C.I.R.P.É.E

Centre Interuniversitaire sur le Risque,
les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **10-38**

Health and Access Effects of New Drugs: Combining Experimental and Non-Experimental Data

Pierre-Carl Michaud
Darius Lakdawalla
Dana Goldman
Neeraj Sood
Ze Cong

Septembre/September 2010

Michaud: Corresponding author. Université du Québec à Montréal, CIRPÉE and Rand
michaud.pierre_carl@uqam.ca
Lakdawalla, Goldman, Sood: University of Southern California
Cong: Amgen GHE

The authors thank in particular Rob Lempert, Bob Brook, Paul Shekelle and Kevin Murphy for their valuable comments. The authors received funding from Pfizer Inc. for this project. All errors are their own.

Abstract:
We propose to combine clinical trial and estimates of behavioral responses in the population to quantify the value of new drug innovations when such values cannot be obtained by randomized experiments alone. New drugs are seen as having two distinct effects on patients. First, they can provide better outcomes for patients currently under treatment, due to better clinical efficacy. Second, they can also provide treatment access to more patients, perhaps by reducing side effects or expanding treatment. We compare these “clinical” and “access” effects using claims data, data on the arrival rate of new drugs, and the clinical trials literature on the effectiveness of these drugs. We find that the effect of new drug introductions on the number of patients treated accounts for a substantial majority of the value created by new drugs.

Keywords: Pharmaceutical innovation, effectiveness, cost-benefit analysis, cancer

JEL Classification: I10, J14, J18, C23

Introduction

There is much debate among policymakers and researchers about the value of pharmaceutical innovation. Some argue that the health improvements generated by spending on research and development more than justify its cost.[1] Others argue that a great deal of pharmaceutical innovation is of an incremental nature that does little to improve patient health, above and beyond the benefits of existing treatments.[2]

To be sure, a substantial body of evidence already exists on these questions. For example, Lichtenberg relates trends in drug launches by disease to trends in disease specific mortality in 52 countries from 1982 to 2001.[3] The results show that new drugs accounted for 40% of the increase in life expectancy during this period. On the other hand, several important papers in the medical literature have concluded from selected clinical trial data that new drugs are not beneficial.[2] The advantage of the first study is its broader applicability but its main disadvantage is that the mechanisms are unclear and the identification strategy depending on very strong assumptions (e.g. the arrival rate of new drugs is exogeneous to disease trends). The advantage of the second is its proximity to the mechanisms that run from new drugs to health, an approach which helps ensure that the relationship between drug arrival and health trends is a causal one.

In this paper, we combine the virtues of these two opposed approaches, by conducting a systematic study of new drug benefits, based on well-defined mechanisms for health improvement. We hope to overcome the problems both of spurious correlation and of narrow specificity. An additional innovation is our consideration of effects on mortality *and* the onset of other diseases. We take this as a second-best but practical approach to deriving such population estimates. When social experiments are too costly, or are

opposed on ethical grounds, we believe this approach can lead to valuable information for decision makers.

Our approach incorporates evidence on clinical efficacy from reviews of the medical literature, as well as the effects of new drug launches on drawing more people into treatment. We refer to these as the "clinical effects" and "access effects" of new drug launches, respectively. A new drug can have therapeutic advantages over existing treatments, in terms of reduced mortality or morbidity risk. This "clinical effect" improves the outcomes of the population already treated using existing therapies.[4] In addition, new drugs often alleviate side effects, or allow the treatment of previously untreatable patients, leading to an "access effect." Where important assumptions need to be made, we deliberately take a conservative approach, hoping our estimates can be tought off as lower bounds on the effect of new drugs.

We begin by specifying a simple conceptual framework for quantifying the benefits of new drugs, and then present estimates of the clinical and access effects, along with the total estimated benefit of new drug introductions. We then provide some context for our results, by simulating the effects of new drug introductions on health for cancer patients.

Conceptual Framework

By definition, patients diagnosed with a particular disease are either treated or untreated. The advent of a new drug can have impacts on the health of both these groups. The "clinical effect" is the impact on the treated, while the "access effect" is the impact on the untreated.

Clinical effect. When a new drug is discovered, treated patients have the option of switching to it and gaining the incremental clinical benefits — if any — of the new drug,

compared to existing treatments. Therefore, the total clinical benefit of a new drug discovery is equal to: the number of treated patients who switch to the new drug, multiplied by the marginal benefit of the new drug compared to the best available alternative. The latter number is the clinical benefit of the drug in a head-to-head trial. When we implement this calculation, we assume that all treated patients eventually switch to a therapy that is superior to alternatives, but not to one that is inferior.

Access effect. A new drug may be indicated for some untreated patients who were previously unable to receive therapy, due to side effects or comorbidities. In this respect, new drug discoveries can have benefits for the untreated population as well. The size of this benefit is equal to the number of untreated patients who switch to the new drug, multiplied by the marginal benefit of the new drug compared to no treatment. The latter is the clinical benefit of the drug compared to a placebo, assuming that the untreated patients already derive the psychological benefits of placebo treatment.

To illustrate the action of the framework above, consider a new drug for the treatment of heart disease. Assume this hypothetical new drug reduces mortality by 10%, compared to the best available existing treatment. Compared with no treatment, however, it leads to a 20% decrease in mortality risk. Now, suppose half of diagnosed heart disease patients take the best-available existing treatment, while the other half is untreated. Since untreated patients are likely to be healthier than the treated, it is not clear who faces higher or lower mortality risks. For simplicity in this example, assume that both treated and untreated patients face the same survival probability. Finally, assume the introduction of the new drug means that half the currently untreated patients

receive access to the new treatment. In our example, the new drug has the following effects:

1. The treated population, or 50% of the diagnosed population, enjoys a 10% reduction in mortality;
2. Half the untreated population, or 25% of the diagnosed population, switches from no treatment to the new treatment, and enjoys a 20% reduction in mortality;
3. Half the untreated population, or 25% of the diagnosed population, continues to remain untreated and derives no benefit.

The average mortality reduction equals: $[25\% * 0\%] + [50\% * 10\%] + [25\% * 20\%] = 10\%$. We now present our methods for estimating each component of this calculation.

Data and Methods

We estimate the health effects of new drugs for the following conditions: heart disease, hypertension, stroke, lung disease, cancer (excluding skin cancer), diabetes, and depression. These broadly defined conditions span the most prevalent health problems in the United States population.

For each of these conditions, one could in principle consider the whole universe of new drugs and calculate an average effect for each of them. This is likely to be a difficult task. However, top-selling drugs are the most likely to have clinical effects, and have been in general the most studied and reviewed. This makes the estimated clinical benefits more reliable. For each of our diseases, therefore, we survey the clinical effects for the top 5 selling drugs in that disease group, and assume *conservatively* that all other drugs outside the top 5 have no therapeutic benefits. Therefore, we estimate the effects

of drugs in two parts: calculate the probability that a new drug will be a "top-seller," and apply the expected therapeutic benefit of a top-selling drug.

List of New Top Selling Drugs

We construct a list of new top-selling drugs from INGENIX, a large, nationwide, longitudinal claims-based database (1997-2004).[5] This database has roughly 8 million person years of observations from over 40 employers and contains detailed information on prescription drug, outpatient and inpatient expenditures. We start with a list of all new drugs approved by the Food and Drug Administration (FDA) from 1995 to 2002. We consider new chemical entities (NCEs) as well as reformulation and recombination drugs, but exclude generics. Next, we map each new drug to at least one particular health condition, based on approved indications for each drug.

We use the above mapping of drugs to health conditions and the INGENIX data on sales of drugs from 1997 to 2004 to rank drugs for each health condition according to revenues in the 2nd year following introduction.[6] We define the top 5 drugs for each health condition as "top selling drugs". The name of the drug, generic name, and the introduction date are presented in Appendix A. We compute the fraction of drugs that were top selling for each health condition by the dividing the number of top selling drugs for each health condition by the total number of drugs approved for each health condition during the period 1995 to 2002.

Calculating Effects on Health

For each of the top selling drugs, we survey the medical literature for clinical trials. When more than one estimate is available, we average the effects found. We found at least one trial reporting an estimate for each health effect.

We searched for the impacts of top-selling drugs on mortality, and on the incidence of all 6 other health conditions under consideration. However, we follow Goldman et al. (2005) in ruling out some causal links, based on expert opinion.[7] For example, we assume that there is a causal link between hypertension and diabetes, but not from hypertension to cancer. We do not investigate the effect of new drugs on recovery or cure rates. Table 1 summarizes results from the survey of the literature for those effects. Appendix C gives detail on the calculation of each estimate. These calculations provide the estimates of (1) the effects of the drug as compared to the best available alternative and (2) the effects of the drug as compared to placebo. As discussed earlier the clinical effect of new drug is simply the estimate in (1) times the size of the treated population.

