HEALTH CARE

Predictive modeling of U.S. health care spending in late life

Liran Einav[1,2], Amy Finkelstein[1,3]*, Sendhil Mullainathan[1,4], Ziad Obermeyer[5]

That one-quarter of Medicare spending in the United States occurs in the last year of life is commonly interpreted as waste. But this interpretation presumes knowledge of who will die and when. Here we analyze how spending is distributed by predicted mortality, based on a machine-learning model of annual mortality risk built using Medicare claims. Death is highly unpredictable. Less than 5% of spending is accounted for by individuals with predicted mortality above 50%. The simple fact that we spend more on the sick—both on those who recover and those who die—accounts for 30 to 50% of the concentration of spending on the dead. Our results suggest that spending on the ex post dead does not necessarily mean that we spend on the ex ante "hopeless."

Only 5% of Medicare beneficiaries in the United States die each year, but one-quarter of Medicare spending occurs in the last 12 months of life (1). This fact is frequently touted as evidence of obvious waste and inefficiency. For example, an article in the *New Yorker* states that "...for most people, death comes only after long medical struggle with an incurable condition—advanced cancer, progressive organ failure..., or the multiple debilities of very old age. In all such cases, death is certain, but the timing isn't" (2). Likewise, the *New York Times* asks, "Does it make sense that older adults in their last year of life consume more than a quarter of Medicare's expenditures...? Are there limits to what Medicare should spend on a therapy prolonging someone's life by a month or two?" (3). In this view, a large share of health care dollars is wasted on small marginal gains for those certain to die within a short period of time (4, 5).

These common interpretations of end-of-life spending flirt with a statistical fallacy: Those who end up dying are not the same as those who were sure to die. Ex post, spending could appear concentrated on the dead, simply because we spend more on sicker individuals who have higher mortality—even if we never spent money on those certain to die within the year.

Empirically, this suggests using predicted mortality, rather than ex post mortality, to assess end-of-life spending. To this end, we draw on rich data from a random sample of almost 6 million Medicare enrollees. We apply machine-learning techniques to generate a prediction of each individual's probability of death in the next 12 months. We then analyze spending by predicted mortality as well as by ex post mortality.

The conceptual distinction between the ex post dead and ex ante dead has been noted previously (6, 7); see also (8) for early empirical analysis. Others have attempted to predict mortality in the Medicare population and have observed that substantial prognostic uncertainty is a challenge for medical care (9–12). Our study combines these themes and examines end-of-life spending from an ex ante perspective.
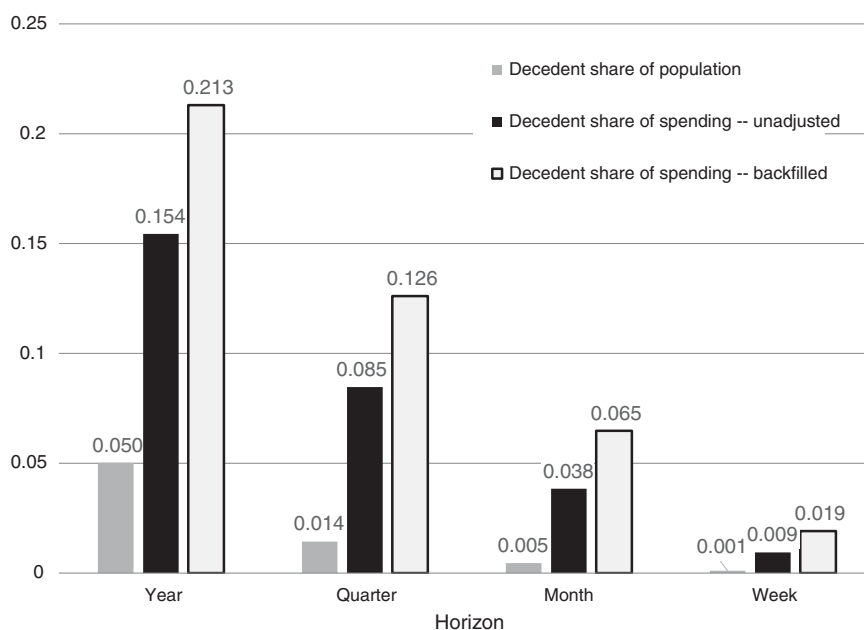
We use Medicare claims data for a random sample of 20% of enrollees. Our main analysis focuses on enrollees alive on 1 January 2008 and continuously enrolled in Medicare in 2007 and all months of 2008 in which they were alive. We observe age; gender; race; Medicaid coverage (a proxy for socioeconomic status); all Medicare claims for inpatient care, outpatient care, and physician services; and all recorded health diagnoses. More details are provided in the supplementary materials, section A.

