generally observed by insurers, positively correlates with owning a mobile phone while negatively correlates with private health. Thus, income is an instance of the role of *channel preference*. If the risk screening effect is due to merely *channel preference*, the insurer's product marketing strategy focusing on a certain income group, such as low-income individuals, would lead to the disappearance of the risk screening effect. A real case is the inclusive insurance products in China that aim to provide insurance to low-income populations.

[6] *Channel preference* can explain/partly explain only *switchers* because it does not affect insurance demand; *New consumers* reflect the impact of digital distribution on insurance demand (insurance purchase probability), which can be explained only by *channel capability*.

To understand their roles, we breakdown D_{ij} and ΔD_{ij} as

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

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

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 Equation (4) is very natural and states that: only when consumers are willing and able to use a digital channel can they be affected by the introduction of digital distribution, and when both channels can be accepted and used by consumers, the utility of channel choice will be affected only by channel features. The roles of *channel capability* and *preference* are presented in Equation (5) where on RHS, the left expression represents the impact of *channel capability* (i.e., $CD_{i1} > CD_{i0}$) and the right expression represents the impact of *channel preference* (i.e., $p_{i1} > p_{i0}$).

Three reasons suggest the relationships of *channel capability* and *preference* with unobserved risk. First, the marginal influence of channel features on the insurance demand is heterogeneous across different risk groups. For example, if there is some advantageous feature of digital distribution that increases insurance demand relative to the offline channel, consumers with high unobserved risk 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. These channel features contribute to the difference in the rigor of implementing underwriting rules between offline and digital distribution, rooted in the fact that imperfect supervision is more problematic for insurers when dealing with offline agents than with machines (i.e., APPs). For example, the commission system absent from digital distribution may motivate offline agents skirt with underwriting rules in pursuit of more sales and commissions.

Third, *channel preference* may negatively correlate with unobserved risk due to the digital divide. Deursen and Dijk (2019) found that even if Internet penetration reaches saturation in a country, differences in individuals' ability or acceptance to use Internet technology (e.g., mobile Apps) still lead to beneficial inequality, creating a widening digital divide[7]. Generally, the low-educated and low-income 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 essence, the heterogeneous *channel preference* reflects different clienteles being catered to 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 7.

3 Data and Empirical Strategies

3.1 Settings

3.1.1 Digital Insurance Distribution

According to statistics from the Insurance Association of China (Securities Journal 2016), among the 55 life insurers that disclosed their domestic Internet insurance business, 18 have launched mobile APPs to sell insurance products. These mobile APPs usually provide the entire service journey from quotation to claim. This history and scale make China's life insurance industry an excellent site to observe the effects of digital distribution adoption. In this article, the investigated life insurer is one of the first insurers to launch the mobile APP channel in China. The mobile APP sells products of various insurance types, allowing our study to cover different insurance types and enabling greater generalization of our research conclusions. The investigated insurance products have consistent plan settings across digital and offline channels, insulating our estimates from the potential bias of unobservable product differences.

[7] In the 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 the second-level digital divide.

The mobile APP can be downloaded for free from mobile application stores. Insurance product information such as application qualification (e.g., insurable ages), liability and exempted liability descriptions, insurance period and optional plan settings (e.g., optional insured amount, additional insurance) are displayed clearly. 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 policy claims.

The insurance application process of traditional offline channels – offline agents and bancassurance – has several notable differences from that of digital distribution channels. The first difference is the obvious absence of the convenience of offline purchasing. The second difference is the time limit on offline channel services. For example, the bancassurance channel has fixed operation times while APPs allow for insurance purchase 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 of offline and digital distribution channels are similar– because offline agents are required by the insurer to let consumers download and register the mobile APP after selling a policy, so that consumers who purchase policies offline can still enjoy the after-sales services on the mobile APP.

3.1.2 Group-Specific Pricing of Life Insurance in China

Life insurers in China generally adopt a group-specific pricing practice offering a premium rate table classifying rates by age and gender for quotation. In Figure A1 of Appendix A, we provide an example representative of this pricing practice with a screenshot of a practical premium rate table of a term life insurance product sold on the insurer's mobile APP. When the insurance period and premium payment term are set, rates only vary with age and gender, implying that age and gender are the dominant risk factors of pricing. This fosters information asymmetry because many other observed or unobserved risk factors are not adjusted into pricing.

To validate the age-gender-specific pricing practice, we regress unit premium on age, gender and their interaction age×gender while fixing the insurance period, rider and premium payment term to examine the risk factors affecting pricing. The specification details and regression results are presented in Appendix B.

The results in Columns (1), (4) and (7) of Table B1, Appendix B show that for all three products, unsurprisingly, unit premiums increase with age. The effect of gender on unit premiums varies with age. Notably, significance of the coefficients of age, gender and age×gender is all very high, with t-statistics over 3.2. R squares are above 0.87, 0.72 and 0.92 for the regressions of the term life, endowment and disease insurance products respectively, manifesting the great explanatory power of age and gender. In Columns (2), (5) and (8), we further add other observed policyholder and policy characteristics including advanced education, financial profession, log insured amount and policy status. Two observations deserve note: on the one hand, most coefficients of age, gender and age×gender as well as R squares change little; on the other hand, none of the additionally added characteristics have significant effects except for the log insured amount with the endowment insurance product (significant only at the 90% confidence level). This suggests that characteristics other than age and gender have no or low impact on pricing. In Columns (3), (6) and (9), we rerun the regressions by further refining age and gender into dummies of age, dummies of gender and interactions of these dummies (totaling 135, 158, 152 dummies for the term life, endowment and disease insurance products, respectively), keeping additionally added characteristics. The R squares increase by 0.003 to 0.03 due to refined age and gender dummies; however, most additionally added characteristics still have no significant effects. Again, this confirms the age-gender-specific pricing practice in which age and gender dominate pricing.

An appropriate identification of the screening on unobserved risk requires controlling for risk factors affecting insurance price (Chaippori and Salanie, 2000). Otherwise, the identification result could partly reflect risk differences resulting from the risk factors that correlate with the channel choice but can be offset by pricing. The age-gender-specific pricing practice in China and the above examinations have clearly demonstrated that the majority of the unit premium variances are attributed to age and gender. Therefore, in our following empirical identification strategies, age dummies, gender dummies and their interactions serve as necessary controls in addition to the dummies of insurance period, rider and premium payment term. Considering that there remains a tiny part of pricing variances accounted for by other factors, we also directly add the unit premium itself as a necessary control.

3.2 The Data

We gathered detailed proprietary data of all purchased policies for three 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, 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 riders. These data allow us to distinguish 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 I reports the descriptive statistics of the main variables by investigated insurance product, based on the policies sold during the periods with both digital and offline distribution. The term life insurance product was sold from 2017 to 2019 via the offline agent channel and was introduced on the insurer’s mobile APP channel on August 15, 2018. It insures against death or disability risk and accumulated a total of 97,495 policies sold, 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 insured until 75 years old with liabilities of death or disability, disease medication, accident medication and severe illness. The APP channel policies account for a share of roughly 17%. In the following empirical tests, the PC-Internet channel policies (roughly 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 both the APP channel and offline bancassurance channel from October in 2016 to May in 2018. This product covers death or disability and severe diseases with 53,296 policies sold and 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 of Appendix A. The ranges of plan settings[8] such as waiting period and insurance period are effectively the same across distribution channels. It is normal that insurers are usually very reluctant to vary plan properties for the same product across channels because variations tend to intensify channel benefit conflicts (Geyskens et al. 2002). The plan setting consistency between digital and offline distribution channels is a crucial advantage of our data, as it avoids potential bias from unobservable product differences that have been shown to be difficult to resolve in prior literature (Brown and Goolsbee 2002).

In addition to the three investigated products, to exploit the treatment event of introducing digital distribution to causally identify its impact on the risk pool, we also source the data of the other purely offline product (control product), which was the only term life insurance product sold during the same period (from 2017 to 2019) as the investigated term life insurance product (treatment product), totaling 87,643 policies. These two products cover the same liability – both insure against only the death or disability risk without any riders, and were both

[8] Insurance policies shares the same premium rate table across different distribution channels.

sold nationwide. The major difference is the insurance period: the control product offers options of 10 or 15 years, while the treatment product offers options of 20 or 30 years. This allows them to differentially cater to consumers with different risk coverage duration needs. Thus, by comparing Columns (5) and (7) in Table I, on average, the control product has a shorter insurance period and lower unit premiums.

There are some additional comparative statistics of interest: the shares of claimed and indemnified policies of the treatment product are both lower than those of the control product, suggesting advantageous risk screening of digital distribution; however, from the share difference between claimed and indemnified policies, the control product claims seem more likely to be rejected, suggesting poorer underwriting risk control for the offline channel. Comparisons of their other plan settings, such as optional premium payment terms, hesitation period and waiting period, are also presented in Table A2 of Appendix A. Interproduct comparisons (e.g., treatment product vs. control product) show basically similar value ranges of plan properties except for the insurance period. Interperiod comparisons (e.g., before vs. after the introduction of digital distribution) show no concurrent changes in plan properties of the treatment product on the introduction date.

In our data, policyholders are not necessarily the insurer person. To ensure the controls of the risk factors of policy pricing in empirical models, we exclude the policies not insuring against policyholders themselves for empirical analyses, which samples 672,562, 23,343, 93,623 and 77,126 policies for the endowment insurance, disease insurance, investigated and control term life insurance products respectively.

Table I: Descriptive Statistics of the Data by Product

	Investigated Products						Control Product	
	Endowment Insurance		Disease Insurance		Term Life Insurance		Term Life Insurance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Outcomes								
Policy Claim	0.035	0.185	0.008	0.089	0.0013	0.036	0.0039	0.058
Policy Compensation	0.013	0.114	0.006	0.076	0.0009	0.029	0.0025	0.023
Indemnity Amount	203.99	13010.12	1823.78	25817.94	169.92	7862.60	130.89	3441.82
Variable of Interest								
Digital Channel Choice	0.166	0.372	0.382	0.486	0.861	0.346	0	0

Controls								
Age	38.922	8.757	37.697	7.711	36.586	7.421	41.222	9.565
Female	0.488	0.500	0.723	0.448	0.566	0.496	0.450	0.498
Financial Profession	0.136	0.343	0.182	0.386	0.386	0.487	0.349	0.477
Advanced Education	0.137	0.344	0.285	0.451	0.251	0.434	0.188	0.391
Cancellation	0.010	0.098	0.001	0.035	0.005	0.068	0.010	0.098
Insurance Period	37.164	9.100	Whole life	Whole life	21.087	3.112	10.988	1.991
Insured Amount	9.428	15.735	48.475	24.370	32.858	18.778	24.767	40.720
Unit Premium	0.138	0.154	0.010	0.008	0.547	0.227	0.415	0.155
Rider	0.381	0.457	0.858	0.349	0	0	0	0
Payment Term	15	0	19.182	2.741	17.766	4.165	13.153	3.437
Observations								
Total	963,244		53,296		97,495		87,643	
Oneself Relation	672,562		23,343		93,623		77,126	
Couple Relation	194,234		7,932		3,103		4,205	
Other Relations	96,448		22,021		769		6,312	

Note: *Female* is an indicator of female policyholders. *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. *Cancellation* indicates whether the policy has been cancelled. *Insured Amount* is in ten thousand Yuan. *Unit Premium* is the yearly premium per unit insured amount in Yuan. *Rider* indicates whether the policy includes additional insurance. *Premium Payment Term* is the period of premium payment in years. *Digital Channel Choice* is a dummy of whether the policy was purchased via the digital channel. *Policy Claim* indicates claimed policies. *Policy Compensation* indicates the policy has been indemnified. *Indemnity Amount* is in Yuan and equals zero for uncompensated policies. Descriptive statistics of the investigated endowment, disease and term life insurance products are based on the total policies sold during the periods with both digital and offline distribution. Descriptive statistics of the control term life insurance product are based on total policies sold from 2017 to 2019. Oneself and couple relations represent the policies purchased to insure against policyholders themselves and the policies purchased by one spouse to insure against the other, respectively. Oneself relation observations are used for the following empirical analyses.