Access Effect: Change in Fraction Untreated

To calculate the access effect, we need to estimate the decrease in the fraction of untreated individuals following the introduction of a new drug. We estimate this effect using prescription claims data from INGENIX. By merging the drug consumption data with data on the introduction date of new drugs (from Appendix B), we get a panel data set of the number of prescriptions consumed monthly for each class before and after the introduction of the top 5 drugs (from 1997.1 to 2004.12).[8] Our strategy is to compute the effect of drug launches on class-level prescriptions, relative to the secular trend in prescriptions for that class. The associated regression model includes the natural logarithm of monthly prescriptions in a class *c* as the dependent variable. The model covariates include class fixed-effects (dummy variables for each class) that control for time-invariant differences in prescriptions for each class and linear class-specific time trends that control for pre-existing time trends in sales.[9]

The key independent variables are a series of dummy variables that take the value of 1 when a new top selling drug has been on the market for: zero to three months, four to six months, seven to twelve months, and more than 12 months, respectively. We use the longest-run effects, twelve months after introduction, to compute the eventual change in fraction of untreated population due to new launches.

Access and Incidence Rates by Conditions

Three more parameters are needed to compute average relative risks from equation 1. First, we need to know, for each condition, the fraction untreated. This is the fraction of diagnosed individuals not taking existing drugs. Second, we need estimates of the mortality risks for those treated and untreated. We use the Health and Retirement Study (HRS) for that purpose. The HRS is a nationally representative longitudinal study of the age 50+ population. It asks about lifetime prevalence of the seven conditions we use as well as the consumption of drugs for those diagnosed with these conditions. It also tracks mortality. Mortality rates from the HRS closely track the corresponding figures from life-tables.[10]

To construct estimates of mortality risk, we use hazard models estimated on the 1992-2002 waves of the HRS. The hazard models include baseline demographics, disease indicators from the previous wave, health-risk behaviors, and age. We use similar models to estimate the risk of onset of health conditions. Appendix D presents the model used along with point estimates and goodness-of-fit tests on the HRS data. Table 3 gives the lifetime prevalence in 2004 of various conditions, the fraction untreated among the diagnosed population and predicted transition rates based on hazard models.

[TABLE 3 HERE]

Of the 54.6% of individuals aged 50+ in 2004 with hypertension, only 11% do not take medication for that condition according to the HRS. A somewhat larger fraction with diabetes and heart disease does not take medication (18.3% and 33.8%). Cancer and stroke are the two conditions with the fewest respondents treated with drugs (77.2% and 63.3% are untreated). The next column presents the estimates of the reduction in the fraction untreated following introduction of a new drug. Finally, predicted incidence rates prior to new drug introduction are roughly similar across groups of treated and untreated patients, but perhaps slightly higher in the treated group. Hence, we explicitly account for the fact that therapeutic benefits of treatment may be lower for the currently untreated, because their condition is likely less severe.

Results

Access Effects

New drugs may increase the proportion of the diagnosed population that receives treatment. This can happen through the amelioration of side effects, expansion of therapeutic eligibility criteria, or greater awareness. Table 2 presents estimation results from multivariate regressions that estimate the effects of top-seller launches on the number of prescriptions within a drug class. Figure 2 presents the key findings based on these regressions.

[TABLE 2 HERE]

[FIGURE 2 HERE]

The results show that the introduction of top-selling chemical entities significantly increases the use of drugs within their drug class. For example, prescriptions in a drug class increase by an average of 7% in the first three months, 18.3% (or an additional

11.3%) by months four to six, 24.5% by months seven to twelve after an NCE introduction to that drug class. These effects are not limited to the first year after the introduction of a NCE. In fact, our estimates show little evidence of decay when considering time frames longer than the first year after introduction (21.3% vs. 24.5%). This long-run effect is statistically significant at the 5% level.[11]

Clinical Effects

New drugs can also improve health by directly reducing mortality or reducing the onset of new conditions. As discussed earlier, we need to compute the clinical effects of new drugs relative to placebo, and relative to the best-available alternative. The results are shown in Table 1.

[TABLE 1 HERE]

The blanks in the table indicate that there is no clinical pathway between the two conditions. Therefore, we restrict these effects to be zero.

Average Effects of Top-Selling Drug Launches

We combine the previous elements — access effects and clinical effects — to construct the effect on health for the average new top-seller introduction. This combines the estimated effects of a top-seller on access and health. Table 4 shows how new top selling drugs for each of the seven diseases we study would reduce mortality and the risk of onset of serious health conditions.

[TABLE 4 HERE]

The results show that new top selling drugs for lung disease, diabetes and cancer would have reduced annual mortality risk by 11%, 9% and 7% respectively. Drugs for other diseases we considered including heart disease, hypertension and stroke had no direct

effect on mortality. However, new top selling drugs for these diseases significantly reduced the risk of onset of new health conditions. For example, new top-selling drugs for heart disease reduced annual risk of suffering a stroke by 8%. Similarly, new top-selling drugs for hypertension reduced the annual risk of suffering from heart disease and stroke by 3.6% and 2.6% respectively. It is important to note that Table 4 presents the health effects of a select group of new drugs – the top sellers within each disease category. However, not every new drug would be a top seller.

Probability of Top-Seller

We have adopted the conservative assumption that any drug that is not a top-seller has no beneficial impacts. Therefore, the access and clinical effects are effectively discounted by the probability that a new drug will be a top-seller. Table 5 presents the results of this calculation.

[TABLE 5 HERE]

The table illustrates the entire time-series of new drug introductions and new top-seller introductions by disease, and presents the resulting probability of top-seller, estimated over the 1995 to 2002 data. The estimated probability ranges from 9% (cancer) to 33% (stroke).

Average Effects of a New Drug Introduction

The next step is to calculate the effect of the average new drug introduction. This combines the probability that the drug will be a top-seller, with the expected health benefits of a top-seller drug. We continue to maintain the conservative assumption that all drugs other than top-sellers have no effects on health.

Table 6 presents the expected effect of a new drug for each of the seven diseases assuming that drugs other than those on our top 5 list had no clinical effects. The numbers reported in this table essentially weight the total effect of top selling drugs with the probability that the drug would be a top seller.

The average new drug has a fairly modest effect on health. New launches in diabetes, lung disease, and cancer lower mortality risk by 2.3%, 1.5%, and 0.7%, respectively. Launches in other categories have no effects. These effects were conservatively estimated. Therefore, the long-run health effects should be viewed as lower bounds on the true benefits from new drugs.

Decomposing Access and Clinical Effects

As discussed earlier, new drugs can improve health through two channels. First they can increase the proportion of people (access effect). Second, they can improve the outcomes of those who switch from old therapies to new therapies (clinical effect). Access effects are often overlooked but might be very important. To illustrate this we estimate the effects of new drugs under 2 scenarios: (1) new drugs have no access effect; (2) new drugs have both clinical effects and access effects (estimates reported in Tables 4 and 6). Comparing estimates from the above scenarios allows us to compute the fraction of the total effect of new drugs that can be attributed to the access effect. Table 7 reports the findings from this analysis.

[INSERT TABLE 7 HERE]

The overall pattern of results is striking. For most new top-selling drugs the access effect accounts for all the health benefits of new drugs. Indeed, with the exception of lung disease and cancer, all the benefits are driven by the access effect. The access

effects are easier to identify empirically than the clinical effects. Therefore, our quantitative approach reveals benefits that are tilted towards these quantifiable components. In most of the clinical trials we reviewed new drugs did not offer any greater clinical benefits compared to existing treatments.

Simulation of the Effects on Population Health

The above estimates show the effects of new drugs on annual mortality risks and risk of onset of new conditions. Such estimates, although useful in their own right, can also be used to simulate impacts on long run population health. To illustrate this, we use a the microsimulation model developed and described in our companion paper on the effects of global pharmaceutical regulation. That model is described more fully in that paper, and in the technical appendix. Briefly, its core is a set of predicted health transition rates which are used to simulate the health transitions that the US population is projected to experience under various scenarios. We apply the estimated impacts of new innovation to those health transitions to calculate the long-run effects on health and health spending of new drug introductions.

