Time aggregation in health insurance deductibles

Long Hong (University of Wisconsin - Madison)
Corina Mommaerts (University of Wisconsin - Madison and NBER)*

August 12, 2022

Abstract

Health insurance plans increasingly pay for expenses only beyond a large annual deductible. This paper explores the implications of deductibles that reset over shorter timespans. We develop a model of insurance demand between two actuarially equivalent deductible policies, in which one deductible is larger and resets annually and the other deductible is smaller and resets biannually. Our model incorporates borrowing constraints, moral hazard, mid-year contract switching, and delayable care. Calibrations using claims data show that the liquidity benefits of resetting deductibles can generate welfare gains of 3-10% of premium costs, particularly for individuals with borrowing constraints.

*Long Hong: long.hong@wisc.edu. Corina Mommaerts: cmommaerts@wisc.edu. We thank Naoki Aizawa, Yu Ding, Cameron Ellis, Chenyuan Liu, Ross Milton, John Mullahy, Andreas Richter, Casey Rothschild, Daniel Sacks, Justin Sydnor, Nicholas Tilipman, Yulya Truskinovsky, and seminar participants at the 2019 AEA meetings, the 2021 ARIA meeting, the University of Wisconsin - Madison, Wayne State University, and the University of Arizona for helpful comments.

1 Introduction

Health insurance contracts in the United States increasingly include sizable annual deductibles as a form of cost-sharing. Deductibles introduce non-linearities in the structure and timing of out-of-pocket expenditures, and unlike deductibles that apply to many other risks, deductibles in health insurance typically apply to cumulative losses over the year. This characteristic, coupled with the fact that most individuals have many health care encounters over the course of a year, introduces a time aggregation dimension whose implications for individual welfare are not well understood. While prior research shows that health care spending is sensitive to non-linearities in health insurance contracts (Brot-Goldberg et al., 2017; Dalton, Gowrisankaran and Town, 2019) and deductibles can reduce the value of insurance under liquidity constraints (Ericson and Sydnor, 2018), it is not obvious how these features interact when deductibles span shorter or longer periods of time. This paper provides the first analysis of the spending and welfare impacts of this overlooked parameter of health insurance design.

Rethinking the design of deductibles is particularly relevant as health insurers increasingly offer high deductible plans. In 2010, only 10% of individuals with single employer-based health insurance had a plan with a deductible over $2,000, while in 2020 26% had such a plan and 24% were offered *only* high-deductible plans (Kaiser Family Foundation, 2020). In the individual market, roughly 90% of enrollees had a high deductible plan in 2015 (Dolan, 2016). These deductibles are large and can pose a significant financial burden for those who need health care: for instance, the average medical deductible in the 2017 federal marketplace was $3,276 for silver plans (Kaiser Family Foundation, 2017*a*), and Rae, Claxton and Levitt (2017) show that only 47% of single households have enough liquid assets to pay this deductible.[1]

To study the effects of alternative deductible timespans in health insurance policies on individual welfare, we build and calibrate a dynamic model of within-year health care and non-health care consumption under uncertainty. Importantly, we introduce an alternative "resetting" deductible policy whose deductible aggregates (and thus resets)

[1]While some individuals qualify for cost-sharing reductions (CSRs), those making over 250% of the federal poverty line are still subject to the full deductible.

over shorter frequencies than the standard deductible that resets annually. To isolate the effect of time aggregation, we hold constant all other features of the insurance contract (e.g., the premium, actuarial value, and length of the contract), and only vary the time over which the deductible aggregates and—to preserve actuarial equivalence—the size of the deductible.

We study two main mechanisms that highlight the trade-offs between a standard (annual) deductible and a resetting deductible. First, resetting deductibles may provide relief for liquidity constrained individuals, but at the cost of higher risk exposure. Deductibles, by nature, front-load out-of-pocket spending toward the beginning of a deductible period, which suggests that individuals who face high borrowing costs may have difficulty smoothing consumption while financing this lumpy spending (Ericson and Sydnor, 2018). Resetting deductibles, by virtue of being smaller in size, provide built-in smoothing. On the other hand, they expose individuals to higher cumulative out-of-pocket exposure due to the fact that they reset and thus might have to be paid more frequently over the course of a year.[2] Which deductible provides more welfare gain thus depends in part on the value of liquidity versus risk protection.

Second, the frequency over which deductibles reset could interact with moral hazard. We investigate two types of moral hazard: ex-post spending moral hazard and timing moral hazard (i.e., claim delay). If individuals "over-consume" medical care once their health spending surpasses their deductible (e.g., Brot-Goldberg et al., 2017; Aron-Dine et al., 2015), all else equal a policy whose deductible resets more frequently could significantly curb ex-post moral hazard. However, again by virtue of being smaller in size, resetting deductibles are more likely to be surpassed than more aggregated, larger deductibles; this could exacerbate moral hazard. Moreover, to the extent that individuals are able to strategically delay care and shift medical costs to times with low out-of-pocket exposure, resetting deductibles might exacerbate timing moral hazard. The overall impact of resetting deductibles on moral hazard is therefore ambiguous, and ultimately an empirical question.

[2]As we will show, actuarially-equivalent biannual deductibles will typically be greater than half the size of annual deductibles, resulting in higher annual out-of-pocket risk exposure.

To quantify these mechanisms as well as the role of exogenous mid-year insurance contract switches (e.g., due to job changes), we use health care claims data from over 16 million individuals in the 2013 Truven Marketscan database and calibrate our model using benchmark figures from the federal Affordable Care Act marketplace and other standard parameters in the literature. Our main demographic group of interest is the population that has low enough income to face potential liquidity issues related to deductibles, but not so low that they qualify for cost-sharing subsidies. We use the calibrated model to quantify an individual's willingness to pay to switch from an annual insurance policy with a year-long deductible to an actuarially equivalent annual policy with a deductible that resets after six months.

Our calibration generates three main results. First, we find that borrowing costs have a first-order effect on the value of a resetting deductible policy. At one extreme in which individuals can costlessly save and borrow, individuals slightly prefer the standard (year-long) deductible policy, as it provides better insurance against large health shocks in multiple periods. At the other extreme in which individuals can neither borrow nor save, individuals strongly prefer the resetting policy: at the calibrated parameter values, they are willing to pay an extra \$270 annually, or 6.3% of their total premiums, for the resetting deductible policy instead of the standard deductible policy.[3] Second, we find that extra medical consumption generated by ex-post moral hazard is similar under the two policies for our empirical distribution. However, the presence of moral hazard amplifies the liquidity benefits of resetting deductibles because moral hazard drives up deductibles; given this, liquidity constrained individuals would be willing to pay an extra \$448 annually, or 10.4% of their total premiums. Third, we find that delayability of care for health shocks dampens the willingness to pay for a resetting deductible because delayability drives up the size of the resetting deductible, which negates its liquidity benefits. We also show that our main results are qualitatively similar when we allow for exogenous contract switching, under health shock distributions with different levels of persistence, shorter period lengths, and policies that additionally have coinsurance arms. Overall, our

[3] We also show that there are welfare gains to resetting deductibles for individuals who must rely on credit card interest rates and payday loan rates, which may more closely resemble loans or payment plans offered to pay medical expenses.

results suggest that the value of a resetting deductible policy depends in large part on the classic trade-off between risk protection, liquidity, and moral hazard.

This paper's central contribution is the first exploration of the spending and welfare consequences of the timespan over which deductibles are defined. This contribution is a relevant addition to several literatures. First, our modeling approach of within-year health spending is motivated by a growing empirical literature on consumer sensitivity to the non-linearities of health insurance contracts. Much of the recent literature has found that individuals respond to "spot" prices more so than expected end-of-year prices (Brot-Goldberg et al., 2017; Dalton, Gowrisankaran and Town, 2019; Abaluck, Gruber and Swanson, 2018; Guo and Zhang, 2019), though not all have come to that conclusion (Aron-Dine et al., 2015; Einav, Finkelstein and Schrimpf, 2015; Campo, 2021). Other work shows evidence of intertemporal substitution for deferrable care (e.g., Cabral (2017) for dental care and Lin and Sacks (2019) in the RAND Health Insurance Experiment). While much of this literature has examined the effect of traditional cost-sharing instruments such as coinsurance rates or the *size* of deductibles, our paper is the first to study the effects of the time aggregation of deductible policies.

Second, our paper contributes to a literature on optimal health insurance contracts. Much of this literature considers the optimal level and mix of various cost-sharing vehicles, including co-insurance, co-pays, deductibles, and out-of-pocket maxima, but hold fixed the basic structure of these vehicles (e.g., Arrow, 1963; Cutler and Zeckhauser, 2000; Ellis, Jiang and Manning, 2015). Closely related to the liquidity results of our paper, Ericson and Sydnor (2018) show through simulation that liquidity constraints can upend the optimality of a straight deductible policy. Our paper further relaxes a particular aspect of deductibles—the timespan over which they aggregate—and shows that this relaxation can provide further welfare gains, especially for liquidity constrained individuals.[4]

Finally, this paper contributes to a small literature on the aggregation of continuous measures over time. Time aggregation underlies many policies and economic models,

[4] A separate optimal contracts literature examines the optimal *contract* length in environments with adverse selection and reclassification risk (Ghili et al., 2021; Atal et al., 2020). These papers hold the length of a period fixed at a year and thus abstract from within-year dynamics, while our paper holds fixed the contract length but explores within-contract aggregation over time.

yet there has been little work understanding the consequences of these aggregation decisions. In a related paper within the context of automobile insurance, Cohen (2006) studies the trade-off between aggregate and per-event deductibles, but focuses on a set of issues more relevant to auto insurance.[5] Other work shows that the periodicity of payments, such as crop insurance premiums (Casaburi and Willis, 2018), paychecks (Parsons and Van Wesep, 2013), and Food Stamp benefits (Shapiro, 2005), is an important consideration for welfare. Our paper suggests that the time aggregation embedded in health insurance deductible policies can also have non-trivial impacts on welfare.

Our analysis raises several interesting questions for further research. One open question is the effect of time aggregation on sorting and selection. Our analysis focuses on liquidity and moral hazard issues, and our simplifications—such as ex-ante homogeneous individuals—assumes away issues related to sorting and adverse selection that may arise when consumers can select from many plans with alternative deductible structures (Marone and Sabety, 2022; Liu, 2021). It would also be interesting to study other dimensions of time aggregation, such as out-of-pocket maxima (which are equivalent to deductibles in our main analysis), as well as other dimensions of smoothing, such as Health Savings Accounts or payment plans.[6] Finally, a more fundamental question is *why* deductibles that reset at shorter frequencies are not offered. Whether this is due to historical precedent, administrative costs, or something else would be an interesting avenue for further research.

Notwithstanding these caveats, our findings have policy implications for the design and effectiveness of health insurance plan offerings. This may be particularly true for plans offered through the Affordable Care Act marketplaces, where high deductible plans are common and often cater to lower income populations who may be liquidity con-

[5] Cohen (2006) shows that per-event deductibles require lower claim verification costs and induce lower ex-ante moral hazard, both of which we abstract from and instead focus on ex-post moral hazard and other mechanisms important to the health insurance context.

[6] An example of an alternative payment plan for deductibles exists in the Netherlands, where deductibles can be paid in monthly installments throughout the year and any excess payments are reimbursed at the end of the year. While this scheme could relieve some of the liquidity issues explored in this paper, it may lead to enforceability issues (e.g., consumers exiting their policies after high early health spending), particularly in the context of much higher deductibles as in the United States.

strained.[7] Given that deductibles are trending into the thousands of dollars, however, these issues extend beyond low-income populations:[8] Rae, Claxton and Levitt (2017) show that fewer than half of single-person households have enough liquid assets to pay a $2,000 deductible. Meanwhile, policymakers continue to encourage the use of high deductible health plans on the individual market. For example, the Centers for Medicare and Medicaid Services in 2018 stated "We would like to encourage issuers to offer HDHPs [high deductible health plans]... as a cost effective option for enrollees" (Department of Health and Human Services, 2018). Designing and introducing alternative deductible structures, such as resetting deductibles, could maintain the use of high deductible policies while alleviating some of the liquidity issues that concern their critics.

The paper proceeds with a discussion of recent trends in health insurance deductibles in Section 2. Section 3 develops a two-period model of health insurance choice and Section 4 describes the calibration and presents results. In Section 5 we explore model extensions. Section 6 concludes.

2 Background on health insurance deductibles

Across a range of contexts, household insurance contracts feature deductibles. For example, the median passenger automobile insurance policy includes a $500 deductible (Barseghyan, Prince and Teitelbaum, 2011), the median homeowner's insurance policy includes a $500 deductible (Sydnor, 2010), the median flood insurance policy includes a $500 deductible (Michel-Kerjan and Kousky, 2010), and the median health insurance policy includes a $1,400 deductible, sometimes in addition to other cost-sharing mechanisms (Kaiser Family Foundation, 2020). One major difference between the first three policies and the health insurance policy is that the first three deductibles are *per-event* while health insurance deductibles are *per-period*.[9] This period is almost always one year, and usually

[7]Our findings also have implications for short-term health insurance plans, which have similarly shorter deductible spans by definition. We leave the analysis of short-term health insurance plans to future work because this market operates in a very different policy and regulatory space.

[8]Moreover, individuals with income less than 250% of the federal poverty line are eligible for cost-sharing subsidies that alleviate some of the burden of these deductibles.