To examine evidence on the *channel capability* mechanism of the risk screening effect, we construct measures of offline insurance search costs with datasets on geography and population. The first dataset is the API service offering the longitude and latitude of insurer branches throughout China in 2019, provided by BAIDU Map[9], 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[10]. Each insurer branch and population grid-cell are matched to the corresponding prefecture according to their coordinates. Moreover, we also gather data on mobile phone ownership per capita for prefectures in China from 2017 to 2019 from the China Urban

[9] Seen on https://lbsyun.baidu.com/
[10] Available on https://www.worldpop.org/

Statistical Yearbooks, to be used for robustness checks. We offer descriptions of related variables in Table A3, 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 to test the association between policy risk and purchase channel choice. Following Dionne et al.'s approach (2010), we use the following baseline specification for each insurance product:

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

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 channel. $\boldsymbol{D}_t$ is a trend term capturing the natural time influence on policy risk. Approaches of this form have been extensively used in the literature testing information asymmetry (e.g., Cohen 2005; Kim et al., 2009; Spindler et al., 2014).

An appropriate identification of the difference in unobserved risk across channels for this test requires controls for characteristics that affect the pricing of policies (Chaippori and Salanie, 2000; Dionne et al., 2001). To achieve this, in Equation (6), $\boldsymbol{X'}_{i,r,t}$ represents a set of controls for policyholder and policy characteristics, including:

Controls of Pricing.— Dummies of policyholder age, a dummy of policyholder gender, interactions of age dummies and gender dummies, and the unit premium.

Controls of Plan Settings.— Dummies of insurance period, dummies of premium payment term, and a dummy of having riders.

Controls of Other Observed Characteristics.— Dummies of policyholder education levels, a dummy indicating employment of the policyholder in the financial industry, log insured amount and a dummy of policy cancellation.

The details of these controls are summarized in Table A4 of Appendix A. $\boldsymbol{X'}_{r,t}$ 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 channels. The regression samples consist of the policies purchased via digital and offline channels for the endowment insurance and disease insurance. While for the term life insurance, the regression sample consists of the policies purchased after the introduction of digital distribution (from August 15, 2018 to December 31, 2019), keeping the investigated sales periods of both digital and offline channels consistent.

Our second estimate examines the effect of the introduction of digital distribution on the risk pool and is performed at the prefecture aggregation level. We construct a difference-in-difference (DID) counterfactual framework utilizing the introduction of digital distribution on August 15, 2018 for the investigated term life insurance product. This DID framework takes the investigated term life insurance product as the treatment product and the other term life insurance product sold only through offline agents[11] during the same period as the control product. As introduced in section 3.2, both products have the same liability and similar plan settings. We aggregate policies by product, prefecture and date. The DID specification can be written as

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

where for product p, prefecture r and date t, $\boldsymbol{AR}_{p,r,t}$ denotes the accident rate[12] of sold policies, $\boldsymbol{Treat}$ is defined as 1 for the treatment product and 0 for the control product, $\boldsymbol{\mathcal{L}}_t$ is defined as 1 after August 15, 2018 and 0 otherwise. $\boldsymbol{X''}_{p,r,t}$ is a vector including the averages of the same

[11] This control term life product meets three criteria for a suitable control group: (1) covering the same liabilities, (2) being sold during the same period and (3) being sold only via the same offline channel as the treatment product. Among all three investigated insurance products, we only identified one product that meets all three criteria for the investigated term life insurance product. This is because term life insurance products are simpler and tend to be homogeneous.
[12] For product p, the accident rate is calculated as the share of the claimed policies sold in prefecture r on date t.

set of policyholder and policy characteristics as the OLS estimate. Specifically, dummy controls are averaged into percentages, for example, the average of the gender dummy indicates the percentage of the female policies among all the policies purchased in prefecture r on date t for product p; while for continuous controls including unit premium and log insured amount, we calculate the average unit premium and insured amount[13] for the policies purchased in prefecture r on date t for product p. $\boldsymbol{X}''_{r,t}$ is a vector comprising the date and prefecture-year interactive fixed effects (also including separate prefecture fixed effects). $\boldsymbol{X}'_{p}$ is a vector of the product fixed effect. Robust standard errors are clustered at the prefecture level. π captures the decrease in the average risk of the treatment product policies purchased after the introduction date relative to the counterfact 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, a key difference in results interpretation from the prior OLS estimate is that, the DID estimated effect reflects the difference in risk pool between reality with digital distribution and the counterfact without digital distribution, instead of the difference in average policy risk between digital and offline channels.

3.4 Descriptive Evidence

First we provide some preliminary descriptive evidence. In the spirit of controlling for group-specific pricing, Figure 1 plots the results of the average accident rates of digital channel policies less the average accident rates of offline channel policies for each investigated product, with fixing the gender and age groups. Obviously, for each five-year age group of each gender for each product, most results are negative with magnitudes seemingly larger for the older policyholders and only a few exceptions above zero (e.g., 46-50 age group of the term life insurance). This figure suggests a reduction in the average policy risk of digital distribution that is not adjusted into pricing.

[13] After calculating the average insured amount, we then take the logarithm of it.

Figure 2 plots the monthly accident rates for the treatment and control term life insurance products. Before the introduction month of digital distribution, the pivotal of the accident rates of the control product is approximately 0.005, significantly higher than the pivotal of the treatment product of approximately 0.003 to 0.004; while once digital distribution is introduced to the treatment product, the pivotal difference between these two products suddenly expands, driven by the notable decrease in the accident rates of the treatment product. This figure suggests that digital distribution is an effective tool to achieve advantageous risk screening and improve the risk profile for the insurer. Our formal DID estimate further excludes the influence of pricing and observed policyholder characteristics.

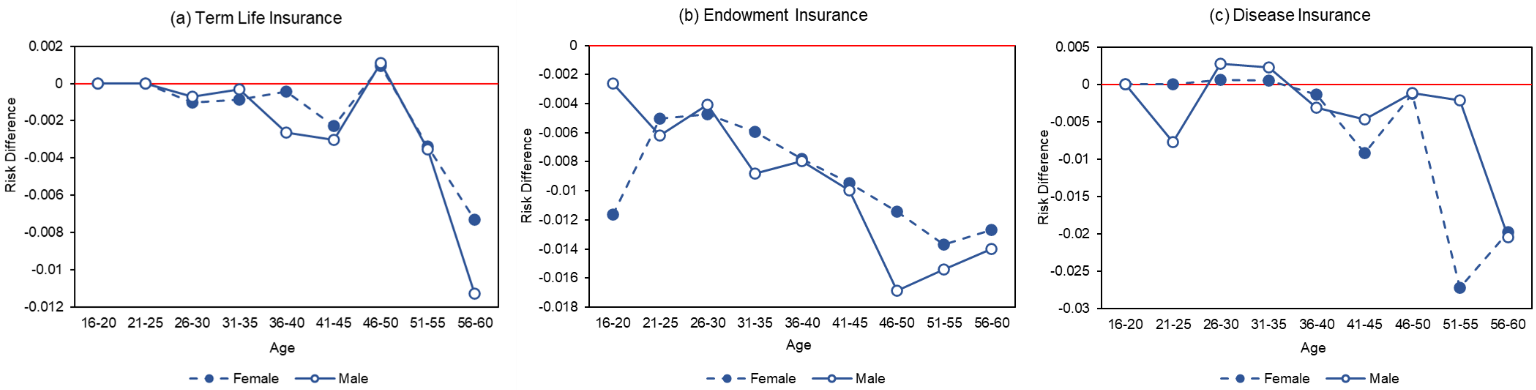


Figure 1: Risk Difference between Digital and Offline Distribution Holding Gender and Age Groups

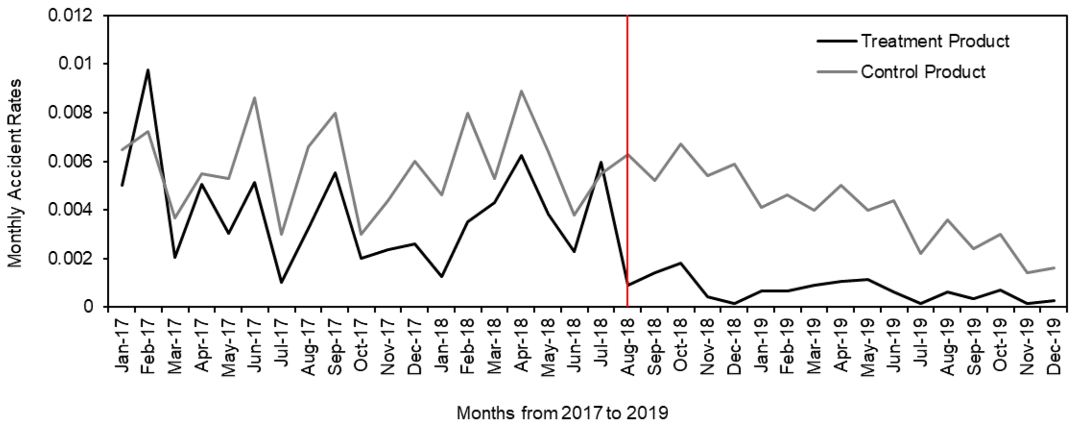


Figure 2: Monthly Accident Rates for Treatment and Control Term Life Insurance Products

4 Results

4.1 OLS Estimates

Table II presents the results of estimating the risk screening effect of the digital distribution on policy risk with the baseline OLS specification. Panels A, B and C present the results for the term life insurance, endowment insurance and disease insurance products respectively. Column (1) employs the full sample and uses only controls of pricing, showing that on average, the accident probabilities of the policies purchased through digital distribution channels are 0.20, 0.32 and 0.76 percent points lower than those of the policies purchased offline for the term life insurance, endowment insurance and disease insurance, respectively. Column (2) further uses full controls and yields very similar results. Columns (3) and (4) respectively exclude cancelled and claim-rejected policies with full controls. We find that again, the results are statistically significant and negative for all three panels, regardless of policy cancellations and claim rejections.

The magnitude of the risk screening effect of digital distribution on policy risk, documented in Table II, 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 channel policies accounts for at least 21% (=0.0027/0.0129) and 54% (=0.0043/0.0080) of the corresponding offline accident rate for the endowment and disease insurance products, respectively. For the term life insurance product, the risk probability reduction magnitude is even comparable to the magnitude of the offline accident rate.

Since our dependent and independent variables are both dummies, we complement Logit regressions using the same empirical specification for robustness. We also include Cloglog regressions given the rarity of claims. Their estimated marginal average effects, presented in Table A5 of Appendix A, are qualitatively consistent with the OLS estimate results.

An insurance product usually covers a package of liabilities insuring different risks, leading to the possibility that the risk screening effect of digital distribution loses universality if it occurs to only a specific risk. To address this concern, for the endowment and disease insurance, we group the policy claims by accident type 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 III 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 7% for the death or disability risk of the endowment insurance. Notably, severe disease of the disease insurance has the largest reduction in accident probability among all risk types. Therefore, the risk screening effect is unconditional on the risk types insured by the investigated products.

Table II: Baseline Results of OLS Estimates

Variables	(1) AC	(2) AC	(3) AC	(4) AC
Panel A. Term Life Insurance				
$digital$	-0.0020***	-0.0019**	-0.0020**	-0.0015**
	(-2.60)	(-2.32)	(-2.41)	(-2.09)
Observations	93,623	93,623	93,206	93,592
Adj. R-squared	0.017	0.018	0.018	0.016
Offline accident rate	0.0018	0.0018	0.0018	0.0013
Panel B. Endowment Insurance				
$digital$	-0.0032***	-0.0031***	-0.0031***	-0.0027***
	(-8.37)	(-8.16)	(-8.06)	(-7.19)
Observations	672,562	672,562	665,247	672,053
Adj. R-squared	0.028	0.029	0.029	0.028
Offline accident rate	0.0137	0.0137	0.0137	0.0129
Panel C. Disease Insurance				
$digital$	-0.0076***	-0.0071***	-0.0071***	-0.0043***
	(-4.16)	(-3.95)	(-4.06)	(-2.76)
Observations	23,343	23,343	23,314	23,290
Adj. R-squared	0.060	0.150	0.149	0.068
Offline accident rate	0.0109	0.0109	0.0109	0.0080
Controls	controls of pricing	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) and (2) employ the full sample, while Columns (3) and (4) excludes cancelled and claim-rejected policies respectively. Column (1) uses only controls of pricing - the policyholder age dummies, gender dummies, dummies of age-gender interactions, and unit premium. Columns (2) to (4) use full controls. 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 III: 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	-0.0003*	-0.0007**	-0.0028***	-0.0001***
	(-1.84)	(-2.24)	(-14.00)	(-5.63)
Observations	663,828	669,584	665,821	663,426
Adj. R-squared	0.006	0.020	0.015	0.002
Panel C. Disease Insurance				
digital	-0.0013**			-0.0062***
	(-2.34)			(-3.61)
Observations	23,102			23,327
Adj. R-squared	0.082			0.133
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: Since the term life insurance insures only one accident type –- the death or disability risk, it is not presented here.