Figure 1 reproduces well-known facts about the concentration of spending at the end of life. We report results for two spending measures. The first, which we refer to as "backfilling," follows the approach of the end-of-life literature (13). For survivors, it measures spending over the relevant time interval from 1 January 2008 going forward; for decedents, it measures spending starting from the date of death in 2008 and going backward over the same length of time. Using this approach, we estimate that the 5% of Medicare beneficiaries who died accounted for 21% of Medicare spending, closely matching prior estimates (13).

This standard analysis suffers from two related biases: We do not know who will die in a given time interval, or when, within that interval, they will die. We therefore also analyze what we refer to as "unadjusted spending," for which we measure spending on all individuals—both survivors and decedents—looking forward from 1 January 2008. Now, the 5% of enrollees who die within the year account for only 15% of spending in that year. But even this analysis assumes that we knew who would die in the next year, an assumption we now investigate.

Our baseline analysis generates annual mortality predictions from the vantage point of 1 January 2008 by using data on enrollee demographics, health care utilization over the prior 12 months—including the level and nature of

Fig. 1. Concentration of spending on the ex post dead. Shown are mortality rates and decedent share of total Medicare spending for various time intervals after 1 January 2008. Data are for the entire baseline sample (*n* = 5,631,168). Spending for survivors is measured in the time interval since 1 January 2008. For decedents, we report two spending measures: backfilled, which measures spending looking backward from the date of death for the length of the relevant interval (for example, for the 1-year measure, we measure spending over the 12 months before death), and unadjusted, which measures spending looking forward over the relevant time interval since 1 January 2008.

[1]National Bureau of Economic Research, Cambridge, MA 02138, USA. [2]Department of Economics, Stanford University, Stanford, CA 94305, USA. [3]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [4]Department of Economics, Harvard University, Cambridge, MA 02138, USA. [5]Department of Emergency Medicine and Health Care Policy, Harvard Medical School, Boston, MA 02115, USA.
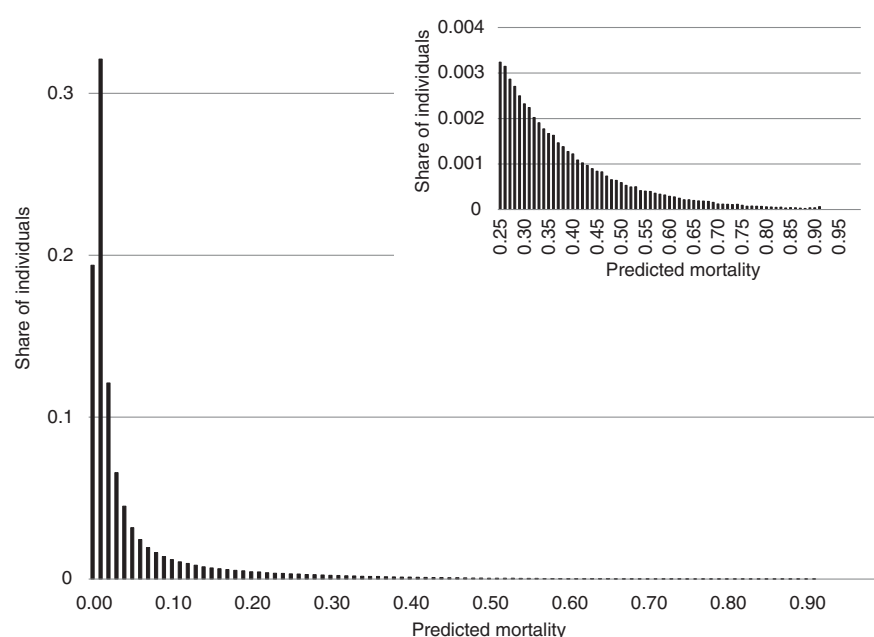*Corresponding author. Email: afink@mit.edu