[9]One theory for this difference is that an "event" in these other policies are relatively well-defined, while in the context of health care an "event" can develop slowly over time and spill over into other "events",

one *calendar* year.[10]

Health insurance deductibles are an increasingly common form of cost-sharing in health insurance plans in the United States. In the past decade, the share of individuals covered by an employer plan with a deductible over $2,000 rose from 10% to 26% (Kaiser Family Foundation, 2020). There has also been a rapid expansion in the use of high-deductible health plans:[11] the share of employers only offering high-deductible plans increased from 7% in 2012 to 24% in 2020 (Towers Watson, 2015; Kaiser Family Foundation, 2020). In plans available through federal and state health insurance exchanges, the average deductible for an individual plan is over $3,000 for silver plans (which account for 67% of plan selections), and almost $6,000 for bronze plans (which make up 22% of plan selections). To help curb this financial exposure, individuals under 250% of the federal poverty line (FPL) are eligible for cost-sharing reductions that lower their effective deductible, but those over 250% of the FPL do not receive any such subsidies.

Alongside this growth in plans with large deductibles has been increased concern about the affordability of these deductible policies. Affordability concerns could be particularly warranted if medical expenditures are concentrated over a short period of time. Figure 1 plots the fraction of individuals who incur a health expenditure shock of at least a given size within a month in our data, among individuals who have not hit a representative deductible.[12] Around 11% of individuals who have not hit the deductible incur a

making it difficult to distinguish between events. Reinsurance policies also have per-period (or "aggregate") deductibles for a similar reason. Relatedly, a prominent exception to this event vs. time period distinction in health insurance is Medicare hospital inpatient (Part A) deductibles, which are per-stay, where "stays" are well-defined events. Health insurance policies often also include co-pays, which are "per-event" and are typically orders of magnitude smaller than the deductibles we are concerned with in this paper. Increasing co-pays in order to decrease the size of deductibles could be an alternative remedy to the problem of within-year consumption smoothing, but would involve a more serious exploration of what qualifies as an "event" that triggers co-pays (e.g., what services?) and would likely imply very different moral hazard consequences. We leave this for future study.

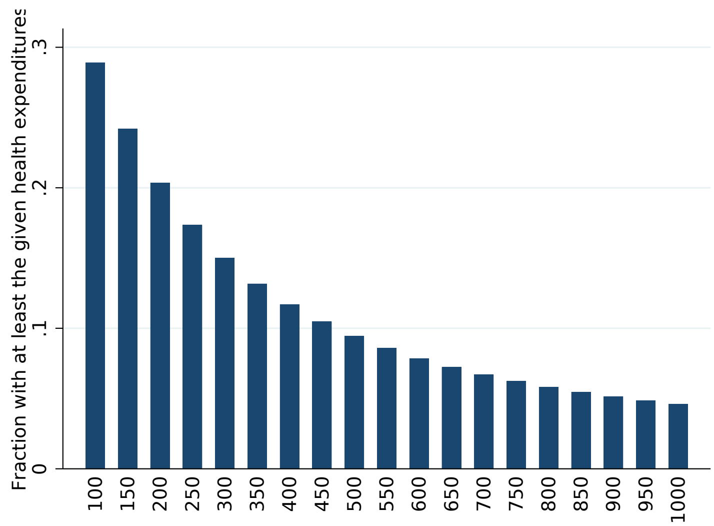
[10]While almost all health insurance contracts—and their corresponding deductibles—span a calendar year, one exception is short-term health insurance plans that provide coverage for a limited amount of time (i.e., less than 365 days). While in theory these policies have parallels with resetting deductible policies, in practice they have many limitations that are beyond the scope of this paper, as they do not face as many regulatory constraints as standard health insurance policies do (e.g., they are not guaranteed renewable and can exclude coverage for pre-existing conditions).

[11]The Internal Revenue Service defines a high deductible health plan as one with a deductible of at least $1,400 for an individual or $2,800 for a family in 2020.

[12]Our MarketScan data is described in more detail in Section 4. We do not know an individual's actual deductible, so we simply assign everyone a deductible of $3,252, which is the size used in our model

health shock of $400 or more in a month, and around 6% incur a health shock of $800 or more. This is in line with Chen et al. (2021), who find that almost one third of individuals with high total health care spending incurred half of their out-of-pocket spending in one day. Given that many households have difficulty coming up with funds of this size (Board of Governors of the Federal Reserve System, 2014), these numbers suggest that this "temporal clustering" of spending could be a major financial burden for liquidity constrained individuals.

Figure 1: Percent of individuals with a given health shock in a month, conditional on being under deductible



Notes: Figure plots the fraction of individuals who have at least a given amount of health expenditures in a month, averaged over all months, among individuals who have not hit a representative deductible of $3,252. Data from the 2013 Truven MarketScan Commercial Claims and Encounters database, as described in Section 4.1.

Health Savings Accounts (HSAs) were created to provide some relief for plans with high deductibles by allowing individuals with such plans to place funds in a tax-preferred savings account to be used for medical spending or retirement, up to a contribution limit (for example, in 2018 the contribution limit was $3,450 for an individual plan). This effectively lowers the price of medical care paid by HSAs by allowing individuals to pay for care using pre-tax dollars. On the other hand, because any other use of these funds prior to retirement incurs a 20% penalty in addition to being taxed, HSAs also introduce illiquidity towards other spending, making them potentially less attractive to liquidity

calibration and very similar to a Silver plan in the ACA marketplaces.

constrained individuals. Since their introduction in 2004, the number of individuals enrolled in an HSA has grown to 22 million in 2017 with total assets of over $45 billion (Devenir Research Team, 2018). However, much of this growth has been concentrated among high-income households, who are less likely to be constrained by the illiquid nature of the accounts (Helmchen et al., 2015).

In sum, the affordability concerns that arise with the increased adoption of high deductible health plans suggests that policies that could provide individuals with some financial relief could increase well-being. This is particularly so given the concentration of expenditures within short periods of time. We next turn to a model of individual demand for health insurance to understand the value of policies that modify the temporal nature of deductibles.

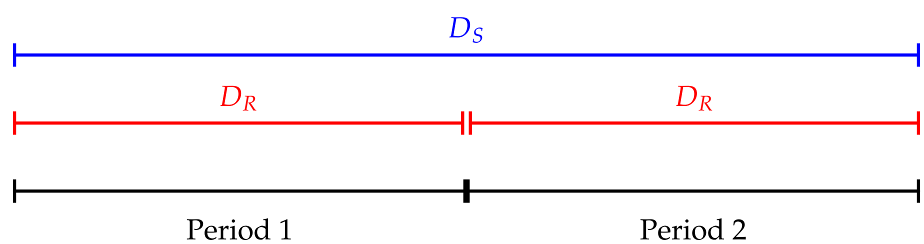
3 Two period model of deductible timespans

To understand the mechanisms through which the time aggregation of a deductible can affect individual welfare, we develop a two-period model of decision-making with uncertainty over medical expenditure shocks and insurance to protect against this uncertainty. Each period, which is meant to capture a six month timespan, a risk-averse individual is subject to health expenditure risk. Individuals are ex-ante homogeneous, thus abstracting from adverse selection concerns. We model two annual insurance policies: one with a "standard" deductible spanning both periods (the standard year-long deductible) and one with a "resetting" deductible that spans one period (six months) before resetting. The model incorporates four main mechanisms that may affect an individual's preference over the two deductibles: (1) borrowing costs, (2) ex-post moral hazard, (3) mid-year exogenous contract switching, and (4) strategic claim delay. We first provide a simple illustrative model with binary health shocks, and then detail the full model, including the preference structure, health risk and insurance contracts, budget constraint, and the full individual problem. Section 5 provides further extensions to the model.

3.1 Illustrative model with binary shocks

To illustrate the concept of time aggregation in health insurance deductibles, we begin with a simplified version of our model with an independent, binary health shock and without moral hazard, mid-year contract switching, or claim delay (all of these features will be relaxed in the full model). Assume an independent large binary health expenditure shock L that occurs with probability $\pi > 0$ in each period.[13] To protect against this risk, an individual purchases an insurance policy that spans two periods of the form (P, D_i) with per-period premiums P and a deductible $D_i < L$ where $i = S$ is a standard deductible policy in which the deductible D_S spans two periods and $i = R$ is a resetting deductible policy in which the deductible D_R resets after each period.[14] Figure 2 depicts the length of time over which a deductible spans: the deductible that does not reset (D_S) spans both periods (blue, top line), while the deductible that resets (D_R) only spans one period (red, middle lines).

Figure 2: Deductible timespan schematic



Notes: Figure depicts the length of time over which a deductible applies and resets over the course of the two periods. D_R in red is the resetting deductible policy in which the deductible resets after each period, while D_S in blue is the standard deductible policy in which the deductible resets after two periods.

In order to isolate the effect of the resetting deductible policy on individual behavior and welfare, we hold the premium P constant and assign the same actuarial value to each policy (P, D_S) and (P, D_R). Thus, for a given standard deductible of size D_S, the size of the resetting deductible D_R must adjust to maintain the same actuarial value, so that the

[13] While this binary shock process is a simplification (and we relax it later), it has also been used in many other papers on similar topics, e.g., Ericson and Sydnor (2018) and Cohen and Einav (2007).

[14] We assume that D_R is constant across the two periods; an interesting extension would be to consider increasing the size of deductibles (see Li, Liu and Yeh (2007) for a related exercise in automobile insurance).

expected insurance payouts over the two periods is the same:

$$\underbrace{2\pi(L - D_R)}_{\text{Expected payout, reset policy}} = \underbrace{\pi^2(2L - D_S) + 2\pi(1-\pi)(L - D_S)}_{\text{Expected payout, standard policy}}$$

Rearranging this equation gives a relationship between the two deductibles of:

$$D_R = \frac{2-\pi}{2} D_S \tag{1}$$

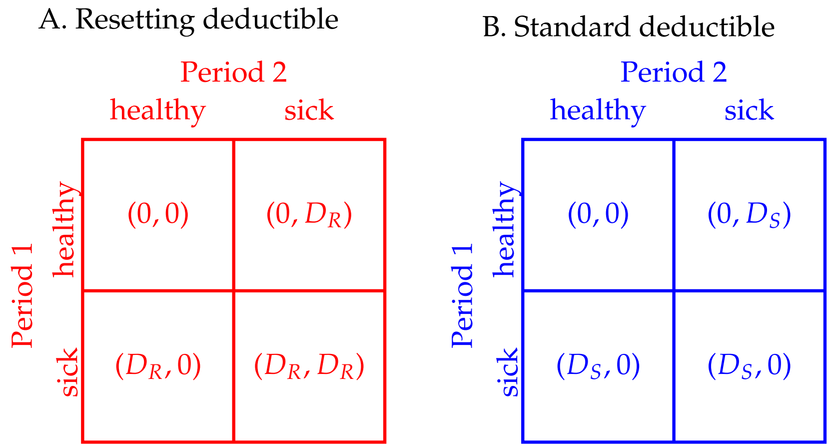
There are two features of this equation of note. First, the resetting deductible is always smaller than the standard deductible. As we will show quantitatively, if individuals are liquidity constrained then they will value this smaller deductible despite its resetting nature because it corresponds to smaller immediate out-of-pocket costs. In addition, it can be shown analytically (through a Jensen's inequality argument) in this binary case that individuals who cannot borrow or save will prefer the resetting deductible policy. Second, despite the fact that the *timespan* of the resetting deductible is half that of the standard deductible, the *size* of the resetting deductible is greater than half that of the standard deductible.[15] As Figure 3 shows, this implies that the worst-case scenario in which an individual receives a health shock in both periods (the bottom-right corner of each schematic) results in higher total out-of-pocket costs for the resetting deductible policy (equal to $D_R + D_R$) than the standard deductible policy (equal to D_S). If individuals strongly value insurance against this worst-case scenario, then they may value the standard deductible more than the resetting deductible. These two effects – liquidity versus insurance – are further explored in the full model below, along with additional mechanisms that affect the value of different deductible timespans.

[15] To see why, rewrite the expected insurer payout as:

$$\pi^2(2L - 2D_R) + 2\pi(1-\pi)(L - D_R) = \pi^2(2L - D_S) + 2\pi(1-\pi)(L - D_S)$$

If $D_R = .5D_S$ then the insurer's losses in the $2\pi(1-\pi)$ states of the world would be too large in the resetting deductible case. If $D_R = D_S$ then the insurer's losses would be too small in the π^2 states of the world in the standard deductible case. This is analogous to the result in Cohen (2006) for aggregate versus per-event deductibles.

Figure 3: Out-of-pocket expenses for resetting and standard deductible policies



Notes: Figure depicts out-of-pocket expenses in the brackets for each cell under the resetting deductible policy (red) and the standard deductible policy (blue) when there is a binary health shock ("healthy" vs. "sick"). The first value in the brackets is the out-of-pocket expense in Period 1 and the second value is the out-of-pocket expense in Period 2.

3.2 Full model

We now characterize the full model, which incorporates a continuous shock distribution and allows for moral hazard, contract switching, and strategic delay.