4.2 Difference-in-Difference Estimates

4.2.1 Pretrends

The key assumption of a standard DID estimate is the parallel pretrends. If the average policy risk of the treatment product exhibited a decreasing trend relative to the control product before the introduction of digital distribution (e.g., due to the insurer's deliberative product selection), the identified effect would be endogenous. To validate this assumption, we use an Event Study specification analyzing the dynamic effects of monthly leads and lags as follows.

$$\boldsymbol{AR}_{p,r,t} = \alpha + \sum_{\substack{m-j=20 \\ m=1,2,3\ldots19}} \tau_{-j}\boldsymbol{Treat}_p \times \boldsymbol{mpre}_{-j} + \sum_{\substack{m-k=20 \\ m=20,21,22\ldots36}} \tau_k \boldsymbol{Treat}_p \times \boldsymbol{mpost}_k + \boldsymbol{X''}_{p,r,t}\boldsymbol{\Gamma} + \boldsymbol{X''}_{r,t}\boldsymbol{\Phi} + \boldsymbol{X'}_p\boldsymbol{\Omega} + \varepsilon_{r,t} \tag{8}$$

where $\boldsymbol{mpre}_{-j}$ and $\boldsymbol{mpost}_k$ are dummies defined as 1 for j months before and k months after August, 2018 respectively. Fixed effects and controls are identical to Equation (7). In operation, $\boldsymbol{mpre}_{-1}$ was dropped to avoid multicollinearity. The point estimates along with their 95% confidence intervals for the coefficients are illustrated in Figure 3. Consistent with the parallel pre-trends assumption, we find that none of the monthly leads have significant effects on the average policy risk; rather, the coefficients of monthly lags fall sharply to be significantly negative with narrowing standard errors after the introduction month. We also note that the risk screening effect persists until the end of 2019, implying that it is unlikely to be merely driven by early adopters of digital distribution.

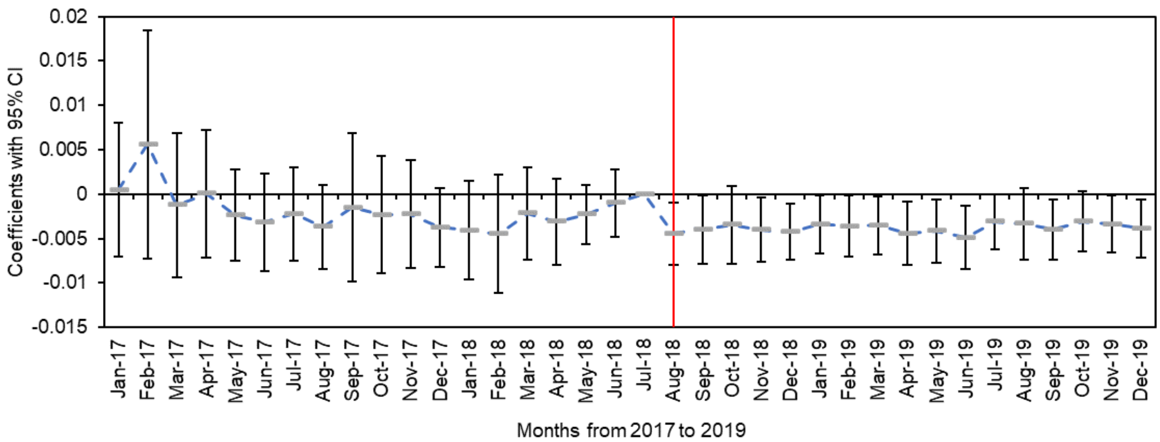


Figure 3: Coefficients of Event Study on the Introduction of Digital Distribution

To further confirm no inherent decrease in the average policy risk of the treatment product relative to the control product, we provide a more direct test on the trend of risk difference using the sample of the policies purchased before the introduction date (from January 1, 2017 to August 14, 2018), as specified in Equation (9).

$$\boldsymbol{AR}_{p,r,t} = \alpha + \gamma \boldsymbol{Treat}_p \times \boldsymbol{D}_t + \boldsymbol{X''}_{p,r,t}\boldsymbol{\Gamma} + \boldsymbol{X''}_{r,t}\boldsymbol{\Phi} + \boldsymbol{X'}_p\boldsymbol{\Omega} + \varepsilon_{r,t} \tag{9}$$

where γ is the coefficient of interest which captures how the risk difference between treatment and control products changes over time before the introduction date. The estimate result, reported in Column (1) of Table IV, is insignificant and around zero, thus falsifying an inherent decreasing pretrend of the average policy risk of the treatment product.

Overall, our evidence suggests there is no pre-existing decreasing trend of the average policy risk of the treatment product. This supports the exclusion of product selection that the treatment product was deliberately selected by the insurer due to some unobserved factor which inherently attracted policies with lower unobserved risk.

4.2.2 Estimate Results

Table IV reports the DID estimate results in Columns (2) and (3), using the full sample and the introduction year sample respectively. They confirm the risk screening effect of digital distribution. Specifically, Column (2) shows that the average accident rate of total policies of the term life insurance reduces by 0.18 percent points after the introduction of digital distribution. This result is similar to the corresponding OLS estimate result in Panel A, Column (1) of Table IV but has a slightly smaller magnitude. The magnitude difference is due to the interpretation difference between OLS and DID estimates as mentioned in Section 3.3 - the DID estimated effect does not include the policy risk difference between the counterfact and offline channel policies.

Column (3) presents a larger reduction in the average accident rate over the shorter term of 2018. Since the full sample incorporates more newer policies (e.g., 2019 policies) than the 2018 sample, this coefficient magnitude difference essentially reflects less substantial risk screening effects for the newer policies than for the older policies. A possible reason is that for new policies, the risk screening of digital distribution is mainly driven by adverse selection, while for old policies, the risk screening is driven not only by adverse selection but also more likely by moral hazard.

Table IV: DID Estimates of the Risk Screening Effect

	Parallel Pre-trend Test	DID Estimates		Falsification Tests			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\boldsymbol{Treat_p \times D_t}$	0.0002						
	(0.54)						
$\boldsymbol{Treat_p \times \mathcal{L}_t}$		-0.0018***	-0.0027**	0.0011	0.0034	-0.0013**	-0.0008
		(-2.97)	(-2.12)	(0.57)	(0.48)	(-2.45)	(-1.19)
$\boldsymbol{Treat_p \times \mathcal{L}_t \times MPO}$						-0.0006**	
						(-2.47)	
Observations	47,996	95,078	41,295	27,321	26,462	95,078	6,267
Adj. R-squared	0.038	0.040	0.039	0.074	0.036	0.042	0.032
Controls	Y	Y	Y	Y	Y	Y	Y
Fixed Effects							
Prefecture-Year	Y	Y	Y	Y	Y	Y	Y
Product	Y	Y	Y	Y	Y	Y	Y
Date	Y	Y	Y	Y	Y	Y	Y

Note: For both treatment and control products, Column (1) uses the policies purchased before the introduction of digital distribution; Columns (2) and (6) use the full sample of 2017-2019 policies; Column (3) uses the policies in 2018; Column (4) uses the policies in 2017 with August 15, 2017 as a virtual introduction date and Column (5) uses the policies in 2019 with August 15, 2019 as a virtual introduction date; Column (7) uses the policies from the prefectures with the 3-year average MPO rates ranking the lowest 10%.

Robustness Checks: A concern of our DID estimates is that there may exist unobserved concurrent factors leading to decreasing policy risk of the treatment product. To exclude this factor, we preform two falsification tests.

Natural Factors.— The first source of concurrent factors is the natural factors, that is, there could be a natural decrease in average policy risk of the treatment product on August 15th in each year. To examine this, we replace the sample with the policies of both products sold in 2017 and 2019 – the year prior and next to the introduction year. The logic is that under the same DID specification, if the policy risk decrease was indeed due to some natural factors, the results would still be significant in these falsification tests. The results presented in Columns (4) and (5) of Table IV were all insignificant and thus allay this suspicion. To be clear, the risk screening effect we find only occurs in the introduction year 2018, not in any other years.

Product Changes.— The other source of concurrent factors is the unobserved product changes, such as an unobserved shift of underwriting rules that accompanies the introduction of digital distribution. To address this, we construct a triple-difference model (DDD model) as a falsification test by adding the prefecture level yearly mobile phone ownership rate (MPO). We obtain the MPO data for the prefectures in our sample from the 2017-2019 China Urban Statistical Yearbooks. The specification of the DDD model is given by

$$\begin{aligned} \boldsymbol{AR}_{p,r,t} = \alpha + \beta \boldsymbol{Treat}_p \times \boldsymbol{\mathcal{L}}_t \times \boldsymbol{MPO} + \pi \boldsymbol{Treat}_p \times \boldsymbol{\mathcal{L}}_t + \rho \boldsymbol{Treat}_p \times \boldsymbol{MPO} \\ + \tau \boldsymbol{\mathcal{L}}_t \times \boldsymbol{MPO} + \boldsymbol{X''}_{p,r,t}\boldsymbol{\Gamma} + \boldsymbol{X''}_{r,t}\boldsymbol{\Phi} + \boldsymbol{X'}_p\boldsymbol{\Omega} + \varepsilon_{p,r,t} \end{aligned} \tag{10}$$

where β is the coefficient of interest and $\boldsymbol{MPO}$ serves as a treatment intensity indicator of the influence of digital distribution. Thus, this DDD model essentially takes low-MPO prefectures as the control group.

The logic is that even if there were unobserved product changes for the digital distribution introduction, these product-related factors should have no relation with prefecture level mobile phone ownership. Therefore, if the estimated risk screening effect indeed comes from these product-related factors instead of the digital distribution introduction, we would expect no difference in the estimated results between low- and high-MPO prefectures. However, the DDD estimate, presented in Column (6) of Table IV, shows that high-MPO prefectures have significantly larger risk screening effects than low-MPO prefectures, which can only be caused by the digital distribution introduction instead of product-related factors.

By the same logic, in Column (7) we use only the policies from the prefectures with the 3-year average MPO rates ranking the lowest 10%[14] as the sample to rerun Equation (7). These prefectures can be regarded as least affected by digital distribution. If there were indeed product-related factors that lead to the average policy risk decrease, the result should still be significantly negative and similar to our main DID estimate in Column (2). However, the estimated coefficient, presented in Column (7), becomes much less substantial and insignificant at the 90% confidence level.

[14] For the 284 prefectures in our sample, we rank them by the average of the MPO rates from 2017 to 2019. The 3-year average MPO rates of the lowest 10% prefectures (28 prefectures) are below 0.49.

5 Consequences of the Risk Screening Effect

This section examines the economic consequences of 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 motivations to utilize their private risk information advantage to behave in a way that is detrimental to the insurer.

A standard test for the effect of information asymmetry is the risk-coverage relation (Cohen and Siegelman 2010). The principle states that information asymmetry incentivizes high-risk consumers to buy more insurance (adverse selection), or that 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 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 OLS estimates. The results, reported in Column (1) of Table V, 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 descriptive evidence in the bottom two rows of each panel. Since the waiting period[15] 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, the percentage of the claims during

[15] 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.