We start the simulation with the 2004 HRS sample of respondents aged 55+. We then predict each respondent's transition probabilities to the various conditions he/she does not already have. Next, we randomly draw the number of top-selling drug introductions using the probabilities calculated in Table 5 and the historical average number of new drugs. If at least one top selling drug is introduced for a particular condition, we apply the appropriate average clinical effect calculated in Table 4 to the predicted probabilities of those diagnosed with the condition. We then use the model to "age" the population

forward, which involves mortality for some respondents, disease onset for others, and no changes for a third set.

To estimate the health benefits from the introduction of new drugs, we consider the following experiment, where the probability of a new top selling drug for cancer doubles permanently from 0.09 to 0.18. We consider how this affects the number of individuals with cancer, as well life-expectancy, from 2005 to 2025.

From our estimates, new cancer drugs only affect survival rates but not the likelihood of being diagnosed with other conditions. Figure 4 shows that doubling the arrival rate of top-selling cancer drugs increases the size of the population with cancer, due to the reduction in cancer mortality. Table 8 shows that, by 2025, population is higher by 242,000 individuals. Average life-expectancy across all 55 year-olds increases by an average of 26.9 days from 2005 to 2025.

Using a value of statistical life-year of $200,000, which is a middle-of-the-road estimate from the research by Viscusi and Aldy,[12] this represents roughly an average benefit of $14,700 per individual. This change can be compared to the change in pharmaceutical profits required to generate this additional innovation. Acemoglu and Linn (2004) estimated that a 1% increase in pharmaceutical sales leads to an approximately 4% increase in the number of new chemical entities annually.[13] This means sales would need to increase by 25% to generate such a gain in population health. Total worldwide sales for cancer drugs in 2005 were $48.3 billion.[14] Hence, the increase in sales necessary for that change in innovation would be $12.1 billion, or $186 per individual aged 55+ in the U.S. in 2004.[15] Since individuals aged 55+ live on average for 23.1 years, this means that to support this permanent increase in innovation, they would

need to pay on average an additional $4.3K over their life-time. This implies a reasonably high rate of return on stimulating pharmaceutical R&D investments, where the benefit is roughly three times the cost.

Of course, there is no feedback in this simulation from changes in health to sales, which in turn may affect future innovation. For example, as the population with cancer grows, so do sales of cancer drugs. This can provide further stimulus for innovation. We are looking at the long-term effect of permanently changing the rate of innovation, without allowing for a behavioral response from the pharmaceutical industry. Our companion paper on the impacts of global pharmaceutical regulation incorporates this feedback loop.

Conclusion

This paper presented a novel methodology to assess the health effects of new drugs. We incorporated clinical evidence along with access effects following the introduction of drugs to calculate average effects on mortality risks and risk of onset of new health conditions. The results show that new drugs for most of the conditions we studied offer substantial health benefits. They either reduce mortality directly or reduce the onset of other serious health conditions. The results also showed that access effects which are often overlooked account for a majority of the health benefits of new drugs.

The estimates we developed could be used as inputs in models of population health. Such models can then be used to analyze various scenarios which may affect pharmaceutical innovation. We analyze the effects of increasing innovation in cancer drugs by 50%. We found that such an increase in innovation offers significant health benefits which more than offset the costs of providing these drugs. We should emphasize

that we have been rather conservative on the health benefits of new drugs. In particular, we have assumed drugs other than those in the top 5 for each condition did not have any effect on population health. When clinical evidence from the literature was not available or not statistically significant for any of the Top 5 drugs, we assumed there was no effect. This allows us to put a lower bound on the health effects. As we have shown, even with such conservative estimates, innovation in cancer drugs appeared to yield substantial benefits relative to their cost.

Notes

1. D. M. Cutler, M. B. McClellan. "Is Technological Change in Medicine Worth It?" *Health Affairs* 20, no. 5 (2001): 11-29.

2. M. Angell, The Truth About the Drug Companies: How They Deceive Us and What to Do About It (New York: Random House, 2004).

3. F. R. Lichtenberg, "Pharmaceutical Innovation, Mortality Reduction, and Economic Growth." In: Murphy K. M., R. H. Topel, eds. Measuring the Gains from Medical Research: An Economic Approach. (Chicago and London: University of Chicago Press, 2003), 74-109.

4. Of course, perfect diffusion is not instantaneous and likely to take some time.

5. Due to the time range of INGENIX dataset, the list was limited to drugs approved by FDA from 1995 to 2002.

6. We deflate expenditures using the overall consumer price index (CPI).

7. D. P. Goldman, B. Shang, J. Bhattacharya, et al. "Consequences of Health Trends and Medical Innovation for the Future Elderly." *Health Affairs* 24 Suppl 2, no. (2005): W5R5-17.

8. We use the USC-5 classification which results in 22 classes for the purpose of estimation.

9. Results were robust to inclusion of more flexible time trend variables.

10. P. Adams, M. Hurd, D. McFadden, A. Merrill, T. Ribeiro. "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status." *Journal of Econometrics* 112, no. (2003): 3-56.

11. In principle, it would be possible to isolate a separate effect for each health conditions since there are generally more than one drug class that treats a condition. The problem is that for some conditions (stroke, diabetes and depression) few classes are used hence leading to identification problems. Estimates from such specifications did not yield differences that were statistically different from each other.

12. W. K. Viscusi, J. E. Aldy. "The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World." *Journal of Risk and Uncertainty* 27, no. 1 (2003): 5-76.

13. D. Acemoglu, J. Linn. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics* 119, no. 3 (2004): 1049-1090.

14. IMS follows the international classification of drugs (ATC) instead of the Redbook classification we have been using in Appendix A. We used the sales for the ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS class which includes all of our Top 5 drugs for Cancer.