3.2.1 Preferences

Individuals choose non-medical consumption c, medical consumption m, and savings a to solve a dynamic problem that maximizes their expected utility over two (six month) periods. Specifically, we use the utility function in Einav et al. (2013) in which non-medical consumption is the numeraire good and there are decreasing returns to medical consumption above a medical loss L, so that per-period utility is:

$$U(c, m - L) = u\left(c + m - L - \frac{1}{2w}(m - L)^2\right)$$

With this utility function, individuals will always choose $m \geq L$ because at any $m < L$, the marginal utility of an extra unit of m is greater than the marginal utility of an extra unit of c. Thus we interpret L as necessary expenditures and any $m > L$ as additional medical care that arises through moral hazard. The severity of moral hazard is dictated by the $w \geq 0$ parameter, where $w = 0$ is the limiting case of no moral hazard. Additional medical consumption ($m > L$) arises when individuals do not have to pay the full price of medical care at the time of purchase. When individuals are fully insured such that an extra dollar of medical consumption m does not impact c, then the first order conditions imply that optimal medical consumption is $m = L + w$ (including when $L = 0$). On the other hand, if individuals must fully pay for medical consumption then they derive higher utility from spending the marginal dollar beyond L on non-medical consumption c than m due to the decreasing return to medical consumption beyond L. In that case, $m = L$. The interpretation of w, then, is the extra medical consumption due to moral hazard.

The implication of this form of moral hazard is that, *ex-post*, individuals value over-consumption of medical goods due to insurance coverage, but since the cost of this extra medical consumption feeds back into their cost of insurance (either through an increase in premium, or, as we will assume below, through an increase in the size of deductibles), *ex-ante* they do not value the over-consumption and thus prefer to minimize moral hazard. Different time aggregations of deductibles may exacerbate or hinder moral hazard, as the change in non-linearity of the contract affects if and when individuals pay full price for their health care.

3.2.2 Health risk and the evolution of health spending

An individual is subject to health shocks L_1 and L_2 in the first and second period, respectively, that evolve stochastically with joint distribution $f(L_1, L_2)$. We assume all individuals have the same underlying health shock distribution.

We allow for two forces that make health *spending* endogenous insofar as it may be greater than that necessitated by the health shock and it may not occur at the same moment as the health shock. The first force, introduced above, is (ex-post) moral hazard. This type of moral hazard arises when individuals do not have to pay the full price of medical care at the time of the care. We assume that individuals only respond to spot prices, not expected end-of-year prices, and thus only over-consume once they have surpassed the deductible. While a fully rational individual would respond to expected end-of-year prices, we shut down this channel for computational tractability when we calibrate the model, and because other work shows that it is likely a more reasonable approximation to reality (Brot-Goldberg et al., 2017).

The second force, which we call "claim delay", allows individuals to delay the timing of treatment (and therefore spending and claiming) of health shocks. One prominent example of this distinction is in dental care, where procedures such as fillings can be performed many months after the advent of the cavity "shock" (Cabral, 2017), but it also applies to many other medical procedures (e.g., hip replacements).[16] Since health insurance policies largely operate on the timing of health *care* (which is manipulable) and not the health *shock* (which is not manipulable), individuals may find it financially beneficial to delay care for certain shocks. To formalize this idea, we allow a fraction q_d of health shocks to be delayable by one period, to capture the notion that some shocks must be cared for immediately (e.g., emergencies) while others can be delayed with minimal consequence to the next period. We allow individuals to choose whether to delay ($d = 1$) the latter type of shocks, and while they will typically prefer to delay care (particularly if delay is costless[17]), there are still very specific scenarios in which individuals may choose not to delay care. These include (1) settings with a high degree of moral hazard, when the ex-post value of overconsuming medical care in both periods outweighs the risk of paying more out of pocket, or (2) settings with liquidity constrained individuals, who may value spreading their medical expenses over two periods rather than bunch all expenses in the second period (as a way to consumption smooth), over the risk of paying more out of pocket.

3.2.3 Health insurance contracts and contract switching

We now characterize the two health insurance contracts under the full model, which includes moral hazard, strategic claim delay, and exogenous contract switching.

At the beginning of the year, individuals face one of two potential annual insurance

[16] Card, Dobkin and Maestas (2008) reports that 7% of the near-elderly delayed care in the last 12 months.

[17] While we assume that it is costless to delay the q_d fraction of claims in our main results, it is of course possible that there are health costs (in the form of pain, for example, or additional eventual health care costs) to delaying care. We explore an extension that includes a cost to delaying care in Section 5.

contracts that have the same per-period premium, the same actuarial value, and a deductible that must be met before insurance pays for health care. The distinction between the two policies is the time over which the deductible resets: the standard deductible policy consists of a deductible that spans the full year (two periods), while a resetting deductible policy consists of a deductible that resets each period and thus spans only one period (i.e., six months) at a time. Because these two policies have the same actuarial value and same premium, the timespan over which the deductible resets means that the *size* of the deductible will also vary over the two policies.

Health insurance deductibles not only span a year, but they almost always span a *calendar* year as opposed to a year from the signing of the insurance contract.[18] This discrepancy may distort the value that individuals place on different insurance policies if mid-year job changes or other life events that are orthogonal to health cause them to abruptly and exogenously change their health insurance mid-year. To formalize this idea, we assume a fraction q_m of individuals abruptly cancel their policy after the first period (i.e., mid-year) and begin a new policy in the second period, with the same deductible structure and same end-date. When this occurs, individuals must sign a new insurance contract with the same deductible end-of-calendar-year end dates and the same deductible sizes, but without the stored health care spending from the first period that had previously gone towards meeting the deductible. Effectively, this means that individuals who have large losses in both periods would have to pay the (higher) deductible *twice* under the standard deductible policy. This feature makes insurance contracts with more time-aggregated deductibles less valuable as the risk of mid-year contract switching increases.

The insurer takes into account the extra and time-varying medical consumption brought about by moral hazard, delay, and contract switching and solves for the standard and resetting deductible sizes in order to break even. The resulting relationship between the resetting and standard deductible sizes is complex, but two features are intuitive. First, as q_d increases, D_R grows closer to D_S, because the ability to delay and thus bunch care in fewer periods means that the size of the resetting deductible must increase. Second, as q_m

[18] Some annual health plans span a fiscal or academic year instead of a calendar year, but the discrepancy between a year at contract signing and a pre-specified year remains.

increases, the size of the standard deductible decreases toward the size of the resetting deductible, because the higher the probability that one must switch contracts mid-year, the higher the probability a standard deductible "resets," which is the defining characteristic of the resetting deductible. The moral hazard effect on deductible size is more nuanced, and we return to it in our quantitative results in Section 4.

3.2.4 Budget constraint

An important feature of individual welfare in a dynamic setting with risk aversion is the ability to smooth consumption over time. As we quantify in the next section, shorter time aggregation of deductibles can help smooth consumption, but otherwise the primary smoothing mechanism is through saving and borrowing. We allow for saving and borrowing to satisfy the budget constraint:

$$a_2 = \begin{cases} R_s(a_1 + Y - P - c_1 - oop_1) & \text{if} \quad a_1 + Y - P - c_1 - oop_1 \geq 0 \\ R_b(a_1 + Y - P - c_1 - oop_1) & \text{if} \quad a_1 + Y - P - c_1 - oop_1 < 0 \end{cases}$$
$$\text{subject to} \quad a_2 \geq -\left[Y - P - \max(oop_2)\right]$$

where Y and P are per-period income and premiums, c_1 is consumption in the first period, oop_t is out-of-pocket medical expenditures in each period t, and a_t are assets in each period.[19] The budget constraint allows for borrowing up to the amount that they would be able to pay back with certainty (i.e., income net of the premium and maximum possible out-of-pocket expenditure) and allows for different interest rates for savings (R_s) and borrowing (R_b). This formulation nests costless saving and borrowing ($R_b = R_s = 1$) as well as borrowing costs ($R_b > R_s$) and an extreme form of liquidity constraints in which individuals are "hand-to-mouth" in that they neither save nor borrow ($R_b = \infty$ and $R_s = 0$).

[19]This formulation could easily be extended to incorporate HSA spending by converting out-of-pocket medical spending to pre-tax dollars, i.e., $(1 - \tau)oop_t$ where τ is the tax rate, which effectively lowers the cost of medical care. We conduct this extension in Section 5.

3.2.5 Individual problem

Given the above ingredients of the model, an individual solves a problem that proceeds in two steps. In the first period, given a deductible regime $D_i \in \{D_R, D_S\}$, the individual chooses first period medical consumption m_1 (and whether to delay the health shock until next period, d, if the shock is delayable) as well as non-medical consumption c_1 and savings for the next period a_2:

$$V_1(a_1, Y, L_1|D_i) = \max_{d, c_1, m_1, a_2} U(c_1, m_1 - (1-d)L_1) + \beta E_{f(L_2|L_1)} V_2(a_2, Y, m_1, d, L_2|D_i) \quad (2)$$

where V_2 is the solution to the following problem in the second period:

$$V_2(a_2, Y, m_1, d, L_2|D_i) = \max_{c_2, m_2} U(c_2, m_2 - (dL_1 + L_2)) \quad (3)$$

and each of the maximization problems are subject to the budget constraint and the law of motion of deductible spending. The state space consists of assets, income, and the health shock in each period; additionally the second period state space includes first period medical spending and whether first period care was delayed.

Because we assume that all individuals have the same (known) underlying health shock distribution $f(L_1, L_2)$, there are no concerns about asymmetric information or adverse selection and thus this setup does not include an initial decision over the deductible regime. If individuals had private information over heterogeneous health shock distributions, riskier individuals would place more value on standard deductible policies given that standard deductibles provide better insurance. On the other hand, if contracts at least partially adjust for risk type by increasing the deductible size, liquidity concerns may push riskier individuals with liquidity constraints back towards resetting deductibles. Thus any bias in our results due to our abstraction from adverse selection is likely to depend on the correlation between an individual's health shock distribution and liquidity position (along with other standard factors like risk aversion).

3.2.6 Willingness to pay for insurance contracts

To measure the value to individuals of the standard versus resetting deductible policies, we calculate the amount of per-period income an individual would be willing to pay (in each period) for the resetting deductible to be indifferent between the standard deductible and the resetting deductible. Specifically, we calculate Z to solve:

$$V_0(a_1, Y|D_S) = V_0(a_1, Y - Z|D_R) \tag{4}$$

where V_0 is the ex-ante value function prior to the revelation of health shocks. A positive value of Z signifies that the individual prefers the resetting deductible over the standard deductible, while a negative value of Z signifies that they prefer the standard deductible.[20]

4 Calibrating the willingness to pay for resetting deductibles

We next calibrate the model and report our willingness-to-pay results for a resetting deductible under various scenarios. We use claims data from the Truven Marketscan database to characterize the health shock distribution, and calibrate the remaining model parameters using standard values from the literature. We then use the parameterized model to characterize the role of liquidity, moral hazard, contract switching, and claim delay in isolation before reporting overall estimates for a representative consumer for which all four mechanisms are at play.

[20] We note that one could imagine an additional cognitive cost to keeping track of a resetting deductible policy. This could be captured by introducing a (monetary) cognitive cost into the budget constraint. The interpretation of this cost is the relative difference in the amount one would have to pay someone to keep track of the details of the deductible policy, or the monetization of the forgone time spent being cognizant of the resetting deductible policy relative to the forgone time spent being cognizant of the standard deductible policy. This cost is mathematically equivalent to a decrease in the willingness to pay for the resetting deductible policy. Put differently, in cases where we estimate that individuals prefer resetting deductibles, the cognitive cost has to be at least as large as the willingness to pay for the resetting deductible policy to overturn the preference for the resetting deductible policy.

4.1 Marketscan health care claims data

An important input into the model is the joint distribution of health care claims over time. We use claims data from the Truven MarketScan Commercial Claims and Encounters database, which includes service use and expenditure data from individuals and families with employer-sponsored coverage in large commercial plans from approximately 350 insurers nationwide.[21] There are a couple limitations of this data for our research question. First, because the data comes from a population with employer-sponsored health insurance, it is not necessarily representative of the health care expenditures of lower income individuals, who may be more likely to have health insurance through an ACA marketplace plan or a public health insurance plan. However, this data underlies the risk adjustment models for plans in the ACA marketplaces (for Medicare and Services, 2016), and thus is thought to serve as a close proxy for spending of households who purchase such plans. Second, the expenditure data may reflect not only health expenditure *needs* but also moral hazard and/or timing manipulation of health expenditures. Given our version of the MarketScan data does not contain plan details for most individuals (such as cost-sharing parameters or individual versus family plans) or claim delay, we cannot perfectly account for these components, so our underlying health shock distribution partially reflects these additional components. We return to this issue in Section 5, where we check robustness to making rough adjustments to the underlying distribution.

Our sample consists of over 16 million individuals aged 26-64 who are continuously enrolled in a health plan for the entirety of 2013 and not enrolled in a capitation plan at any point during the year. We calculate total health care spending in the first six months and second six months of the year based on service dates, corresponding to the first and second periods of the model, respectively. Total health care spending is defined as the sum of expenditures paid by the insurance plan and expenditures paid out-of-pocket, such as deductibles, co-insurance, and co-pays, and includes inpatient services, outpatient services, and outpatient pharmaceutical claims.

Table 1 reports summary statistics of the distribution of per-period (i.e., six month)

[21] We obtained this data through an agreement with the National Bureau of Economic Research.