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

Table V: Examining Consequences of the Risk Screening Effect

Variables	(1) AC	(2) Ind	(3) $LnInd$	(4) $Loss\ Ratio$
Panel A. Term Life Insurance				
$digital \times cov$	-0.0001			
	(-0.24)			
$digital$	-0.0012	-144.35**		-0.0032**
	(0.45)	(-2.09)		(-2.63)
Observations	93,623	93,623		93,623
Adj. R-squared	0.032	0.015		0.013
Digital Channel Claims in Waiting Period	0.0000			
Offline Channel Claims in Waiting Period	0.0455			
Panel B. Endowment Insurance				
$digital \times cov$	-0.0010***			
	(-7.99)			
$digital$	-0.0002	-133.38**	-0.1952***	-0.0274***
	(-0.63)	(-1.99)	(-3.49)	(-3.30)
Observations	672,562	672,562	8,668	672,562
Adj. R-squared	0.027	0.024	0.398	0.002
Digital Channel Claims in Waiting Period	0.0000			
Offline Channel Claims in Waiting Period	0.0014			
Panel C. Disease Insurance				
$digital \times cov$	-0.0761***			
	(-10.01)			
$digital$	0.9807***	-1143.83**	-0.1274	-0.0599**
	(10.01)	(-2.32)	(-1.02)	(-2.33)
Observations	23,343	23,343	156	23,343
Adj. R-squared	0.136	0.069	0.860	0.052
Digital Channel Claims in Waiting Period	0.0682			
Offline Channel Claims in Waiting Period	0.0833			
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 only 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 potential 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 the indemnity amount of each policy (equal to 0 for uncompensated policies) and repeating OLS analyses. The results are shown in Column (2) of Table V. As expected, they are significant and negative for all three panels, implying an average decrease in indemnity of 144, 133 and 1144 Yuan for the term life, endowment and disease insurance products, respectively.

To understand the magnitude of the indemnity difference between digital and offline distribution, in Column (3) of Table V, we further replace the outcome variable with logarithmic indemnity and limit the samples only to compensated policies, with additional dummy controls of accident types for the endowment and disease insurance[16]. The limited sample reduces the coefficient significance for the disease insurance. The results show that even among compensated policies, the average indemnity for policies of digital distribution is still lower than offline policies by 17.7% (=$e^{-0.1952}$-1) for endowment insurance and by 12.0% (=$e^{-0.1274}$-1) for disease insurance, strengthening evidence supporting the risk screening effect.

To test the difference in profitability between digital and offline channels, we use the loss ratio as the outcome variable calculated by $\frac{Indemnity}{Premiums\ Paid}$ for each policy, where the denominator is the accumulated premiums received by the insurer. The higher the loss ratio, the lower the profitability. The results of the OLS estimates, reported in Column (4) of Table V, are significantly negative for all three panels, with the largest loss ratio reduction for the disease insurance. Overall, our evidence shows lower indemnity and higher profitability of digital distribution as a result of the risk screening effect.

[16] Here the data of the term life insurance product is not used for analysis due to very few compensated policies.

6 Who Contributes to the Risk Screening Effect?

In this section, we disentangle the extensive margin and intensive margin of the risk screening effect. Customers of the digital distribution channel are (i) *new consumers* who would not have purchased insurance and (ii) *switchers* who switched from offline channels. Correspondingly, the risk screening effect should originate from average risk changes from these two consumer sources relative to offline sales only. A distinction between these two consumer sources is important to determine whether the risk screening of digital distribution comes from attracting consumers not priorly covered (extensive margin) or from cannibalizing lower risk customers from the offline channel (intensive margin).

To make the decomposition logic more explicit, let $\boldsymbol{R}_0$ denote the average risk of offline consumers in the counterfact where there exists no digital distribution, $\boldsymbol{R}_d$ denote the average policy risk of digital channel consumers in reality 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 channel consumers can be decomposed as

$$\boldsymbol{RD} = \boldsymbol{R}_d - \boldsymbol{R}_f = (\boldsymbol{R}_d - \boldsymbol{R}_0) + (\boldsymbol{R}_0 - \boldsymbol{R}_f) \tag{11}$$

On RHS, the left expression is the average risk decrease of digital channel consumers, which is elicited by both *new consumers* and *switchers* from offline channels; the right expression is the average risk increase of retained offline consumers, which is elicited by *switchers*.

Next, we further decompose $(\boldsymbol{R}_d - \boldsymbol{R}_0)$ by *new consumers* and *switchers*. Notate that relative to the counterfact, the actual number of total policies increases by $(\boldsymbol{K} - 1)$, while the actual number of offline policies decreases by $(1 - \boldsymbol{K}_f)$. Then $\boldsymbol{K}$ is the ratio between the actual number of total policies and the counterfactual policy number, and $\boldsymbol{K}_f$ is the ratio between the actual number of offline policies and the counterfactual policy number. Through calculations (details seen in Appendix D), the contribution in ratio to $(\boldsymbol{R}_d - \boldsymbol{R}_0)$ from *switchers* is

$$\frac{\boldsymbol{K}_f}{\boldsymbol{K}-\boldsymbol{K}_f}\cdot\frac{\boldsymbol{R}_0-\boldsymbol{R}_f}{\boldsymbol{R}_d-\boldsymbol{R}_0} \tag{12}$$

Equation (12) shows that *switchers* 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 counterfact.

Thus far, we have decomposed the risk screening effect into a risk increase of offline policies and risk decreases of *new consumers* and *switchers*. To quantify this decomposition, we further exploit the previous DID framework to estimate $(\boldsymbol{R}_d-\boldsymbol{R}_0)$, $\boldsymbol{K}_f$ and $\boldsymbol{K}$. Their specifications and used samples are elaborated below and summarized in Table VI.

Estimating $\boldsymbol{RD}$.— We rerun Equation (6) but use the same fixed effects as the DID estimates to ensure consistency. The result reported in Column (1) of Table VII similarly shows a 0.20 percent points lower accident probability for digital distribution.

Estimating $(\boldsymbol{R}_d-\boldsymbol{R}_0)$.— Keep in mind that (i) $\boldsymbol{R}_d$ is the average risk of digital channel policies and (ii) the treatment group of our previous DID estimate is a mixture of offline and digital channel policies of the treatment product. Hence, to separate out the risk difference between the digital channel policies of the treatment product and the counterfact, we use the same control group, but drop out the offline policies purchased after the introduction date for the treatment group. After doing so, the treatment group consists of two components, offline policies before the introduction date and digital channel policies after the introduction date. Using this new sample to rerun Equation (7), π captures the average policy risk difference between digital channel policies and the counterfact, which corresponds to $(\boldsymbol{R}_d-\boldsymbol{R}_0)$. The result, presented in Column (2) of Table VII, shows that the introduction of digital distribution leads to an average decrease of 0.18 percent points (roughly 90% of $\boldsymbol{RD}$) in the accident probability.

Estimating $\boldsymbol{K}_f$ and $\boldsymbol{K}$.— Both estimates are based on Equation (7) but replace the dependent variable with the logarithmic daily number of purchased policies of each product in each prefecture. The specification can be written as

$$\boldsymbol{LnPolicy}_{p,r,t} = \alpha + \pi \boldsymbol{Treat}_p \times \boldsymbol{\mathcal{L}}_t + \boldsymbol{X''}_{p,r,t}\boldsymbol{\Gamma} + \boldsymbol{X''}_{r,t}\boldsymbol{\Phi} + \boldsymbol{X'}_p\boldsymbol{\Omega} + \varepsilon_{p,r,t} \quad (13)$$

where $\boldsymbol{LnPolicy}_{p,r,t}$ denotes the log number of purchased policies for product p, prefecture r and date t. The only estimate difference between $\boldsymbol{K}_f$ and $\boldsymbol{K}$ the is the sample. To estimate $\boldsymbol{K}$, the growth in total policy number, we use the full sample of the policies of both control and treatment products; while to estimate $\boldsymbol{K}_f$, the decrease in the number of offline policies, we drop the digital channel policies of treatment product from the full sample. Columns (3) and (4) in Table VII report the results of estimating $\boldsymbol{K}_f$ and $\boldsymbol{K}$ respectively. The introduction of digital distribution contributed to 122% (=$e^{0.7953}$-1) growth in the total number of policies and a slight drop in the number of offline policies by 3% (=$e^{-0.0269}$-1). From Equation (12), calculations show that nearly 91% (=1-8.7%) 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.

Taken together, most of the risk screening effect – over 81% (=90%×91%) - derives from the attracted *new consumers* with lower risk, which suggests an increased coverage of underinsured low-risk consumers and thus an improvement on market efficiency. For robustness, we also redo the above analysis using the short-term sample of 2018. The contribution ratio of *new consumers* shown in Table A6 of Appendix A is similar to that shown in Table VII.

Table VI: Samples and Specifications for Estimating $(\boldsymbol{R}_d - \boldsymbol{R}_0)$, $\boldsymbol{K}$ and $\boldsymbol{K}_f$

Estimates	**Treatment Product Sample**	**Control Product Sample**	**Specification**
Main DID Estimate	full policies	full policies	Equation (7)
$(\boldsymbol{R}_d - \boldsymbol{R}_0)$ Estimate	offline policies before the introduction date and digital channel policies after the introduction date	full policies	Equation (7)
$\boldsymbol{K}$ Estimate	full policies	full policies	Equation (13)
$\boldsymbol{K}_f$ Estimate	only offline policies both before and after the introduction date	full policies	Equation (13)

Note: The second and third columns show the policies used for the treatment and control products, respectively, for each estimate.

Table VII: Decomposed Consumer Sources of the Risk Screening Effect Using the Full Sample

	Estimating $\boldsymbol{RD}$	*Estimating* $(\boldsymbol{R}_d -$	*Estimating* $\boldsymbol{K}_f$	*Estimating* $\boldsymbol{K}$

		$\boldsymbol{R_0}$)		
Variables	(1)	(2)	(3)	(4)
$\boldsymbol{Treat \times \mathcal{L}_t}$		-0.0018***	-0.0269*	0.7953***
		(-2.90)	(-1.71)	(14.29)
$\boldsymbol{digital}$	-0.0020**			
	(-2.20)			
Observations	93,623	92,884	66,725	95,078
Adj. R-squared	0.033	0.040	0.556	0.499
$(\boldsymbol{R_d - R_0})/\boldsymbol{RD}$		90.0%		
$\dfrac{\boldsymbol{K_f}}{\boldsymbol{K - K_f}} \cdot \dfrac{\boldsymbol{R_0 - R_f}}{\boldsymbol{R_d - R_0}}$				8.7%
Controls	Y	Y	Y	Y
Fixed Effects				
Prefecture-Year	Y	Y	Y	Y
Product	N	Y	Y	Y
Date	Y	Y	Y	Y

Note: Column (1), the OLS estimate, uses the same sample as the baseline OLS specification; Column (2) 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 from 2017 to 2019; Column (3) uses the sample of the control product policies and offline policies of the treatment product from 2017 to 2019; Column (4) uses all policies of the treatment product and control product from 2017 to 2019. The second row from the bottom of Column (2) is calculated by 0.0018/0.0020; the last row of Column (4) is calculated by $[e^{-0.0269}/(e^{0.7953} - e^{-0.0269})]\times[(0.0020 - 0.0018)/0.0018]$.

7 Understanding What Drives the Risk Screening Effect

In this section, we examine the mechanisms of the risk screening effect 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 third examination focuses on *channel preference* by typically studying how advanced education correlates with the digital divide and policy risk. Finally, we compare the magnitudes of the roles between *adverse selection* and *moral hazard* as well as between *channel preference* and *capability*.

7.1 *Channel Capability*: Search Cost

Extant literature has noted and emphasized the important role of information frictions, such as information complexity (Bhargava and Manoli, 2015), insurance knowledge (Domurat et al., 2021) and automatic enrollment (Shepard and Wagner, 2022), in risk selection behaviors. Search cost is also an information friction that occurs everywhere in an economy and can cause substantial consumer welfare loss (Jolivet and Turon 2019). As implied in the settings, a remarkable advantage of digital insurance distribution is the lower search cost than that of offline insurance distribution. In theory, low-risk consumers should be more sensitive to search cost reduction than high-risk consumers due to the higher margin of insurance demand. This spurs the enrollment of low-risk consumers, particularly underinsured low-risk consumers via digital distribution. In Appendix C, we present theoretical details on this prediction with a search-based model.

To test the above prediction that the reduced search cost of digital distribution leads to policy risk screening, we examine whether the risk screening effect diminishes with lower offline insurance search costs. Offline insurance search costs are proxied in two ways: the population weighted average distance to the nearest insurer branch and the correlation between the population distribution and the insurer branch distribution.

