Online Appendix
Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts

Benjamin Handel[*]

August 29, 2013

Abstract

This online appendix provides supporting analysis for the primary manuscript 'Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts' published in the *American Economic Review*. Appendix A describes the cost model setup and estimation. Appendix B describes the choice model estimation algorithm in greater detail. Appendix C discusses a robust check allowing for moral hazard, i.e. a price elasticity of demand for medical expenditures. Appendix D discusses different interpretations of and foundations for inertia. Appendix E presents additional supporting analyses.

---

[*]Department of Economics, University of California at Berkeley. Contact: handel@berkeley.edu.

Online Appendix A: Cost Model Setup and Estimation

This appendix describes the details of the cost model, which is summarized at a high-level in section 3. The output of this model, F_{kjt}, is a family-plan-time specific distribution of predicted out-of-pocket expenditures for the upcoming year. This distribution is an important input into the choice model, where it enters as a family's predictions of its out-of-pocket expenses at the time of plan choice, for each plan option. We predict this distribution in a sophisticated manner that incorporates (i) past diagnostic information (ICD-9 codes) (ii) the Johns Hopkins ACG predictive medical software package (iii) a novel non-parametric model linking modeled health risk to total medical expenditures using observed cost data and (iv) a detailed division of medical claims and health plan characteristics to precisely map total medical expenditures to out-of-pocket expenses.[1] The level of precision we gain from the cost model leads to more credible estimates of the choice parameters of primary interest (e.g. inertia).

In order to most precisely predict expenses, we categorize the universe of total medical claims into four mutually exclusive and exhaustive subdivisions of claims using the claims data. These categories are (i) hospital and physician (ii) pharmacy (iii) mental health and (iv) physician office visit. We divide claims into these four specific categories so that we can accurately characterize the plan-specific mappings from total claims to out-of-pocket expenditures since each of these categories maps to out-of-pocket expenditures in a different manner. We denote this four dimensional vector of claims C_{it} and any given element of that vector $C_{d,it}$ where $d \in D$ represents one of the four categories and i denotes an individual (employee or dependent). After describing how we predict this vector of claims for a given individual, we return to the question of how we determine out-of-pocket expenditures in plan j given C_{it}.

Denote an individual's past year of medical diagnoses and payments by ξ_{it} and the demographics age and sex by ζ_{it}. We use the ACG software mapping, denoted A, to map these characteristics into a predicted mean level of health expenditures for the upcoming year, denoted θ:

$$A : \xi \times \zeta \to \theta$$

In addition to forecasting a mean level of total expenditures, the software has an application that predicts future mean pharmacy expenditures. This mapping is analogous to A and outputs a prediction λ for future pharmacy expenses.

We use the predictions θ and λ to categorize similar groups of individuals across each of four claims categories in vector in C_{it}. Then for each group of individuals in each claims category, we use the actual ex post realized claims for that group to estimate the ex ante distribution for each individual under the assumption that this distribution is identical for all individuals within the cell. Individuals are categorized into cells based on different metrics for each of the four elements of C:

Pharmacy:	λ_{it}
Hospital / Physician (Non-OV):	θ_{it}
Physician Office Visit:	θ_{it}
Mental Health:	$C_{MH,i,t-1}$

For pharmacy claims, individuals are grouped into cells based on the predicted future mean phar-

---

[1] Features (iii) and (iv) are methodological advances. We are aware of only one previous study that incorporates diagnostic information in cost prediction for the purposes of studying plan choice (Carlin and Town (2009)) in a structural setup. Recent work by Einav et al. (2013) use this type of framework as well.

macy claims measure output by the ACG software, λ_{it}. For the categories of hospital / physician (non office visit) and physician office visit claims individuals are grouped based on their mean predicted total future health expenses, θ_{it}. Finally, for mental health claims, individuals are grouped into categories based on their mental health claims from the previous year, $C_{MH,i,t-1}$ since (i) mental health claims are very persistent over time in the data and (ii) mental health claims are uncorrelated with other health expenditures in the data. For each category we group individuals into a number of cells between 8 and 10, taking into account the trade off between cell size and precision. The minimum number of individuals in any cell is 73 while almost all cells have over 500 members. Thus since there are four categories of claims, each individual can belong to one of approximately 10^4 or 10,000 combination of cells.

Denote an arbitrary cell within a given category d by z. Denote the population in a given category-cell combination (d, z) by I_{dz}. Denote the empirical distribution of ex-post claims in this category for this population $\hat{G_{I_{dz}}}(\cdot)$. Then we assume that each individual in this cell has a distribution equal to a continuous fit of $\hat{G_{I_{dz}}}(\cdot)$, which we denote G_{dz}:

$$\varpi : \hat{G_{I_{dz}}}(\cdot) \rightarrow G_{dz}$$

We model this distribution continuously in order to easily incorporate correlations across d. Otherwise, it would be appropriate to use $G_{I_{dz}}$ as the distribution for each cell.

The above process generates a distribution of claims for each d and z but does not model correlations over D. It is important to model correlation over claim categories because it is likely that someone with a bad expenditure shock in one category (e.g. hospital) will have high expenses in another area (e.g. pharmacy). We model correlation at the individual level by combining marginal distributions $G_{idt} \forall$ d with empirical data on the rank correlations between pairs (d, d').[2] Here, G_{idt} is the distribution G_{dz} where $i \in I_{dz}$ at time t. Since correlations are modeled across d we pick the metric θ to group people into cells for the basis of determining correlations (we use the same cells that we use to determine group people for hospital and physician office visit claims). Denote these cells based on θ by z_θ. Then for each cell z_θ denote the empirical rank correlation between claims of type d and type d' by $\rho_{z_\theta}(d, d')$. Then, for a given individual i we determine the joint distribution of claims across D for year t, denoted $H_{it}(\cdot)$, by combining i's marginal distributions for all d at t using $\rho_{z_\theta}(d, d')$:

$$\Psi : G_{iDt} \times \rho_{z_{\theta_{it}}}(D, D') \rightarrow H_{it}$$

Here, G_{iDt} refers to the set of marginal distributions $G_{idt} \forall d \in D$ and $\rho_{z_{\theta_{it}}}(D, D')$ is the set of all pairwise correlations $\rho_{z_{\theta_{it}}}(d, d') \forall (d, d') \in D^2$. In estimation we perform Ψ by using a Gaussian copula to combine the marginal distribution with the rank correlations, a process which we describe momentarily.

The final part of the cost model maps the joint distribution H_{it} of the vector of total claims C over the four categories into a distribution of out of pocket expenditures for each plan. For each of the three plan options we construct a mapping from the vector of claims C to out of pocket expenditures OOP_j:

$$\Omega_j : C \rightarrow OOP_j$$

This mapping takes a given draw of claims from H_{it} and converts it into the out of pocket expenditures an individual would have for those claims in plan j. This mapping accounts for plan-specific

---

[2]It is important to use rank correlations here to properly combine these marginal distribution into a joint distribution. Linear correlation would not translate empirical correlations to this joint distribution appropriately.

features such as the deductible, co-insurance, co-payments, and out of pocket maximums listed in table A-2. I test the mapping Ω_j on the actual realizations of the claims vector C to verify that our mapping comes close to reconstructing the true mapping. Our mapping is necessarily simpler and omits things like emergency room co-payments and out of network claims. We constructed our mapping with and without these omitted categories to insure they did not lead to an incremental increase in precision. We find that our categorization of claims into the four categories in C passed through our mapping Ω_j closely approximates the true mapping from claims to out-of-pocket expenses. Further, we find that it is important to model all four categories described above: removing any of the four makes Ω_j less accurate. Figure A-1 shows the results of one validation exercise for PPO_{250}. The top panel reveals that actual employee out-of-pocket spending amounts are quite close to those predicted by Ω_j, indicating the precision of this mapping. The bottom panel repeats this mapping when we add out of network expenses as a fifth category. The output in this case is similar to that in the top panel without this category, implying that including this category would not markedly change the cost model results.

Once we have a draw of OOP_{ijt} for each i (claim draw from H_{it} passed through Ω_j) we map individual out of pocket expenditures into family out of pocket expenditures. For families with less than two members this involves adding up all the within family OOP_{ijt}. For families with more than three members there are family level restrictions on deductible paid and out-of-pocket maximums that we adjust for. Define a family k as a collection of individuals i_k and the set of families as K. Then for a given family out-of-pocket expenditures are generated:

$$\Gamma_j : OOP_{i_k,jt} \rightarrow OOP_{kjt}$$

To create the final object of interest, the family-plan-time specific distribution of out of pocket expenditures $F_{kjt}(\cdot)$, we pass the claims distributions H_{it} through Ω_j and combine families through Γ_j. $F_{kjt}(\cdot)$ is then used as an input into the choice model that represents each family's information set over future medical expenses at the time of plan choice. Eventually, we also use H_{it} to calculate total plan cost when we analyze counterfactual plan pricing based on the average cost of enrollees. Figure A-2 outlines the primary components of the cost model pictorially to provide a high-level overview and to ease exposition.