Table 1: Summary statistics of the distribution of health care expenditures

	Health care spending ($)	
	First half of year	Second half of year
Mean	2,969	3,313
Standard deviation	12,372	13,818
Percent with zero expenditure	0.210	0.202
25th percentile	75	80
Median	550	587
75th percentile	1,992	2,179
90th percentile	5,963	6,670
Mean number of claims	20	21
Correlation of spending between periods	0.413	
Biannual mean	3,141	
Biannual standard deviation	11,016	
Average number of claims per month	3	
Number of enrollees	16,351,864	

Notes: Table reports biannual (six month) total health care expenditures (unweighted, in 2017 dollars). Data are 2013 health care claims from the Truven Marketplace database, restricted to individuals 26-64 who are not in a capitated plan and do not switch plans over the course of 2013.

health care expenses in our sample, inflated to 2017 dollars. Average health expenditures over the first six months total $2,969, with standard deviation $12,372, though the distribution is highly skewed as shown by the median of $550. The average number of health care claims within the first six months is 20, though a significant fraction of individuals (around 20%) have zero expenditures in a given period. There is also persistence in expenditures over time: the correlation over the two periods of health care expenditures is 0.41. We incorporate this data into our model by constructing an empirical joint distribution of health expenditures over the two periods $f(L_1, L_2)$, discretized into 30 bins of equal probability.

4.2 Calibrated parameters

Table 2 reports our calibrated parameter values and their source. We set the interest rate for savings to 0% (or $R_s = 1.0$) and the discount factor to 1.0 (or $\beta = 1$) to isolate our main mechanisms of interest. Following much of the health literature, we use CARA preferences in which $U(x) = 1 - \exp(-\alpha x)$ with coefficient $\alpha = 0.0004$ from Handel

(2013). We set the premium for health insurance to the average benchmark silver plan in the ACA exchanges in 2017, which was \$359 monthly, or \$2,154 biannually (Kaiser Family Foundation, 2017*b*). We set initial assets to \$0 and annual income to \$30,150 (250% of the Federal Poverty Level in 2017), which was the cutoff for cost-sharing reductions (CSRs) in the health insurance marketplace (though with CARA utility, there are no income effects).[22] We thus view our calibration as pertaining to individuals who are lower income but do not qualify for CSRs.

Table 2: Calibrated parameters

Symbol	Parameter definition	Value	Source
t	Length of a period	6 mo.	
α	CARA coefficient	0.0004	Handel (2013)
β	Discount factor	1.0	
R_s	Interest rate, saving	1.0	
R_b	Interest rate, borrowing	{1.0; 1.015; 1.095; 2.236}	3% interest rate; credit card 20% APR; payday loan rates (Ericson and Sydnor, 2018)
w	Moral hazard parameter	\$896	30% of average biannual health shock (Einav et al., 2013)
q_m	Prob. mid-year job change	0.08	Bjelland et al. (2011)
q_d	Prob. shock is delayable	0.40	Cabral (2017)
P	Premium	\$2,154	Average second lowest-cost silver plan premium (Kaiser Family Foundation, 2017*b*)
Y	Income	\$15,075	250% Federal Poverty Level in 2017
a_1	Initial assets	\$0	

Notes: Biannual (six month) rates shown. The four values of R_b correspond to different specifications in Figure 5.

We supplement these parameters with parameters that dictate the additional forces described in Section 3. To capture the fact that borrowing can be expensive, we provide a range of interest rates for borrowing to approximate low borrowing costs (3% annual interest rate), borrowing from a credit card at 20% APR, and payday loan rates of 400% APR (Ericson and Sydnor, 2018), and convert these annual rate to biannual rates. To capture overconsumption of medical care that arises with insurance, we convert the moral hazard parameter estimated in Einav et al. (2013) of 30% of annual health shocks into a biannual value of \$896. To capture the fact that some workers switch jobs – and therefore

[22]This group is still relatively sizable: for example, in 2014, the percent of the population with income between 200 and 300 percent of the Federal Poverty Level was 16% (U.S. Census Bureau, 2014). While individuals with income between 200-250% of the Federal Poverty Level receive cost-sharing subsidies, they are modest.

health plans – in the middle of the year, we use the estimate of 4% quarterly employment-to-employment job flows from Bjelland et al. (2011), converted to a biannual probability of 8%. Finally, we allow a fraction of shocks to be delayable, and set the probability of delayability to 40% from Cabral (2017).

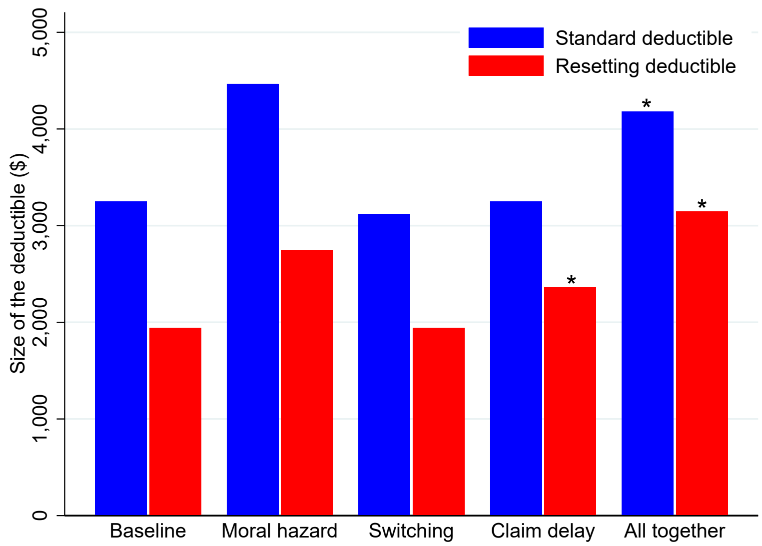
4.3 Willingness to pay: the role of liquidity

We use the calibrated parameters in Table 2 to estimate the willingness to pay for the resetting deductible over the standard deductible. We begin by analyzing the role of liquidity. To do this, we shut down moral hazard ($w = 0$), contract switching ($q_m = 0$), and care delay ($q_d = 0$), and vary R_b. The sizes of the deductibles in this case are given in the leftmost blue and red bars in Figure 4, and correspond to a standard deductible of \$3,252 (which is very similar to the average deductible in a silver plan in the 2017 federal marketplace) and a resetting deductible that is 60% the size of the standard deductible.

Using this relationship, we examine the willingness to pay for the resetting deductible under different liquidity environments. For the case in which individuals cannot save or borrow, which corresponds to $R_s = 0$ and $R_b = \infty$, the per-period willingness to pay for the resetting deductible is $Z = \$135$, or 6.3% of the premium, as the right-most bar in Figure 5 shows. For the case in which individuals are able to save and borrow with no interest ($R_s = R_b = 1.0$), the opposite result emerges: individuals who can easily smooth slightly prefer the standard deductible. The left-most bar of Figure 5 shows that they value the resetting deductible at only -\$5 over the standard deductible.

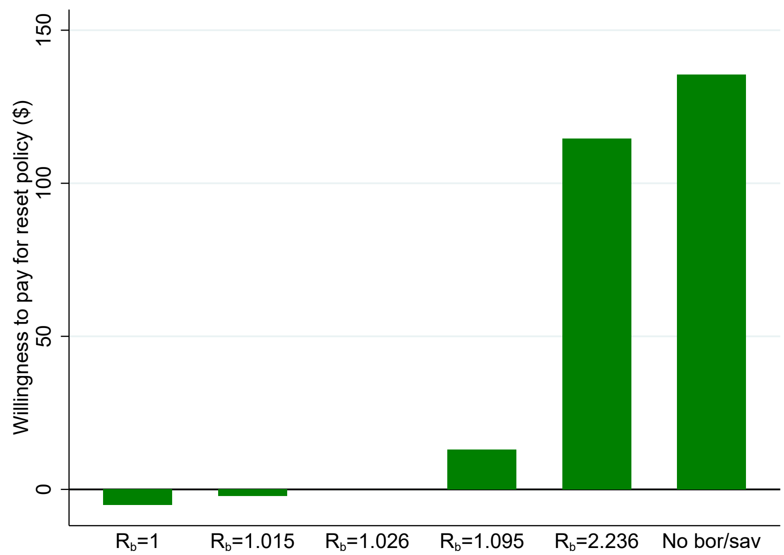
The reason that borrowing constrained individuals prefer the resetting deductible while unconstrained individuals prefer the standard policy relates directly to the trade-off between insurance and liquidity. The standard deductible provides better insurance against the worst-case state of the world (e.g., large shocks in both periods, where $D_S < 2D_R$). On the other hand, the resetting deductible provides an alternative form of liquidity to individuals by, in essence, breaking up the deductibles into smaller but potentially more frequent payments. Unconstrained individuals can transfer resources between states of the world on their own, and thus value the better insurance of the standard

Figure 4: Size of deductible policies under different environments



Notes: Figure presents the size of the standard (annual) deductible D_S in blue and the size of the resetting (biannual) deductible D_R in red, under four scenarios: from left to right, (1) the baseline scenario with no moral hazard ($w = 0$), no contract switching ($q_m = 0$), and no claim delay (q_d), (2) the scenario with only moral hazard, (3) the scenario with only contract switching, (3) the scenario with only claim delay, and (4) the scenario with all three of moral hazard, switching, and delay. * indicate small differences in deductible size, due to endogenous claim delay, between individuals with no borrowing costs and individuals who cannot borrow or save (respectively: $2,379 vs $2,359 for the resetting deductible under claim delay; $4,204 vs $4,179 for the standard deductible in the "all together" scenario; $3,163 vs $3,149 for the resetting deductible in the "all together" scenario).

Figure 5: Welfare gain of a resetting deductible policy under different liquidity assumptions



Notes: Figure presents Z, the per-period willingness-to-pay for a reset policy, under five liquidity cases: from left to right, (1) benchmark saving and borrowing ($R_b = R_s = 1.0$), (2) low borrowing costs ($R_b = 1.015$), (3) credit card borrowing costs ($R_b = 1.095$), (4) payday loan borrowing costs ($R_b = 2.236$), which in practice is equivalent to saving but no borrowing, and (5) no borrowing or savings ($R_s = 0$ and $R_b = \infty$) using the other calibrated parameters in Table 2.

deductible, while constrained individuals rely on the insurance policy to smooth across states of the world, and thus value the less "lumpy" resetting deductible at the cost of slightly worse insurance. This is reminiscent to the result in Ericson and Sydnor (2018) that liquidity constrained individuals prefer seemingly dominated plans with lower deductibles and higher (and more frequent) premium payments.

To further understand this feature that resetting deductibles provide alternative liquidity, take the extreme case of 100% probability of a very large shock in both periods that surpasses both deductibles (note that this implies $D_R = D_S/2$). While unconstrained individuals are indifferent between the two policies (because they are actuarially equivalent and it is costless to save and borrow), constrained individuals strongly prefer the resetting deductible policy (beyond any insurance it provides, since in this case there is no uncertainty) because it provides smoothing of health costs across time periods.

We next consider the case of individuals who can save and borrow, but at higher borrowing costs. Specifically, we now allow individuals to save at rate $R_s = 1.0$ and borrow at rate $1.0 < R_b < \infty$, where $R_b > R_s$ captures that it is more costly to borrow than it is beneficial to save. The middle four specifications in Figure 5 report willingness-to-pay estimates for low borrowing costs ($R_b = 1.015$, equivalent to a 3% annual interest rate), borrowing costs that make individuals indifferent between the resetting deductible and the standard deductible ($R_b = 1.026$), higher borrowing costs ($R_b = 1.095$), which correspond to average credit card borrowing rates at 20% APR (Ericson and Sydnor, 2018), and very high borrowing costs ($R_b = 2.236$), which correspond to payday loan rates at 400% APR.[23] As it becomes more expensive to borrow, the willingness to pay for a resetting deductible policy increases and becomes positive. At credit card rates, the per-period willingness to pay for the resetting deductible policy is \$13 and at payday loan rates is \$115.

In sum, this baseline model shows that preferences over the deductible timespan depends on the extent of liquidity constraints. For individuals who are not constrained, standard (more time-aggregated) deductible timespans provides better insurance, while resetting (less time-aggregated) deductibles are valuable to liquidity constrained individuals because they are smaller in size and thus provide implicit smoothing.

4.4 Willingness to pay: the role of moral hazard

We next explore the willingness to pay for the resetting deductible policy in an environment with moral hazard, as well as the differential effect of the two deductible policies in curbing moral hazard. As described in Section 3, moral hazard arises when individuals value medical consumption beyond the health shock L, but to a lesser degree than non-medical consumption, and thus only over-consume medical care if they are not paying the full price of medical consumption. However, the cost of this over-consumption feeds back into their *ex-ante* insurance costs in the form of higher premiums or higher deductibles (we model the latter in this paper). Because individuals do not value this over-consumption ex-ante, moral hazard is welfare decreasing.

[23] We offer this range of borrowing costs to capture the heterogeneity in households' financial situations: while some households have access to (cheap) savings, others must appeal to loans from credit cards, retirement savings, or unpaid medical debt, which all come with varying implicit interest rates (Pollitz et al., 2014).