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 (14). 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.

$$\boldsymbol{SC}_r = \frac{\sum_{k\in r} \boldsymbol{dist}_{r,k} \times \boldsymbol{pop}_{r,k}}{\sum_{k\in r} \boldsymbol{pop}_{r,k}} \tag{14}$$

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 $\boldsymbol{SC}$. Columns (1) to (3) of Table VIII report the results. These columns respectively show that an average population weighted distance to the nearest insurer's branch, the nearest and top three nearest branch of local life insurers (presented in the last row under each panel) enlarges the risk probability reduction of digital distribution by 0.34 (=-0.0033×1.0382×100), 0.46 (=-0.0138×0.3292×100) and 0.46 (=-0.0107×0.4291×100) percent points for the term life insurance, and by 0.08 (=-0.0007×1.1182×100), 0.15 (=-0.0041×0.3566×100) and 0.16 (=-0.0034×0.4628×100) percent 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 (2) of Table II as a basis. Simple calculations show that the search cost mechanism can account for 26% to 52% of the risk screening effect for the endowment insurance and even a larger share for the term life insurance product. 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 likely 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 (15).

$$\boldsymbol{SC}_r = \frac{Covariance(\boldsymbol{dist}_{r,k}, \boldsymbol{pop}_{r,k})}{\boldsymbol{AD}_r \times \boldsymbol{AP}_r} \tag{15}$$

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 the population, the more insurer branches there are, then the distance to the insurer branch should negatively correlate with the population density in that prefecture. On the contrary, the less negative the correlation, the higher the distribution bias between the population and insurer branch, indicating a higher offline search cost. Therefore, the smaller $\boldsymbol{SC}$ is in Equation (15), the lower the 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 interest. The results, reported in Columns (4) to (6) in Table VIII, 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.

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

Table VIII: Examining the Search Cost Mechanism

	Distance to Insurer Branch			Correlation between Distributions of Population and Insurer Branch		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	$\boldsymbol{AC}$	$\boldsymbol{AC}$	$\boldsymbol{AC}$	$\boldsymbol{AC}$	$\boldsymbol{AC}$	$\boldsymbol{AC}$
Panel A. Term Life Insurance						
$\boldsymbol{digital} \times \boldsymbol{SC}$	-0.0033***	-0.0138**	-0.0107**	-0.0072*	-0.0083**	-0.0077**
	(-2.99)	(-2.06)	(-2.17)	(-1.87)	(-2.08)	(-2.12)
Observations	93,623	93,623	93,623	93,623	93,623	93,623
Adj. R-squared	0.019	0.019	0.019	0.018	0.018	0.018
Mean of $\boldsymbol{SC}$, Obs.	1.0398	0.3296	0.4297	-0.4428	-0.5507	-0.5296
Panel B. Endowment Insurance						
$\boldsymbol{digital} \times \boldsymbol{SC}$	-0.0007*	-0.0041**	-0.0034**	-0.0058***	-0.0090***	-0.0088***
	(-1.84)	(-2.26)	(-2.31)	(-2.87)	(-3.68)	(-3.71)
Observations	672,562	672,562	672,562	672,562	672,562	672,562
Adj. R-squared	0.028	0.029	0.029	0.029	0.029	0.029
Mean of $\boldsymbol{SC}$, Obs.	1.1182	0.3566	0.4628	-0.4133	-0.5152	-0.4947
Controls	Y	Y	Y	Y	Y	Y
Fixed Effects						
Prefecture-Year	Y	Y	Y	Y	Y	Y
Month	Y	Y	Y	Y	Y	Y
Day-in-Month	Y	Y	Y	Y	Y	Y
Day-in-Week	Y	Y	Y	Y	Y	Y

Note: The search cost proxy uses 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. Mean of $\boldsymbol{SC}$ denote the average population weighted distance to the nearest insurer's branch, the nearest and top three nearest branch of local life insurers across policies.

7.2 *Channel Capability*: Rigor of Implementing Underwriting Rules

Another reason for the risk screening effect is the difference in the rigor of implementing underwriting rules between digital and offline distribution. There are two channel features that may enhance this difference. First, most offline agents in the Chinese life insurance industry are employed by insurers, are not independent and are renumerated on policy commissions. It is possible that they do not strictly comply with underwriting rules in pursuit of more sales and commissions, leading to underwriting 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. Second, offline agents have higher discretion, making it possible for their underwriting decisions to be affected by personal relations (or other factors), generating greater volatility in underwriting risk control. For example, friendship between offline agents and unqualified consumers may lead to the wrong approval of underwriting, which, however, is unlikely to occur when consumers face machines (e.g., APPs). In a broader sense, irrespective of discretion differences or commission incentives, their roles are essentially the consequences of imperfect supervision, which matters for insurers when dealing with offline agents but is absent when dealing with machines.

If the aforementioned channel features exist, claims of offline policies should be more likely to be rejected than those of digital channel policies due to ineligibility for insurance. Based on this tenet, we investigate the effect of digital distribution on claim rejection.

We sample 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 Columns (1) and (3) of Table IX, showing that the rejection probability of the offline policies is 3.55 and 11.13 percent points higher than that of digital channel policies for the endowment and disease insurance, respectively. We also exclude the claims rejected for reasons unrelated to ineligibility and redo the same analyses. The results are presented in Columns (2) and (4) of Table IX. We find that they are qualitatively consistent but insignificant for the disease insurance possibly due to the limited subsample size.

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 less rigorous implementation of underwriting rules by offline agents, resulting in underwriting more ineligible consumers with higher 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 compensation on behalf of policyholders than direct underwriters such as digital distribution. There are two explanations for 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 lower quality services (Eckardt and Räthke 2010; Focht et al. 2013), such as misleading sales.

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

	Endowment Insurance		Disease Insurance	
	(1)	(2)	(3)	(4)
Variables	$\boldsymbol{Rej}$	$\boldsymbol{Rej}$	$\boldsymbol{Rej}$	$\boldsymbol{Rej}$
$\boldsymbol{digital}$	-0.0355***	-0.0184***	-0.1113**	-0.0529
	(-5.30)	(-4.39)	(-2.03)	(-1.30)
Observations	9,178	8,841	322	244
Adj. R-squared	0.096	0.085	0.218	0.139
Rejection for Ineligibility on Digital Channels	0.3333		0.2174	
Rejection for Ineligibility on Offline Channels	0.4364		0.2727	
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 each insurance product, Columns (1) and (3) both use all claimed policies, Columns (2) and (4) both drop the claim rejections due to the reasons unrelated to ineligibility. Here, the term life insurance is not analyzed due to its small size of claimed policies. For the same reason, Columns (3) and (4) only control Province-Year fixed effects with robust standard errors clustered at the province level. While other columns keep Prefecture-Year fixed effects with robust standard errors clustered at the prefecture level.

7.3 *Channel Preference*

Our last examination is based on *channel preference*. Although there are many potential factors that determine the ability and acceptance to use digital technology, we choose education level for two reasons. First, education level is a typical risk characteristic not adjusted into the unit premium by the insurer. This is evidenced by the insignificant coefficients of advanced education in Table B1 of Appendix B, which holds for all three investigated products. Second, there is much literature supporting the positive relationship between education and internet use (e.g., 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

$$\boldsymbol{digital}_{i,r,t} = \alpha + \beta \boldsymbol{edu}_{i,r,t} + \theta \boldsymbol{D}_t + \boldsymbol{X'}_{i,r,t}\boldsymbol{\Gamma} + \boldsymbol{X'}_{r,t}\boldsymbol{\Omega} + \varepsilon_{i,r,t} \tag{16}$$

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

where $\boldsymbol{edu}_{i,r,t}$ is the dummy of advanced education. The controls (except education) and fixed effects are the same as in Equation (6).

The results of Equations (16) and (17) are presented in Table X. As anticipated, after controlling for pricing and all observed characteristics, advanced education has a significantly positive relationship with the digital channel choice and a significantly negative relationship with policy risk. Specifically, 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 products respectively. Taken together, we show that consumers with advanced education prefer to use the digital channel while having lower policy risk than those with lower education levels. Education is not adjusted into pricing but still observed by the insurer. In Appendix E.II, we also use income, a typical risk characteristic unobserved by the insurer, as an instance to verify the role of *channel preference* based on a survey data.

Table X: Relationships between Channel Choice, Risk and Advanced Education

	Term Life Insurance		Endowment Insurance		Disease Insurance	
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	$digital$	AC	$digital$	AC	$digital$	AC
edu	0.0100***	-0.0007***	0.2538***	-0.0043***	0.0486***	-0.0013*
	(3.19)	(-2.71)	(10.04)	(-13.07)	(5.70)	(-1.70)
Observations	93,623	93,623	672,562	672,562	23,343	23,343
Adj. R-squared	0.641	0.018	0.163	0.028	0.280	0.052
Controls	Y	Y	Y	Y	Y	Y
Fixed Effects						
Prefecture-Year	Y	Y	Y	Y	Y	Y
Month	Y	Y	Y	Y	Y	Y
Day-in-Month	Y	Y	Y	Y	Y	Y
Day-in-Week	Y	Y	Y	Y	Y	Y

Note: for each product, the left and the right columns respectively use the specification of Equations (13) and (14).

7.4 Further Discussion

Adverse Selection vs. *Moral Hazard.*— In theory, the engagement between digital distribution and information asymmetry can also be decomposed by adverse selection and moral hazard. The reason for engagement with adverse selection is obviously manifested in *channel capability*; while *Channel preference* is likely to produce engagement with moral hazard, because the ability or acceptance to use digital channels not only affects channel choice but also may correlate with health habits. For example, people with higher abilities to use digital channels are also more able to learn the free health knowledge sharing over the mobile internet, which may improve their health habits and moral hazard behaviors better than those with lower abilities. However, due to data limitations, we cannot quantitatively achieve this decomposition which also makes sense for understanding risk screening mechanisms. In this regard, a qualitative inference could be that our estimated effects reflect more severe adverse selection than moral hazard because compared to the insurance period, the maturity of policies in our data is relatively short; for instance, the term life insurance policies sold on the digital channel were generally 2 to 3 years old as of the data collection. Such a short maturity makes the risk screening effect less likely to be part of a moral hazard story because it takes time for moral hazard behaviors to effectively change risk.

Channel Preference vs. *Channel Capability.*— A natural question following the above tests that verify the roles of *channel capability* and *preference* is which of them plays a larger role? It is difficult to clearly decompose them, but one way to gain an insight is to separate out the influence of *channel preference* in the case where the ability or acceptance to use digital channels is less likely to correlate with policy risk. Note that the purchase channel choice is made by policyholders, while the policy risk is up to the insured. Then a reasonable expectation is that when the policyholder is the insured, policy risk does correlate with the ability or acceptance to use digital channels; however, when the policyholder is not the insured, this correlation is likely to be weak or even absent, because the digital ability or acceptance of policyholders is almost irrelevant to the status of another person – the insured. Under this assumption, the risk screening effect should come from both *channel preference* and *capability* when the policyholder is the insured; while the risk screening effect should mainly come from *channel capability* when the policyholder and the insured differ. Therefore, we can qualitatively assess the magnitude of *channel capability* and *preference* by comparing the risk screening effects between oneself and non-oneself relation policies – the effect difference should reflect the role of *channel preference*.

Such analyses require controls of the risk characteristics of the insured for non-oneself relation policies. To overcome the data limitation that we only have policyholder characteristics, we limit the non-oneself relation policies to the couple-relation policies of the endowment insurance. The unique product setting of the endowment insurance - covering the insured until 75 years old - allows insurance period to reflect the insured's age, and couple-relation allows the policyholder's gender to reflect the insured's gender. In this way, we ensure the same controls of risk factors of pricing for OLS estimates of both oneself- and couple-relation policies and thus their results are comparable.

Using the same specification as in Equation (6), we report the results in Table E1, Appendix E.I. Column (1) shows an even larger risk screening effect for couple-relation policies than the result in Column (2), Panel B of Table II. To test the difference significance, we redo the analysis for the sample combining oneself- and couple-relation policies with adding an interaction between channel choice and couple-relation (specification details seen in Appendix E). The interaction has an insignificantly negative coefficient, indicating no significant difference in risk screening effect between oneself- and couple-relation policies. Regardless of negative coefficients or insignificant difference, they imply that the role of *channel preference* is negligible and much smaller than that of *channel capability*. Therefore, our documented risk screening effect has little dependence on the heterogeneity in people's ability or acceptance to use digital technology. This is consistent with our finding in section 6 that the risk screening effect mainly comes from the extensive margin.

8 Implication and Conclusion

Using a unique dataset of the term life, endowment and disease insurance products sold via 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 has important economic consequences by reducing information asymmetry, lowering the average indemnity and increasing the profitability of digital distribution.