care and its trajectory—and health diagnoses and their trajectories over the prior 12 months. This produces thousands of potential predictors. We use an ensemble (of random forest, gradient boosting, and LASSO)—a standard and popular machine-learning technique—to generate mortality predictions. To avoid overfitting, we randomly split the data into a "training" subsample, for which we develop the prediction algorithm, and a "test" subsample, for which we apply the resulting algorithm to generate predicted mortalities. All subsequent results are for this test subsample, which is one-third of our original sample. How we construct the potential mortality predictors and the prediction algorithm is described in detail in the supplementary materials, section B. It shows that predicted mortality varies in sensible ways with individual characteristics and that our algorithm's performance is comparable to other recent mortality-prediction endeavors.

Figure 2 shows the distribution of annual mortality predictions and illustrates one of our key findings: There is no sizable mass of people for whom death is certain (or even near certain) within the year. The 95th percentile of predicted annual mortality is only about 25%. Less than 10% of those who end up dying within the year have an annual mortality probability above 50%.
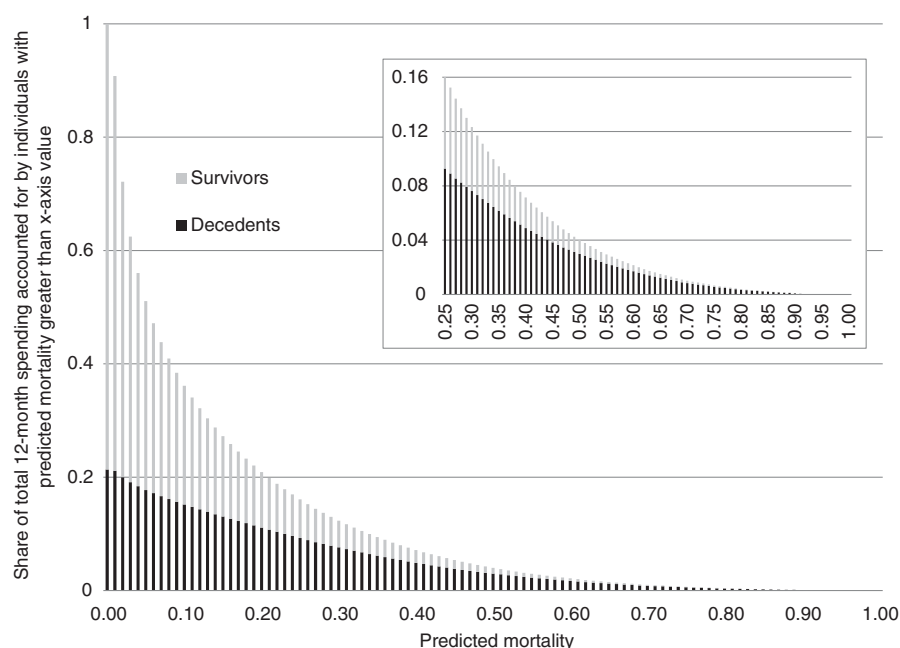
Figure 3 shows that, relatedly, individuals with high predicted mortality account for only a small share of total spending. For example, the highest-risk percentile, those with predicted mortality above 46% percent, accounts for under 5% of total spending, and 45% of these individuals are survivors. To capture a group of decedents who account for at least 5% of total spending, we must set a threshold of predicted mortality of 39% or higher. These results are based on the backfilled measure of decedent spending; when using the unadjusted measure, spending on decedents is even lower, so that a smaller share of spending above each mortality prediction threshold is accounted for by decedents.