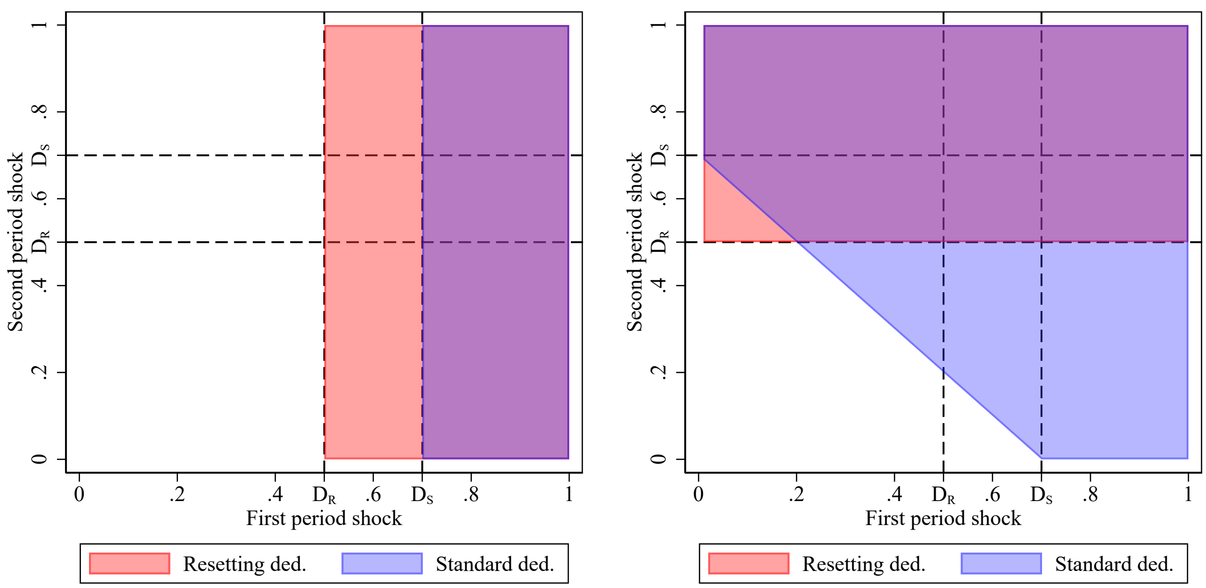
Moral hazard manifests differently in the standard and resetting deductible policies. Figure 6 shows the region of the joint distribution of health shocks in which overconsumption occurs under the resetting and standard deductible policies. Panel (a) shows overconsumption in the first period and Panel (b) shows overconsumption in the second period. In each panel, the x-axis is the first period health shock and the y-axis is the second period health shock, both over a normalized health shock distribution from 0 to 1. D_R and D_S denote example resetting and standard deductible sizes, respectively, and red shading indicates areas in which overconsumption occurs under the resetting deductible policy and blue shading indicates areas in which overconsumption occurs under the standard deductible policy. These figures show that in the first period, there is more extra medical consumption under the resetting deductible, because both shocks that are above D_S induce overconsumption but also shocks between D_R and D_S induce overconsumption. In the second period, the same amount of overconsumption occurs under the resetting deductible policy, while overconsumption under the standard deductible policy covers a much wider area of the joint distribution. In total, which policy generates more overconsumption depends on the empirical joint distribution of health shocks.

Our empirical joint distribution of health shocks suggests that overconsumption is almost exactly the same size under the two policies: individuals consume \$351 more under the standard deductible policy and \$358 more under the resetting deductible policy (see Appendix Figure 1). While this is the case for our empirical distribution, Section 5 shows that the persistence of the health shock distribution can play an important role for the relative overconsumption of the two policies. In particular, lower persistence increases extra consumption under the standard deductible policy, and vice versa for higher persistence.

Consumers ultimately must pay for this additional medical consumption, and because we assume a constant premium P across all specifications, the deductible sizes adjust in response. The second set of bars in Figure 4 shows that the standard deductible rises from \$3,252 without moral hazard to \$4,467 with moral hazard, and the resetting deductible increases from \$1,943 without moral hazard to \$2,748 with moral hazard (both around a 40% increase).

Figure 7 reports the willingness to pay estimates for the baseline case with no moral

Figure 6: Overconsumption due to moral hazard, stylized example



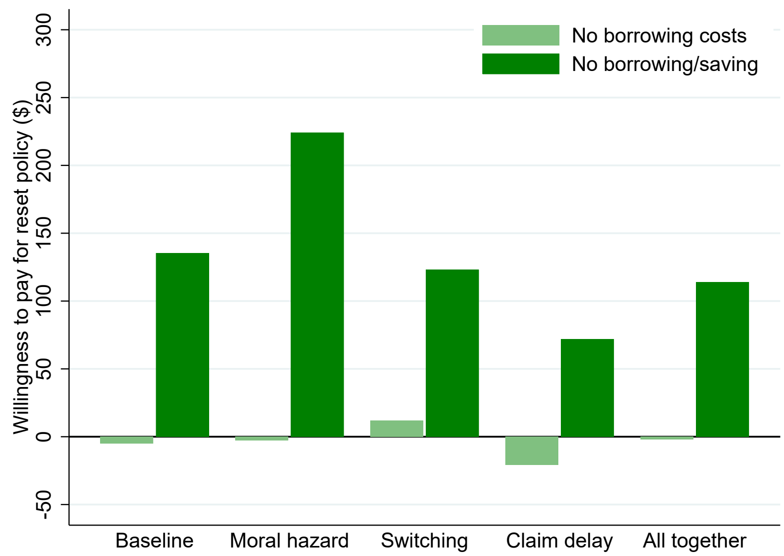
(a) First period overconsumption

(b) Second period overconsumption

Notes: Figure depicts regions of the first period (x-axis) by second period (y-axis) health shock space in which extra medical consumption occurs due to moral hazard for a stylized health distribution and deductible contracts. D_R and D_S are hypothetical deductible sizes. Red shading indicates areas in which overconsumption occurs under the resetting deductible policy and blue shading indicates areas in which overconsumption occurs under the standard deductible policy. Panel (a) graphs the region in which overconsumption occurs in the first period and panel (b) graphs the region in which overconsumption occurs in the second period.

hazard in the first set of bars (with no borrowing costs vs no borrowing or saving) on the left and with moral hazard in the second set of bars (again with no borrowing costs vs no borrowing or saving). With no borrowing costs, the addition of moral hazard changes the per-period willingness-to-pay for the resetting deductible policy from -$5 to -$3, and adding in the inability transfer funds across time increases it further to $224. This corresponds to a willingness to pay of an additional 10.4% in premiums. Thus, resetting deductible policies can be even more valuable as a tool to relieve borrowing constraints under moral hazard environments.

Figure 7: Welfare gain of a resetting deductible policy, by environment



Notes: Figure presents Z, the per-period willingness to pay for the resetting deductible policy, under five environments and within each environment, costless borrowing (light green) and no borrowing or saving (dark green). From left to right, the sets of bars are (1) baseline (i.e., no moral hazard, contract switching, or claim delay), (2) moral hazard only, (3) contract switching only, (4) claim delay only, and (5) all mechanisms in (2)-(4). Other calibrated parameters are in Table 2.

4.5 Willingness to pay: the role of contract switching

The next environment we consider is one in which individuals change health insurance plans mid-way through the year due to events orthogonal to their health, such as job changes. As discussed in Section 3, we model this as an exogenous probability of the same contract restarting at the end of the first period. This mechanically has no effect on the resetting deductible policy, but for the standard deductible policy this effectively induces a reset of the deductible after the first period, but at the size of the standard deductible. The third set of bars in Figure 4 shows the sizes of the deductibles when there is an 8% chance of mid-year plan switching (as estimated in Bjelland et al. (2011)), and confirms that the resetting deductible is the same size as in the baseline case. The standard deductible is slightly smaller than the baseline standard deductible ($3,119 instead of $3,252) because contract switches make some individuals pay the standard deductible twice, and to remain actuarially equivalent, the size must adjust downward.[24]

The third set of bars in Figure 7 shows that when borrowing is costless, the willingness to pay for the resetting deductible policy becomes positive. This is because the standard deductible policy is now more similar in nature to the resetting deductible policy since it can unexpectedly reset. In the case of no borrowing or saving, the willingness to pay for the resetting deductible policy is slightly less positive than the baseline case because the standard deductible is slightly smaller in size, which reduces the liquidity value of the resetting deductible.

There are two caveats to these results. First, our calibrated probability of contract switching is quite low at 8% biannually. If the probability of switching contracts was instead 50%, individuals with no borrowing costs would more strongly prefer the resetting deductible at a willingness-to-pay of $32 per period. Second, it is unclear whether insurance companies account for the fact that some individuals switch plans mid-year into their pricing (Ericson, Geissler and Lubin, 2018). If insurers do not price in switching, the willingness to pay for the resetting deductible policy (with 8% switching) would increase to $45 for individuals with no borrowing costs and $166 for those who cannot borrow or save.

4.6 Willingness to pay: the role of strategic claim delay

The final environment we consider is one in which some health shocks do not necessarily require immediate medical attention (e.g., dental care or some orthopedic surgeries).

[24] At the limit of 100% switching, $D_S = D_R$.

Since health insurance policies – and therefore deductible spending – depend on the timing of health *care* (which is manipulable), not the health *shock* (which is not manipulable), individuals may find it financially beneficial to delay care for such shocks. For instance, in the resetting deductible case, it is typically financially advantageous to bunch care into one period and only pay the deductible once. As discussed in Section 3, we model this phenomenon as a probability q_d of a shock being delayable, and we allow individuals to choose whether to delay shocks when presented with the option. In our two period model, this probability only applies to the first period, while all shocks must receive care in the second period.

Because delay does not affect insurer payouts in the standard deductible policy relative to the baseline model, the standard deductible is equivalent to that of the baseline case. The resetting deductible, on the other hand, is slightly larger than that of the baseline case (roughly \$2,379 vs \$1,943 in Figure 4) because the ability to shift expenditures to a later period raises the probability of only paying the resetting deductible once.[25] To maintain actuarial equivalence, the size of the resetting deductible must therefore increase.

The willingness to pay for the resetting deductible policy decreases in this environment compared to the baseline case, because as the size of the resetting deductible grows in response to individuals delaying care, the more the resetting policy resembles the standard policy. Individuals with no borrowing costs will always choose to delay under the resetting deductible, and are indifferent under the standard deductible. Individuals who cannot borrow or save, however, will never delay under the standard deductible and sometimes choose not to delay under the resetting deductible, in order to spread their health care costs over the two periods. The crucial factors in this choice under the resetting deductible are the size of the initial shock and the subsequent expectation of the second period shock. For small initial shocks with a low expectation of $L_1 + L_2 > D_R$, it is preferable to pay L_1 in the first period and L_2 in the second period rather than $L_1 + L_2$ in the second period, for consumption smoothing gains. For large initial shocks with a high

[25]With the ability to delay, the resetting deductible is \$2,379 for individuals who do not have borrowing costs and \$2,359 for individuals who cannot borrow or save. The discrepancy arises because individuals who cannot borrow or save sometimes prefer not to delay in order to spread potential health care costs between two periods; this results in a deductible that is closer to the baseline case.

expectation of $L_1 + L_2 > D_R$, it is preferable to delay L_1 in order to cap total spending at D_R. Overall, for individuals who cannot borrow or save D_R still grows in size relative to the baseline case, which reduces the willingness to pay for the resetting deductible policy.

Thus, by increasing the size of the resetting deductible, the possibility of claim delay counteracts the beneficial liquidity effects of resetting deductibles. This negative welfare effect is similar in spirit to the more traditional moral hazard channel discussed in Section 4.4 in which ex-post optimal behavior (i.e., once shocks are realized) is distinct from ex-ante welfare. Moreover, if claim delay introduces additional costs (e.g., from additional pain or future health care costs incurred by not caring for the shock immediately), there are potentially additional negative welfare effects of resetting deductibles (as explored in an extension in Section 5).

4.7 Willingness to pay: putting it all together

The final set of main results in Figures 4 and 7 report the deductible sizes and willingness to pay for the resetting deductible policy in an environment in which all features (moral hazard, mid-year contract switching, and claim delay) are present. In this case, both deductibles are larger than the baseline case (predominantly due to moral hazard) but the standard deductible is slightly smaller than in the moral hazard case because mid-year contract switching exerts downward pressure it. On the other hand, the resetting deductible is slightly larger than in the moral hazard case because claim delay exerts upward pressure on it. Overall, the per-period willingness to pay for the resetting deductible policy is similar to the baseline case under both costless borrowing and no borrowing or saving environments ($-2 and $114, respectively).[26]

In sum, our results suggest that while resetting deductibles may have small or even negative welfare benefits under some environments (e.g., when individuals can strategically delay medical care), they may have relatively large welfare gains under other environments, such as for individuals with borrowing constraints.

[26]See Appendix Figure 2 for the range of willingness to pay estimates across various interest rates.

5 Extensions

In this section, we explore several extensions to our main analysis.

5.1 Persistence of the health shock distribution

The willingness to pay for a resetting deductible policy may be particularly sensitive to the persistence of health shocks across the two periods, as it may have liquidity and moral hazard implications. To test how persistence affects our main results, we modify our empirical health shock distribution by making the distribution slightly more persistent, slightly less persistent, and independent.[27]

Table 3 reports the effects of persistence in the case of moral hazard (see Appendix Figure 3 for the full set of results under an independent shock distribution). With more persistent distributions, the standard deductible size increases because individuals are more likely to generate large health shocks in both periods, which is costly to the insurer. On the other hand, overconsumption of medical care under the standard deductible policy decreases because individuals are mechanically less likely to reach a higher deductible (which then triggers overconsumption).[28] The willingness to pay for the resetting deductible policy is not much affected by persistence for individuals without borrowing constraints, but increases with persistence for individuals with borrowing constraints, largely because of the change in deductible sizes.