In addition, we decompose consumer sources of the risk screening effect. We find that the risk screening of digital distribution mainly comes from improvement on market efficiency. At least 81% of the risk screening effect is attributed lower-risk, *new consumers* who were previously not covered, with only a small part attributable to the risk increases in offline policies due to crowded-out, low-risk *switchers*.

We theoretically and empirically show three mechanisms of the risk screening effect from the roles of *channel capability* and *channel preference*. First, advantageous channel features of digital distribution, such as reduced search costs, have higher marginal incentives to the insurance demand of low-risk consumers than to those of high-risk consumers. Second, channel features that relate to risk control, such as commission institution, may distort the rigor of implementing underwriting rules. Third, the ability or acceptance to use digital channels may positively correlate with unobserved policy risk via the risk characteristics not adjusted into pricing, such as advanced education. Our evidence also suggests that *channel capability* plays a dominantly larger role than that of *channel preference* and therefore the risk screening effect persists regardless of digital divide.

There are three significant implications of this article. First, an actuarial implication is that channel strategies should be taken as a factor of pricing for insurance products due to the risk profile difference. Our finding of lower unobserved policy risk for digital distribution than for offline distribution provides a new explanation for the well-documented phenomenon that the price of internet insurance products is usually set lower than that of homogeneous offline products (Brown and Gollsbee, 2002; Pauly et al., 2023; Chen et al., 2023). Second, 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, so we can likely apply similar principles to the associated effects on the stock market. Third, this article indicates lowering search costs as a new way to mitigate adverse selection for insurance and other industries suffering from adverse selection. Our findings suggest that measures to reduce search costs, such as digital distribution, improve the risk profile by creating higher incentives for low-risk consumers.

Two data-related caveats of this article deserve note. First, the data used in this article are sourced from a large Chinese life insurer and the persistence of the estimated effects for insurers in other countries remains an open question. Second, since digital sales of insurance products is in its early stages, the limited maturity of the claim tail is inevitable and not an isolated case. This makes our estimated risk screening effect more likely to be inferred as adverse selection than as moral hazard, which leaves the engagement between digital distribution and moral hazard for future research when the data will have accumulated for longer.

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.

Aguiar, L., & Waldfogel, J. (2018). As streaming reaches flood stage, does it stimulate or depress music sales?. International Journal of Industrial Organization, 57, 278-307.

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

Bakos, Y., & Brynjolfsson, E. (2000). Bundling and Competition on the Inter-net. Marketing science, 19(1), 63-82.

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.

Chen, A., Chen, Y., Murphy, F., Xu, W., & Xu, X. (2023). How does the insurer's mobile application sales strategy perform? *Journal of Risk and Insurance Review*, 1– 33.

Chiappori, P. A., & Salanie, B. (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy*, *108*(1), 56-78.

Cohen, A. (2005). Asymmetric information and learning: Evidence from the automobile insurance market. *Review of Economics and statistics*, *87*(2), 197-207.

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.

Dionne, G., Gouriéroux, C., & Vanasse, C. (2001). Testing for evidence of adverse selection in the automobile insurance market: A comment. *Journal of Political Economy*, *109*(2), 444-453.

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.

Kim, H., Kim, D., Im, S., & Hardin, J. W. (2009). Evidence of asymmetric information in the automobile insurance market: dichotomous versus multinomial measurement of insurance coverage. *Journal of Risk and Insurance*, *76*(2), 343-366.

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.

Pauly, M. V., Herring, B., & Song, D. (2006). Information technology and consumer search for health insurance. *International Journal of the Economics of Business*, *13*(1), 45-63.

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.

Spindler, M., Winter, J., & Hagmayer, S. (2014). Asymmetric information in the market for automobile insurance: Evidence from Germany. *Journal of Risk and Insurance*, *81*(4), 781-801.

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.

Appendix

Appendix A

Tables

Table A1: Comparing Plan Settings between Digital and Offline Distribution Channels

	Digital distribution		Offline distribution	
Plan Settings	Minimum	Maximum	Minimum	Maximum
Panel A. Treatment Term Life Insurance Product				
Waiting period (in days)	180	180	180	180
Hesitation period (in days)	20	20	20	20
Age	18	65	18	65
Gender	0	1	0	1
Insurance period (in years)	20	30	20	30
Face value of insured amount (in thousand Yuan)	5.2	10000	1.1	10000
Additional liability	0	0	0	0
Payment term (in years)	10	20	10	20
Panel B. Control Term Life Insurance Product				
Waiting period (in days)	-	-	180	180
Hesitation period (in days)	-	-	14	14
Age	-	-	18	64
Gender	-	-	0	1
Insurance period (in years)	-	-	10	15
Face value of insured amount (in thousand Yuan)	-	-	2.4	10000
Additional liability	-	-	0	0
Payment term (in years)	-	-	10	20
Panel C. Endowment Insurance				
Waiting period (in days)	180	180	180	180
Hesitation period (in days)	20	20	20	20
Age	18	67	18	70
Gender	0	1	0	1
Insurance period (in years)	20	57	20	57
Face value of insured amount (in thousand Yuan)	2.5	2500	2.5	3000
Additional liability	0	1	0	1
Payment term (in years)	20	20	20	20
Panel D. Disease Insurance				
Waiting period (in days)	365	365	365	365
Hesitation period (in days)	14	14	14	14
Age	19	64	19	69
Gender	0	1	0	1
Insurance period (in years)	Whole life	Whole life	Whole life	While life
Face value of insured amount (in thousand Yuan)	12.8	1200	39.3	1200
Additional liability	0	1	0	1
Payment term (in years)	10	20	10	20

Note: Descriptive statistics of the investigated endowment, disease and term life insurance products are based on the total policies sold during the periods with both digital and offline distribution. Descriptive statistics of the control term life insurance product are based on total policies sold from 2017 to 2019.

Table A2: Interperiod Comparisons of Plan Settings of Offline Treatment Term Life Insurance Policies

Plan Settings	Before Introduction Date		After Introduction Date	
	Minimum	Maximum	Minimum	Maximum
Waiting period (in days)	180	180	180	180
Hesitation period (in days)	20	20	20	20
Age	20	65	18	65
Gender	0	1	0	1
Insurance period (in years)	20	30	20	30
Face value of insured amount (in thousand Yuan)	1.2	9000	1.1	10000
Additional liability	0	0	0	0
Payment term (in years)	10	20	10	20

Note: This table can be connected to Panel A in Table A1 to compare plan settings among three policy groups of the treatment product: offline policies before the introduction date, offline policies post the introduction date and digital channel policies.

Table A3: Description of Prefecture Level Data

Variables	Mean	Std. Dev.	Obs.
Distance to the insurer's nearest branch per capita (in 10 kilometers)	1.974	1.449	326
Distance to the nearest local life insurer branch per capita (in 10 kilometers)	0.583	0.348	326
Average distance to the top 3 nearest local life insurer branches per capita (in 10 kilometers)	0.745	0.420	326
Correlation between population distribution and the insurer's nearest branch distribution	-0.309	0.191	326
Correlation between population distribution and the nearest local life insurer branch distribution	-0.451	0.157	326
Correlation between population distribution and the distribution of the average distance to the top 3 nearest local life insurer branches	-0.427	0.159	326
Mobile Phone Ownership Per 100 Population	1.099	0.335	978

Note: The measures from the first three rows that index offline insurance search cost are calculated by Equation (14) and the measures from rows 4 to 6, calculated by Equation (15). The mobile phone ownership is the number of mobile phones per capita for each prefecture in each year from 2017 to 2019. Since all these measures are prefecture level data, here their means and standard errors are also reported at the prefecture level.

Table A4: Summary of Controls of OLS and DID Specifications

Controls	Description	OLS			DID
		Term Life Insurance	Endowment Insurance	Disease Insurance	
Age	Dummies of policyholder age for OLS; Percentages of each age across policies per day, prefecture and product for DID.	45	53	52	45
Female	A dummy of the female policyholder for OLS; A percentage of female policyholders across policies per day, prefecture and product for DID.	1	1	1	1
Age×Female	Interactions of age and gender dummies for OLS; Percentages of each interaction across policies per day, prefecture and product for DID.	89	104	99	89
Unit Premium	Unit premium of each policy for OLS; Average unit premium across policies per day, prefecture and product for DID.	1	1	1	1
Insurance Period	Dummies of insurance period for OLS; Absorbed by product fixed effect for DID	2	38	-	-
Payment Term	Dummies of payment term for OLS; Percentages of each payment term across policies per day, prefecture and product for DID.	2	-	2	2
Rider	A dummy indicating policies with riders for OLS; A percentage of policies with riders per day, prefecture and product for DID.	-	1	1	-
Financial Profession	A dummy indicating work-in-finance policyholders for OLS; A percentage of policies with work-in-finance policyholders across policies per day, prefecture and product for DID.	1	1	1	1
Education Levels	Dummies of policyholder education levels for OLS; Percentages of each education level across policies per day, prefecture and product for DID.	10	10	10	10
Log Insured Amount	Logarithmic insured amount for OLS; A logarithm of the average unit premium across policies per day, prefecture and product for DID.	1	1	1	1

Note: This table shows the number of controls corresponding to OLS and DID specifications. The treatment and control term life insurance products have no riders, the endowment insurance product has a unified premium payment term and the disease insurance product insures whole life. The insurance period dummies of term life insurance products have been absorbed by the product fixed effect in the DID specification.

Table A5: Results of Logit and Cloglog Regressions

	Logit Regressions				Cloglog Regressions			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	AC	AC	AC	AC	AC	AC	AC	AC
Panel A. Term Life Insurance								
$digital$	-1.1041***	-1.3412***	-1.3412***	-1.1973***	-1.0952***	-1.3365***	-1.3365***	-1.1955***
	(-3.02)	(-3.65)	(-3.65)	(-3.06)	(-3.05)	(-3.69)	(-3.69)	(-3.09)
Observations	93,623	93,623	93,206	93,592	93,623	93,623	93,206	93,592
Marginal Effect	-0.0025***	-0.0030***	-0.0030***	-0.0026***	-0.0025***	-0.0031***	-0.0031***	-0.0027***
	(-3.04)	(-3.67)	(-3.68)	(-3.02)	(-3.05)	(-3.69)	(-3.69)	(-3.08)
Panel B. Endowment Insurance								
$digital$	-0.3418***	-0.2875***	-0.2850***	-0.2476***	-0.3366***	-0.2820***	-0.2795***	-0.2432***
	(-7.87)	(-6.56)	(-6.41)	(-5.66)	(-7.86)	(-6.54)	(-6.40)	(-5.66)
Observations	672,562	672,562	665,247	672,053	672,562	672,562	665,247	672,053
Marginal Effect	-0.0046***	-0.0039***	-0.0038***	-0.0032***	-0.0046***	-0.0039***	-0.0039***	-0.0032***
	(-7.88)	(-6.57)	(-6.42)	(-5.67)	(-7.88)	(-6.55)	(-6.40)	(-5.66)
Panel C. Disease Insurance								
$digital$	-0.9833***	-0.4447**	-0.4504**	-0.4267*	-0.9725***	-0.4161**	-0.4187**	-0.4019**
	(-5.85)	(-2.06)	(-2.08)	(-1.71)	(-5.96)	(-2.02)	(-2.03)	(-2.53)
Observations	23,343	23,343	23,314	23,290	23,343	23,343	23,314	23,290
Marginal Effect	-0.0112***	-0.0039**	-0.0038***	-0.0029*	-0.0114***	-0.0039**	-0.0039**	-0.0034**
	(-6.00)	(-2.04)	(-2.07)	(-1.70)	(-6.03)	(-2.01)	(-2.02)	(-2.56)
Controls	controls of pricing	Y	Y	Y	controls of pricing	Y	Y	Y
Fixed Effects								
Province-Year	Y	Y	Y	Y	Y	Y	Y	Y
Month	Y	Y	Y	Y	Y	Y	Y	Y
Day-in-Month	Y	Y	Y	Y	Y	Y	Y	Y
Day-in-Week	Y	Y	Y	Y	Y	Y	Y	Y