We note that the decision to do the cost model by grouping individuals into cells, rather then by specifying a more continuous form, has costs and benefits. The cost is that all individuals within a given cell for a given type of claims are treated identically. The benefit is that our method produces local cost estimates for each individual that are not impacted by the combination of functional form and the health risk of medically different individuals. Also, the method we use allows for flexible modeling across claims categories. Finally, we note that we map the empirical distribution of claims to a continuous representation because this is convenient for building in correlations in the next step. The continuous distributions we generate very closely fit the actual empirical distribution of claims across these four categories.

Cost Model Identification and Estimation. The cost model is identified based on the two assumptions of (i) no moral hazard / selection based on private information and (ii) that individuals within the same cells for claims d have the same ex ante distribution of total claims in that category. Once these assumptions are made, the model uses the detailed medical data, the Johns Hopkins predictive algorithm, and the plan-specific mappings for out of pocket expenditures to generate the the final output $F_{kjt}(\cdot)$. These assumptions, and corresponding robustness analyses, are discussed at more length in the main text.

Once we group individuals into cells for each of the four claims categories, there are two statistical

components to estimation. First, we need to generate the continuous marginal distribution of claims for each cell z in claim category d, G_{dz}. To do this, we fit the empirical distribution of claims $G_{I_{dz}}$ to a Weibull distribution with a mass of values at 0. We use the Weibull distribution instead of the log-normal distribution, which is traditionally used to model medical expenditures, because we find that the log-normal distribution overpredicts large claims in the data while the Weibull does not. For each d and z the claims greater than zero are estimated with a maximum likelihood fit to the Weibull distribution:

$$\max_{(\alpha_{dz}, \beta_{dz})} \Pi_{i \in I_{dz}} \frac{\beta_{dz}}{\alpha_{dz}} (\frac{c_{id}}{\alpha_{dz}})^{\beta_{dz}-1} e^{-(\frac{c_{id}}{\alpha_{dz}})^{\beta_{dz}}}$$

Here, $\hat{\alpha_{dz}}$ and $\hat{\beta_{dz}}$ are the shape and scale parameters that characterize the Weibull distribution. Denoting this distribution $W(\hat{\alpha_{dz}}, \hat{\beta_{dz}})$ the estimated distribution $\hat{G_{dz}}$ is formed by combining this with the estimated mass at zero claims, which is the empirical likelihood:

$$\hat{G_{dz}}(c) = \begin{cases} G_{I_{dz}}(0) & \text{if } c = 0 \\ G_{I_{dz}}(0) + \frac{W(\hat{\alpha_{dz}}, \hat{\beta_{dz}})(c)}{1 - G_{I_{dz}}(0)} & \text{if } c > 0 \end{cases}$$

Again, we use the notation $\hat{G_{iDt}}$ to represent the set of marginal distributions for i over the categories d: the distribution for each d depends on the cell z an individual i is in at t. We combine the distributions $\hat{G_{iDt}}$ for a given i and t into the joint distribution H_{it} using a Gaussian copula method for the mapping Ψ. Intuitively, this amounts to assuming a parametric form for correlation across $\hat{G_{iDt}}$ equivalent to that from a standard normal distribution with correlations equal to empirical rank correlations $\rho_{z_{\theta_{it}}}(D, D')$ described in the previous section. Let $\Phi^i_{1|2|3|4}$ denote the standard multivariate normal distribution with pairwise correlations $\rho_{z_{\theta_{it}}}(D, D')$ for all pairings of the four claims categories D. Then an individual's joint distribution of non-zero claims is:

$$\hat{H_{i,t}}(\cdot) = \Phi_{1|2|3|4}(\Phi_1^{-1}(\hat{G_{id_1t}}), \Phi_2^{-1}(\hat{G_{id_2t}}), \Phi_3^{-1}(\hat{G_{id_3t}}), \Phi_4^{-1}(\hat{G_{id_4t}}))))$$

Above, Φ_d is the standard marginal normal distribution for each d. $\hat{H}_{i,t}$ is the joint distribution of claims across the four claims categories for each individual in each time period. After this is estimated, we determine our final object of interest $F_{kjt}(\cdot)$ by simulating K multivariate draws from $\hat{H}_{i,t}$ for each i and t, and passing these values through the plan-specific total claims to out of pocket mapping Ω_j and the individual to family out of pocket mapping Γ_j. The simulated $F_{kjt}(\cdot)$ for each k, j, and t is then used as an input into estimation of the choice model.

Table A-1 presents summary results from the cost model estimation for the final choice model sample, including population statistics on the ACG index θ, the Weibull distribution parameters $\hat{\alpha_{dz}}$ and $\hat{\beta_{dz}}$ for each category d, as well as the across category rank correlations $\rho_{z_{\theta_{it}}}(D, D')$. These are the fundamentals inputs used to generate F_{kjt}, as described above, and lead to very accurate characterizations of the overall total cost and out-of-pocket cost distributions (validation exercises which are not presented here).

Online Appendix B: Choice Model Estimation Algorithm Details

This appendix describes the details of the choice model estimation algorithm. The corresponding section in the text provided a high-level overview of this algorithm and outlined the estimation assumptions we make regarding choice model fundamentals and their links to observable data.

We estimate the choice model using a random coefficients simulated maximum likelihood approach similar to that summarized in Train (2009). The simulated maximum likelihood estimation approach has the minimum variance for a consistent and asymptotically normal estimator, while not being too computationally burdensome in our framework. Since we use panel data, the likelihood function at the family level is computed for a *sequence* of choices from t_0 to t_2, since inertia implies that the likelihood of a choice made in the current period depends on the choice made in the previous period. The maximum likelihood estimator selects the parameter values that maximize the similarity between actual choices and choices simulated with the parameters.

First, the estimator simulates Q draws from the distribution of health expenditures output from the cost model, F_{kjt}, for each family, plan, and time period. These draws are used to compute plan expected utility conditional on all other preference parameters. It then simulates S draws for each family from the distributions of the random coefficients γ_k and δ_k, as well as from the distribution of the preference shocks $\epsilon_j(Y_k)$.[3]

We define the set of parameters θ as the full set of ex ante model parameters (before the S draws are taken):

$$\theta \equiv (\mu, \beta, \sigma_\gamma^2, \mu_\delta(Y_k), \sigma_\delta(Y_k), \alpha, \mu_{\epsilon_J}(Y_k), \sigma_{\epsilon_J}(Y_k), \eta_0, \eta_1, \eta_2).$$

We denote θ_{sk} one draw derived from these parameters for each family, including the parameters constant across draws:

$$\theta_{sk} \equiv (\gamma_k, \delta_k, \alpha, \epsilon_{JT}, \eta_0, \eta_1, \eta_2)$$

Denote θ_{Sk} the set of all S simulated draws for family k. For each θ_{sk} the estimator then uses all Q health draws to compute family-plan-time-specific expected utilities U_{skjt} following the choice model outlined in earlier in section 3. Given these expected utilities for each θ_{sk}, we simulate the probability of choosing plan j in each period using a smoothed accept-reject function with the form:

$$Pr_{skt}(j = j^*) = \frac{\left(\frac{\frac{1}{-U_{skj^*t}}(\cdot)}{\Sigma_J \frac{1}{-U_{skjt}}(\cdot)}\right)^\tau}{\Sigma_{\hat{j}}\left(\frac{\frac{1}{-U_{sk\hat{j}t}}(\cdot)}{\Sigma_J \frac{1}{-U_{skjt}}(\cdot)}\right)^\tau}$$

This smoothed accept-reject methodology follows that outlined in Train (2009) with some slight modifications to account for the expected utility specification. In theory, conditional on θ_{sk}, we would want to pick the j that maximizes U_{kjt} for each family, and then average over S to get final choice probabilities. However, doing this leads to a likelihood function with flat regions, because for small changes in the estimated parameters θ, the discrete choice made does not change. The smoothing function above mimics this process for CARA utility functions: as the smoothing parameter τ becomes large the smoothed Accept-Reject simulator becomes almost identical to the

---

[3]As noted in the main text in Section 3, we assume that the family-plan-time specific error terms ϵ_{kjt} are distributed i.i.d. normal for each j with zero mean and variance $\sigma_{\epsilon_j}(Y_k)$. Since Y_k is binary we make no additional assumptions on how these variances relate to Y_k. We normalize the value of ϵ_{250}, the preference shock for PPO_{250}, to zero for each realization of Y_k, and estimate the preference shock variances for the other two plans relative to PPO_{250}. Section 3 also discusses the specification and paramaterization of the random coefficients δ and γ.

true Accept-Reject simulator just described, where the actual utility-maximizing option is chosen with probability one. By choosing τ to be large, an individual will always choose j^* when $\frac{1}{-U_{kj^**t}} > \frac{1}{-U_{kjt}} \forall j \neq j^*$. The smoothing function is modified from the logit smoothing function in Train (2009) for two reasons (i) CARA utilities are negative, so the choice should correspond to the utility with the lowest absolute value and (ii) the logit form requires exponentiating the expected utility, which in our case is already the sum of exponential functions (from CARA). This double exponentiating leads to computational issues that our specification overcomes, without any true content change since both models approach the true Accept-Reject function.