5.2 Extension to shorter periods

We next extend the model to define utility over shorter periods: instead of two six-month periods, we model six two-month periods.[29] Utility is defined over the period, but the

[27] Specifically, we redistribute 5% of the mass away from the diagonal and distribute the mass equally among all other parts of the joint distribution in our less persistent specification and we redistribute 13% of the mass to the diagonal in our more persistent specification.

[28] Additionally, more overconsumption will occur in the second period under an independent shock distribution than a persistent shock distribution because there are more instances in which an individual is under the deductible in the first period (and thus not overconsuming) and over the deductible in the second period.

[29] Increasing the number of periods exponentially increases computational burden and hence we refrain from monthly periods. Additionally, we use a discretized grid of 10 health shocks; since our main results

Table 3: Sensitivity of results to health shock persistence under moral hazard

Model description	Deductible size ($)		Overconsumption (annual, $)		Willingness to pay (per period, $)	
	D_S	D_R	Standard	Resetting	No bor. costs	No bor/sav
Independent	3769	2748	406	358	3	167
Less persistence	4385	2748	356	358	-2	217
Empirical	4467	2748	351	358	-3	224
More persistence	4673	2748	340	358	-5	244

Notes: Table presents deductible sizes, annual overconsumption due to moral hazard, and per-period willingness to pay for the resetting deductible policy based on an environment with moral hazard and under different levels of persistence of the health shock distribution.

resetting deductible still spans six months and resets after six months and the standard deductible still spans 12 months (see Appendix Figure 4 for the modified schematic in the case of monthly periods). This extension may be important if the dynamics of utility or consumption are more appropriately captured at shorter frequencies rather than biannual frequencies (as in our baseline model) or annual frequencies (as in much of the health insurance literature and beyond, though there are extensions, e.g., Cronin (2019) and Campo (2021)).[30] To keep all else constant between the two models, we disaggregate the empirical health expenditure distribution to two-month arrival rates for both versions of the model, and then re-aggregate the shocks in the six-month model (which may matter computationally due to the discretization of the health expenditure distribution). We also convert the premium and income level into equally divided periodic levels.

Table 4 reports willingness to pay (scaled to biannual amounts to be comparable to our main results) under our two main liquidity environments and without moral hazard, contract switching, or claim delay, and shows that the willingness to pay for the resetting deductible is very similar across the two models. This suggests that the length of a "period" within the model is not the main force behind our empirical findings that resetting deductibles are beneficial for those with borrowing constraints.

are very robust to this smaller health shock grid, we do not expect this to affect our results.

[30]This issue is more generally applicable in settings with separable, discrete-time utility functions, which assume that individuals do not care about consumption fluctuations within a period (or could smooth them out easily on their own) but do so across periods. In our context, the accuracy of this feature of the utility function likely depends on realities about when medical bills can be paid without consequence and how substitutable medical and non-medical consumption are across time.

Table 4: Willingness to pay for resetting deductible (biannualized), shorter periods

Model description	No bor. costs	No bor/sav
Two six-month periods	0	120
Six two-month periods	0	107

Notes: Table presents willingness to pay for the resetting deductible policy under no borrowing costs (first column) and no borrowing or saving (second column) for different period lengths. Premium and income are equally divided between periods to be the same annual amount as in the baseline case, and health shocks are disaggregated to two-month frequencies. The first row is slightly different from the baseline case because of this disaggregation and because the health shock distribution is discretized to 10 points instead of 30 throughout the rest of the paper.

5.3 Three-armed policy

Next we consider a policy that includes three cost-sharing "arms": a deductible arm, a coinsurance arm after the deductible in which the individual pays a percentage of the health care costs above the deductible, and a full insurance arm once an out-of-pocket maximum is reached, at which point the individual has no further out-of-pocket costs.[31] In the "two-armed" policies we have considered thus far (a deductible arm and a full insurance arm), the deductible is equivalent to the out-of-pocket maximum. To investigate how the addition of a coinsurance arm affects our results, we return to the baseline two-period model and now characterize the policies as (P,D_i, co, M) with per-period premium P, deductible D_i where $i = S$ and $i = R$ are the standard deductible and resetting deductible, respectively, co is the coinsurance rate after the deductible, and M is the out-of-pocket maximum that spans the two periods.[32] We set $P = \$2,154$ as in the baseline model, $co = 0.2$, and solve for the deductible sizes D_R and D_S that satisfy the insurer break-even condition under various values of M.[33] Table 5 shows that deductible sizes are much smaller than in the main analysis (because the presence of coinsurance shifts costs to individuals), which leads to smaller willingness-to-pay estimates. Nevertheless,

[31] See Appendix Figure 5 for a schematic of the three-armed policy.

[32] In principle we could assume an out-of-pocket maximum for the standard policy that spans two periods and an out-of-pocket maximum for the resetting policy that spans one period to mimic the differences between the standard and resetting deductible. We chose to use a single out-of-pocket maximum for two reasons: first, distinct out-of-pocket maxima for each deductible adds an additional free parameter and it is not obvious how to pin it down, and second, using a single out-of-pocket maximum keeps the focus strictly on the time aggregation of deductibles.

[33] The average coinsurance rate for silver plans in the federal Marketplace in 2016 was around 25%, and the average out-of-pocket maximum was $6,160 (Kaiser Family Foundation, 2017*a*).

the pattern remains the same: individuals without the ability to borrow prefer shorter deductible spans, while those without borrowing costs prefer the standard deductible.

Table 5: Willingness to pay for resetting deductible, three-armed policy

			Willingness to pay	
Model description	D_S	D_R	No bor costs	No bor/sav
$M = 5,000$	2,482	1,516	-3	70
$M = 6,000$	2,211	1,312	-7	45
$M = 7,000$	2,002	1,177	-7	26

Notes: Table shows willingness to pay for a three-armed policy with a resetting deductible compared to a three-armed policy with a standard deductible, both with an annual out-of-pocket maximum as denoted in each row and a coinsurance rate of 20% after the deductible is reached.

5.4 Other extensions

Finally, we report the effects of other adjustments to our main model. One set of adjustments modifies the health shock distribution to account for various additional features. In Appendix Figure 6, we show that our main results are slightly muted but largely unchanged when we adjust the underlying health shock distribution to (approximately) correct for the fact that observed health care expenditures include the underlying shock plus moral hazard and/or timing manipulation. Appendix Figure 7 shows the effect of reducing out-of-pocket costs by 20% to proxy the effect of Health Savings Accounts (HSAs), which allow individuals to use tax-free dollars for medical care. Results are again slightly muted — which is expected given that overall healthcare expenditure risk is lower — but overall similar to the main results.

A second set of adjustments involves modifications to calibrated parameters. We show robustness to CRRA preferences ($U(x) = \frac{x^{1-\gamma}}{1-\gamma}$) with $\gamma = 4.5$ in Appendix Figure 8, which corresponds closely to our CARA value. We explore a range of incomes, and show that the willingness to pay is especially high for lower income individuals who cannot borrow or save. Appendix Figure 9 reports the willingness to pay for the resetting deductible policy and the corresponding deductibles across a range of delay probabilities q_d, a range of mid-year job switching q_m, and a range of moral hazard parameters w, and shows that the results are both stable and in line with economic intuition.

Finally, in Appendix Table 1 we add an additional cost for individuals who choose to delay care, in the model with only claim delay. In the first alternative specification, individuals incur an additional cost of 5% of the health shock in the *first* period to proxy for the monetary value of something like living an additional period with hip pain. Note that this is not a health *care* cost so it does not factor into the insurance policy. In the second alternative specification, individuals incur an additional health *care* cost of 5% of the health shock in the *second* period to capture potential additional health care costs of delay (if, for example, the shock is exacerbated when not treated immediately). This is a health care cost, so it potentially affects deductible sizes.

A cost of delay has two effects: (1) it can reduce the percent of individuals who delay in the resetting deductible case, which dampens the deductible size effect and thus increases the willingness to pay for the resetting deductible, but (2) it increases costs directly through the delay cost, which decreases the willingness to pay for the resetting deductible. These offsetting effects are evident particularly in the no borrowing/saving case, for which the willingness to pay increases slightly when a 5% health cost is added in the first period (Panel B) but decreases when a 5% cost is added to health care costs in the second period (Panel C).

6 Discussion and conclusions

This paper analyses the extent to which an unexplored feature of health insurance plans—the time aggregation of deductibles—affects the financial well-being of individuals. Specifically, we build and calibrate a dynamic model of within-year health care and consumption choice under uncertainty, and use it to calculate the willingness to pay for alternative deductible time aggregation policies. We show that, in lieu of an annual deductible, an actuarially-equivalent deductible that is smaller in size but resets after six months can generate welfare gains for liquidity constrained populations. On the other hand, resetting deductibles are inherently worse insurance for worst-case health shock scenarios. Thus, whether an individual prefers a resetting deductible policy depends in large part on the trade-off between risk-protection, liquidity, and to a smaller extent, moral hazard, mid-year contract switching, and the delayability of care.

Resetting deductibles are not the only policy tool to combat issues related to liquidity, moral hazard, and the other issues we explore. One set of tools involves financial assistance such as payment plans, which would allow health care costs to be smoothed over a longer period of time. Another set of tools involves rethinking traditional cost-sharing tools (Ericson and Sydnor, 2018). While we explore a "three-armed" policy that incorporates co-insurance into a resetting deductible policy, an interesting alternative policy could more heavily incorporate co-payments in lieu of (or in addition to) deductibles. Co-payments, however, are defined per event rather than over time, and in health care an "event" may not always be easily defined, which raises complex implementation questions. In addition, an excessive number of co-payments within a short period of time could run up against out-of-pocket maxima, which mirrors the question of time aggregation for deductibles.

Our model explores the effect of changing the timespan over which a deductible resets, but holds the timespan of the insurance contract itself fixed. A parallel literature finds large welfare gains of long-term health insurance contracts, which accrue by solving selection and reclassification risk issues (Ghili et al., 2021; Atal et al., 2020). While we abstract from these issues and instead focus on the higher-frequency issues of liquidity and moral hazard, analyzing asymmetric information and selection issues raised by resetting deductibles would be an interesting extension. In particular, it would be fruitful to explore the extent to which correlation between an individual's health risk and their liquidity position creates adverse selection in a market with a resetting deductible plan. Additionally, given the empirical reality of "policy churn" (Diamond et al., 2018), more work on the design of short-term health insurance would help inform policy. This study takes a first step by studying the disaggregation of a particular—and increasingly prominent—component of health insurance contracts.

References

Abaluck, Jason, Jonathan Gruber, and Ashley Swanson. 2018. "Prescription drug use under Medicare Part D: A linear model of nonlinear budget sets." *Journal of Public Economics*, 164: 106–138.

Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen. 2015. "Moral hazard in health insurance: Do dynamic incentives matter?" *Review of Economics and Statistics*, 97(4): 725–741.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review*, 53(5): 941–973.

Atal, Juan Pablo, Hanming Fang, Martin Karlsson, and Nicolas R Ziebarth. 2020. "Long-term health insurance: Theory meets evidence." National Bureau of Economic Research.

Barseghyan, Levon, Jeffrey Prince, and Joshua C Teitelbaum. 2011. "Are risk preferences stable across contexts? Evidence from insurance data." *American Economic Review*, 101(2): 591–631.

Bjelland, Melissa, Bruce Fallick, John Haltiwanger, and Erika McEntarfer. 2011. "Employer-to-employer flows in the United States: Estimates using linked employer-employee data." *Journal of Business & Economic Statistics*, 29(4): 493–505.

Board of Governors of the Federal Reserve System. 2014. "Report on the Economic Well-Being of US Households in 2013." Washington, DC: Federal Reserve Board.

Brot-Goldberg, Zarek C, Amitabh Chandra, Benjamin R Handel, and Jonathan T Kolstad. 2017. "What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics." *The Quarterly Journal of Economics*, 132(3): 1261–1318.

Cabral, Marika. 2017. "Claim timing and ex post adverse selection." *The Review of Economic Studies*, 84(1): 1–44.

Campo, Cecilia S Diaz. 2021. "Dynamic Moral Hazard in Nonlinear Health Insurance Contracts."

Card, David, Carlos Dobkin, and Nicole Maestas. 2008. "The impact of nearly universal insurance coverage on health care utilization: evidence from Medicare." *American Economic Review*, 98(5): 2242–58.

Casaburi, Lorenzo, and Jack Willis. 2018. "Time versus State in Insurance: Experimental Evidence from Contract Farming in Kenya." *American Economic Review*, 108(12): 3778–3813.

Chen, Steven, Paul R Shafer, Stacie B Dusetzina, and Michal Horný. 2021. "Annual Out-Of-Pocket Spending Clusters Within Short Time Intervals: Implications For Health Care Affordability." *Health Affairs*, 40(2): 274–280.

Cohen, Alma. 2006. "The disadvantages of aggregate deductibles." *Topics in Economic Analysis & Policy*, 6(1).