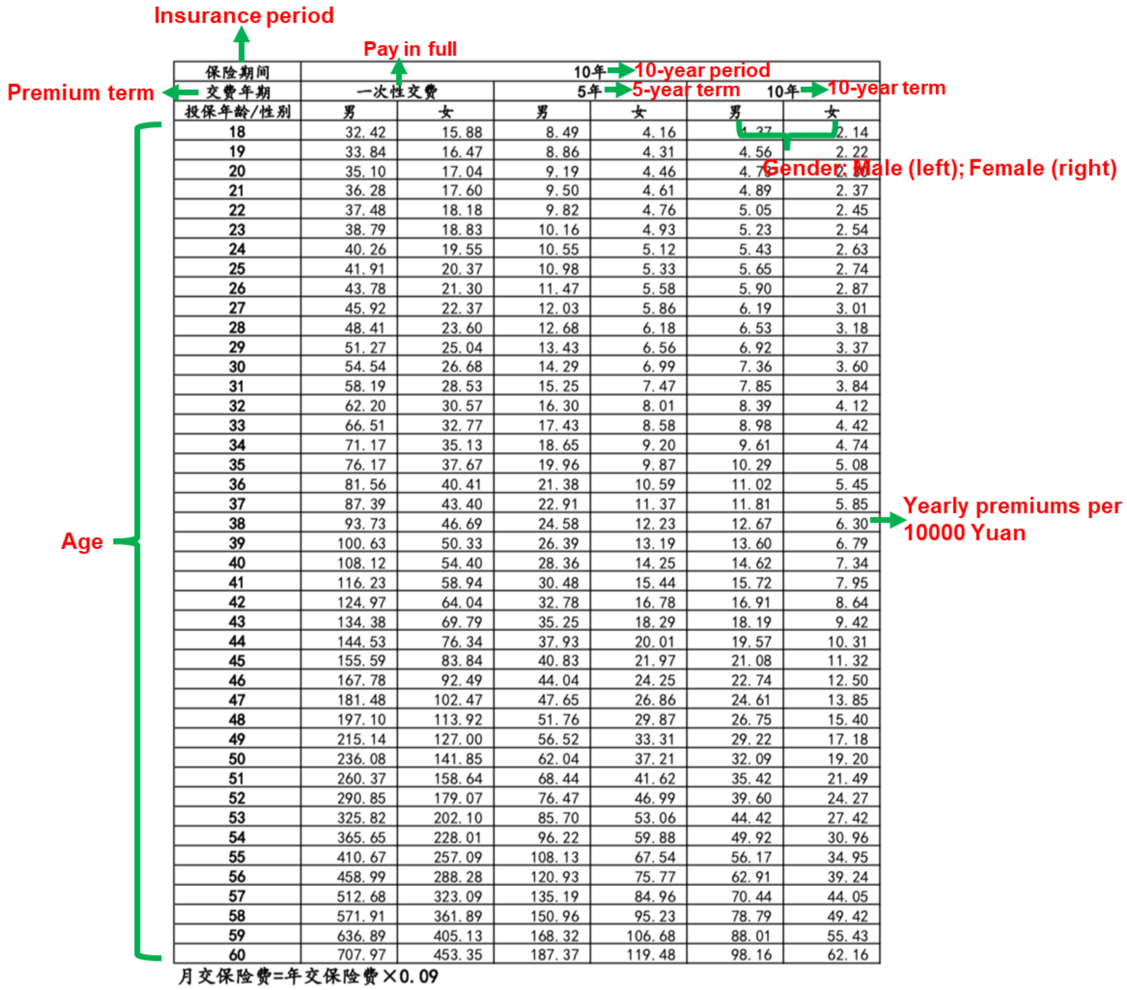
Note: In this table, all regressions use fixed Province-Year effects instead of Prefecture-Year effects, otherwise many observations would be dropped due to singleton in Logit and Cloglog regressions. Columns (1), (2), (5) and (6) employ the full sample, Columns (3) and (7) exclude cancelled policies, Columns (4) and (8) exclude claim-rejected policies respectively. Columns (1) and (5) use only controls of pricing – the policyholder age dummies, gender dummies, dummies of age-gender interactions, and unit premium. Other columns use full controls. Robust z-statistics in parentheses with standard errors clustered at the province level and *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Decomposed Consumer Sources of the Risk Screening Effect Using Only 2018 Policies

	Estimating $\boldsymbol{RD}$	*Estimating* $(\boldsymbol{R_d} - \boldsymbol{R_0})$	*Estimating* $\boldsymbol{K_f}$	*Estimating* $\boldsymbol{K}$
Variables	(1)	(2)	(3)	(4)
$\boldsymbol{Treat} \times \boldsymbol{\mathcal{L}_t}$		-0.0028**	-0.0391*	0.8735***
		(-2.49)	(-1.89)	(14.92)
$\boldsymbol{digital}$	-0.0030**			
	(-2.00)			
Observations	50,962	39,995	27,041	41,295
Adj. R-squared	0.010	0.031	0.478	0.423
$(\boldsymbol{R_d} - \boldsymbol{R_0})/\boldsymbol{RD}$		93.3%		
$\frac{\boldsymbol{K_f}}{\boldsymbol{K} - \boldsymbol{K_f}} \cdot \frac{\boldsymbol{R_0} - \boldsymbol{R_f}}{\boldsymbol{R_d} - \boldsymbol{R_0}}$				4.8%
Controls	Y	Y	Y	Y
Fixed Effects				
Prefecture-Year	Y	Y	Y	Y
Product	N	Y	Y	Y
Date	Y	Y	Y	Y

Note: Column (1), the OLS estimate, uses the same OLS specification as in Equation (6) for treatment product policies after the introduction date in 2018; Column (2) 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 (3) uses the sample of the control product policies and offline policies of the treatment product in 2018; Column (4) uses all policies of the treatment product and control product in 2018. The second row from the bottom of Column (2) is calculated by 0.0028/0.0030; The last row of Column (4) is calculated by $[e^{-0.0391}/(e^{0.8735} - e^{-0.0391})] \times [(0.0030 - 0.0028)/0.0028]$. Robust t-statistics in parentheses with standard errors clustered at the prefecture level and *** p<0.01, ** p<0.05, * p<0.1.

Figures



保险期间	10年					
交费年期	一次性交费		5年		10年	
投保年龄/性别	男	女	男	女	男	女
18	32.42	15.88	8.49	4.16	4.37	2.14
19	33.84	16.47	8.86	4.31	4.56	2.22
20	35.10	17.04	9.19	4.46	[illegible]	[illegible]
21	36.28	17.60	9.50	4.61	4.89	2.37
22	37.48	18.18	9.82	4.76	5.05	2.45
23	38.79	18.83	10.16	4.93	5.23	2.54
24	40.26	19.55	10.55	5.12	5.43	2.63
25	41.91	20.37	10.98	5.33	5.65	2.74
26	43.78	21.30	11.47	5.58	5.90	2.87
27	45.92	22.37	12.03	5.86	6.19	3.01
28	48.41	23.60	12.68	6.18	6.53	3.18
29	51.27	25.04	13.43	6.56	6.92	3.37
30	54.54	26.68	14.29	6.99	7.36	3.60
31	58.19	28.53	15.25	7.47	7.85	3.84
32	62.20	30.57	16.30	8.01	8.39	4.12
33	66.51	32.77	17.43	8.58	8.98	4.42
34	71.17	35.13	18.65	9.20	9.61	4.74
35	76.17	37.67	19.96	9.87	10.29	5.08
36	81.56	40.41	21.38	10.59	11.02	5.45
37	87.39	43.40	22.91	11.37	11.81	5.85
38	93.73	46.69	24.58	12.23	12.67	6.30
39	100.63	50.33	26.39	13.19	13.60	6.79
40	108.12	54.40	28.36	14.25	14.62	7.34
41	116.23	58.94	30.48	15.44	15.72	7.95
42	124.97	64.04	32.78	16.78	16.91	8.64
43	134.38	69.79	35.25	18.29	18.19	9.42
44	144.53	76.34	37.93	20.01	19.57	10.31
45	155.59	83.84	40.83	21.97	21.08	11.32
46	167.78	92.49	44.04	24.25	22.74	12.50
47	181.48	102.47	47.65	26.86	24.61	13.85
48	197.10	113.92	51.76	29.87	26.75	15.40
49	215.14	127.00	56.52	33.31	29.22	17.18
50	236.08	141.85	62.04	37.21	32.09	19.20
51	260.37	158.64	68.44	41.62	35.42	21.49
52	290.85	179.07	76.47	46.99	39.60	24.27
53	325.82	202.10	85.70	53.06	44.42	27.42
54	365.65	228.01	96.22	59.88	49.92	30.96
55	410.67	257.09	108.13	67.54	56.17	34.95
56	458.99	288.28	120.93	75.77	62.91	39.24
57	512.68	323.09	135.19	84.96	70.44	44.05
58	571.91	361.89	150.96	95.23	78.79	49.42
59	636.89	405.13	168.32	106.68	88.01	55.43
60	707.97	453.35	187.37	119.48	98.16	62.16

月交保险费=年交保险费×0.09

Figure A1: A Screenshot of a Practical Premium Rate Table for Quotations of Term Life Insurance

Note: The table in this figure shows the premium quotations per 10000 Yuan of insured amount for a term life insurance product on sale via the insurer's mobile APP. It can be seen that the premium quotations are classified by age and gender with fixed insurance period (10-year) and payment term (pay-in-full, 5- or 10-year).

Appendix B

To validate the age-gender-specific pricing practice, we examine the risk factors affecting unit premiums for each product with the below empirical specification at the individual policy level,

$$\boldsymbol{UP}_{i,r,t} = \alpha + \beta \boldsymbol{age}_{i,r,t} + \theta \boldsymbol{gender}_{i,r,t} + \sigma \boldsymbol{age}_{i,r,t} \times \boldsymbol{gender}_{i,r,t} + \boldsymbol{IP}_{i,r,t} + \boldsymbol{PT}_{i,r,t} + \boldsymbol{Rider}_{i,r,t} + \varepsilon_{i,r,t}$$

where i, r and t index policyholder, prefecture and purchase date, respectively. $\boldsymbol{UP}$ denotes the unit premium of the policy. $\boldsymbol{age}$ denotes the policyholder age, $\boldsymbol{gender}$ takes 1 for female policyholders and otherwise 0, with $\boldsymbol{age} \times \boldsymbol{gender}$ as their interaction. $\boldsymbol{IP}$ is a vector including dummies of insurance period and $\boldsymbol{PT}$, a vector including dummies of premium payment term. $\boldsymbol{Rider}$ is a dummy of whether the policy has riders. Robust standard errors are clustered at the prefecture level. We use the same samples as in the baseline OLS estimates in Column (2), Table II.

The results in Column (1) of Table B1 show that unsurprisingly, unit premiums are positively correlated with age for all three products, but have correlations with gender which vary with age. Specifically, females always have lower unit premiums than males for the term life insurance; for the endowment insurance, unit premiums of the female are lower under 36 (=0.0072/0.00021) but higher above 36 years old than the male; while for the disease insurance, unit premiums of the female are higher under 35 (=0.0046/0.00013) but lower above 35 years old than the male. Notably, the significance of their coefficients is all very high with t-statistics above 3.2. The great explanatory power of age and gender makes the R squares as substantial as above 0.87, 0.72 and 0.92 for the term life, endowment and disease insurance products, respectively. In Column (2), we further add into regressions additional observed policyholder and policy characteristics including advanced education, financial profession, log insured amount and policy status. Specifically, $\boldsymbol{edu}$ is a dummy of whether the policyholder has an education level of or above undergraduate; $\boldsymbol{finance}$ is a dummy of whether the policyholder works in the financial industry; $\boldsymbol{cov}$ denotes logarithmic insured amount; $\boldsymbol{cancel}$ is a dummy indicating policy cancellation. As shown, most coefficients of age, gender and age×gender as well as R squares change little, while none of additionally added characteristics have significant effects except for the log insured amount with the endowment insurance (only significant at the 90% confidence level).

In Column (3), we redo the estimate in Column (2) by further refining age and gender into dummies of age, dummies of gender and interactions of these dummies (totaling 135, 158, 152 dummies for the term life, endowment and disease insurance products respectively). The R squares increase by 0.003 to 0.03 due to refined dummies of age and gender, however, most additional characteristics still have no significant effects. Again, this confirms the age-gender-specific pricing practice where age and gender dominate pricing.

