Executive Summary of

Upcoding: Evidence from Medicare on Squishy Risk Adjustment

by Michael Geruso & Timothy Layton

Key Findings:

- **Medicare Advantage (MA) codes more intensely than Traditional Medicare (TM).** MA plans generate risk scores for their enrollees that are on average 6.4% larger in the first year of MA enrollment than what the same enrollees would have generated under TM.
- **MA coding intensity ratchets up over time.** We find that the TM/MA difference reaches 8.7% by the second year of MA enrollment and continues to grow into the third year.
- **Integrated plans code most intensely.** Diagnosis codes ultimately originate from provider visits. We find provider-owned plans generate 16% higher risk scores for the same patients compared to TM.
- **Upcoding leads to MA overpayments.** The coding differential leads to billions of dollars in MA overpayments. Even if the coding difference were entirely due to TM under-coding and MA correctly (medically appropriately) coding, the end result is an overpayment to MA plans.

Contact: Prof. Michael Geruso
University of Texas at Austin and NBER
mike.geruso@austin.utexas.edu

Background: Today, seniors and other Medicare eligibles choose between the public, Traditional Medicare (TM) option and a private PPO or HMO plan in the Medicare Advantage (MA) program. A key consideration with respect to contracting with private MA plans is how to set the level of payments. In practice, the federal government makes capitated monthly payments to insurance carrier for each enrolled Medicare beneficiary.

To keep insurers from attempting to avoid the sickest patients, these capitated payments are *risk adjusted* so that insurers enrolling sicker, costlier patients receive higher payments from the regulator. In the absence of such risk-adjusted payments, plans would design benefit schedules and target marketing activities in a way that attempted to screen out higher cost enrollees. For example, plans might offer relatively poor quality coverage for cardiac, diabetes, or cancer care, because good coverage for these conditions would attract sick patients. When risk adjustment works well, insurers would compete to attract such patients, rather than attempting to avoid them.

We challenge the assumption that diagnoses are objective and observable.

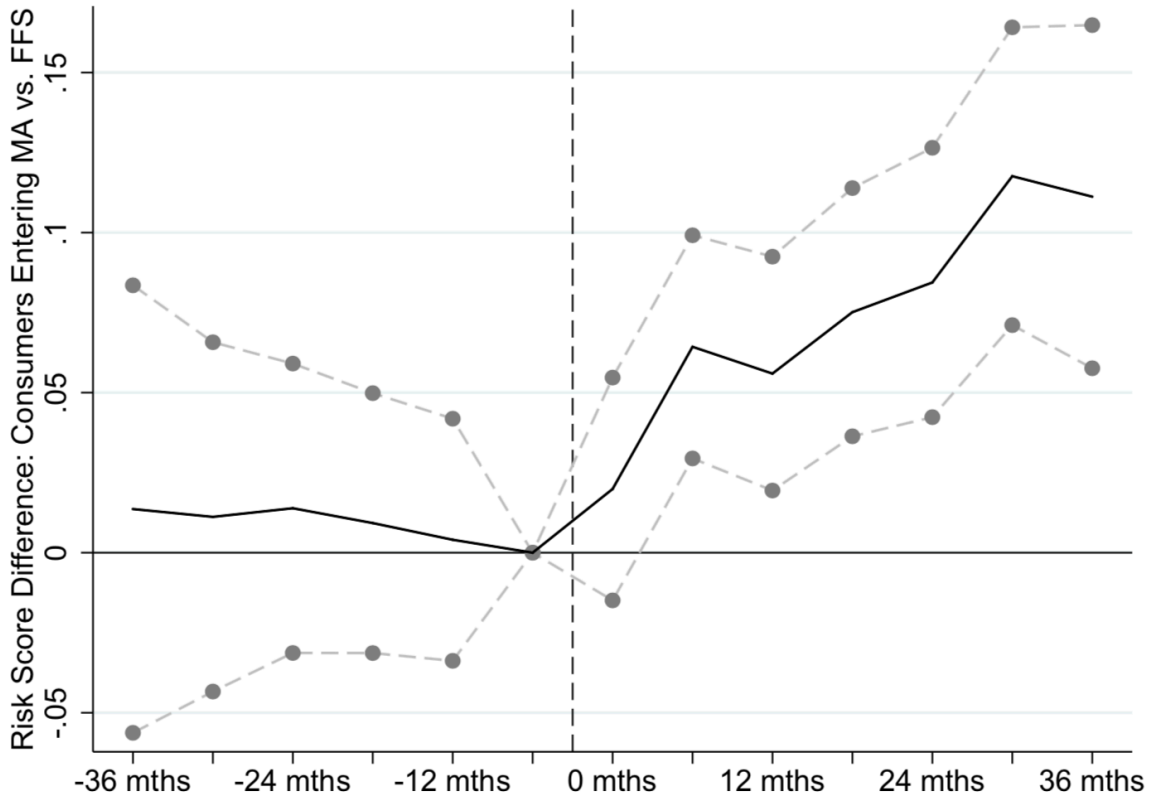
Risk-adjustment payments are based on the diagnosed conditions that are recorded in a patient's healthcare claims. For example, a diagnosis of *diabetes with acute complications* in a patient's recent medical claims history would increase in the typical capitated payment to the insurer by about $3,400 per year.