Cohen, Alma, and Liran Einav. 2007. "Estimating risk preferences from deductible choice." *American Economic Review*, 97(3): 745–788.

Cronin, Christopher J. 2019. "Insurance-Induced Moral Hazard: A Dynamic Model of Within-Year Medical Care Decision Making Under Uncertainty." *International Economic Review*, 60(1): 187–218.

Cutler, David M, and Richard J Zeckhauser. 2000. "The anatomy of health insurance." In *Handbook of health economics*. Vol. 1, 563–643. Elsevier.

Dalton, Christina M, Gautam Gowrisankaran, and Robert J Town. 2019. "Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D." *The Review of Economic Studies*.

Department of Health and Human Services. 2018. "Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2019." *Federal Register*, 83(74): 16930–17071.

Devenir Research Team. 2018. "2017 Year-End HSA Market Statistics & Trends Exective Summary." Devenir Group.

Diamond, Rebecca, Michael J Dickstein, Timothy McQuade, and Petra Persson. 2018. "Take-up, drop-out, and spending in ACA marketplaces." National Bureau of Economic Research.

Dolan, Rachel. 2016. "High-Deductible Health Plans." *Health Affairs Health Policy Brief*.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2015. "The response of drug expenditure to nonlinear contract design: Evidence from Medicare Part D." *The Quarterly Journal of Economics*, 130(2): 841–899.

Einav, Liran, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen. 2013. "Selection on moral hazard in health insurance." *American Economic Review*, 103(1): 178–219.

Ellis, Randall P, Shenyi Jiang, and Willard G Manning. 2015. "Optimal health insurance for multiple goods and time periods." *Journal of Health Economics*, 41: 89–106.

Ericson, Keith Marzilli, and Justin R Sydnor. 2018. "Liquidity Constraints and the Value of Insurance." National Bureau of Economic Research.

Ericson, Keith M. Marzilli, Kimberley H. Geissler, and Benjamin Lubin. 2018. "The Impact of Partial-Year Enrollment on the Accuracy of Risk-Adjustment Systems: A Framework and Evidence." *American Journal of Health Economics*, 4(4): 454–478.

for Medicare, Centers, and Medicaid Services. 2016. "March 31, 2016 HHS-Operated Risk Adjustment Methodology Meeting." Discussion Paper.

Ghili, Soheil, Benjamin Handel, Igal Hendel, and Michael D Whinston. 2021. "Optimal long-term health insurance contracts: characterization, computation, and welfare effects." *Cowles Foundation Discussion Paper*.

Guo, Audrey, and Jonathan Zhang. 2019. "What to expect when you are expecting: Are health care consumers forward-looking?" *Journal of Health Economics*, 67: 102216.

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

Helmchen, Lorens A, David W Brown, Ithai Z Lurie, and Anthony T Lo Sasso. 2015. "Health savings accounts: growth concentrated among high-income households and large employers." *Health Affairs*, 34(9): 1594–1598.

Kaiser Family Foundation. 2017*a*. "Cost-Sharing for Plans Offered in the Federal Marketplace for 2017."

Kaiser Family Foundation. 2017*b*. "State Health Facts."

Kaiser Family Foundation. 2020. "Employer Health Benefits: 2020 Annual Survey." Henry Kaiser Family Foundation.

Li, Chu-Shiu, Chwen-Chi Liu, and Jia-Hsing Yeh. 2007. "The Incentive Effects of Increasing Per-Claim Deductible Contracts in Automobile Insurance." *Journal of Risk and Insurance*, 74(2): 441–459.

Lin, Haizhen, and Daniel W Sacks. 2019. "Intertemporal substitution in health care demand: Evidence from the RAND Health Insurance Experiment." *Journal of Public Economics*, 175: 29–43.

Liu, Chenyuan. 2021. "Sorting on Plan Design: Theory and Evidence from the ACA." working paper.

Marone, Victoria R, and Adrienne Sabety. 2022. "When Should There Be Vertical Choice in Health Insurance Markets?" *American Economic Review*, 112(1): 304–42.

Michel-Kerjan, Erwann O, and Carolyn Kousky. 2010. "Come rain or shine: Evidence on flood insurance purchases in Florida." *Journal of Risk and Insurance*, 77(2): 369–397.

Parsons, Christopher A, and Edward D Van Wesep. 2013. "The timing of pay." *Journal of Financial Economics*, 109(2): 373–397.

Pollitz, Karen, Cynthia Cox, Kevin Lucia, and Katie Keith. 2014. "Medical debt among people with health insurance." *Kaiser Family Foundation*.

Rae, Matthew, Gary Claxton, and Larry Levitt. 2017. "Do Health Plan Enrollees Have Enough Money to Pay Cost Sharing?" *Kaiser Family Foundation Issue Brief*.

Shapiro, Jesse M. 2005. "Is there a daily discount rate? Evidence from the food stamp nutrition cycle." *Journal of Public Economics*, 89(2): 303 – 325.

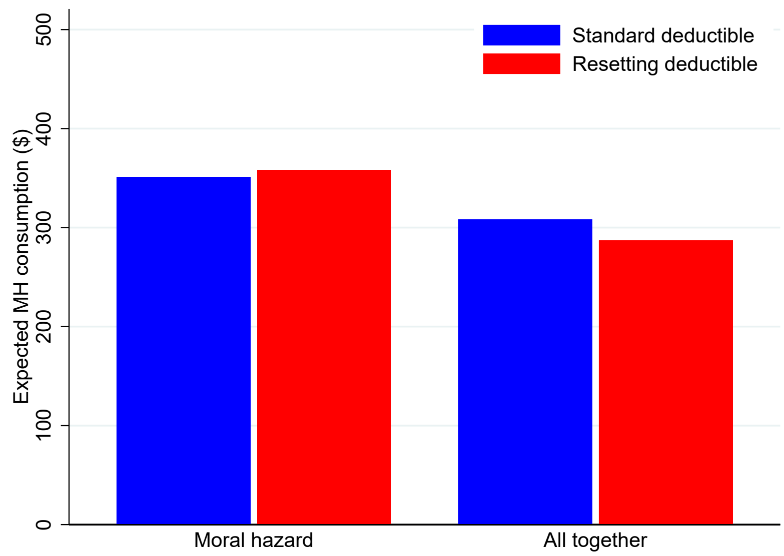
Sydnor, Justin. 2010. "(Over) insuring modest risks." *American Economic Journal: Applied Economics*, 2(4): 177–99.

Towers Watson. 2015. "High-Performance Insights: Best Practices in Health Care." *20th Annual Towers Watson/National Business Group on Health Employer Survey on Purchasing Value in Health Care*.

U.S. Census Bureau. 2014. "POV-01. Age and Sex of All People, Family Members and Unrelated Individuals." https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pov/pov-01.2014.html, Accessed: 2022-06-18.

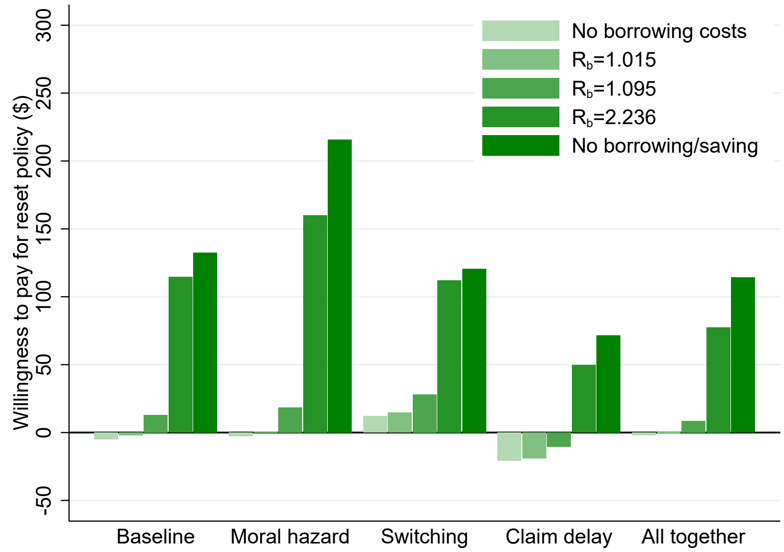
Appendix figures and tables

Appendix Figure 1: Overconsumption due to moral hazard, by environment



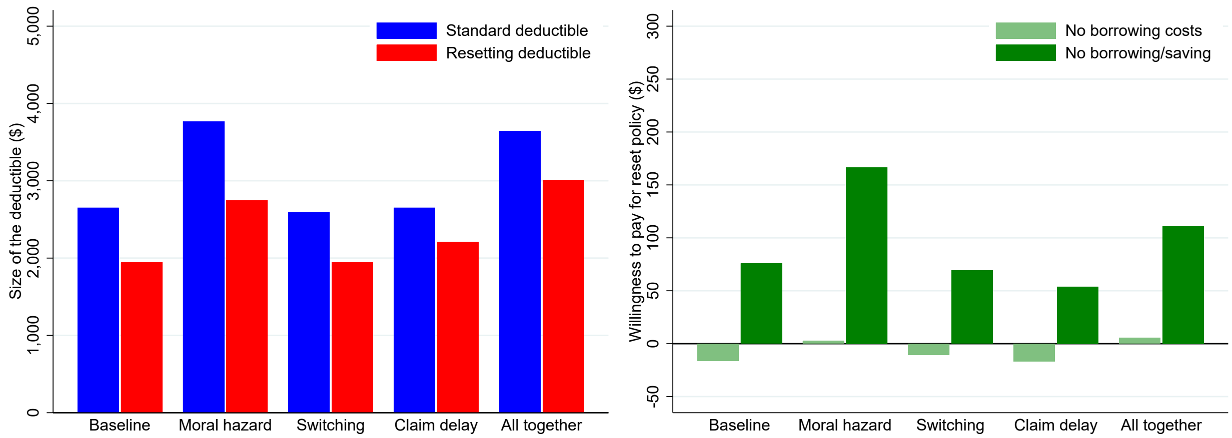
Notes: Figure presents the total extra medical consumption over the two periods due to moral hazard. The blue bars denote extra consumption under the standard deductible policy and red bars denote extra consumption under the resetting deductible policy. The bars on the left correspond to the environment with moral hazard only and the bars on the right correspond to the environment with moral hazard, contract switching, and claim delay.

Appendix Figure 2: Welfare gain of a resetting deductible policy, by environment and extra interest rates



Notes: Figure presents Z, the per-period willingness to pay for the resetting deductible policy, under five environments and within each environment, costless borrowing (lightest green), no borrowing or saving (darkest green), and a range of R_b values in between. From left to right, the sets of bars are (1) baseline (i.e., no moral hazard, contract switching, or claim delay), (2) moral hazard only, (3) contract switching only, (4) claim delay only, and (5) all mechanisms in (2)-(4). Other calibrated parameters are in Table 2.

Appendix Figure 3: Deductibles and willingness to pay under independent shock distribution

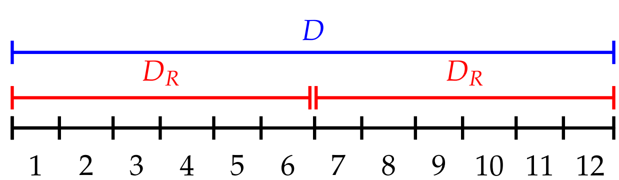


(a) Deductible sizes

(b) Willingness to pay for resetting ded. policy

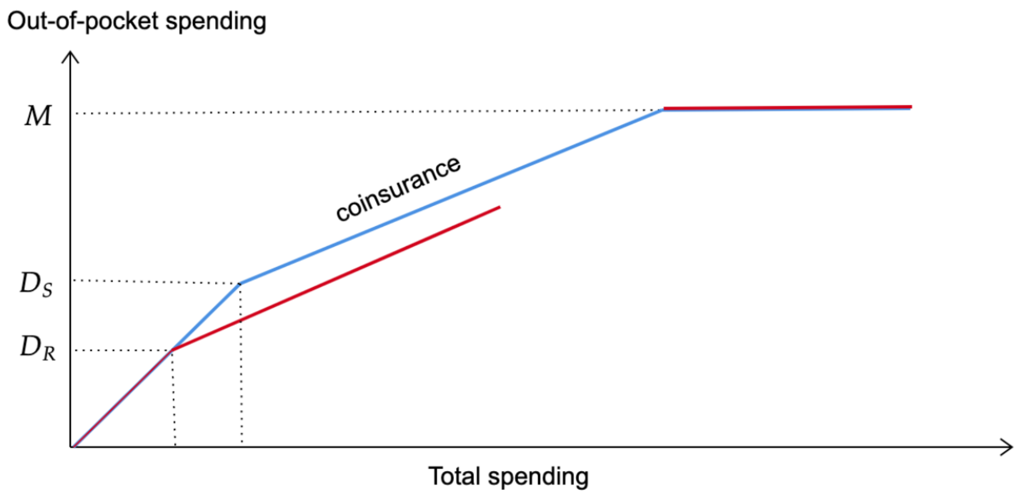
Notes: Left figure presents the size of the standard (annual) deductible D_S in blue and the size of the resetting (biannual) deductible D_R in red, under four scenarios: from left to right, (1) the baseline scenario with no moral hazard ($w = 0$), no contract switching ($q_m = 0$), and no claim delay (q_b), (2) the scenario with only moral hazard, (3) the scenario with only contract switching, (3) the scenario with only claim delay, and (4) the scenario with all three of moral hazard, switching, and delay. Right figure presents the per-period willingness to pay for the resetting deductible policy for individuals who can costlessly borrow in light green and who cannot borrow in dark green for the same scenarios as in Panel (a).