Denote any sequence of three choices made as j^3 and the set of such sequences as J^3. In the limit as τ grows large the probability of a given j^3 will either approach 1 or 0 for a given simulated draw s and family k. This is because for a given draw the sequence (j_1, j_2, j_3) will either be the sequential utility maximizing sequence or not. This implicitly includes the appropriate level of inertia by conditioning on previous choices within the sequential utility calculation. For example, under θ_{sk} a choice in period two will be made by a family k only if it is optimal conditional on θ_{sk}, other preference factors, and the inertia implied by the period one choice. For all S simulation draws we compute the optimal sequence of choices for k with the smoothed Accept-Reject simulator, denoted j_{sk}^3. For any set of parameter values θ_{Sk} the probability that the model predicts j^3 will be chosen by k is:

$$\hat{P_k^{j^3}}(\theta, F_{kjt}, X_k^A, X_k^B, H_k, Y_k) = \Sigma_{s \in S} \mathbf{1}[j^3 = j_{sk}^3]$$

Let $\hat{P_k^{j^3}}(\theta)$ be shorthand notation for $\hat{P_k^{j^3}}(\theta, F_{kjt}, X_k^A, X_k^B, H_k, Y_k)$. Conditional on these probabilities for each k, the simulated log-likelihood value for parameters θ is:

$$SLL(\theta) = \Sigma_{k \in K} \Sigma_{j^3 \in J^3} d_{kj^3} \ ln\hat{P_k^{j^3}}$$

d_{kj^3} is an indicator function equal to one if the actual sequence of decisions made by family k was j^3. The maximum simulated likelihood estimator (MSLE) is the value of θ in the parameter space Θ that maximizes $SLL(\theta)$. In the results presented in the text, we choose $Q = 100$, $S = 50$, and $\tau = 4$, all values large enough such that the estimated parameters vary little in response to changes.

Online Appendix C: Moral Hazard Robustness analysis

In the text we discuss a robustness specification that investigates the cost model assumption of no moral hazard. To do this we necessarily make some substantial simplifying assumptions: for a full structural treatment of moral hazard in health insurance utilization see, e.g., Cardon and Hendel (2001), Einav et al. (2013) or Kowalski (2012). We implement the moral hazard robustness check by adjusting the output of the cost model to reflect lower total utilization in the less comprehensive plans (and vice-versa). The intent is to show that even when including price elasticities that are quite large relative to those found in the literature, the model output for inertia and risk preferences does not change substantially. While the specification addresses moral hazard, it also sheds light on whether our estimates are sensitive to consumers having a reasonable level of private information above and beyond the detailed medical data we observe.

We operationalize this test in the following steps. First, we find the implied total spending changes across plans in the population for a price elasticity of -1.3, well higher than the range of -0.1 to -0.4 that represents most of the literature (for further discussion, see Chandra et al. (2010)). To do this we perform a back of the envelope calculation for the arc-elasticity of demand with

respect to price, following the prior work of Einav et al. (2013) and Newhouse (1993). For this calculation, we use the average share of out-of-pocket spending for each plan as the price, and total medical expenditures as the quantity. For the three plans we study, the empirical shares of out-of-pocket spending are 15.5%, 20.9%, and 23.4% going from most to least comprehensive. We use these prices together with the average total spending of \$13,331 in PPO_{-1} at t_{-1} as the basis for this calculation. The formula for the arc-elasticity is:

$$Elasticity = \frac{(q_2 - q_1)/(q_2 + q_1)}{(p_2 - p_1)/(p_2 + p_1)}$$

We use the conservative elasticity of -1.3 and solve for the corresponding total cost changes this price response implies (here, p_j are the empirical shares of out-of-pocket spending for each plan and q_j is implied total spending in plan j). Solving for q_2 as a percentage of q_1 implies an, approximately, 25% reduction in total spending moving from PPO_{250} to PPO_{500}, a 33% reduction in total spending moving from PPO_{250} to $PP0_{1200}$, and a 10% reduction in total spending moving from PPO_{500} to PPO_{1200}. We then apply these reductions (or increases when moving into more comprehensive plans) and adjust the output of the cost model F_{kjt} according to the potential plan being chosen and the previous plan the family incurred costs in. These new 'moral hazard adjusted' F_{kjt} then are input into the choice model, which is otherwise specified and estimated as in the text. The results from the analysis are presented in column 3 of table 5 in the main text and suggest that the initial assumption of no moral hazard does not markedly change our estimates of inertia or risk preferences.

It is important to point out that this exercise does not explicitly model the value of additional health spending, which the literature does through a non-linear budget constraint model where the family trades off the value of extra spending with the price of medical care (see Cardon and Hendel (2001), Einav et al. (2013) or Kowalski (2012)). Moral hazard here is captured purely by differences in spending. For a family that chooses PPO_{500} or PPO_{1200} in the data and is considering switching to PPO_{250}, our moral hazard cost wedge serves as an upper bound for the expected utility difference between these two plans and the more comprehensive plan, since their actual choice implies they value the increase in medical services less than the corresponding overall utility gain from a different financial lottery. The reverse is not true: for consumers considering switching to a less comprehensive plan our out-of-pocket cost wedge is not an upper bound value of utility differences between two plans. However, we view this as a conservative approach in this case, since the high elasticity we've chosen together with the differences in the marginal prices of care between the plans (and the resulting implications for value of medical care foregone) make it unlikely that the value wedge between prospective plans is larger in reality than in this specification.

Online Appendix D: Sources of Inertia

In the empirical framework inertia enters the expected utility function as a fixed monetary cost paid for switching to an alternative plan option relative to the incumbent plan. This is similar in spirit to a tangible search or switching cost, both potential sources of inertia. However, it is likely that the substantial inertia documented in this work arises from a variety of sources tied to different underlying microeconomic fundamentals. This appendix discusses some prominent alternative explanations in detail. We separate sources of inertia linked primarily to explicit, tangible, costs and those linked to factors that implicitly lead to losing money even though they don't lead to direct costs. While we believe that identifying between different sources of inertia is crucial for determining which specific policies will be effective in reducing inertia, we do not believe that

empirically differentiating between these sources would change the nature of our main results or primary lessons. For further discussion of this, see sections 3 and 5 in the text.

Transactions Costs: Explicit Cost. Transactions costs are the most commonly cited potential source of switching costs in the theoretical and empirical literatures (see e.g. Farrell and Klemperer (2007) for the former and Shum (2004) or Dube et al. (2010) for the latter). Tangible switching costs, such as those from transaction costs, contribute directly to inertia by favoring incumbent plan options versus alternative options. In our setting, transaction costs could result from, e.g., the time and hassle costs of (i) researching alternative health plan options or (ii) actually switching plans. It is clear that transaction costs have the potential to be quite large in our setting. Health plans are complicated objects to choose between, often described in long documents with substantial legalese, while the process of actually making a choice can require multiple non-trivial and costly actions. Our empirical choice framework represents inertia as a transaction cost, following prior empirical work that quantifies switching costs. See sections 3 an 5 in the text for further discussion of this modeling choice.

Learning Costs: Explicit Cost. Learning costs are another potential direct cost from switching plans (see e.g. Klemperer (1995)). Learning costs occur when, after purchasing a product, a consumer has to put in time and effort to learn about how to use its various features. With learning costs, once you have purchased a product, you generally don't incur them again when moving back to that product for a second time. Here, learning costs are likely only a small component of the inertia we estimate since the three PPO plan options have the exact same network of providers and the only real product-specific learning that could occur relates to the Health Savings Account (HSA) option in PPO_{1200}.

Product Compatibility: Explicit Cost. Learning costs and product compatibility, another potential direct cost of switching plans contributing to inertia, would both be important issues in cases where consumers had to switch medical providers and disrupt continuity of care when switching health plan. Compatibility costs occur when product-specific investments are made that increase the value of a product relative to an otherwise similar product (see e.g. Klemperer (1995)). In our setting, since the PPO plan options we study have the same network of providers, there are no compatibility issues because relationship-specific investments can be transferred seamlessly across these choices. This implies that the inertia we estimate is a lower bound relative to inertia in a broader plan setting where a consumer has to switch providers to switch health plan. We believe that studying costs and inertia from provider lock-in are important avenues for future work.