15. This assumes the U.S. population over age 55+ bears the entire cost of this increase in sales.

Figure 1 Effects of a New Drug

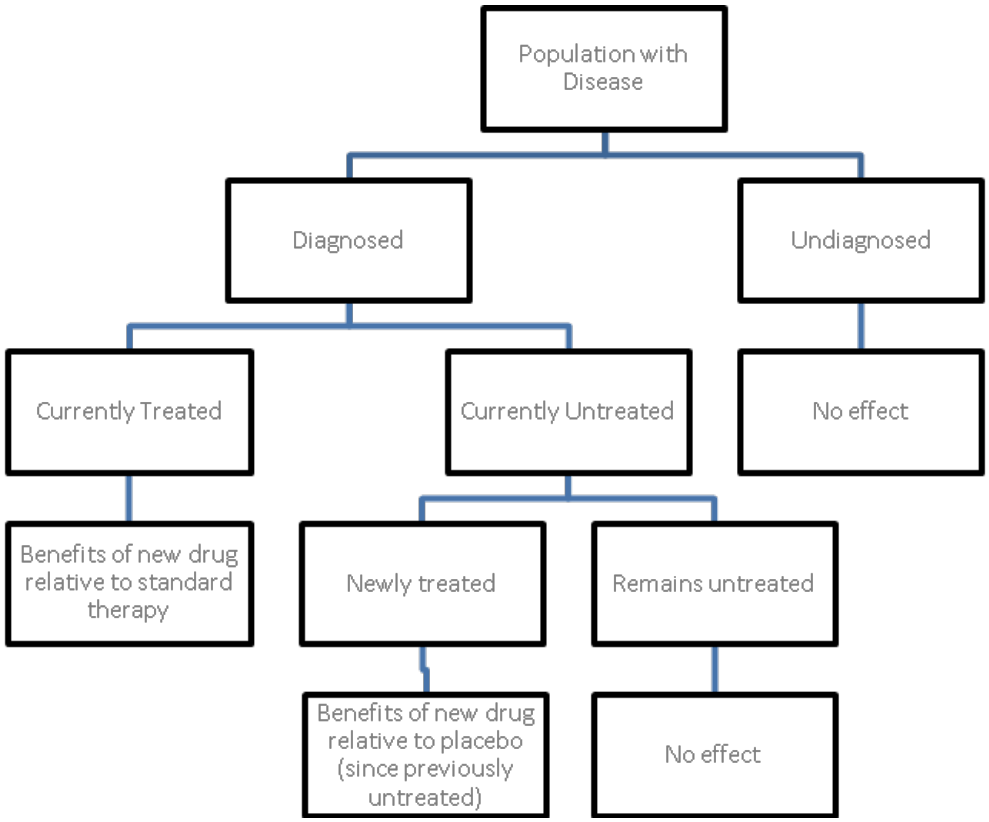
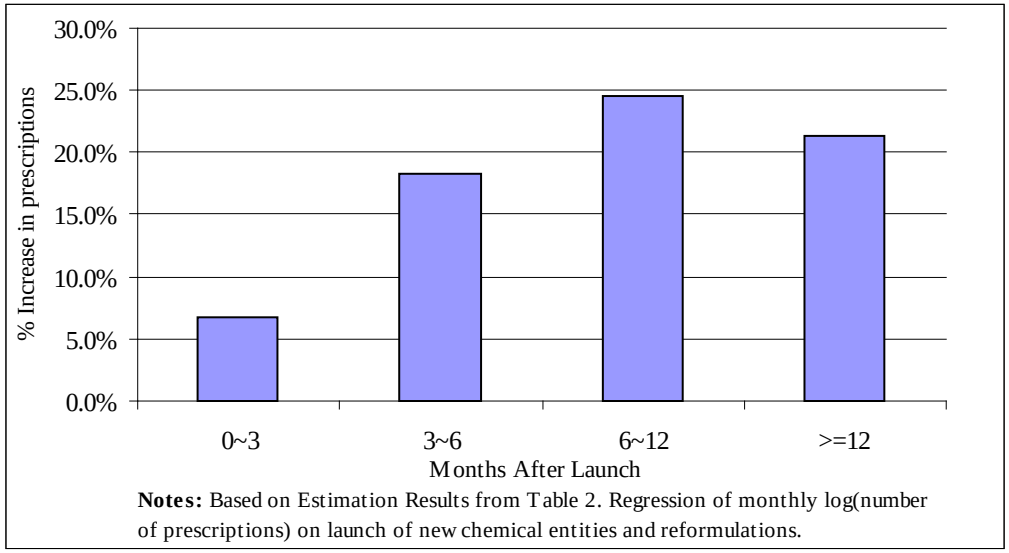
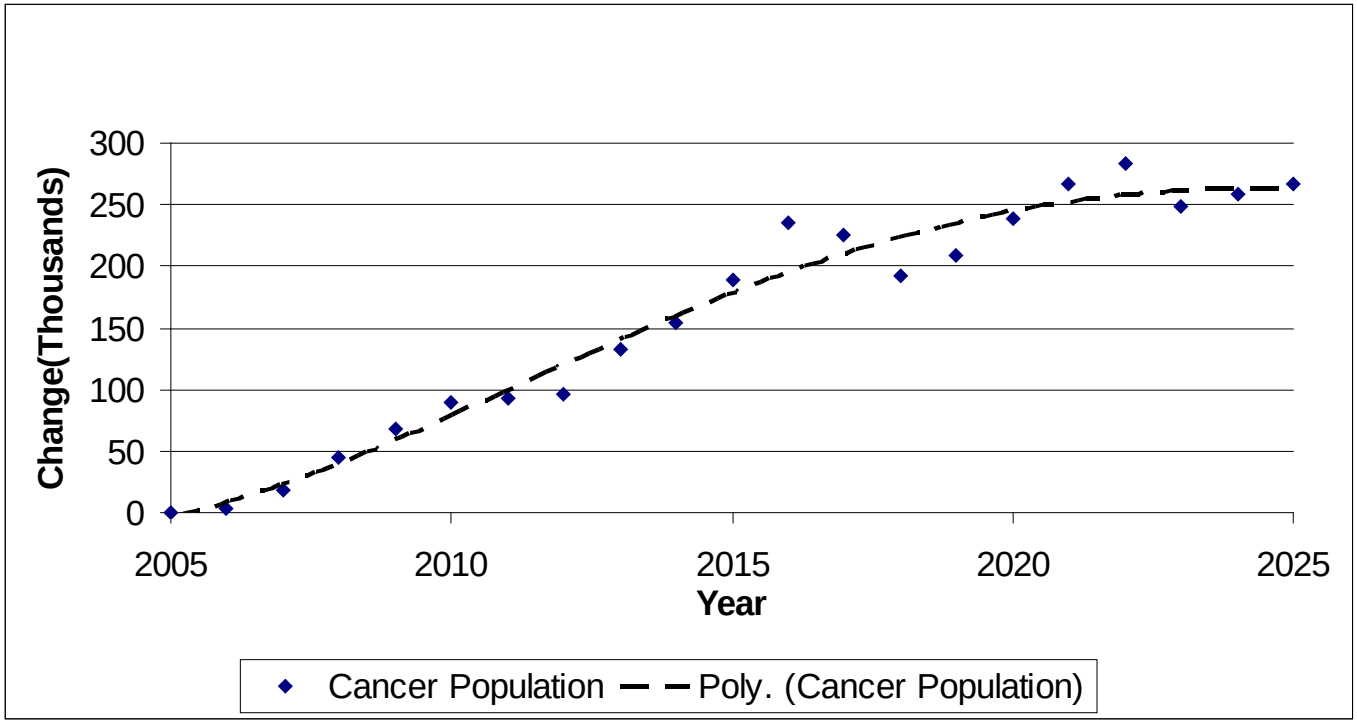


Figure 2 Access Effect



Notes: Based on Estimation Results from Table 2. Regression of monthly log(number of prescriptions) on launch of new chemical entities and reformulations.

Figure 3 Health Effects of Increased Innovation for Cancer Drugs



Notes: Serie smoothed using simulation results from Table 8.

Table 1 Summary of Clinical Effects Found in Medical Literature

	Relative Risk (RR) for New Drug Divided by a Control Group, where control is P=Placebo, E=Existing Treatment															
	Heart disease		Hypertension		Stroke		Lung Disease		Diabetes		Cancer		Depression		Mortality	
Condition treated	RR(P)	RR(E)	RR(P)	RR(E)	RR(P)	RR(E)	RR(P)	RR(E)	RR(P)	RR(E)	RR(P)	RR(E)	RR(P)	RR(E)	RR(P)	RR(E)
heart disease					0.475	1							1	1	1	1
hypertension	0.643	1			0.729	1							1	1	1	1
stroke													1	1	1	1
lung disease													1	1	1	0.796
diabetes	0.690	1	1	1	0.533	1							1	1	0.52	1
cancer					1	1							1	1	0.837	0.728
depression															1	1

Notes: See Appendix C for details on the calculations. This matrix assumes a set of causal clinical mechanisms described in Goldman et al. (2004). Empty cell implies that there is no causal clinical mechanism in the model.

Table 2 Access Effect Regression Results

	Coefficient	P-Value	
Launched 0~3 Months	0.067	0.417	Number of observations: 2711
Launched 3~6 Months	0.183	0.095	R-Square: 0.973
Launched 6~12 Months	0.245	0.032	Time period: Jan 1997~Dec 2004
Launched >=12 Months	0.213	0.042	

Notes: OLS regression of log(number of prescriptions). Standard errors allow for clustering at the USC-5 class level.

Table 3 HRS Disease Prevalence, Drug Usage, and Predicted Incidence Rate

condition treated	prevalence	fraction untreated before introduction	reduction in fraction untreated	Predicted Incidence Rate Prior to Introduction For: Heart disease		Hypertension		Stroke		Depression		Mortality	
				untreated	treated	untreated	treated	untreated	treated	untreated	treated	untreated	treated
heart disease	0.253	0.338	0.141					0.026	0.030	0.028	0.031	0.037	0.047
hypertension	0.546	0.110	0.110	0.048	0.050			0.022	0.023	0.024	0.024	0.026	0.027
stroke	0.080	0.632	0.078							0.039	0.040	0.063	0.069
lung disease	0.102	0.467	0.114							0.040	0.047	0.047	0.053
diabetes	0.176	0.183	0.174	0.068	0.071	0.074	0.081	0.032	0.032	0.029	0.030	0.041	0.040
cancer	0.141	0.772	0.049					0.022	0.022	0.026	0.021	0.040	0.037
mental	0.165	0.426	0.122									0.032	0.029

Notes: Calculations from HRS 2004 data. Sample weights used.