A natural question is whether these results would change if we had better predictions, for example, made with higher quality data such as electronic medical records. The available evidence, although limited, suggests that, relative to using only (detailed) claims data, the incremental predictive power obtained from electronic medical records (*14*) or subjective physician predictions (*15*, *16*) is relatively small. Moreover, such data are arguably less relevant for national policy, which needs to be based on standardized, nationally available data.

There is also the possibility of better prediction algorithms. Indeed, some cutting-edge machine-learning methods (*17*, *18*) do better in select patient groups. To study how a hypothetical, better predictor might plausibly affect our results, we produce an artificial "oracle" predictor by adjusting predicted probabilities toward realized outcomes (i.e., increasing predictions for the dead and lowering them for survivors); our hypothetical predictor is thus a weighted average of our actual predictor and the realized outcome (death occurs
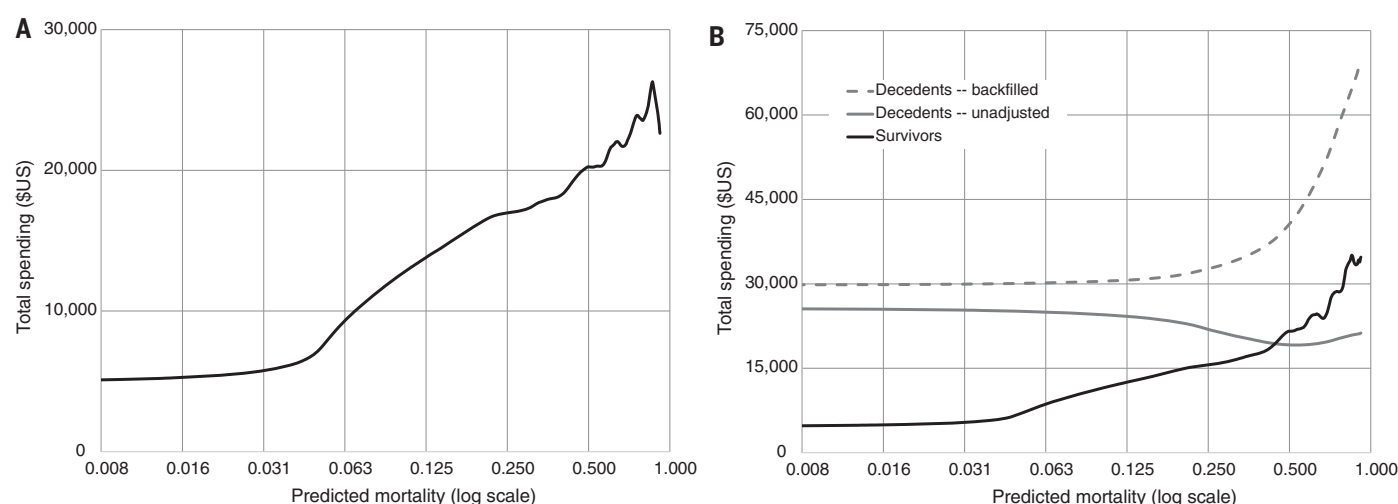

Fig. 2. Distribution of predicted mortality. The distribution of predicted annual mortality from 1 January 2008 is shown. Data are from the test subsample (*n* = 1,877,168). The inset provides more detail about the corresponding section of the distribution.