Search and Re-Optimization: Explicit and Implicit Cost. This model, and much prior work on inertia and underlying switching costs, presumes that consumers have perfect knowledge of product characteristics when it is time to make a choice. In the transaction costs section above, we noted that research costs are a potentially important source of inertia, implying that there may be imperfect knowledge of plan options. In this case, a more sophisticated model could treat the decision process as a two-stage process where (i) consumers pay some tangible search cost to gain information about alternative options and (ii) potentially change plans (and pay any relevant transaction costs) conditional on searching. Search costs are different than transaction costs because ex ante beliefs about product characteristics are important to the decision to re-optimize: a consumer could have a small search / re-optimization cost but believe that they have little to gain from acquiring information. In our setting, where prices changed substantially from t_0 to t_1, consumer beliefs may have been centered around status-quo prices, implying that a small search cost could

dissuade them from re-optimizing despite the large amount of money they had to gain (leading to a large impact / estimate of inertia). In this case, our inertia estimates can be interpreted as money that consumers leave on the table as a result of this decision process. Note that since there is low switching for t_2 as well, under this model consumers in our environment also do not learn about the potential to gain a lot of money from switching over time, implying incorrect beliefs can be quite persistent. Search and re-optimization costs in this two stage framework lead to both the direct costs of acquiring information, and the implicit cost of money left on the table if they don't re-optimize but their ex ante beliefs about market changes are biased. See Cebul et al. (2011) for further empirical analysis of search costs in health insurance markets.

Psychological Costs, Procrastination, and Inattention: Implicit Cost. There is a substantial empirical literature studying consumer inertia in the presence of default options and psychological costs of switching, broadly defined (see e.g. Madrian and Shea (2001), Samuelson and Zeckhauser (1988), or Thaler and Sunstein (2008)). This explanation is also cited in the brand loyalty literature in marketing (see e.g. Dube et al. (2009)). There are a multitude of reasons discussed in this prior work for why inertia might be present above and beyond explicitly rational considerations such as search, transaction, learning, and product compatibility costs. For example, procrastination, combined with inattention, could lead to enrolling in the default option in a health insurance setting, even in the presence of an open enrollment deadline. While it is beyond the scope of this work to list all relevant potential explanations for inertia related to psychological costs and bounded rationality, we recognize that understanding the contributions of such factors to overall inertia is an interesting topic for future work. While we list these factors as implicit costs, since they may not lead to easily identifiable direct costs in the standard economic framework, some portion could reasonably be explicit as well.

Online Appendix E: Supporting Analysis

This appendix contains supporting evidence for the analysis in the main text, including additional description of the data and environment as well as additional investigation of the empirical model and counterfactual inertia reduction analysis. The appendix is organized into short sections corresponding to the implications of each piece of supporting evidence and references related material in the main text where relevant.

Health Plan Characteristics. Table A-2, also referenced in section 1 of the main text, presents the detailed characteristics of the PPO plans offered at the firm over time. After the deductible is paid, PPO_{250} has a coinsurance rate of 10% while the other two plans have rates of 20%, implying they have double the marginal price of post-deductible claims. Out-of-pocket maximums indicate the maximum amount of medical expenditures that an enrollee can pay post-premium in a given plan. These are larger the less comprehensive the plan is and vary with income tier. For example, for PPO_{250} the out-of-pocket maximum for the lowest income consumers is $3,000 per year, while for those in the highest income tier it is $8,000 per year. For PPO_{1200} the lowest income consumers have an $8,000 out-of-pocket maximum while the highest income consumers have a $10,000 maximum. Both PPO_{250} and PPO_{500} have the same flat-fee copayment structures for pharmaceuticals, emergency room visits, and physician office visits, while in PPO_{1200} these apply to the deductible and coinsurance. The coinsurance for mental health expenditures post-deductible is 50% in all three plans. Note that pharmacy, ER, and physician office visit copayments do not apply to the out-of-pocket maximums in PPO_{250} and PPO_{500}. Consumers in PPO_{1200} receive free preventive

care (e.g. regular check-ups), are ineligible for a standard health care FSA but are eligible for the linked Health Savings Account (HSA). The characteristics described in the table are for in-network purchases: the plans also have out-of-network payment policies, which we do not present or model. For all three plans, the out-of-network coverage policies are relatively generous, with deductibles, coinsurance rates, and out-of-pocket maximums that are approximately double the corresponding in network values for each plan. Since all three plans have the same exact networks of providers, and cover the same services and drugs, the out-of-network characteristics are quite similar across the plans. Note that only 2% of realized total expenditures are out-of-network. Finally, the first column describes PPO_{-1}, the PPO plan offered by the firm before the menu change at t_0. This plan is most similar to PPO_{250} offered after the menu change, though there are some clear differences in financial characteristics (in addition to the largely different marketing that accompanied the plan change).

Our cost model, described in detail in Online Appendix A, incorporates a detailed representation of each plan's out-of-pocket mapping in order to accurately capture health risk. We divide claims into medical categories that are treated differently in plan financial characteristics, as described in table A-2, and make some necessary simplifying assumptions. See the cost model appendix for a discussion and validation of the total cost to out-of-pocket mappings for each plan.

Health Plan Choices. Table A-3, also discussed in section 1 of the main text, presents a detailed breakdown of the pattern of employee choices over time before and after the menu change. In t_{-1}, 39% of employees enroll in PPO_{-1}, 47% enroll in one of the four HMO options, and 14% waive coverage. At t_0, 46% of employees choose one of the three new PPO options, with 25% choosing PPO_{250}. 37% choose either of the two remaining HMO plans while 16% waive coverage. The table also presents clear evidence that the nest of PPO options and nest of HMO options are quite horizontally differentiated from one another by examining consumer health plan transitions over time. An individual who switches plans from a PPO option is much more likely to choose another PPO option than to choose an HMO option. Of the 2,757 employees enrolled in PPO_{-1} in year t_{-1} who also enroll in any plan at t_0, only 85 (3%) choose an HMO option at t_0. In reverse, despite the expansion of PPO options and reduction of HMO options, only 15% of employees who chose an HMO option in t_{-1}, and choose any plan at t_0, switch to a PPO option. This suggests that restricting the set of choices to PPO options should not lead to biased parameters within that population. See the table for further details on plan enrollment and plan transitions over time.

Basic Specification for Choice Model. Table A-4 presents the estimated parameter results for a basic version of the primary choice model presented in section 3 of the main text. The basic model estimates inertia, risk preferences, and persistent tastes for PPO_{1200} but includes only very limited observable heterogeneity related to demographics and linked decisions for both inertia and risk preferences. The estimates in this basic specification are similar in magnitude to those in the primary specification, and give a sense of what a more coarse but simpler specification delivers. The table can be thought of as an extension to Table 5 in the main text, which presents our primary choice model estimates along with four robustness checks.

Interpretation of Risk Preference Estimates. As the CARA risk preference coefficients presented in table 5 in the main text are difficult to interpret, we follow Cohen and Einav (2007) and analyze these estimates in a more intuitive manner in table A-5. The table presents the value X that would make an individual of average age and income with our estimated risk preferences

indifferent between (i) inaction and (ii) accepting a gamble with a 50% chance of gaining $100 and a 50% chance of losing $X. Thus, a risk neutral individual will have $X = \$100$ while an infinitely risk averse individual will have X close to zero. The top section of the table presents the results for the primary specification: X is $94.6 for the median individual, implying moderate risk aversion relative to other results in the literature, shown below in the table. X is $92.2 for the 95th percentile of γ and $91.8 for the 99th, so preferences don't exhibit large heterogeneity in the context of the literature. The bottom section of the table presents the results for the robustness check that studies log-normal risk preference heterogeneity (presented in column 4 of Table 5) instead of normal heterogeneity. X is $96.3 for the median individual, X is $91.1 for the 75th percentile, $72.8 for the 95th percentile, and $50.9 for the 99th percentile. Thus, the log-normal specification leads to similar median and mean risk aversion, but has a fatter right tail of individuals with substantial risk aversion.

Reduced Inertia with No Plan Re-Pricing: Additional Analysis. This section presents additional details for the 'naive' counterfactual analysis where inertia is reduced from the baseline estimates but health plan premiums are held fixed as observed in the data. Consumers may switch to a new health plan as a result of lower inertia, but this selection does not feed back into prices on the basis of new enrollee cost profiles. In this context, the policy intervention can only increase welfare since prices are by definition unchanged and the policy helps consumers make weakly better decisions relative to the benchmark case. See the main text for discussion in the context of related work.