Table 4 Average Risk Reduction in Mortality and Disease Onset of a Top Selling Drugs

	Average Risk Reduction Following Introduction of New Blockbuster Drug							
condition treated	heart	hypertension	stroke	lung disease	diabetes	cancer	depression	mortality
heart			6.7%				0.0%	0.0%
hypertension	3.8%		2.9%				0.0%	0.0%
stroke							0.0%	0.0%
lung disease							0.0%	11.5%
diabetes	5.2%	0.0%	8.1%				0.0%	8.5%
cancer			0.0%				0.0%	6.6%
depression								0.0%

Table 5 Probability of a New Top Selling Drug for Each Health Condition, 1998-2002

New drugs	Heart		Hypertension		Stroke		Lung Disease		Diabetes		Cancer		Depression	
	total	top	total	top	total	top	total	top	total	top	total	top	total	top
1995	8	0	8	0	4	0	4	0	2	0	8	1	0	0
1996	5	1	6	1	1	0	8	1	2	0	10	0	2	1
1997	4	1	6	1	3	2	7	0	3	1	6	0	3	1
1998	6	1	4	1	1	0	5	0	0	0	7	1	2	1
1999	1	0	1	0	2	1	5	0	4	2	7	2	4	1
2000	4	1	2	0	2	1	4	3	6	2	5	0	1	0
2001	5	0	2	1	1	1	7	1	1	0	4	1	2	0
2002	5	1	3	1	1	0	5	1	2	0	6	0	3	1
Total	38	5	32	5	15	5	45	6	20	5	53	5	17	5
Average	4.75	0.625	4	0.625	1.875	0.625	5.625	0.75	2.5	0.625	6.625	0.625	2.125	0.625
Fraction top		0.13		0.16		0.33		0.13		0.25		0.09		0.29

Notes: Information on new chemical entities, new formulation, and new combination drugs taken from the FDA websites. The FDA lists indications for each drug which were then mapped to our set of conditions. The top-selling drugs are identified in Appendix A according to revenues, two years after introduction according to the INGENIX database.

Table 6 Expected Effect of a New Drug

	Expected Risk Reduction Following Introduction of a New Drug							
condition treated	heart	hypertension	stroke	lung disease	diabetes	cancer	depression	mortality
heart			2.2%				0.0%	0.0%
hypertension	0.5%		1.0%				0.0%	0.0%
stroke							0.0%	0.0%
lung disease							0.0%	1.5%
diabetes	0.7%	0.0%	2.7%				0.0%	2.1%
cancer			0.0%				0.0%	0.6%
depression								0.0%

Table 7 Fraction of Access Effect out of Total Effect

	Fraction of Access Effect							
condition treated	heart	hypertension	stroke	lung disease	diabetes	cancer	depression	mortality
heart			100.0%				100.0%	100.0%
hypertension	100.0%		100.0%				100.0%	100.0%
stroke							100.0%	100.0%
lung disease							100.0%	0.0%
diabetes	100.0%	100.0%	100.0%				100.0%	100.0%
cancer			100.0%				100.0%	12.1%
depression								100.0%

Table 8 Health Effects of Increased Innovation for Cancer Drugs

	Simulation outcomes(millions)								
	Diagnosed with Cancer			Population			Remaining Life-Expectancy (at age is 55)		
year	base	high prob	effect	base	high prob	effect	base	high prob	effect
2005	9.894	9.894	0.000	65.088	65.111	0.023	23.12	23.13	0.010
2006	10.445	10.476	0.003	67.357	67.366	0.009	23.15	23.21	0.059
2007	10.916	10.934	0.019	69.504	69.574	0.070	23.19	23.26	0.071
2008	11.293	11.337	0.045	71.585	71.672	0.086	23.27	23.38	0.116
2009	11.647	11.715	0.068	73.673	73.757	0.084	23.33	23.46	0.136
2010	11.965	12.055	0.090	75.764	75.857	0.094	23.33	23.47	0.143
2011	12.246	12.338	0.092	77.891	77.971	0.080	23.37	23.48	0.117
2012	12.466	12.562	0.096	80.037	80.163	0.126	23.40	23.51	0.109
2013	12.747	12.880	0.133	82.175	82.324	0.149	23.44	23.52	0.077
2014	12.976	13.130	0.154	84.409	84.580	0.171	23.41	23.48	0.074
2015	13.218	13.408	0.189	86.623	86.802	0.180	23.46	23.49	0.035
2016	13.404	13.638	0.235	88.792	89.015	0.222	23.43	23.48	0.051
2017	13.700	13.926	0.226	91.019	91.225	0.206	23.40	23.44	0.038
2018	13.962	14.155	0.192	93.134	93.363	0.229	23.42	23.44	0.019
2019	14.215	14.423	0.209	95.281	95.523	0.242	23.43	23.47	0.039
2020	14.476	14.714	0.238	97.371	97.605	0.234	23.37	23.48	0.109
2021	14.727	14.995	0.268	99.318	99.550	0.232	23.39	23.44	0.055
2022	15.009	15.292	0.283	101.123	101.370	0.246	23.38	23.43	0.051
2023	15.367	15.616	0.249	102.751	103.010	0.259	23.35	23.43	0.074
2024	15.641	15.899	0.258	104.227	104.541	0.314	23.32	23.41	0.081
2025	15.891	16.158	0.266	105.658	105.900	0.242	23.34	23.42	0.083

Notes: Simulation repeated 30 times. Remaining life expectancy is life years divided by population age 55 alive in that year. Probability a new cancer drug is a blockbuster drug is 0.09 in base scenario and 0.18 in the high probability scenario.

Appendix Table: Top 5 Drugs by Health Condition

Heart Disease	Hypertension	Stroke	Lung Disease	Diabetes	Cancer	Depression
LIPITOR	CARTIA XT	PLAVIX	ADVAIR DISKUS	ACTOS	GLEEVEC	LEXAPRO
ZETIA	TRACLEER	AGGRENOX	FLOVENT	AVANDIA	CASODEX	PAXIL CR
PLAVIX	ACCURETIC	AGRYLIN	BIAXIN XL	REZULIN	TEMODAR	CELEXA
CARTIA XT	LOTREL	ARIXTRA	AUGMENTIN XR	GLUCOPHAGE XR	XELODA	EFFEXOR XR
WELCHOL	VERELAN PM	INNOHEP	ZYVOX	GLUCOVANCE	AROMASIN	WELLBUTRIN SR

Appendix A Mapping from Drug Class to Health Conditions

Heart Disease	Hypertension	Stroke	Lung Disease	Diabetes	Cancer	Depression
Antibiot, Penicillins	Cardiac, ACE Inhibitors	Thrombolytic Agents, NEC	Antibiot, Penicillins	Antidiabetic Ag, Sulfonylureas	Antibiot, Antifungals	Psychother, Antidepressants
Antihyperlipidemic Drugs, NEC	Cardiac, Beta Blockers	Antiplatelet Agents, NEC	Vaccines, NEC	Antidiabetic Agents, Insulins	Antiemetics, NEC	Antimanic Agents, NEC
Cardiac Drugs, NEC	Vasodilating Agents, NEC	Coag/Anticoag, Anticoagulants	Antibiot, Cephalosporin & Rel.	Antidiabetic Agents, Misc	Antineoplastic Agents, NEC	
Cardiac, ACE Inhibitors	Cardiac, Alpha-Beta Blockers		Antibiot, Tetracyclines		Folic Acid & Derivatives, NEC	
Cardiac, Antiarrhythmic Agents	Hypotensive Agents, NEC		Antibiotics, Misc		Gonadotrop Rel Horm Antagonist	
Cardiac, Beta Blockers	Cardiac, Calcium Channel		Antituberculosis Agents, NEC		Immunosuppressants, NEC	
Cardiac, Cardiac Clycosides	Symnpatholytic Agents, NEC		Sulfonamides & Comb, NEC		Interferons, NEC	
Cardiac, Cardiac Glycosides	Sympatholytic Agents, NEC		Sulfones, NEC		Blood Derivatives, NEC	
Hemorrheologic Agents, NEC	Diuretics, Loop Diuretics		Tuberculosis, NEC		Antibiot, Aminoglycosides	
Vasodilating Agents, NEC	Diuretics, Potassium-Sparing		Antibiot, B-Lactam Antibiotics			
Blood Derivatives, NEC	Diuretics, Thiazides & Related		Antibiot, Erythromycn&Macrolid			
Cardiac, Calcium Channel	Cardiac Drugs, NEC		Anticholinergic, NEC			
Diuretics, Loop Diuretics			Autonomic, Nicotine Preps			
Diuretics, Potassium-Sparing						
Diuretics, Thiazides & Related						
Thrombolytic Agents, NEC						
Antiplatelet Agents, NEC						
Coag/Anticoag, Anticoagulants						

Source: Web search and expert opinion. Printouts of the sources are available upon requests.