Appendix Figure 4: Deductible timespan schematic, monthly periods



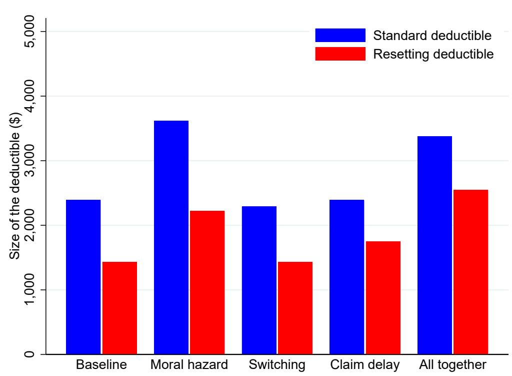
Notes: Figure depicts the length of time over which a deductible applies and resets. The black lines and numbers denote periods. D_R in red is the reset policy in which the deductible resets after six periods, while D_S in blue is the standard policy in which the deductible resets after twelve periods.

Appendix Figure 5: Three-armed policy schematic

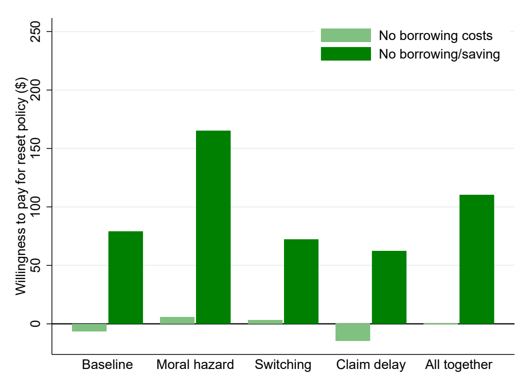


Notes: Figure depicts the three-armed policy as described in Section 5.

Appendix Figure 6: Deductibles and willingness to pay when the health shock is net of moral hazard and claim delay



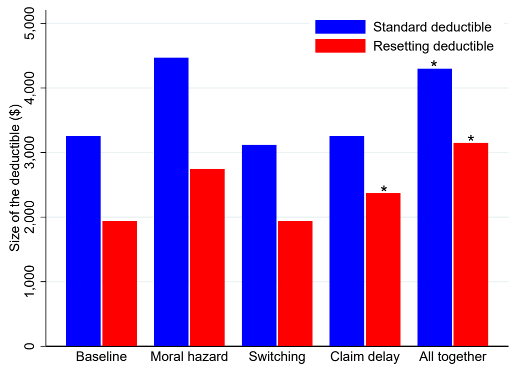
(a) Deductible sizes



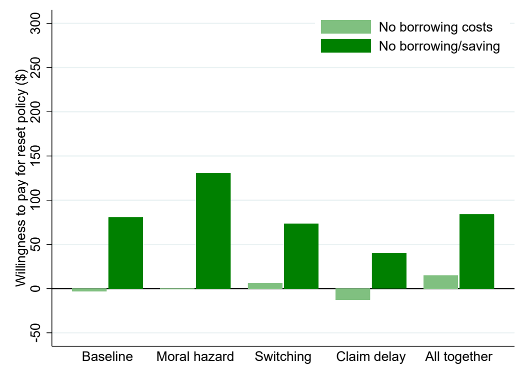
(b) Willingness to pay for resetting ded. policy

Notes: Left figure presents deductible sizes and right figure presents the per-period willingness to pay for the resetting deductible policy when the underlying health shock distribution is roughly purged of moral hazard and claim delay. We subtract moral hazard using the following (rough) rules: denoting L_{1e} and L_{2e} as the health expenditures observed in the data and L_1 and L_2 as the recovered health shocks, if $L_{1e} \geq D_S + w$ then $L_1 = L_{1e} - w$ and $L_2 = \max\{L_{2e} - w, 0\}$; if $L_{1e} < D_S + w$ and $L_{1e} + L_{2e} \geq D_S + w$ then $L_2 = \max\{L_{2e} - w, 0\}$. We then adjust for claim delay by drawing a 40% random sample of the subset of claims for which $L_{1e} = 0$ and $L_{2e} > 0$, and shifting $0.5L_{2e}$ to L_{1e}.

Appendix Figure 7: Deductibles and willingness to pay when health costs are tax-free



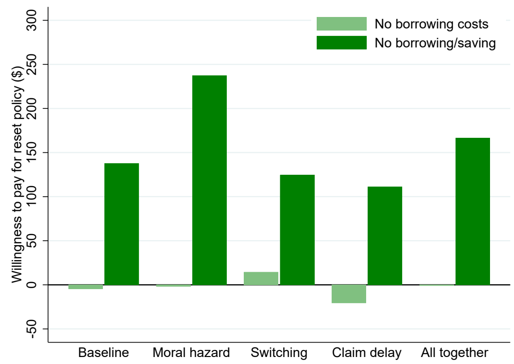
(a) Deductible sizes



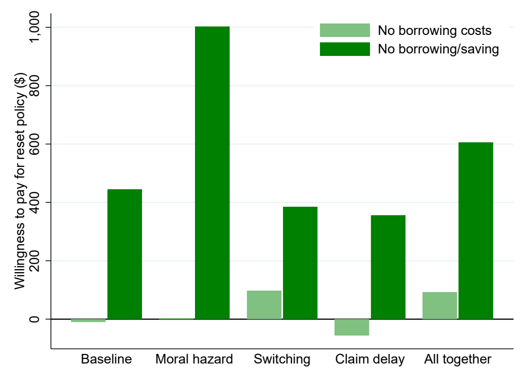
(b) Willingness to pay for resetting ded. policy

Notes: Left figure presents deductible sizes and right figure presents the per-period willingness to pay for the resetting deductible policy when out-of-pocket health expenditures are tax-free, as proxied by a 20% reduction in out-of-pocket costs.

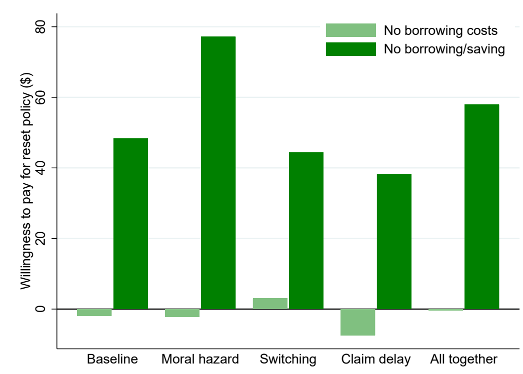
Appendix Figure 8: Welfare gain of a resetting deductible policy, CRRA, by income



(a) Income Y = 250% FPL (baseline)



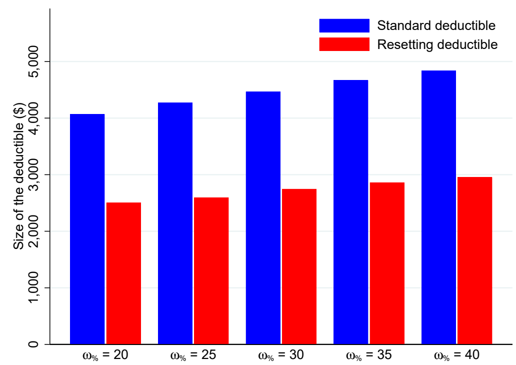
(b) Income Y = 150% FPL



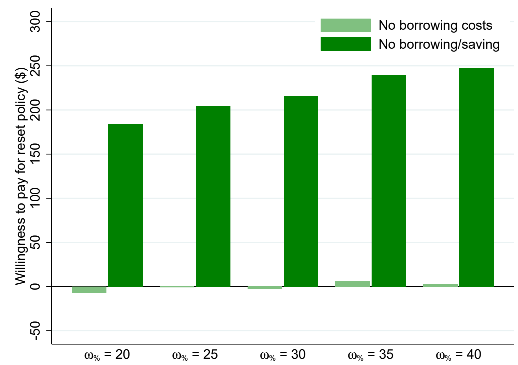
(c) Income Y = 500% FPL

Notes: Each figure presents Z, the per-period willingness to pay for the resetting deductible policy for CRRA utility with risk aversion parameter of 4.5, under five environments and within each environment, costless borrowing (light green) and no borrowing or saving (dark green). From left to right, the sets of bars are (1) baseline (i.e., no moral hazard, contract switching, or claim delay), (2) moral hazard only, (3) contract switching only, (4) claim delay only, and (5) all mechanisms in (2)-(4). Other calibrated parameters are in Table 2 except the utility function, which is now CRRA with parameter $\gamma = 4.5$ ($U(x) = \frac{x^{1-\gamma}}{1-\gamma}$) and income Y, which is denoted in the subtitles.

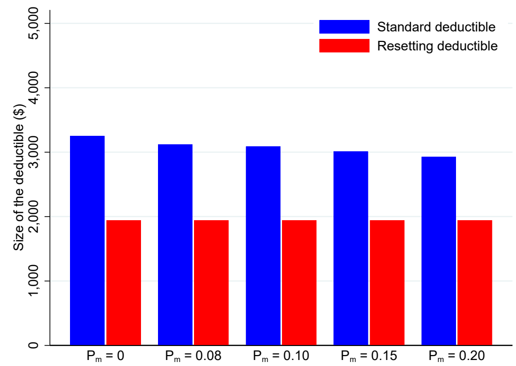
Appendix Figure 9: WTP and deductibles by varying parameter values



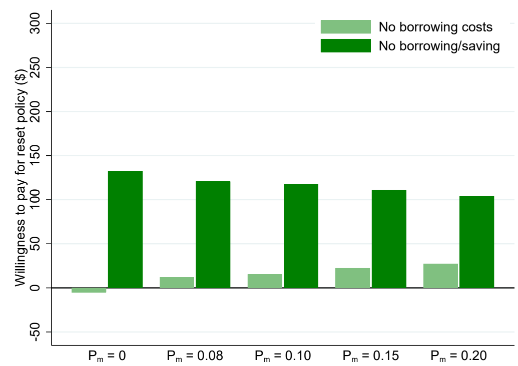
(a) Ded., varying moral hazard



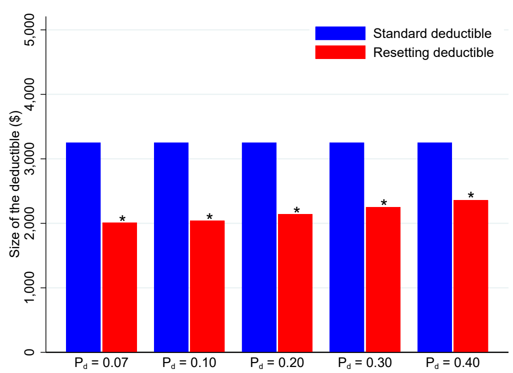
(b) WTP, varying moral hazard



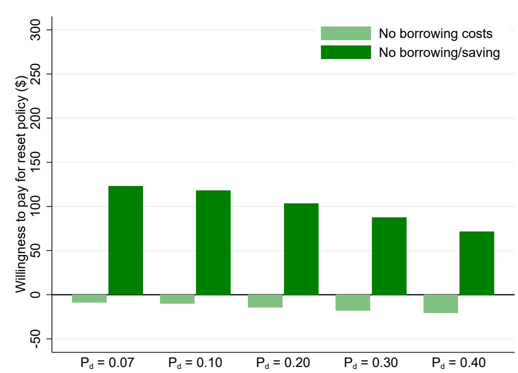
(c) Ded., varying prob. of switching



(d) WTP, varying prob. of switching



(e) Ded., varying prob. of delayability



(f) WTP, varying prob. of delayability

Notes: Left figures present deductible sizes and right figures present the per-period willingness to pay for the resetting deductible policy under the moral hazard environment with varying moral hazard parameters w (subfigures (a) and (b)), under the mid-year job switching environment with varying probability of mid-year job switching q_m (subfigures (c and (d)), and under the endogenous claim delay environment with varying probability of delayability q_d (subfigures (e) and (f)).

Appendix Table 1: Willingness to pay for resetting deductible, claim delay case with various costs of delay

Liquidity environment:	No borrowing costs		No borrowing/saving	
Deductible policy:	Standard	Resetting	Standard	Resetting
Panel A: Zero cost of delay				
Percent that delay (%)	0	80	0	40
Deductible size ($)	3252	2379	3252	2359
Willingness to pay ($)	—	-20.8	—	71.6
Panel B: 5% health cost in 1st period				
Percent that delay (%)	0	76.7	0	33.3
Deductible size ($)	3252	2319	3252	2292
Willingness to pay ($)	—	-35.2	—	72.9
Panel C: 5% health care cost in 2nd period				
Percent that delay (%)	0	80	0	33.3
Deductible size ($)	3252	2506	3252	2466
Willingness to pay ($)	—	-62.4	—	32.3

Notes: Table reports the percent of individuals who delay (where we assume they do not delay if they are indifferent), the deductible sizes, and willingness to pay for the resetting deductible policy, in an environment with endogenous claim delay but not moral hazard or exogenous mid-year contract switching. The 5% cost corresponds to 5% of the first period health shock, and is applied as a non-health care cost to the first period in Panel B and as a health care cost to the second period in Panel C.