Figure A-3 presents market share and average cost results for different levels of reduced inertia (corresponding to different levels of Z). Results are presented for years t_1 and t_2 for several cases ranging from no inertia to full inertia. The top panel reveals that, as Z decreases towards 0, consumers adjust more readily to the observed price changes and are more likely to enroll in PPO_{500} (recall relative prices for PPO_{250} went up in t_1). For the case where $Z = \frac{1}{4}$ (75% of inertia removed) 913 employees enroll in PPO_{500} at t_1, a 44% increase over the benchmark model with full inertia where 639 consumers choose that plan. For the cases of $Z = \frac{1}{2}$ and $Z = 0$, t_1 enrollments in PPO_{500} are 780 (21% increase) and 1,052 (65% increase) respectively. Moving forward to t_2, for $Z = \frac{1}{4}$, there are 1,010 enrollees in PPO_{500} relative to 702 in the benchmark case (a 44% increase). Almost all of the switchers towards PPO_{500} would have continued enrollment in PPO_{250} in the benchmark case. The figure also reveals that switching to and from PPO_{1200} as the result of reduced inertia is limited, due to the horizontal differentiation δ resulting from the health savings account and linked features. The lower panel reveals that these different enrollment patterns over time imply different plan average costs. For the family coverage tier average costs for PPO_{250} increase relative to those in PPO_{500} as enrollment in the former declines relative to the latter, implying that the people who switch out of PPO_{250} into PPO_{500} are healthier than those who do not switch. This result is similar for other coverage tiers and suggests that, in the primary analysis with endogenous plan re-pricing, PPO_{250} premiums will become more expensive, potentially leading to even more selection against that plan in subsequent periods.

Table A-6 presents the welfare impact of moving from the benchmark environment with full inertia to the case where $Z = \frac{1}{4}$. At t_2, the mean per employee certainty equivalent increase is $114. For those who switch plans in the counterfactual environment relative to the benchmark case, the mean benefit is $196 (for those who do not switch, the change is zero by definition). The policy intervention improves welfare by 5.8% of total employee premium contributions, and by 2.5% relative to the total employee spending benchmark (the table also reports the % changes for the certainty equivalent at stake benchmark). The analogous percentage changes for those who switch plans are

10.0% and 4.4% relative to these respective metrics. These numbers are similar to, but slightly larger than, the impact of the reduced inertia during t_1. The positive welfare impact of improved individual-level choices in the environment without plan re-pricing is similar to prior results in the empirical literature on choice inadequacy, but stands in stark contrast to the generally negative welfare results in our analysis with endogenous plan re-pricing, in section 5 in the main text.

Reduced Inertia with Plan Re-Pricing: Additional Analysis. This section presents additional details for the main counterfactual analysis where inertia is reduced from our baseline estimates and health plan premiums re-adjust as enrollment patterns change. Figure A-4 presents market shares and average costs for $PP0_{500}$ and PPO_{250} in years t_1, t_2, and t_4 as a function how effectively inertia is reduced (Z). We report results for values of Z equal to 0,.25,.5, and 1. It is clear from the top panel that as Z decreases from 1, and inertia is further reduced, enrollment in PPO_{250} declines at the expense of enrollment in PPO_{500}. The bottom panel reveals that as Z decreases and enrollment increases, the relative average costs (and total premiums) of PPO_{250} increase, implying that as inertia is reduced we find higher incremental adverse selection.

Table A-7 studies the profile of maximum direct inertial costs incurred for different Z over time. The 'maximum' here corresponds to the case where the inertia we estimate comes purely from tangible / direct costs, such as transaction, learning, or search costs. Costs incurred reflect both (i) costs per switcher and (ii) the number of switchers: the former decreases as inertia is reduced (by definition) while the latter increases as inertia is reduced. The mean per employee per year maximum inertia cost incurred is \$185 for the baseline case and \$188 for the case where $Z = 0.75$ (note that this calculation covers the entire population, not switchers only). This means that the product of the number of switchers and maximum inertia costs incurred is slightly larger as inertia is reduced from the baseline, implying that the increased number of switchers makes up for the decrease in inertia costs per switcher. For $Z = .5$ and $Z = .25$ the mean per employee per year maximum inertia costs incurred are \$142 and \$83 respectively, while $Z = 0$ by definition implies no costs. These patterns are similar for each individual year, though the overall level of maximum inertia costs incurred is higher in t_1 and t_2 than in later years as the market moves towards a steady state. This analysis supports table 8 in the main text, which investigates the welfare treatment of inertia itself, studying the cases ranging from (i) inertia is purely a direct / tangible cost to the case where (ii) inertia leads to monetary losses but is not directly costly to overcome.

References

Cardon, James and Igal Hendel, "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," *The RAND Journal of Economics*, 2001, *32* (3), 408–427.

Carlin, Caroline and Robert Town, "Adverse Selection, Welfare, and Optimal Pricing of Employer Sponsored Health Plans," 2009. University of Minnesota Working Paper.

Cebul, Randall, James Rebitzer, Lowell Taylor, and Mark Votruba, "Unhealthy Insurance Markets: Search Frictions and the Cost and Quality of Health Insurance," *The American Economic Review*, 2011, *101* (5), 1842–1871.

Chandra, Amitabh, Jonathan Gruber, and Robin McKnight, "Patient Cost-Sharing and Hospitalization Offsets in the Elderly," *American Economic Review*, 2010, *100* (1), 193–213.

Cohen, Alma and Liran Einav, "Estimating Risk Preferences from Deductible Choice," *The American Economic Review*, 2007, *97* (3), 745–788.

Dube, Jean-Pierre, Gunter Hitsch, and Peter Rossi, "Do Switching Costs Make Markets More Competitive," *Journal of Marketing Research*, 2009, *46*, 435–445.

_ , _ , **and** _ , "State Dependence and Alternative Explanations for Consumer Inertia," *The RAND Journal of Economics*, 2010, *41* (3), 417–445.

Einav, Liran, Amy Finkelstein, Stephen Ryan, Paul Schrimpf, and Mark Cullen, "Selection on Moral Hazard in Health Insurance," *American Economic Review*, 2013, *103* (1), 178–219.

Farrell, Joseph and Paul Klemperer, "Coordination and Lock-In: Competition with Switching Costs and Network Effects," *Handbook of Industrial Organization Chapter 31 Editors M. Armstrong and R. Porter*, 2007, pp. 1967–2072.

Gertner, Robert, "Game Shows and Economic Behavior: Risk-Taking on 'Card Sharks'," *Quarterly Journal of Economics*, 1993, *108* (2).

Holt, Charles and Susan Laury, "Risk Aversion and Incentive Effects," *American Economic Review*, 2002, *92* (5).

Klemperer, Paul, "Competition When Consumers Have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade," *The Review of Economic Studies*, 1995, *62* (4), 515–539.

Kowalski, Amanda, "Estimating the Tradeoff Between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance," May 2012. NBER working paper 18108.

Madrian, Brigitte and Dennis Shea, "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 2001, *141* (4), 1149–1187.

Newhouse, Joseph, *Lessons from the RAND Health Insurance Experiment*, Harvard University Press, 1993.

Samuelson, William and Richard Zeckhauser, "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, 1988, *1*, 7–59.

Shum, Matthew, "Does Advertising Overcome Brand Loyalty? Evidence From Breakfast Cereals," *Journal of Economics and Management Strategy*, 2004, *13*, 241–272.

Sydnor, Justin, "Over(insuring) Modest Risks," *American Economic Journal: Applied Economics*, 2010, *2* (4), 177–199.

Thaler, Richard and Cass Sunstein, *Nudge*, Yale University Press, 2008.