Appendix B Top 5 Drugs by Health Condition

(1) Heart Disease

	Drug Name	Class	Active Ingredient	Innovation Type	Introduction Date
1	LIPITOR	Antihyperlipidemic Drugs, NEC	Atorvastatin Calcium	New Ingredient	1996.12
2	ZETIA	Antihyperlipidemic Drugs, NEC	Ezetimibe	New Ingredient	2002.10
3	PLAVIX	Antiplatelet Agents, NEC	Clopidogrel Bisulfate	New Ingredient	1997.11
4	CARTIA XT	Calcium Channel Blocker	Diltiazem Hydrochloride	New Formulation	1998.7
5	WELCHOL	Anti-hyperlipidemic, NEC	Colesevelam Hydrochloride	New Ingredient	2000.5

(2) Hypertension

	Drug Name	Class	Active Ingredient	Innovation Type	Introduction Date
1	CARTIA XT	Cardiac, Calcium Channel	Diltiazem Hydrochloride	New Formulation	1998.7
2	TRACLEER	Vasodilating Agents, NEC	Bosentan	New Ingredient	2001.11
3	BENICAR	Cardiac Drugs, NEC	Olmesartan Medoxomil	New Ingredient	2002.4
4	AVAPRO	Cardiac Drugs, NEC	Irbesartan	New Ingredient	1997.9
5	DIOVAN	Cardiac Drugs, NEC	Valsartan	New Ingredient	1996.12

(3) Stroke

	Drug Name	Class	Active Ingredient	Innovation Type	Introduction Date
1	PLAVIX	Antiplatelet Agents, NEC	Clopidogrel Hydrochloride	New Ingredient	1997.11
2	AGGRENOX	Antiplatelet Agents, NEC	Dipyridamole + Aspirin	New Combination	1999.11
3	AGRYLIN	Antiplatelet Agents, NEC	Anagrelide Hydrochloride	New Ingredient	1997.3
4	ARIXTRA	Coag/Anticoag, Anticoagulants	Fondaparinux Sodium	New Ingredient	2001.12
5	INNOHEP	Coag/Anticoag, Anticoagulants	Tinzaparin Sodium	New Ingredient	2000.7

(4) Lung Disease

	Drug Name	Class	Active Ingredient	Innovation Type	Introduction Date
1	ADVAIR DISKUS	Adrenals & Comb, NEC	Fluticasone Propionate+ Salmeterol Salmeterol Xinafoate	New Combination	2000.8
2	FLOVENT	Adrenals & Comb, NEC	Fluticasone Propionate	New Formulation	1996.3
3	BIAXIN XL	Antibiot, Erythromycn&Macrolid	Clarithromycin	New Formulation	2000.3
4	AUGMENTIN XR	Antibiot, Penicillins	Amoxicillin + Clavulanate	New Formulation	2002.9
5	ZYVOX	Antibiotics, Misc	Linezolid	New Ingredient	2000.4

(5) *Diabetes*

	Drug Name	Class	Active Ingredient	Innovation Type	Introduction Date
1	ACTOS	Antidiabetic Agents, Misc	Pioglitazone Hydrochloride	New Ingredient	1999.7
2	AVANDIA	Antidiabetic Agents, Misc	Rosiglitazone Maleate	New Ingredient	1999.5
3	REZULIN	Antidiabetic Agents, Misc	Withdrawn	New Ingredient	1997.1
4	GLUCOPHAGE XR	Antidiabetic Agents, Misc	Metformin Hydrochloride	New Formulation	2000.10
5	GLUCOVANCE	Antidiabetic Ag, Sulfonylureas	Glyburide + Metformin Hydrochloride	New Combination	2000.7

(6) Cancer

	Drug Name	Class	Active Ingredient	Innovation Type	Introduction Date
1	GLEEVEC	Antineoplastic Agents, NEC	Imatinib Mesylate	New Ingredient	2001.5
2	CASODEX	Antineoplastic Agents, NEC	Bicalutamide	New Ingredient	1995.10
3	TEMODAR	Antineoplastic Agents, NEC	Temozolomide	New Ingredient	1999.8
4	XELODA	Antineoplastic Agents, NEC	Capecitabine	New Ingredient	1998.4
5	AROMASIN	Antineoplastic Agents, NEC	Exemestane	New Ingredient	1999.10

(7) Depression

	Drug Name	Class	Active Ingredient	Innovation Type	Introduction Date
1	LEXAPRO	Psychother, Antidepressants	Escitalopram Oxalate	New Indication	2002.8
2	PAXIL CR	Psychother, Antidepressants	Paroxetine Hydrochloride	New Formulation	1999.2
3	CELEXA	Psychother, Antidepressants	Citalopram Hydrovromide	New Ingredient	1998.7
4	EFFEXOR XR	Psychother, Antidepressants	Venlafaxine Hydrochloride	New Formulation	1997.1
5	WELLBUTRIN SR	Psychother, Antidepressants	Buproprion Hydrochloride	New Formulation	1996.10

Appendix C Details and Reference for Calculation of Clinical Effects in Table 1

Causal Link					
From	To	Drug	Control Group	Reference	Calculation
heart	stroke	LIPITOR	Placebo	Schwartz, G. G., Olsson Ag, Ezekowitz Md, et al. Effects of atorvastatin on early recurrent ischemic events in acute coronary syndromes the MIRACL study a randomized controlled trial. Journal of the American Medical Association. 2001;285(13):1711-1718.	translate RRR into annual RRR by assuming RRR to be conatant over years
hypertenstion	heart	LIPITOR	Placebo	Sever, P. S., Dahlof, B., Poulter, N. R., et al. Prevention of coronary and stroke events with atorvastatin in hypertensive patients who have average or lower-than-average cholesterol concentrations, in the Anglo-Scandinavian Cardiac Outcomes Trial--Lipid Lowering Arm (ASCOT-LLA): a multicentre randomised controlled trial. Lancet. 2003;361(9364):1149-1158.	translate RRR into annual RRR by assuming RRR to be conatant over years
hypertension	stroke	LIPITOR	Placebo	Sever, P. S., Dahlof, B., Poulter, N. R., et al. Prevention of coronary and stroke events with atorvastatin in hypertensive patients who have average or lower-than-average cholesterol concentrations, in the Anglo-Scandinavian Cardiac Outcomes Trial--Lipid Lowering Arm (ASCOT-LLA): a multicentre randomised controlled trial.[comment]. Lancet. 2003;361(9364):1149-1158.	translate RRR into annual RRR by assuming RRR to be conatant over years
diabetes	heart	LIPITOR	Placebo	Colhoun, H. M., Betteridge, D. J., Durrington, P. N., et al. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): Multicentre randomised placebo-controlled trial. Lancet. 2004;364(9435):685-696.	translate RRR into annual RRR by assuming RRR to be conatant over years
diabetes	stroke	LIPITOR	Placebo	Colhoun, H. M., Betteridge, D. J., Durrington, P. N., et al. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): Multicentre randomised placebo-controlled trial. Lancet. 2004;364(9435):685-696.	translate RRR into annual RRR by assuming RRR to be conatant over years
lung disease	mortality	ZYVOX	Vancomycin	ZYVOX label at FDA: http://www.fda.gov/cder/foi/label/2005/021130s008,009,021131s009,010,021132s008,009lbl.pdf	RRR=cure rate in treatment group/ cure rate in control group
diabetes	mortality	ACTOS or AVANDIA	Placebo	Michael Sheehan, Current Therapeutic Options in Diabetes Mellitus, Clinical Medicine and Research, 2003, Vol. 1, No. 3 Kay-Tee Khaw, Nicholas Wareham, Robert Luben, Sheila Bingham, Suzy Oakes, Ailsa Welch,Nicholas Day. Glycated haemoglobin, diabetes, and mortality in men in Norfolk cohort of European Prospective Investigation of Cancer and Nutrition (EPIC-Norfolk).British Medical Journal. Volume 322(7277), 6 January 2001, pp 15-18.	Both lower Hemoglobin A1C levels by 1.5 percentage points,a 1 % increase in HbA1c leads a relative risk increase of 1.28 relative to placebo, so RRR= 1/(1.28)(1.5))= 0.52.
cancer	mortality	TEMODAR+radiotherapy	radiotherapy	Stupp et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastmoa, New England Journal of Medicine, March 10, 2005	RRR=average of RRR in each clinical trial, which is translated tinto annual RRR by assuming RRR to be conatant over years
		TAXOTERE+XELODA	TAXOTERE	FDA drug label of XELODA: http://www.fda.gov/cder/foi/label/2005/020896s016lbl.pdf	
cancer	mortality	GLEEVEC	Interferon-α+Cytarabine	Roy L, Guilhot J, Krahnke T et al. Survival Advantage from Imatinib Compared with the Combination Interferon-α plus Cytarabine in Chronic-phase Chronic Myelogenous Leukemia: Historical Comparison Between Two Phase 3 Trials. Blood. 2006;108:1478-1484.	translate RRR into annual RRR by assuming RRR to be conatant over years

Notes: all the values not available from clinical literature are imputed as 1, i.e., assuming no clinical effect.