Fig. 3. Concentration of spending by ex ante mortality. For each level of predicted annual mortality (*x* axis), the share of total annual Medicare spending accounted for by individuals with predicted mortality of that value or greater is shown. Each bar stacks the share accounted for by decedents (black) and by survivors (gray), so that the height of the bar represents total annual Medicare spending accounted for by individuals (decedents and survivors) with predicted mortality of that value or greater. All results use the backfilled measure of decedent spending. All data are from the test subsample. The inset provides more detail about the corresponding section of the distribution.

or does not). If we put a weight of 0.1 on the realized outcome, this generates an area under the curve (AUC) of 0.963—a level of algorithm performance well above any in the literature—but our results do not qualitatively change: Individuals with predicted mortality above 47% still

only account for 5% of total spending. This happens because, at low baseline mortality rates (i.e., annual mortality rate of 5%), models can be extremely good at identifying those at high risk (i.e., AUC can be extremely high), but the highest percentiles can still have modest absolute rates

Fig. 4. Spending by predicted mortality. (**A**) Kernel density of total Medicare spending in the 12 months after 1 January 2008 against predicted annual mortality. (**B**) Kernel density of Medicare spending separately for survivors and decedents. Spending measures are as defined in Fig. 1. All data are from the test subsample.

of predicted mortality (under 50%). As a result, there is little concentration of spending on individuals with high absolute rates of predicted mortality. More details are provided in the supplementary materials, section C.

Nor do our conclusions change when we view the prediction task from an arguably more "decision-relevant" time point: when potentially costly medical treatment decisions are made, at hospital admission. In the supplementary materials, section D, we reestimate the prediction algorithm to generate 12-month mortality predictions at the time of hospital admission for the subsample of individuals admitted to the hospital during 2008. We use the same predictors, now measured in the 12 months before admission, as well as the admitting diagnosis. Even from the vantage point of admission to the hospital, where annual mortality is about 20%, the 95th percentile of annual death probabilities is still only 67%. Less than 4% of those who end up dying in the subsequent year have a predicted mortality above 80% at the time of admission. Even if we zoom in further on the subsample of individuals who enter the hospital with metastatic cancer—63% die over the subsequent 12 months, but they account for only 7% of annual Medicare deaths—we find that only 12% of decedents have an annual predicted mortality of more than 80%. Qualitatively similar findings hold if we look at mortality in the month, rather than year, after hospital admission.

Figure 4 shows the distribution of spending by predicted mortality and illustrates another key finding: A large share of the concentration of spending at the end of life can be explained by the concentration of spending on the sick. Decedents have higher predicted mortality than survivors and, as Fig. 4A shows, spending is increasing in predicted mortality. This simple observation goes a long way toward explaining the concentration of spending at the ex post end of life.

Figure 4B shows the relationship between spending and predicted mortality separately for

subsequent decedents and survivors. Using these estimates, we find that survivors randomly sampled from the decedents' distribution of predicted mortality spend about twice as much on health care as a randomly sampled survivor. As a result, 30 to 50% of the concentration of spending on decedents relative to survivors would be eliminated (depending on whether the unadjusted or backfilled spending measure was used), simply by accounting for the fact that spending is higher on those with higher mortality risk. More details are provided in the supplementary materials, section E.

However, Fig. 4B also shows that, even for individuals with the same predicted mortality probability, spending is higher for those who subsequently die, particularly for individuals with the lowest predicted mortality. This may be because of ex ante differences across patients that our current prediction algorithm does not incorporate, or it may be related to the process by which individuals die or even the basic mechanics of death. More work is needed to fully understand why death remains expensive, even conditional on mortality risk.

In sum, although spending on the ex post dead is very high, we find there are only a few individuals for whom, ex ante, death is near certain. Moreover, a substantial component of the concentration of spending at the end of life is mechanically driven by the fact that those who end up dying are sicker, and spending, naturally, is higher for sicker individuals. Of course, we do not—and cannot—rule out individual cases where treatment is performed on an individual for whom death is near certain. But our findings indicate that such individuals are not a meaningful share of decedents.

These findings suggest that a focus on end-of-life spending is not, by itself, a useful way to identify wasteful spending. Instead, researchers must focus on quality of care for very sick patients—identifying the impact of specific health

care interventions on survival rates and, just as importantly, on palliation of symptoms; such research should focus not just on averages but also on potentially heterogeneous impacts across different individuals (*19–21*).