Train, Kenneth, *Discrete Choice Methods with Simulation, 2nd ed.*, Cambridge University Press, 2009.
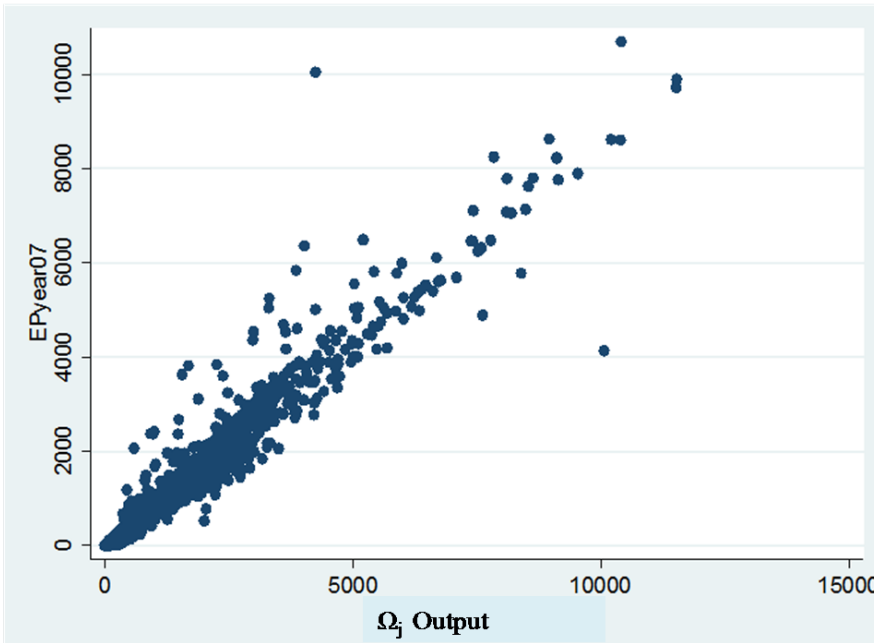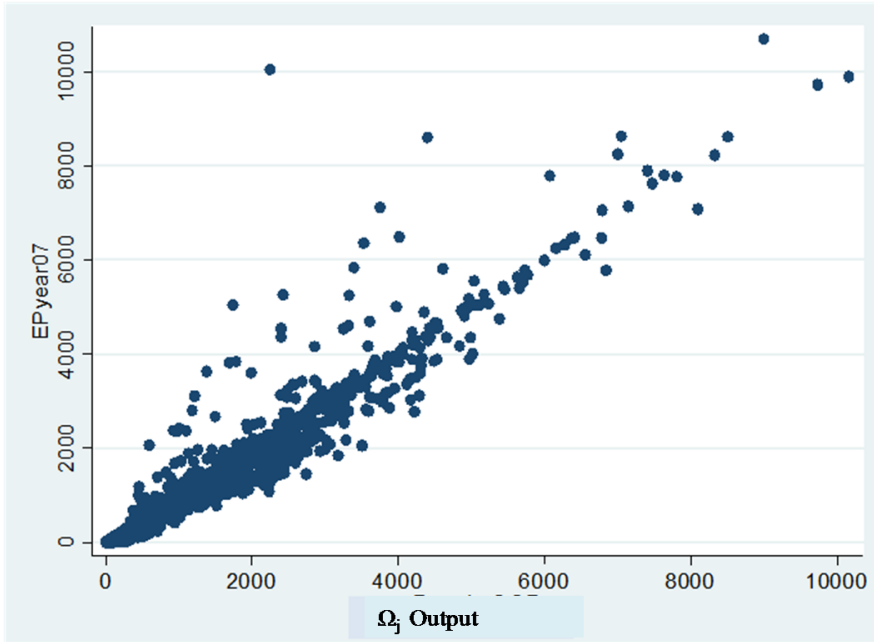
Figure A-1: This figure validates the mapping Ω_j that translates the vector of total claims C into plan specific out of pocket expenditures. The two charts show Ω for PPO_{250}, with predicted out-of-pocket spending on the x-axis and actual out-of-pocket spending on the y-axis. The top chart is the mapping actually used where claims are categorized into four categories (i) hospital and outpatient (ii) pharmacy (iii) mental health and (iv) physician office visit. Ideally, we want all points to be on the 45 degree line where the actual employee paid out of pocket equals the model predicted out of pocket. The plot is condensed around the 45 degree line so we believe this our mapping is close to the true mapping. The bottom figure adds out of network expenses to the mapping as a fifth category and does not materially improve upon the mapping used.

Cost Model Estimation Structure
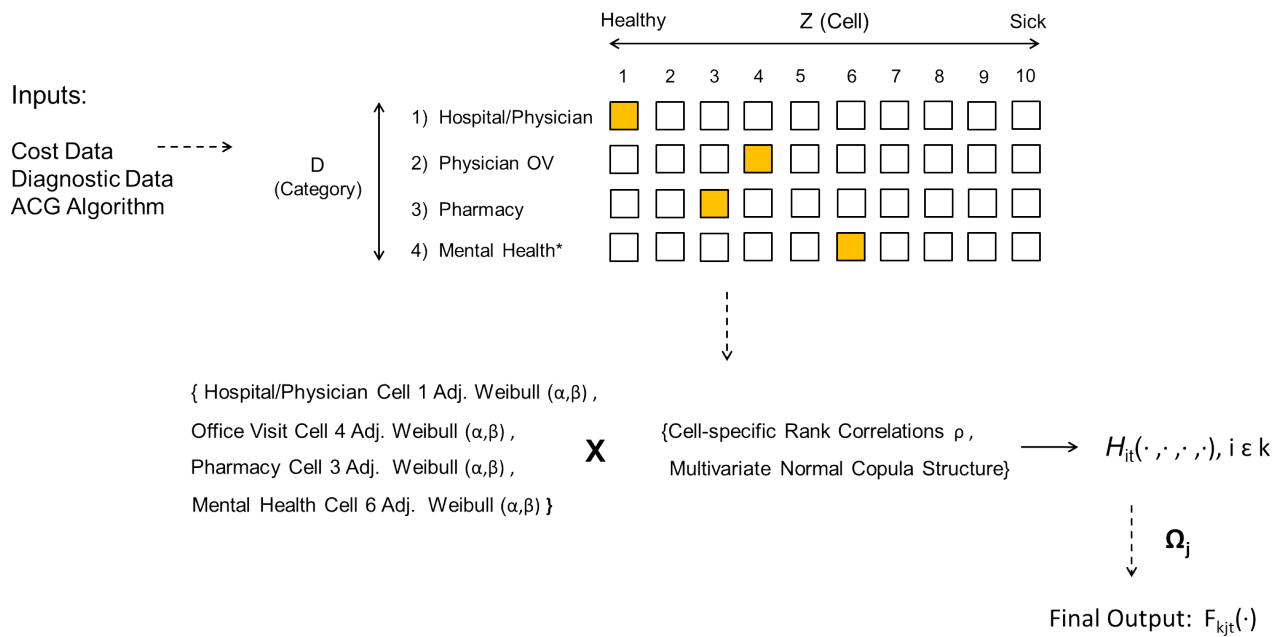
Figure A-2: This figure outlines the primary steps of the cost model described in Online Appendix A. It moves from the initial inputs of cost data, diagnostic data, and the ACG algorithm to the final output F_{kjt} which is the family, plan, time specific distribution of out-of-pocket expenditures that enters the choice model for each family. The figure depicts an example individual in the top segment, corresponding to one cell in each category of medical expenditures. The last part of the model maps the expenditures for all individuals in one family into the final distribution F_{kjt}.

Partial Equilibrium Information Provision
Market Share

Partial Equilibrium Information Provision
Plan Family Average Cost
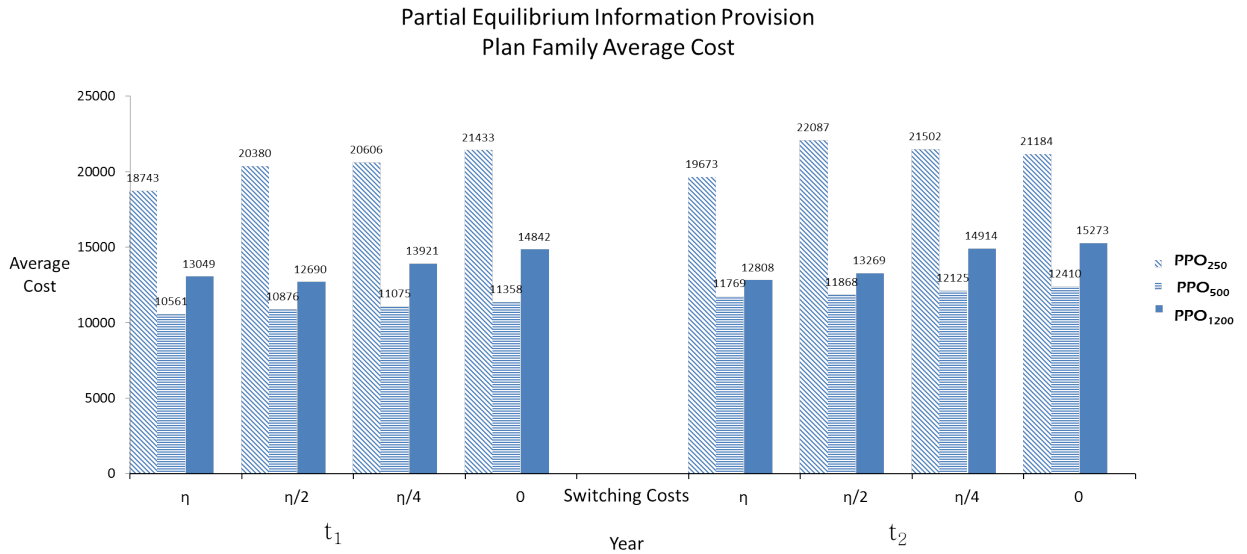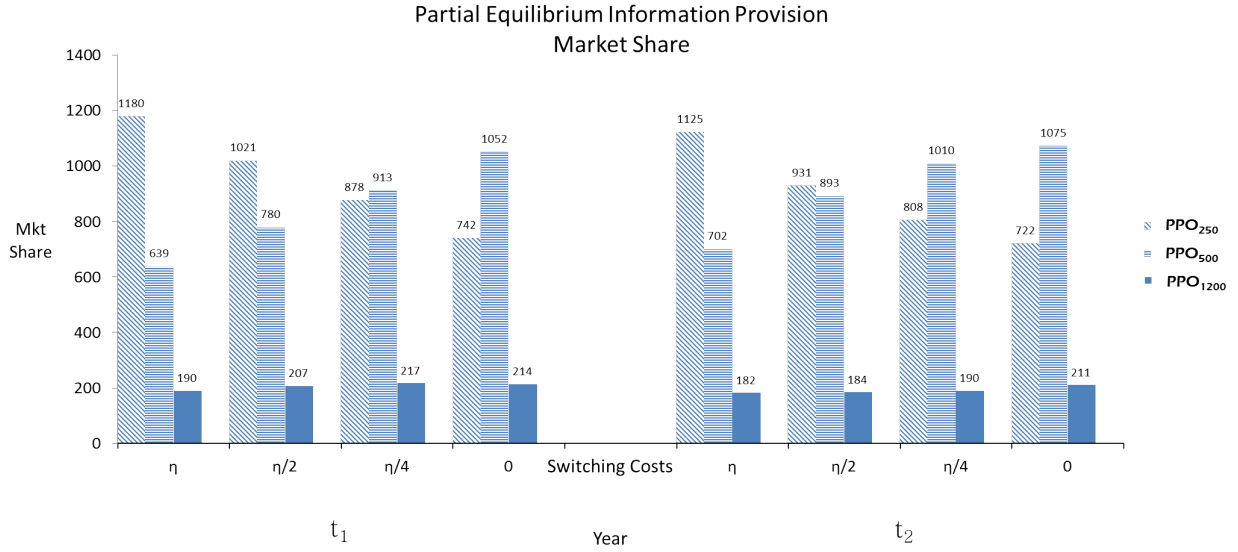
Figure A-3: This figure describes the new plan enrollment levels and average costs under the 'naive' inertia reduction counterfactual where prices are held fixed as observed in the data. The market shares are presented for all coverage tiers combined, while the average costs are presented for the family coverage tier (employee + spouse + dependent). The figure studies these quantities as a function of the effectiveness of the policy to reduce inertia, where η signifies no reduction in inertia from our estimates and 0 signifies no inertia. In general, the market share of PPO_{250} is decreasing in both t_1 and t_2 as the policy to reduce inertia becomes more effective. The relative average cost of this plan increases as the market share declines, implying that when we allow for endogenous plan re-pricing there will be increased adverse selection on the margin.