Appendix D Health Transition Model

This appendix describes the methodological approach used for the estimation of the transition model as well as the simulation of future health transitions. We use the observed (reported) medical history of respondents in the Health and Retirement Study to infer incidence rates as a function of prevailing health conditions, age and other socio-demographic characteristics (sex, race, risk factors such as obesity and smoking). The data from the Health and Retirement Study consists of a series of record of disease prevalence, recorded roughly every 2 years, from 1992 to 2002. Since incidence can only be recorded every two years, we use a discrete time hazard model.

The estimation of such model is complicated by three factors. First, the report of conditions is observed at irregular intervals (on average 24 months but varying from 18 to 30) and interview delay appears related to health conditions. Second, the presence of persistent unobserved heterogeneity (frailty) could contaminate the estimation of dynamic pathways or "feedback effects" across diseases. Finally, because the HRS samples is from a population of respondents aged 50+, inference is complicated by the fact that spells are left-censored, some respondents are older than 50 when first observed and have health conditions for which we cannot establish the age of onset.

Since we have a stock sample from the age 50+ population, each respondent goes through an individual specific series of intervals. Hence, we have an unbalanced panel over the age range starting from 50 years old. Denote by j_{i0} the first age at which respondent *i* is observed and j_{iT_i} the last age when he is observed. Hence we observe incidence at ages $j_i = j_{i0},..., j_{iT_i}$. Record as $h_{i,j_i,m}$ =1 if the individual has condition *m* as of age j_i. We assume the individual specific component of the hazard can be decomposed in a time invariant and variant part. The time invariant part is composed of the effect of observed characteristics x_i and permanent unobserved characteristics specific to disease *m* , $\eta_{i,m}$. The time variant part is the effect of previously diagnosed health conditions $h_{i,j_i-1,-m}$, (other than the condition *m*) on the hazard.[2] We assume an index of the form $z_{m,j_i} = x_i\beta_m + h_{i,j_i-1,-m}\gamma_m + \eta_{i,m}$. Hence, the latent component of the hazard is modeled as

$$\begin{aligned} &h^*_{i,j_i,m} = x_i\beta_m + h_{i,j_i-1,-m}\gamma_m + \eta_{i,m} + a_{m,j_i} + \varepsilon_{i,j_i,m}, \\ &m = 1,...,M,\ j_i = j_{i0},..., j_{iT_i},\ i = 1,...,N \end{aligned} \quad . \qquad \text{(D.1)}$$

We approximate a_{m,j_i} with an age spline. After several specification checks, a node at age 75 appears to provide the best fit. This simplification is made for computational reasons since the joint estimation with unrestricted age fixed effects for each condition would imply a large number of parameters..

Diagnosis, conditional on being alive, is defined as

[2] With some abuse of notation, $j_i - 1$ denotes the previous age at which the respondent was observed.

$$h_{i,j_i,m} = \max(I(h^*_{i,j_i,m} > 0), h_{i,j_i-1,m})$$
$$m = 1,...,M,\ j_i = j_{i0},...,j_{iT_i},\ i = 1,...,N \qquad \text{(D.2)}$$

As mentioned in the text we consider 7 health conditions to which we add functional limitation (disability) and mortality. Each of these conditions is an absorbing state. The same assumption is made for ADL limitations, the measure of disability we use. The occurrence of mortality, censors observation of diagnosis for other diseases in a current year. Mortality is recorded from exit interviews and tracks closely the life-table probabilities.

Interview Delays

As we already mentioned, time between interviews is not exactly 2 years. It can range from 18 months to 30 months. Hence, estimation is complicated by the fact that intervals are different for each respondents. More problematic is that delays in the time of interview appears related to age, serious health conditions and death (Adams et al., 2003). Hence a spurious correlation between elapsed time and incidence would be detected when in fact the correlation is due to delays in interviewing or finding out the status of respondents who will be later reported dead. To adjust hazard rates for this, we follow Adams et al. (2003) and include the logarithm of the number of months between interviews, $\log(s_{i,j_i})$ as a regressor.

Unobserved Heterogeneity

The term $\varepsilon_{i,j_i,m}$ is a time-varying shock specific to age j_i. We assume that this last shock is Type-1 extreme value distributed, and uncorrelated across diseases.[3] Unobserved difference η_{im} are persistent over time and are allowed to be correlated across diseases $m = 1,...,M$. However, to reduce the dimensionality of the heterogeneity distribution for computational reasons, we consider a nested specification. We assume that heterogeneity is perfectly correlated within nests of conditions but imperfectly correlated across nests. In particular, we assume that each of first 7 health conditions (heart disease, hypertension, stroke, lung disease, diabetes, cancer and mental illness) have a one-factor term $\eta_{im} = \tau_m \alpha_{iC}$ where τ_m is a disease specific factor-loading for the common individual term α_{iC}. We assume disability and mortality have their own specific heterogeneity term α_{iD} and α_{iM}. Together, we assume that the triplet $(\alpha_{iC}, \alpha_{iD}, \alpha_{iM})$ has some joint distribution that we will estimate. Hence, this vector is assumed imperfectly correlated. We use a discrete mass-point distribution with 2 points of support for each dimension (Heckman and Singer, 1984). This leads to $K=8$ potential combinations.

Likelihood and Initial Condition Problem

Parameters $\theta_1 = (\{\beta_m, \gamma_m, \mu_m, \tau_m\}_{m=1}^M, F_\alpha)$, where F_α are the parameters of the discrete distribution, can be estimated by maximum likelihood. Given the extreme value

[3] The extreme value assumption is analogous to the proportional hazard assumption in continuous time.

distribution assumption on the time-varying unobservable (a consequence of the proportional hazard assumption), the joint probability of all time-intervals until failure, right-censoring or death conditional on the individual frailty is the product of Type-1 extreme value univariate probabilities. Since these sequences, conditional on unobserved heterogeneity, are also independent across diseases, the joint probability over all disease-specific sequences is simply the product of those probabilities.

For a given respondent with frailty $\alpha_i = (\alpha_{iC}, \alpha_{iD}, \alpha_{iM})$ observed from initial age j_{i0} to a last age j_{T_i}, the probability of the observed health history is (omitting the conditioning on covariates for notational simplicity)

$$l_i^{-0}(\theta; \alpha_i, h_{i,j_{i0}}) = \left[\prod_{m=1}^{M-1} \prod_{j=j_{i1}}^{j_{T_i}} P_{ij,m}(\theta; \alpha_i)^{(1-h_{ij-1,m})(1-h_{ij,M})} \right] \times \left[\prod_{j=j_{i1}}^{j_{T_i}} P_{ij,M}(\theta; \alpha_i) \right] \qquad \text{(D.3)}$$

We make explicit the conditioning on $h_{i,j_{i0}} = (h_{i,j_{i0},0}, ..., h_{i,j_{i0},M})'$, we have no information on health prior to this age.

To obtain the likelihood of the parameters given the observables, it remains to integrate out unobserved heterogeneity. The complication is that $h_{i,j_{i0},-m}$, the initial condition in each hazard is not likely to be independent of the common unobserved heterogeneity term which needs to be integrated out. A solution is to model the conditional probability distribution $p(\alpha_i \mid h_{i,j_{i0}})$. Implementing this solution amounts to including initial prevalence of each condition at baseline each hazard. Therefore, this allows for permanent differences in the probability of a diagnosis based on baseline diagnosis on top of additional effects of diagnosis on the subsequent probability of a diagnosis. The likelihood contribution for one respondent's sequence is therefore given by

$$l_i(\theta; h_{i,j_{i0}}) = \sum_k p_k l_i(\theta; \alpha_k, h_{i,j_{i0}}) \qquad \text{(D.4)}$$

where the p_k are probabilities for each combination of points of support α_k $k=1,\ldots,K$. The BFGS algorithm is used to maximize the log sum of likelihood contributions in equation (12) over the admissible parameter space.

Clinical Restrictions

Although statistically speaking, all elements of γ_m for all diseases should be unrestricted, it is likely that some of these estimates will reflect associations rather than causal effects because they help predict future incidence. Although we control for various risk factors, it is likely to that we do not observe some factors which are correlated with other diseases. In Medical terms however, some of these effects might be ruled improbable and we use results from the Medical literature to guide restrictions to impose on the elements of the γ_m.

We use a set of clinical restrictions proposed by Goldman et al. (2005) based on expert advice. It turns out that these restrictions are not rejected in a statistical sense one we include initial conditions and unobserved frailty.