An implicit assumption in this type of risk-adjusted contracting arrangement between the federal government and private insurance carriers is that diagnoses are fixed, observable properties of the enrollee. In this paper, we challenge the assumption that diagnosis coding is objective and show how patients' reported diagnoses tend to vary with the plan in which they are enrolled.

Methods: The key challenge in identifying coding intensity differences between FFS and MA, or within the MA market segment across competing insurers, is that upcoding is difficult to disentangle from adverse selection. An insurer might report an enrollee population as having higher-than-average risk scores either because the consumers who choose the insurer's plan are in worse health (selection) or because for the same individuals, the plan's benefit design and coding practices result in higher risk scores (upcoding). We develop several approaches to separate coding intensity differences from selection or sorting. In our most transparent exercise, we track risk scores of individual consumers as they transition from an employer or commercial health plan at age 64 to Medicare at age 65.

Results: The figure tracks risk scores of Massachusetts consumers as they age into Medicare at 65. The plot compares the risk scores of two groups: Those that will eventually choose MA and those that will eventually choose TM. The solid line plots the difference in risk scores between the two groups over time, including in the 36 months before they actually make the MA/TM choice. The horizontal axis is time relative to Medicare enrollment at age 65.

The figure shows that during the years prior to Medicare enrollment when both groups were enrolled in similar employer and commercial plans, level differences in



coding intensity were stable. Following Medicare enrollment, however, the difference in coding intensity between the MA and FFS groups spiked upward, providing transparent visual evidence of a coding intensity effect of MA. By the end of the third year of MA enrollment, MA and TM risk scores have diverged by 10% or more for the average enrollee.

By the end of the third year in MA, an enrollee's reported risk score will be more than 10% higher than it would have been in TM.

In complementary findings, we use a national sample and find 6.4% MA/TM coding difference in the first year, which ratchets up to 8.7% by the second year of MA enrollment. To put these numbers in some perspective, a 6.4% risk score difference in the population overall could be generated by 6% of all consumers in a market becoming paraplegic, 11% developing Parkinson's disease, or 39% becoming diabetic. While these effects would be implausibly large if they reflected actual changes in population health, they are plausible when viewed as reflecting only differential coding behavior.

In other findings, we show that coding intensity varies significantly according to the contractual relationship between the physician and the insurer. Fully vertically integrated (i.e., provider owned) plans generate 16% higher risk scores for the same patients compared to FFS, nearly triple the effect of non-integrated plans. This suggests that the cost of aligning physician incentives with insurer objectives may be significantly lower in vertically integrated firms.

Implications: Our findings have specific implications for the Medicare program as well as broader implications for the regulation of private insurance markets. Medicare is the costliest public health insurance program in the world and makes up a significant fraction of US government spending. The size of the overpayment due to manipulable coding is striking. Absent a coding correction, our estimates for our study period imply excess payments of around $10.2 billion to Medicare Advantage plans annually, or about $650 per MA enrollee per year.

The MA coding difference results in billions of dollars in overpayments to MA plans.

In fact, in 2010, toward the end of our study period, CMS began deflating MA risk payments due to concerns about upcoding, partially counteracting these overpayments. Today, CMS deflates scores by 5.9%. Given our findings that coding differentials exceed 10% in the long run, this is too small an adjustment.

Recommendations: One simple-to-implement improvement over the current system would be to apply one coding deflation factor to plan-reported diagnoses and but no deflation factor to the demographic components of the risk score that originate from administrative data, such as age, sex, and disability status. Our results show that the demographic components are not manipulable. By applying zero deflation to the demographic portion and a higher deflation to the diagnoses portion, the deflation would better target the coding activity of concern.

Another simple improvement would be for risk scores to be based on a several-year look-back period for diagnoses, rather than the current 1-year window. In current practice, only diagnoses from a one-year look-back enter the risk adjustment formula. Changing this could decrease the

differential in consistent reporting of valid diagnoses that seems to currently exist between MA and TM, and which contributes to the overall coding difference.

Finally, a more complicated but important potential reform applies to the audit process. Currently, CMS audits submissions of diagnoses from MA plans only for the purpose of determining whether diagnoses were legally submitted. Given that much of the upcoding we document is likely to be legal rather than fraudulent, audits that focus instead on the question of whether a given diagnosis would have been submitted under FFS could be helpful in assessing the proper deflation factors to combat overpayments. This could be done, for example, by assessing which diagnoses are established by MA plans solely via chart review (an activity that would not occur in FFS) and not captured in any claim. Such audits could help reduce the excess payments in MA, even if they would not address the marginal incentives to overcommit resources to coding.