Full Equilibrium Information Provision
Market Share as Function of Switching Costs

Mkt Share

1600 — 1400 — 1200 — 1000 — 800 — 600 — 400 — 200 — 0

t_1: η 1165, 672; $\eta/2$ 988, 866; $\eta/4$ 932, 928; 0 746, 1113

t_2: η 1054, 800; $\eta/2$ 815, 1071; $\eta/4$ 699, 1185; 0 612, 1278

t_4: η 857, 998; $\eta/2$ 445, 1446; $\eta/4$ 351, 1519; 0 482, 1389

Switch Cost — Year — Switch Cost — Year

PPO_{250}
PPO_{500}

Full Equilibrium Information Provision
Plan Family AC as Function of Switching Costs

Average Cost

30000 — 25000 — 20000 — 15000 — 10000 — 5000 — 0

t_1: η 17849, 10520; $\eta/2$ 18443, 10575; $\eta/4$ 19093, 10671; 0 19998, 11150

t_2: η 18508, 11467; $\eta/2$ 19193, 11459; $\eta/4$ 21746, 11704; 0 23045, 12695

t_4: η 19245, 11411; $\eta/2$ 21953, 11908; $\eta/4$ 25840, 13387; 0 21736, 12903

Switch Cost — Year — Switch Cost — Year

PPO_{250}
PPO_{500}

Figure A-4: This figure shows market shares and average costs for $PP0_{500}$ and PPO_{250} in years t_1, t_2, and t_4, as a function of Z, in the environment with endogenous plan pricing. We report results for values of Z equal to 0,.25,.5, and 1. It is clear from the top panel that as Z decreases and inertia is further reduced, enrollment in PPO_{250} declines at the expense of enrollment in PPO_{500}. The bottom panel reveals that as Z decreases and enrollment increases, the average costs (and total premiums) of PPO_{250} increase relative to those of PPO_{500} in each year. This is indicative of increased relative premiums for comprehensive insurance and increased adverse selection.

Final Sample
Cost Model Output

	Overall	PPO$_{250}$	PPO$_{500}$	PPO$_{1200}$
Individual Mean (Median)				
Unscaled ACG Predictor				
Mean		1.42	0.74	0.72
Median		0.83	0.37	0.37
Pharmacy: Model Output				
Zero Claim Pr.	0.35 (0.37)	0.31 (0.18)	0.40 (0.37)	0.42 (0.37)
Weibull α	1182 (307)	1490 (462)	718 (307)	596 (307)
Weibull β	0.77 (0.77)	0.77 (0.77)	0.77 (0.77)	0.77 (0.77)
Mental Health				
Zero Claim Pr.	0.88 (0.96)	0.87 (0.96)	0.90 (0.96)	0.90 (0.96)
Weibull α	1422 (1295)	1447 (1295)	1374 (1295)	1398 (1295)
Weibull β	0.98 (0.97)	0.99 (0.97)	0.98 (0.97)	0.98 (0.97)
Hospital / Physician				
Zero Claim Pr.	0.23 (0.23)	0.21 (0.23)	0.26 (0.23)	0.26 (0.23)
Weibull α	2214 (1599)	2523 (1599)	1717 (1599)	1652 (1599)
Weibull β	0.58 (0.55)	0.59 (0.55)	0.55 (0.55)	0.55 (0.55)
($> \$40,000$) Claim Pr.	0.02 (0.01)	0.02 (0.01)	0.01 (0.01)	0.01 (0.01)
Physician OV				
Zero Claim Pr.	0.29 (0.20)	0.26 (0.20)	0.33 (0.46)	0.34 (0.46)
Weibull α	605 (553)	653 (553)	517 (410)	529 (410)
Weibull β	1.15 (1.14)	1.15 (1.14)	1.15 (1.14)	1.14 (1.14)
Correlations				
Rank Correlation Hospital-Pharm.	0.28 (0.34)	0.26 (0.32)	0.31 (0.34)	0.32 (0.34)
Rank Correlation Hospital-OV	0.73 (0.74)	0.72 (0.74)	0.74 (0.74)	0.74 (0.74)
Rank Correlation Pharm.-OV	0.35 (0.41)	0.33 (0.37)	0.38 (0.41)	0.39 (0.41)

Table A-1: This table describes the output of the cost model in terms of the means and medians of individual level parameters, classified by the plan actually chosen. These parameters are aggregated for these high-level groups but have more micro-level groupings, which are the primary inputs into our cost projections in the choice model. Weibull α, Weibull β, and Zero Claim Probability correspond to the cell-specific predicted total individual-level health expenses as described in more detail in Online Appendix A.

PPO Plan Characteristics

	PPO_{-1}	PPO_{250}	PPO_{500}	PPO_{1200}
Basic Characteristics				
Individual Deductible (Family)	250*	250	500	1200
	(750)	(750)	(1500)	(2400)
Coinsurance Rate	10%	10%	20%	20%
Physician Office Visit Copay	20	25	25	NA
Emergency Room Copay	100	100	100	NA
Mental Health Coinsurance	50%	50%	50%	50%
Prescription Drug Copay	5/25/45**	5/25/45**	5/25/45*	NA
	(10/45/65)	(10/50/75)	(10/50/75)	NA
Individual OOP Max (Family)***				
Income Tier 1 (< \$41K)	2000	1000	1500	2000
	(6000)	(3000)	(4500)	(6000)
Income Tier 2-3 (\$41K-\$124K)	2000	2000	3000	4000
	(6000)	(5000)	(7000)	(8000)
Income tier 4-5 (> \$124K)	2000	3000	4000	5000
	(6000)	(8000)	(9000)	(10000)
Other Features				
Free Preventative Care	No	No	No	Yes
Traditional FSA Eligible	Yes	Yes	Yes	No
HSA Eligible	No	No	No	Yes

* PPO_{-1} has an inpatient deductible of 150 per admission

** Prescription copay max of 1500 per person.
  Copays for 30-day supply. 90-day supply copay in parentheses.
  Copays increase in t_2 by approx. 20%

*** Office visit and pharmacy claims only apply to OOP max for PPO_{1200}

Table A-2: This table describes the financial characteristics for each PPO option available at the firm over time, both before and after the menu change at t_0. For most medical expenses, an individual pays the full amount until he reaches the yearly plan deductible, after which he pays the coinsurance rate for all further medical expenses. Once an individual spends up to the out-of-pocket maximum, he pays no further general medical expenses. Pharmacy products and physician office visits only apply to the deductible and coinsurance for PPO_{1200}; all other plans have fixed co-payments for these services. Mental health services apply to all plan deductibles but not the out-of-pocket maximum for each plan. Out-of-pocket maximums vary with income, presumably for equity considerations. This chart does not include out-of-network plan characteristics which account for only 2% of total expenses and don't vary substantially across the three plans. These three plans have the same provider network and cover the same medical services.