Table D.1 Clinical Restrictions

	hazard at (t)								
prevalence t-1	heart	blood pressure	stroke	lung disease	diabetes	cancer	mental	disability	mortality
heart			x				x	x	x
blood pressure	x		x				x	x	x
stroke							x	x	x
lung disease							x	x	x
diabetes	x	x	x				x	x	x
cancer			x				x	x	x
mental							x	x	x
disability							x	x	x

Notes: x denotes a parameter which is allowed to be estimated.

Descriptive Statistics and Estimation Results

For estimation, we construct an unbalanced panel from pooling all cohorts together. We delete spells if important information is missing (such as the prevalence of health conditions). Hence, in the final sample, a sequence can be terminated because of death, unknown exit from the survey (or non-response to key outcomes), or finally because of the end of the panel.

In each hazard, we include a set of baseline characteristics which capture the major risk factors for each condition. We consider education, race & ethnicity, marital status, gender and behaviors such as smoking and obesity. Finally, as discussed previously, we also include a measure of the duration between interviews in month. The average duration is close to 2 years. Table D.2 gives descriptive statistics at first interview.

Table D.2 Baseline Characteristics in Estimation Sample

Characteristics (at first interview)	N	mean	std. dev.	min	max
age in years	21302	64.1	11.2	50	103
less than high school		0.350	0.477	0	1
some college education		0.346	0.476	0	1
black		0.140	0.347	0	1
hispanic		0.068	0.251	0	1
married		0.703	0.457	0	1
male		0.431	0.495	0	1
ever smoked		0.591	0.492	0	1
obese (BMI>30)		0.210	0.407	0	1
duration between interviews (in months), averaged over all waves		23.4	2.8	1.8	30.9

Notes: All HRS Cohorts (HRS, AHEAD, CODA, War Babies)

Estimates of the hazard models are presented in Table D.3. Estimates can be interpreted as the effect on the log hazard.

Table D.3 Estimates with Heterogeneity and Clinical Restrictions

	Heart disease	Blood pressure	Stroke	Lung disease	Diabetes	Cancer	Mental	Disability	Mortality
prevalence t-1	pe	pe	pe	pe	pe	pe	pe	pe	pe
heart			-0.212 *				0.037	-0.160 *	0.599 **
blood pressure	0.033		0.042				0.169 *	-0.115	0.426 **
stroke							-0.172	0.240 *	0.864 **
lung disease							-0.225	0.185	1.152 **
diabetes	0.062	0.346 **	0.043				-0.422 **	-0.086	0.634 **
cancer			-0.204				-0.141	0.222 **	1.428 **
mental								0.336 **	0.740 **
disability							0.199 **		0.840 **
prevalence t=0									
heart		0.076	0.483 **	0.395 **	0.190 **	0.143 **	0.272 **	0.518 **	-0.220 **
blood pressure	0.358 **		0.418 **	0.046	0.578 **	0.082	0.099	0.356 **	-0.277 **
stroke	0.030	0.238 **		-0.229	0.012	0.086	0.466 **	0.425 **	-0.418 **
lung disease	0.511 **	0.006	0.396 **		0.014	0.301 **	0.841 **	0.627 **	-0.509 **
diabetes	0.540 **	0.014	0.584 **	0.014		-0.069	0.705 **	0.711 **	0.005
cancer	0.191 **	0.050	0.244	0.259 **	-0.023		0.308	-0.128	-1.037 **
mental	0.335 **	0.208 **	0.422 **	0.594 **	0.171 *	0.012		0.512 **	-0.581 **
disability	0.330 **	0.109	0.152	0.423 **	0.127	0.057	0.478 **		-0.065
demographics									
age <75	0.042 **	0.021 **	0.071 **	0.019 **	0.013 **	0.044 **	-0.007 *	0.035 **	0.030 **
age >75	0.038 **	-0.022 **	0.055 **	-0.004	-0.044 **	-0.023 **	0.038 **	0.143 **	0.112 **
black	-0.268 **	0.336 **	0.153 *	-0.363 **	0.210 **	-0.092	-0.225 **	0.447 **	0.244 **
hispanic	-0.441 **	0.095	-0.199	-0.582 **	0.420 **	-0.370 **	0.194 **	0.424 **	-0.073
male	0.336 **	-0.109 **	0.063	-0.146 **	0.364 **	0.366 **	-0.458 **	-0.195 **	0.420 **
ever smoked	0.176 **	-0.009	0.255 **	1.040 **	0.102 *	0.257 **	0.187 **	0.210 **	0.344 **
obese (BMI>30)	0.196 **	0.350 **	0.106	0.059	1.065 **	0.027	-0.032	0.552 **	-0.273 **
high school	-0.169 **	-0.091 *	-0.137 *	-0.356 **	-0.274 **	-0.024	-0.334 **	-0.430 **	-0.029
college	-0.191 **	-0.146 **	-0.252 **	-0.581 **	-0.312 **	0.088	-0.461 **	-0.586 **	-0.168 **
log(time since l.w.)	0.996 **	1.224 **	1.242 **	1.015 **	1.296 **	1.063 **	1.102 **	0.614 **	6.547 **
constant	-6.573 **	-6.121 **	-8.592 **	-7.164 **	-8.016 **	-7.859 **	-6.121 **	-4.433 **	-26.079 **
point 1	0	0	0	0	0	0	0	0	0
point 2	-1.353 **	-1.353 **	-1.353 **	-1.353 **	-1.353 **	-1.353 **	-1.353 **	-2.164 **	-2.176 **
Loading Factor	1	0.625 **	1.637 **	1.085 **	0.678 **	0.244 **	1.308 **	1	1
Probability estimates									
point	p(1,1,1)	p(1,1,2)	p(1,2,1)	p(1,2,2)	p(2,1,1)	p(2,1,2)	p(2,2,1)	p(2,2,2)	
Probability	0.193 **	0.082 **	0	0.024 **	0.085 **	0	0.530 **	0.087 **	
loglike/N	-3.632								

To judge the fit of the model we perform a goodness-of-fit exercise. To do that, we re-estimate the model on a sub-sample and keep part of the sample for evaluating the fit. We randomly select observations from the original HRS cohort with probability 0.5 and simulate outcomes for this cohort starting from observed 1992 outcomes. Table D.4 gives the observed frequencies as well as the predicted ones. Predicted and observed frequencies are quite close to each other in 2002.

Table D.4 Goodness-of-Fit

Prevalence Rate (Independent Draws)

	heart		pressure		stroke		lung	
year	data	sim	data	sim	data	sim	data	sim
1992	0.117	0.120	0.347	0.344	0.027	0.027	0.061	0.062
1994	0.138	0.147	0.376	0.393	0.030	0.037	0.074	0.074
1996	0.157	0.172	0.401	0.437	0.039	0.046	0.078	0.087
1998	0.176	0.196	0.435	0.478	0.047	0.055	0.088	0.097
2000	0.199	0.218	0.480	0.516	0.056	0.063	0.093	0.106
2002	0.236	0.241	0.527	0.551	0.064	0.072	0.109	0.113
# cond.	825	853	1843	1949	224	254	380	399
	diabetes		cancer		disability		mental	
year	data	sim	data	sim	data	sim	data	sim
1992	0.104	0.108	0.058	0.058	0.053	0.056	0.072	0.072
1994	0.121	0.130	0.065	0.076	0.094	0.116	0.090	0.095
1996	0.137	0.150	0.078	0.093	0.160	0.164	0.104	0.115
1998	0.151	0.169	0.093	0.111	0.197	0.205	0.118	0.133
2000	0.169	0.185	0.107	0.125	0.224	0.237	0.131	0.148
2002	0.199	0.201	0.125	0.141	0.248	0.264	0.154	0.160
# cond.	695	711	436	500	867	934	540	566
	no conditions		1 cond		2 cond		3 cond.+	
	data	sim	data	sim	data	sim	data	sim
year								
1992	0.475	0.476	0.317	0.311	0.136	0.138	0.072	0.075
1994	0.422	0.390	0.328	0.329	0.149	0.166	0.101	0.115
1996	0.371	0.323	0.326	0.333	0.168	0.191	0.134	0.153
1998	0.324	0.270	0.321	0.326	0.191	0.211	0.163	0.192
2000	0.279	0.230	0.312	0.315	0.215	0.229	0.194	0.226
2002	0.231	0.199	0.295	0.299	0.235	0.240	0.238	0.263

Incidence Rate

	mortality	
year	data	sim
1992	0.000	0.000
1994	0.014	0.009
1996	0.014	0.012
1998	0.016	0.014
2000	0.019	0.017
2002	0.019	0.020

Goodness-of-Fit test			
Prevalence rates		4.05	0.774
(dF = 7)			
Np	3539		
Nu	3500		

Notes: Simulation for HRS 1992 subsample (N=4131)