Table B1: Testing for the Risk Factors Affecting Unit Premium

	Term Life Insurance			Endowment Insurance			Disease Insurance		
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\boldsymbol{age}$	0.0010***	0.0009***		0.0023***	0.0023***		0.0006***	0.0006***	
	(11.14)	(9.22)		(57.48)	(57.42)		(99.13)	(98.49)	
$\boldsymbol{gender}$	-0.0130***	-0.0098***		-0.0071***	-0.0071***		0.0046***	0.0046***	
	(-4.32)	(-3.27)		(-8.82)	(-8.85)		(23.28)	(23.12)	
$\boldsymbol{edu \times gender}$	-0.0002***	-0.0001***		0.0002***	0.0002***		-0.0001***	-0.0001***	
	(-5.38)	(-3.44)		(10.89)	(11.05)		(-23.56)	(-23.73)	
$\boldsymbol{edu}$		0.0007	0.0007		0.0085	0.0056		0.00003	-0.00001
		(0.95)	(-0.99)		(1.47)	(1.31)		(1.34)	(-0.04)
$\boldsymbol{finance}$		-0.0030	-0.0025		0.0007	0.0002		-0.00002	-0.00001
		(-0.86)	(-0.7)		(1.03)	(0.17)		(-1.11)	(-0.52)
$\boldsymbol{cancel}$		-0.0073	-0.0069		-0.0317	-0.0327		(0.0007)	0.0010
		(-1.01)	(-0.93)		(-1.36)	(-1.57)		(0.52)	(0.70)
$\boldsymbol{cov}$		0.0012	0.0012		-0.0688*	-0.0683*		-0.0004	-0.0003
		(1.38)	(1.32)		(-1.88)	(-1.76)		(-1.57)	(-1.42)
Controls	Dummies of insurance period, rider and premium payment term.								
Observations	93,623	93,623	93,623	672,562	672,562	672,562	23,343	23,343	23,343
Adj. R-squared	0.876	0.877	0.880	0.722	0.722	0.732	0.922	0.923	0.952

Note: Robust t-statistics in parentheses with standard errors clustered at the prefecture level and *** p<0.01, ** p<0.05, * p<0.1.

Appendix C

This appendix theoretically supports the role of search cost reduction of digital distribution into the risk screening effect. 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 and Ψ 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[17] 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[18]. 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

$$\mathrm{U} = u(y - \lambda\Phi L) - u(y) + \alpha\Phi q^{RA} u(L) \tag{A1}$$

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

$$E(\mathrm{U}) = \int_{\lambda} max\{\mathrm{U}, 0\}\, d\Psi(\lambda) \tag{A2}$$

Prior to receiving product information, the consumer does not know the actual loading

[17] 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.
[18] This is without loss of generality as the following proof does not rely on variations in the loss rate.

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[19]

$$E(\mathrm{U}) > s \tag{A3}$$

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

$$\mathrm{U}(q^{RA} \mid \lambda = \lambda^r) > 0 \tag{A4}$$

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

$$\mathrm{Prob}(\varphi(s, q^{RA}) > 0) \tag{A5}$$

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

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 (A5), the expected policy risk of this income group is

$$E\left(q^{RA} \mid q^{RA} > \varphi_q^{-1}(s)\right) \tag{A6}$$

[19] Equation (A6) 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.

Obviously, Equation (A6) 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.

Appendix D

This appendix calculates how much the offline crowded-out consumers contribute to the average risk decreases of the digital distribution channel.

Let $\boldsymbol{N}_d$ denote the actual number of digital channel policies and $\boldsymbol{N}_0$ denote the total number of policies in the counterfact. Then relative to the counterfact, the actual total number of policies increases by $(\boldsymbol{K} - 1)$ among which the actual number of offline policies $\boldsymbol{N}_f$ decreases by $(1 - \boldsymbol{K}_f)$. Using $\boldsymbol{R}_{new}$ to denote the average policy risk of new consumers and $\boldsymbol{R}_{switch}$ to denote the average policy risk of switchers, we have

$$\boldsymbol{N}_d \cdot \boldsymbol{R}_d = [\boldsymbol{N}_d - (\boldsymbol{N}_0 - \boldsymbol{N}_f)] \cdot \boldsymbol{R}_{new} + (\boldsymbol{N}_0 - \boldsymbol{N}_f) \cdot \boldsymbol{R}_{switch} \tag{A7}$$

$$\boldsymbol{N}_0 \cdot \boldsymbol{R}_0 = (\boldsymbol{N}_0 - \boldsymbol{N}_f) \cdot \boldsymbol{R}_{switch} + \boldsymbol{N}_f \cdot \boldsymbol{R}_f \tag{A8}$$

where $\boldsymbol{N}_f = \boldsymbol{K}_f \cdot \boldsymbol{N}_0$ and $\boldsymbol{N}_d = (\boldsymbol{K} - \boldsymbol{K}_f) \cdot \boldsymbol{N}_0$. On the RHS in Equation (A7), the left component shows the contribution of the new consumers to the average risk decrease of digital distribution $(\boldsymbol{R}_d - \boldsymbol{R}_0)$ and the right component, the contribution of switchers who switched from offline. Plugging Equations (A7) and (A8) into Equation (11) in the main text can induce

$$\frac{(\boldsymbol{N}_0 - \boldsymbol{N}_f) \cdot (\boldsymbol{R}_{switch} - \boldsymbol{R}_0)}{\boldsymbol{N}_d \cdot (\boldsymbol{R}_d - \boldsymbol{R}_0)} = \frac{\boldsymbol{K}_f}{\boldsymbol{K} - \boldsymbol{K}_f} \cdot \frac{\boldsymbol{R}_0 - \boldsymbol{R}_f}{\boldsymbol{R}_d - \boldsymbol{R}_0} \tag{A9}$$

The LHS of Equation (A9) indicates the contribution in ratio to $\boldsymbol{R}_d - \boldsymbol{R}_0$, the average risk decreases of digital channel policies, from switchers, and the RHS is Equation (13) in the main text.

Appendix E

I. Qualitative Assessment on Magnitudes of *Channel Capability* and *Preference*

In this appendix, we present details and results of the qualitative assessment on the magnitudes of the roles of *channel capability* and *preference*. This assessment relies on two assumptions: first, the purchase channel choice is made by the insurance purchaser – policyholder; second, the policyholder's ability or acceptance to use digital channels has no relation with the insured's risk when they are not the same person. Then the difference in risk screening effect between oneself and non-oneself relation policies should reflect the role of channel preference under the same empirical specification.

We firstly do the OLS estimate of Equation (6) for the couple-relation policies of the endowment insurance, as shown in Column (1) of Table E1. Note that the insurance period reflects the insured's age and the policyholder's gender reflects the insured's gender in this sample. Therefore, by using policyholder gender dummies, insurance period dummies and interactions of gender and insurance period dummies, the controls are essentially identical to the OLS estimate of oneself-relation policies (Column (2), Panel B of Table II). The result has a larger magnitude compared with that of oneself-relation policies.

In Column (2), we use the following specification for the sample combining both oneself- and couple-relation policies.

$$\begin{aligned} \boldsymbol{AC}_{i,r,t} = \alpha + \vartheta \boldsymbol{digital}_{i,r,t} \times \boldsymbol{couple} + \beta_1 \boldsymbol{digital}_{i,r,t} + \beta_2 \boldsymbol{couple} \\ + \theta \boldsymbol{D}_t + \boldsymbol{X'}_{i,r,t} \boldsymbol{\Gamma} + \boldsymbol{X'}_{r,t} \boldsymbol{\Phi} + \varepsilon_{i,r,t} \end{aligned} \tag{A10}$$

where $\boldsymbol{couple}$ is a dummy indicating the couple-relation policy and ϑ is of interest which captures the difference in risk screening effect between oneself and non-oneself relation policies. ϑ, however, is negative but insignificant, indicating no significant risk screening effect difference.

Table E1: Comparing the Magnitudes of the Roles between *Channel Capability* and *Preference*

Variables	(1) $\boldsymbol{AC}$	(2) $\boldsymbol{AC}$
$\boldsymbol{digital}$	-0.0053***	-0.0033***
	(-8.28)	(-8.51)
$\boldsymbol{digital} \times \boldsymbol{couple}$		-0.0013
		(-1.38)
Observations	194,234	866,796
Adj. R-squared	0.025	0.027
Controls	Y	Y
Fixed Effects		
Prefecture-Year	Y	Y
Month	Y	Y
Day-in-Month	Y	Y
Day-in-Week	Y	Y

Note: Column (1) uses only the couple-relation policies of the endowment insurance and Column (2) uses the sample combining both oneself- and couple-relation policies. Robust t-statistics in parentheses with standard errors clustered at the prefecture level and *** p<0.01, ** p<0.05, * p<0.1.

II. Testing for *Channel Preference* with CGSS Data

In this appendix, we use the 2021 CGSS (Chinese General Social Survey) data publicly provided by NSRC of Renmin University of China[20] to support the examination on *channel preference*. We present evidence that income, as a risk characteristic generally not observed by insurers, negatively correlates with health risk while positively correlates with the ability to use internet technology.

This survey collects individual data by interviewing 8,148 subjects randomly sampled from 320 communities in 19 provinces of China during June to September in 2021. The data includes not only demographics and financial status, but also health assessment and life styles by asking questions with two- or five-point scales. We use log income as the independent variable and use scale answers to a number of health-related and internet-use-related questions as outcome variables. The general form of the empirical specification is

$$Y_i = \alpha + \beta \boldsymbol{lnincome}_i + Controls + FE + \varepsilon_i$$

where Y_i is individual i's scale answer, $\boldsymbol{lnincome}_i$ is the logarithmic income, $Controls$ is a

[20] CGSS data is publicly available on request at http://cgss.ruc.edu.cn/English/Home.htm.

vector including demographic dummies of gender, age, interactions of age and gender, education level, nationality, religion and political status. FE is a vector including fixed effects of community and interview date. β is of interest. Robust standard errors are clustered at the community level. Records in this survey data with missing variables are excluded in regression samples. We only report the results of the OLS linear probability models. The Logit or Ordered Logit models yield very similar results.

We exploit five questions on individual health and five questions on internet-use habits in this survey. Details on the content of questions and answers are provided in Table E2. The scales in these answers generally follow the rule that the larger the point, the better the health status and the more use of Internet.

Estimated results are reported in Table E3. Panel A shows that for all five questions on health status, the coefficients are significantly positive, indicating the higher the income, the better the health; The significantly positive results of Panel B also indicate the higher the income, the more use of Internet. Thus, income leads to a negative relationship between health risk and the ability to use digital channels.

Table E2: Details on Used Questions and Answers in 2021 CGSS

Outcome Variables	Questions	Answers
Questions on Health Status		
Height	What's your height?	Number
Weight	What's your weight?	Number
Health Status	How do you feel like your current health status?	Points 1 to 5 from very unhealthy to very healthy
Chronic Disease	Do you have chronic diseases?	Point 0 for Yes and 1 for No
Cognitive Ability	What do you think of your ability to understand Mandarin?	Points 1 to 5 from very bad to very good
Questions on Internet Use		
Internet Media	How often do you use internet media (including mobile apps)?	Points 1 to 5 from never to very frequent
Information	Is internet media your most important information source?	Point 1 for Yes and 0 for No
Leisure Activity	How often do you surf the Internet in your leisure time?	Points 1 to 5 from never to everyday
Ownership	Do you own a mobile phone?	Point 1 for Yes and 0 for No
Internet Use	Have you surfed the internet in the past 6 months (including mobile apps)?	Point 1 for Yes and 0 for No

Table E3: Testing for Channel Preference with CGSS Survey Data

Variable	(1)	(2)	(3)	(4)	(5)
			Panel A. Correlation between Income and Health		
	Height	Weight	Health Status	Chronic Disease	Cognitive Ability
lnincome	0.3180***	0.6749**	0.1153***	0.0253*	0.0410***
	(3.43)	(2.08)	(7.09)	(1.89)	(3.09)
Observations	5,768	5,806	5829	1,931	5,830
Adj. R-squared	0.563	0.351	0.272	0.429	0.357
			Panel B. Correlation between Income and Internet Use		
	Internet Media	Information	Leisure Activity	Ownership	Internet Use
lnincome	0.1158***	0.0303***	0.1310***	0.0108***	0.0301***
	(6.58)	(5.55)	(6.68)	(3.38)	(5.87)
Observations	5,826	5,692	5,813	5,829	5,822
Adj. R-squared	0.551	0.515	0.534	0.240	0.525

Note: The results in this table are yielded from OLS estimates. Robust t-statistics in parentheses with standard errors clustered at the community level and *** p<0.01, ** p<0.05, * p<0.1.