Choice Behavior

	t_{-1}	t_0	t_1
PPO_{250}	-	2,199 (25%)	1,937 (21%)
PPO_{500}	-	998 (11%)	1,544 (18%)
PPO_{1200}	-	876 (10%)	824 (9%)
HMO_1	2,094 (25%)	2,050 (23%)	2,031 (22%)
HMO_2	701 (8%)	1,273 (14%)	1,181 (13%)
PPO_{-1}	3,264 (39%)	-	-
HMO_3	668 (8%)	-	-
HMO_4	493 (6%)	-	-
Waive	1,207 (14%)	1,447 (16%)	1,521 (17%)

Switching Behavior

t_0 plan / t_{-1} plan	PPO_{-1}	All HMO
PPO_{250}	1,710	194
PPO_{500}	570	118
PPO_{1200}	392	147
HMO_1	49	1,703
HMO_2	36	943

t_1 plan / t_0 plan	PPO_{250}	PPO_{500}	PPO_{1200}	HMO_1	HMO_2
PPO_{250}	1,732	14	14	21	8
PPO_{500}	129	774	112	31	31
PPO_{1200}	17	11	577	12	15
HMO_1	10	7	8	1,694	22
HMO_2	9	6	5	6	983

Table A-3: This table describes plan choice at the firm between years t_{-1} and t_1. The top panel documents the number of employees who chose each plan in each of these three years. Before the menu change about 39% of employees enrolled in a PPO and 47% in an HMO. After the menu change these numbers are essentially reversed with 47% in a PPO and 35% in an HMO. The middle panel studies the choice behavior of all employees at the firm who were enrolled in any plan in both of the years t_{-1} and t_0. Each column corresponds to the plan an employee was in at t_{-1} while each row corresponds to the plan an employee was in at t_0. It is clear that, when the menu of plans changed for t_0, most employees in PPO_{-1} moved to one of the new PPO options while most employees enrolled in an HMO at t_{-1} still re-enroll in an HMO at t_0. The bottom panel presents the analogous chart for all employees at the firm enrolled in a plan both in years t_0 and t_1. The vast majority of t_0 PPO enrollees who switch plans at year t_1 choose another PPO option at t_1. These panels together reveal significant horizontal differentiation between the nest of PPOs and nest of HMOs.

Empirical Model Results: Basic Model

Parameter	Basic Model
Inertia - Single, η_0	1779
	(72)
Inertia - Family, $\eta_0 + \eta_2$	2354
	(62)
Inertia - Total Pop. Mean, η **[Pop. Standard Deviation]**	2087 [286]
Risk Aversion Mean - Intercept , μ_γ	$3.12 * 10^{-4}$
	$(1.1 * 10^{-5})$
Risk Aversion Mean - Income ,β	$4.21 * 10^{-5}$
	$(3.0 * 10^{-6})$
Risk Aversion Std. Deviation , σ_γ	$1.19 * 10^{-4}$
	$(8.0 * 10^{-6})$
CDHP - Single - RC Mean , δ	-3665
	(218)
CDHP - Single - RC Variance, σ_δ	1283
	(104)
CDHP - Family - RC Mean , δ	-4847
	(204)
CDHP - Family - RC Variance, σ_δ	1733
	(99)
High Total Cost - PPO_{250}, α	937
	(175)
ϵ_{500} - Single	60
	(67)
ϵ_{1200} - Single	872
	(122)
ϵ_{500} - Family	180
	(94)
ϵ_{1200} - Family	888
	(104)

Table A-4: This table presents the estimated parameter results for the basic version of the primary choice model presented in section 3 of the main text. The table can be thought of as an extension to Table 5, which presents our primary choice model estimates along with four robustness checks. The basic model estimates inertia, risk preferences, and persistent tastes for PPO_{1200} but includes only very limited observable heterogeneity for both inertia and risk preferences. The estimates in the basic specification are similar in magnitude to those in the primary specification.

Risk Preference Analysis

	Absolute Risk Aversion	Interpretation
Normal Heterogeneity		
Mean / Median Individual	$4.22 * 10^{-4}$	95.9
25th percentile	$2.95 * 10^{-4}$	97.1
75th percentile	$5.49 * 10^{-4}$	94.8
95th percentile	$7.31 * 10^{-4}$	93.1
99th percentile	$8.59 * 10^{-4}$	92.1
Log normal Heterogeneity		
Mean	$9.82 * 10^{-4}$	91.0
25th percentile	$1.53 * 10^{-4}$	98.1
Median	$3.85 * 10^{-4}$	96.3
75th percentile	$9.72 * 10^{-4}$	91.1
95th percentile	$3.70 * 10^{-3}$	72.8
99th percentile	$9.30 * 10^{-3}$	50.9
Comparable Estimates		
Cohen and Einav (2007) Benchmark Mean	$3.1 * 10^{-3}$	76.5
Cohen and Einav (2007) Benchmark Median	$3.4 * 10^{-5}$	99.7
Gertner (1993)	$3.1 * 10^{-4}$	97.0
Holt and Laury (2002)	$3.2 * 10^{-2}$	21.0
Sydnor (2010)	$2.0 * 10^{-3}$	83.3

Table A-5: This table examines the risk preference estimates from the empirical results presented in table 5 in the main text. The first section of the table is for the normally distributed risk preference estimates in the Primary specification, where the age and income coefficients are evaluated at the median values of those variables. The second section is for the model with log-normally distributed preferences studied in column 4 of table 5. The interpretation column is the value X that would make someone indifferent about accepting a 50-50 gamble where you win $100 and lose X versus a status quo where nothing happens. Our estimates are similar under both specifications with the exception that the log normal model predicts a fatter tail with higher risk aversion. These estimates are in the middle of the (wide) range found in the literature and show moderate risk aversion (relative to this literature) except at the tails in the log-normal model where consumers are quite risk averse.

Reduced Inertia and No Plan Re-Pricing
Welfare Analysis

	t_1	t_2
Mean Δ CEQ		
Population	$96	$114
Switchers Only	$175	$196
Mean Welfare Change: % Total Premiums		
Mean Employee Premium (MEP)	$2,067	$1,954
Welfare Change Population	4.6%	5.8%
Welfare Change Switchers	8.5%	10.0%
Mean Welfare Change: % Total Emp. Spending		
Mean Total Emp. Spending	$4,373	$4,486
Welfare Change Population	2.2%	2.5%
Welfare Change Switchers	4.0%	4.4%
Mean Welfare Change: % $\|CEQ\|$		
Mean Total $\|CEQ\|$	$6,694	$6,773
Welfare Change Population	1.4%	1.7%
Welfare Change Switchers	2.6%	2.9%

Table A-6: This table presents the welfare results of the 'naive' counterfactual inertia reduction analysis. We present the mean per employee per year dollar change in certainty equivalents and corresponding percentage welfare changes resulting from the policy reducing inertia to $.25\eta$ from η. We present three alternative welfare benchmarks to assess the % impact of this these certainty equivalent changes. These metrics divide the change in certainty equivalent from the policy intervention by mean (i) total employee premiums (ii) total employee spending and (iii) the absolute value of the certainty equivalent loss. Note that since all figures are losses the certainty equivalent absolute value is larger than the total spending figure. Since we hold plan prices fixed in this exercise, the welfare changes must be positive when consumers make better individual-level decisions.

Endogenous Plan Re-Pricing
Maximum Direct Inertia Cost Incurred

		η	$.75\eta$	$.5\eta$	$.25\eta$	0
t_1	Inertia Cost/ Switcher	1,982	1,502	1,010	506	0
	Switcher %	14%	19%	21%	23%	32%
	Avg. Inertia Pop.	291	284	216	118	0
t_2	Inertia Cost/ Switcher	1,959	1,488	1001	506	0
	Switcher %	14%	17%	20%	23%	25%
	Avg. Inertia Pop.	269	252	201	117	0
t_4	Inertia Cost/ Switcher	1,973	1,506	982	480	0
	Switcher %	7%	10%	11%	12%	15%
	Avg. Inertia Pop.	144	156	109	56	0
t_6	Inertia Cost / Switcher	1,942	1,451	962	483	0
	Switcher %	6%	8%	10%	12%	16%
	Avg. Inertia Pop.	110	109	90	60	0
Avg. t_1-t_6	Inertia Cost / Switcher	1,963	1,489	988	493	0
	Switcher %	9%	13%	14%	17%	20%
	Avg. Inertia Pop.	185	188	142	83	0

Table A-7: Table A-7 investigates the maximum extent of inertia costs consumers experience from switching plans under the range of settings where $Z = 1$ (full inertia) to $Z = 0$ (no inertia). This 'maximum' here corresponds to the case where the inertia we estimate comes entirely from tangible / direct costs that must be overcome to switch plans. As Z decreases, maximum inertia costs incurred are generally decreasing, though they increase from \$185 per population member to \$188 per population member when moving from $Z = 1$ to $Z = 0.75$. This reflects the fact that as Z decreases, maximum inertial costs per switcher are lower (by definition) but more people actually switch plans due to reduced inertia. The table presents maximum inertia costs incurred broken down across year t_1 to t_6, with the bottom section (also in table 8 in the main text) averaging across all six years after t_0. This analysis relates to the analysis of welfare treatment of inertia, included in the bottom part of table 8 in the main text.