REFERENCES AND NOTES

1. G. F. Riley, J. D. Lubitz, *Health Serv. Res.* **45**, 565–576 (2010).
2. A. Gawande, "Letting go," *New Yorker*, 2 August 2010; www.newyorker.com/magazine/2010/08/02/letting-go-2.
3. E. Porter, "Rationing health care more fairly," *New York Times*, 21 August 2012; www.nytimes.com/2012/08/22/business/economy/rationing-health-care-more-fairly.html.
4. E. J. Emanuel, L. L. Emanuel, *N. Engl. J. Med.* **330**, 540–544 (1994).
5. Medicare Payment Advisory Commission, "Report to the Congress: Selected Medicare Issues," June 1999; https://babel.hathitrust.org/cgi/pt?id=mdp.39015046749704;view=1up;seq=3.
6. A. A. Scitovsky, *Milbank Q.* **83**, 825–841 (2005).
7. J. P. Newhouse, *Health Aff. (Millwood)* **12** (suppl. 1), 152–171 (1993).
8. A. S. Detsky, S. C. Stricker, A. G. Mulley, G. E. Thibault, *N. Engl. J. Med.* **305**, 667–672 (1981).
9. M. M. Desai, S. T. Bogardus Jr., C. S. Williams, G. Vitagliano, S. K. Inouye, *J. Am. Geriatr. Soc.* **50**, 474–481 (2002).
10. J. J. Gagne, R. J. Glynn, J. Avorn, R. Levin, S. Schneeweiss, *J. Clin. Epidemiol.* **64**, 749–759 (2011).
11. J. Makar, M. Ghassemi, D. M. Cutler, Z. Obermeyer, *Int. J. Mach. Learn. Comput.* **5**, 192–197 (2015).
12. L. C. Yourman, S. J. Lee, M. A. Schonberg, E. W. Widera, A. K. Smith, *JAMA* **307**, 182–192 (2012).
13. C. Hogan, J. Lunney, J. Gabel, J. Lynn, *Health Aff. (Millwood)* **20**, 188–195 (2001).
14. S. K. Inouye *et al.*, *Med. Care* **41**, 70–83 (2003).
15. P. Glare *et al.*, *BMJ* **327**, 195–198 (2003).
16. J. R. Lakin *et al.*, *JAMA Intern. Med.* **176**, 1863–1865 (2016).
17. A. Elfiky *et al.*, bioRxiv 204081 [Preprint]. 18 October 2017.
18. A. Rajkomar *et al.*, arXiv:1801.07860 [cs.CY] (11 May 2018).
19. M. D. Aldridge, A. S. Kelley, *Am. J. Public Health* **105**, 2411–2415 (2015).
20. A. S. Kelley *et al.*, *Health Serv. Res.* **52**, 113–131 (2017).
21. A. S. Kelley, E. Bollens-Lund, K. E. Covinsky, J. S. Skinner, R. S. Morrison, *J. Palliat. Med.* **21**, 44–54 (2018).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6396/1462/suppl/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S9

Tables S1 to S6

References (22–39)

Science

Predictive modeling of U.S. health care spending in late life

Liran Einav, Amy Finkelstein, Sendhil Mullainathan and Ziad Obermeyer

End-of-life health care spending
In the United States, one-quarter of Medicare spending occurs in the last 12 months of life, which is commonly seen as evidence of waste. Einav *et al.* used predictive modeling to reassess this interpretation. From detailed Medicare claims data, the extent to which spending is concentrated not just on those who die, but on those who are expected to die, can be estimated. Most deaths are unpredictable; hence, focusing on end-of-life spending does not necessarily identify "wasteful" spending.
Science, this issue p. 1462

ARTICLE TOOLS	http://science.sciencemag.org/content/360/6396/1462
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2018/06/27/360.6396.1462.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/360/6396/1459.full
REFERENCES	This article cites 26 articles, 5 of which you can access for free http://science.sciencemag.org/content/360/6396/1462#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions