

Making Medicare Advantage a Middle-Class Program

Jacob Glazer
Tel Aviv University
Boston University

Thomas G. McGuire
Harvard Medical School

December 3, 2011

Abstract: This paper studies the role of Medicare's premium policy in sorting beneficiaries between traditional Medicare (TM) and managed care plans in the Medicare Advantage (MA) program. Beneficiaries vary in their demand for care. TM fully accommodates demand but creates a moral hazard inefficiency. MA rations care but disregards some elements of the demand. We describe an efficient assignment of beneficiaries to these two options, and argue that efficiency requires an MA program oriented to serve the large middle part of the distribution of demand: the “middle class.” Current Medicare policy of a “single premium” for MA plans cannot achieve efficient assignment. We characterize the demand-based premium policy that can implement the efficient assignment of enrollees to plans. If only a single premium is feasible, the second-best policy involves too many of the low-demand individuals in MA and a too low level of services relative to the first best. We identify approaches to using premium policy to revitalize MA and improve the efficiency of Medicare.

Acknowledgements: The authors are grateful to the National Institute of Aging for support through P01 AG032952, The Role of Private Plans in Medicare, J. Newhouse, PI. This paper grew out of discussions with Rhema Vaithianathan. Chris Afendulis, Kate Baicker, Mike Chernew, Richard Frank, Nolan Miller, Joe Newhouse, Steve Pizer, Zirui Song, and Alan Zaslavsky provided helpful comments on an earlier draft. Guest editor David Dranove and two reviewers redirected our initial analysis into more fruitful directions. We thank members of the BU/Harvard/MIT Health Economics Workshop and seminar participants at Duke University and the University of Pennsylvania for helpful discussion. The opinions and conclusions in this paper are the authors' alone.

1. Introduction

Beneficiaries in Medicare, the federal health insurance program for the elderly and disabled, have for some time chosen between two major options: traditional Medicare (TM) and a set of private health insurance plans, including managed care plans, offered under Medicare Part C. Presently, only about 25% of beneficiaries elect a Part C plan. The objectives of Part C, since 2003 known as Medicare Advantage (MA),¹ are to expand health insurance options for beneficiaries while taking advantage of economies of managed care to save money for Medicare. Achieving these dual objectives requires that Medicare pay an MA plan less than what beneficiaries would cost Medicare in TM but more than cost for the beneficiaries in MA, leaving some savings to share with beneficiaries in the form of lower premiums/better coverage in MA to attract them to an MA plan. Research and policy have focused on the “risk adjustment” of Medicare payments to pay more for the sick and less for the healthy joining the MA plans.² In spite of improvements in risk adjustment technology and many other policy reforms, Part C has yet, however, to save Medicare money (McGuire, Newhouse and Sinaiko, 2011). As part of the Affordable Care Act (ACA), Medicare payments to MA plans are being cut, and, based on experience, plan and beneficiary exit will follow;³ but, based on the same experience, these cuts are unlikely to move the MA program into the black for Medicare.⁴

We argue that a major contributor to the chronically poor performance of Part C is the inefficient sorting of beneficiaries between MA and TM caused by Medicare’s premium policy. In this paper, we shift analytic focus away from plan payment and risk adjustment and to the premiums and the incentives faced by heterogeneous beneficiaries when they elect MA or TM. To address normative questions around beneficiary choice between MA and TM, we propose a formulation of which beneficiaries *should* be in MA

¹ The Medicare Modernization and Improvement Act of 2003 renamed Part C plans as Medicare Advantage plans. Previously, plans were called Medicare+Choice plans.

² For reviews of the literature on risk adjustment, see Newhouse (2002) or Van de Ven and Ellis (2000). Broader themes regarding risk and variation are covered in Breyer, Bundorf and Pauly (2012).

³ For recent analysis of payment changes from the ACA and plan response showing up in anticipation of these changes, see Afendulis et al. (2011).

⁴ In addition to refining risk adjustment, Medicare has raised and lowered the overall level of payment to MA plans a number of times in the past 25 years depending on whether increasing access to MA plans or saving money were the dominant policy concern. McGuire, Newhouse and Sinaiko (2011) describe these changes and the plan and beneficiary response.

and in TM. Somewhat surprisingly, in light of the policy attention to Medicare and the Part C program, the Medicare policy literature says little about what socially efficient sorting looks like. The fundamental answer to who should be in MA underlies the title of the paper. We argue that MA should, from the standpoint of social welfare, draw beneficiaries from the thick central part of the distribution of preferences for health care, and in this sense make MA a “middle class” program.

In a simple model of beneficiary choice, we show that the present premium policy leads to a set of beneficiaries in MA that is far off from those who should be in MA. Medicare (and social welfare) would be better served by a larger, better-financed MA program, targeted to the middle class. To get there requires a redesign of premium policy.

Our approach is based on the observation that beneficiaries vary in their demand for health care for many reasons. Most attention has been directed to the heterogeneity related to “health status” and the risk adjustment technology designed to deal with it (Pope et al., 2004). Health status-based risk adjustment explains a small share (10% or less) of the individual variation in health care spending, partly because health status is difficult to measure and predict. Another reason, and one that we call attention to here, is that factors other than health status – income, education, “taste” more generally⁵ – also influence demand for health care, and therefore choice of plan.

From the standpoint of the beneficiary, anticipated demand for health care, together with the premiums for TM and MA, determine the best plan option. Premiums for TM are described below, and depend on the circumstances of the beneficiary. MA plans choose the premium beneficiaries pay (subject to Medicare regulation), and this premium is the same for all beneficiaries. There is a fundamental problem with this approach. *Any* single premium for MA does a poor job of sorting beneficiaries between

⁵ Income effects, for example, have been studied in health care. Cross-sectional studies generally report a positive income elasticity of demand that is less than one. The Rand Health Insurance Experiment found income elasticities of between .1 and .2 (Newhouse, 1993). Studies using longitudinal variation in income find much larger elasticities, generally classifying health care as a “luxury good” with income elasticities exceeding 1.0. See Fogel (2008). Borger, Rutherford and Won (2008) reviewed over twenty papers and settled on a unit income elasticity to use in their simulation model. In a recent study, Acemoglu et al. (2009) use oil price shocks to estimate the income effect on demand for health care at .7. Those with higher income tend to be in better health, so adequate controls for health status are necessary to identify income effects. The same is true for education. The better-educated tend to be healthier, but once health status is controlled for, education increases health care demand.

MA and TM. We argue that some form of premium discrimination by non-health status factors affecting demand is necessary to rescue MA from its chronically poor performance. After our analysis, we discuss some ways to change premium policy in TM as well as MA to better achieve Medicare objectives and economic efficiency.

2. Traditional Medicare and Medicare Advantage

2.1 Program Descriptions

At age 65, most Americans become eligible for Medicare.⁶ If beneficiaries do not elect an MA plan, they are automatically enrolled in Part A of Medicare at no cost to them.⁷ Part A is financed largely by a payroll tax shared by employees and employers (Kaiser Family Foundation, (KFF) 2008). Part A covers inpatient hospital services, some post-hospital stays in nursing facilities and home health care, and hospice care, but requires considerable beneficiary cost sharing. The most significant beneficiary cost sharing is the deductible (\$1,132 for 2011) per hospital episode (“benefit period”). Beneficiaries may also enroll in Part B, which covers doctors’ visits, other ambulatory services and some drugs administered in physician offices. The standard Part B premium for 2011, which applies to new enrollees, is \$115.40 per month. Many enrollees pay only \$96.40 per month because of hold-harmless provisions applying to social security payments from which the Part B premium is deducted (CMS, 2010). Very few (3%) of beneficiaries pay a higher premium because of high individual or family income, and some (17%) have premiums all or partly covered by Medicaid.⁸ Part B premiums cover only about 25 percent of Medicare's cost of Part B, the balance being paid for by general revenues (Kaiser Family Foundation, 2008). The vast majority of beneficiaries in TM enroll in Part B. Beneficiary cost sharing in Part B includes an annual deductible of \$162 in 2011 and a 20 percent coinsurance on Medicare allowed charges. Since 2006, beneficiaries may also join a Part D plan covering prescription drug costs. Part D plans receive about 75 percent of their federal revenue from general revenues, are offered by

⁶ Medicare also provides health insurance for qualified disabled beneficiaries below age 65. These beneficiaries may also choose to join the same MA plans on the same terms as the elderly beneficiaries.

⁷ Beneficiaries with short Medicare work histories and not married to a beneficiary with a long work history may pay a Part A premium but this applies to about 1% of beneficiaries CMS (2011).

⁸ In addition, there are penalties for delaying enrollment in Part B in the form of higher premiums (CMS, 2010).

private insurers, and vary in coverage. Part D premiums are set by a bidding procedure, the average premium for a stand-alone drug plan (taken as part of TM) was \$38/month in 2011 (MedPAC, 2011b, Section 10). Low-income beneficiaries receive a premium subsidy (and lower cost sharing). Beneficiaries with Part A, and the optional Parts B and D, are considered to be in “traditional Medicare.”

Most beneficiaries in TM avoid cost sharing in Parts A and B with medigap or some other supplemental coverage. Medicaid pays cost sharing for eligible low-income beneficiaries. Some employers buy wrap-around coverage for retirees. Finally, most beneficiaries in neither of these groups buy medigap policies to cover some or all of the cost sharing. The average monthly premium for the most popular medigap policy (Plan F) was \$167 in 2009 (MedPAC, 2011a). Given the pervasiveness of supplemental coverage, we think of TM as traditional health insurance with low cost sharing, with a premium for beneficiaries equal to the Part B premium plus what they pay for Part D and supplemental coverage.

Virtually all hospitals and practicing physicians accept Medicare payment, giving beneficiaries wide choice of providers. Medicare and its regional intermediaries make broad coverage decisions but do not interfere with (aka “manage”) physician and patient choice of treatment. Health care in TM has been criticized as being uncoordinated and costly (Newhouse, 2002). TM contends with cost issues primarily by using its monopsony power to pay physicians and hospitals roughly 20-30 percent less than private plans on average (MedPAC, 2008).⁹

The Medicare Modernization Act of 2003 (MMA) created MA to replace the short-lived Medicare+Choice version of Part C. MA plans are private, must cover all Part A and B benefits, and may supplement these benefits by reduced cost sharing or coverage for additional services not part of TM, such as vision or dental care (Gold, 2008).¹⁰ MA plans may or may not include drug coverage. Those that do are referred to

⁹ In traditional Medicare, physicians are paid for each procedure according to a fee schedule. Hospitals are paid according to the diagnosis-related group (DRG) in which a patient is classified at discharge. The hospital payment system is partly “prospective,” embodies some incentives to the hospital to economize on resources during the hospital stay. For an overview of Medicare payment policies applying to physicians, hospitals and health plans, See Newhouse (2002).

¹⁰ Gold and her colleagues at Mathematica Policy Research have tracked policy, enrollment, plan types and other data on Part C for a number of years in a useful series of publications (Gold, 2009 and Gold et al., 2004).

as MA-PD (i.e., “Prescription Drug”) plans. In total, 11 million beneficiaries, or 24% of all, were enrolled in an MA plan in 2010 (KFF, 2010a).

The MMA created new plan types within MA and the higher payments mandated in the legislation awakened dormant plan types established earlier. We distinguish between what we consider to be bona fide managed care plans and others. We regard Health Maintenance Organizations (HMOs), the oldest and largest plan type, and the mostly tightly managed, along with Local Preferred Provider (PPOs) and the small number of Provider-Sponsored Organizations (PSOs) as managed care. The other plan types, notably the Regional PPOs and Private Fee-for-Service (PFFS) plans, are not in this category.¹¹ Our model of MA plan behavior laid out below applies to managed care plans only; 78 percent of Medicare beneficiaries in MA were in bona fide managed care plans in 2010.¹² Coverage in MA managed care plans is more comprehensive than in TM, with smaller (or no) deductibles, less cost sharing, and some coverage of additional services. Beneficiaries in MA plans cannot purchase medigap coverage.

Beneficiaries in MA plans must pay the Part B premium (which goes to Medicare).¹³ MA plans can charge a premium above this. Many plans, however, are “zero premium” plans, meaning they charge nothing above the Part B premium. A small percentage of plans reduce the beneficiary Part B premium obligation. More than half of beneficiaries in HMO plans pay zero premium. Most beneficiaries are in MA-PD plans. The average monthly premium for an MA-PD plan (across all plan types) was \$51 per month, unweighted by enrollment (KFF (2010b)).

Medicare payments to plans are based on a Medicare “benchmark” rate set for each county and the plan's “bid.” The benchmark is based on the maximum of the CMS estimate of Medicare costs for a typical beneficiary in Parts A and B, a minimum or “floor” payment, and the past payment rate for the county trended forward at national average Medicare cost growth rates. In the past several years, benchmarks have generally exceeded expected FFS costs in a county (GAO, 2011). For 2011, benchmarks were frozen at 2010 levels. Beginning in 2012, the ACA phases in a new formula by

¹¹ Part C also includes Special Needs Plans (SNPs) intended for beneficiaries in long-term care and with certain chronic diseases.

¹² Government Accountability Office (GAO) (2011), p.14.

¹³ Thus, the higher Part B premium charged to a small number of higher-income beneficiaries is like a tax, and does not affect the relative cost facing beneficiaries in MA and TM.

which the benchmark is a percentage of the estimated FFS spending based on a county's level of spending relative to other counties. For counties in the lowest spending quartile the new formula pays 115% of the benchmark, and this ranges down to 95% of the benchmark for the highest quartile counties (GAO, 2011). In most counties, these changes are expected to reduce Medicare payments (Afendulis et al., 2011).

A plan's bid is supposed to be the plan's estimate of what regular benefits from Part A and B would cost the plan for a beneficiary of average health status. If a plan bids above the benchmark, the plan is paid the benchmark. If the bid falls below the benchmark, Medicare pays the bid plus 75% of the difference between the bid and the benchmark. The idea is that Medicare gets 25% of any "savings" (difference between benchmark and bid) and the plan must use the balance of the difference to reduce beneficiary premiums or provide additional coverage and benefits. In 2010, the average plan bid was 98% of the benchmark, but HMO forms were the only plan type with an average below the benchmark (94%) (GAO, 2011). All others were greater than 100% of the benchmark. As a final step in figuring plan payments, Medicare then applies the CMS-HCC risk adjustment formula reflecting the measured health status of actual enrollees.

Figure 1 summarizes the typical 2011 monthly premium costs for a Medicare beneficiary in TM and MA. The beneficiary would pay the Part B premium to Medicare in either case. She buys a Part D free-standing drug plan for \$38/month and a medigap plan for \$167. If the beneficiary were to join a typical "zero-premium MA-PD plan, she pays nothing per month over Part B.

2.2 Previous Research

A large literature in public finance is concerned with the public provision of private goods. Health care is a private good financed collectively for reasons, among others, related to equity. In Besley and Coate (1991), rich and poor consumers demand a good that could be provided by government with an equal consumption requirement, or left to the market.¹⁴ Some redistribution to the poor can be achieved by providing a base level to everyone financed collectively by progressive taxes, and letting the rich opt out

¹⁴ Government provision can give a private good the "equal consumption" characteristic of a public good, as is the case in government provision of health insurance, education, and other services.

to the private system. Within a health care context, researchers study resource allocation in public systems, and how a private sector fringe affects public costs and efficiency (Barros and Siciliani, 2012). Grassi and Ma (2008) analyze the interaction between a budget-constrained public health care system that allocates by rationing and a private sector that accommodates demand but may set prices. Consumers differ according to wealth and cost. The analogy to our paper is that our MA plan has an equal consumption feature, and the consumer has the option of choosing a TM plan that accommodates heterogeneity. In our analysis of Medicare, however, the alternative system, TM, is also publicly financed and subject to regulation.

A focus on the premiums paid by beneficiaries recalls Rothschild and Stiglitz (1976) who recognized the premium as a powerful selection device. In a Medicare context, the emphasis in the literature has been on skimping on services as a tool for selection.¹⁵ Other papers stress that plan choice of benefits and premiums can serve beneficiary heterogeneity, as well as being a device for selection.¹⁶ The idea, as expressed by Pizer, Frakt and Feldman (2003), is that “Beneficiaries who highly value certain benefits can search for a plan that offers those benefits and pay the marginal premium that corresponds to their choice.” One can imagine Medicare risk-adjusted plan payments as a voucher, with plan-set premiums sorting beneficiaries by tastes for additional services.¹⁷ However, as shown in Bundorf et al. (2009) and Glazer and McGuire (2011), except in special circumstances, a single premium for each plan does not sort beneficiaries efficiently among plans.¹⁸

¹⁵ The early literature made overall cost comparisons between managed care and traditional Medicare. The more recent literature on service-level selection studies categories of expenditures. See Ellis and McGuire (2007) for review of some papers.

¹⁶ Another use of the premium in the context of private health insurance is for an employer to recover some of the inframarginal surplus health insurance benefits confer on high-demand employees (Miller, 2005).

¹⁷ Restructuring the “choice architecture” in Medicare – limiting and clarifying the options – may be necessary for choice to serve beneficiaries. See McWilliams et al (2011) for recent evidence for how too many choices may lead to ineffective decisions, at least for the substantial share of the elderly who have some impairment in decision making.

¹⁸ Town and Liu (2003), following methods proposed by Berry (1994), use market share and premium (price) data to estimate consumer surplus and profits associated with Part C plans during the 1990s, when Part C rode the success of managed care throughout the US health insurance market, and plan and beneficiary participation was growing rapidly. They found that demand for HMO plans was inelastic to the premium, implying ample surplus to be divided between consumers and HMO plans. Beneficiaries had diverse tastes, but costs were assumed to be uniform, ruling out any selection issues by assumption.

Bundorf et al. (2009) show, in their theoretical analysis, that any uniform price for health plans will generally not lead patients to efficiently sort themselves across plans. They also estimate the efficiency loss due to this inefficient pricing in the context of a small group market in California. Glazer and McGuire (2011) model managed competition with demand heterogeneity to consider plan payment and enrollee premium policies in relation to efficiency and fairness. Specifically, this paper studies how to implement a “Silver” and “Gold” health plan efficiently and fairly in a managed competition context. The assumption shared with these two papers and the current one is that demand for health services and health insurance is driven by both tastes and need. The current paper introduces demand heterogeneity in the context of Medicare, with beneficiaries faced with a choice between TM and MA. This specific application allows us to demonstrate the fundamental inefficiencies in current Medicare policy, and to identify new approaches to address these inefficiencies.

3. Utility and Willingness to Pay for MA and TM

Beneficiaries are heterogeneous in their demand for health care. A beneficiary’s valuation of the quantity of health care, x , is positively related to a preference parameter, θ . Specifically, we assume that a beneficiary characterized by θ values health care at $V(x, \theta)$, with $V_x > 0$, $V_{xx} < 0$, $V_{xxx} \leq 0$, $V_\theta > 0$, $V_{x\theta} > 0$, and as a normalization, we assume $V_{\theta\theta} = 0$.¹⁹ We also assume that $\theta \in [\underline{\theta}, \bar{\theta}]$, $F(\theta)$ is the distribution function and $f(\theta)$ is the density function of θ . We assume that $f(\theta)$ has full support over the interval $[\underline{\theta}, \bar{\theta}]$. Quantity x is measured in dollars so the cost of x is 1. Efficient x for beneficiary θ is $x^*(\theta)$, the solution to $V_x(x^*, \theta) = 1$.

In order to focus on heterogeneity in demand from sources other than health status, we interpret θ as representing elements affecting demand for health care that the MA plan does not or cannot accommodate. Such elements could include income, education, and other dimensions affecting taste, so long as these elements are not recognized and responded to by the plan. In our model, individuals differ only in θ ,

¹⁹ Most of our assumptions about the shape of $V(x, \theta)$ are intuitive and conventional. The assumption that $V_{xxx} \leq 0$ needs some discussion and this will be carried out after our presentation of Proposition 1, where it comes into play.

implying that demand varies only because of taste parameters. We make this assumption because we regard the clinicians in the managed care MA plan as trained and interested in gearing treatment to health status (more resources go to a patient with greater need), but not in accommodating differences in “taste” for health care. The “equal consumption” assumption about MA plans refers to heterogeneity with respect to taste.

The economic literature on managed care is concerned with rationing in the presence of heterogeneity in demand, which is most often associated with health status differences.²⁰ Rationing can avoid moral hazard problems and improve efficiency, but an equal consumption feature of rationing creates inefficiency by homogenizing services to those with heterogeneous preferences. Essentially, this paper is about how to use the efficiency gains MA plans can create to best advantage among Medicare beneficiaries with heterogeneous preferences for health care. We now describe more specifically our models of MA and TM.

In MA there are no copayments. Since (by assumption) individuals differ only on dimensions not accommodated by the MA plan, quantity in MA is the same for everyone in the plan at x^{MA} . Willingness to pay (or utility measured in dollars) for MA is simply:

$$U^{\text{MA}}(x^{\text{MA}}, \theta) = V(x^{\text{MA}}, \theta) \quad (1)$$

U^{MA} is positive for every value of θ and, given our normalization on $V_{\theta\theta}$, for a given x^{MA} , it increases with θ at a constant rate.

We assume, in contrast to MA, TM does accommodate beneficiary demand due to heterogeneity in θ . In TM, beneficiaries pay a copayment $0 < c < 1$ for each unit of x , and choose quantity to maximize utility by setting the marginal valuation of x equal to c :

$$V_x(x, \theta) = c \quad (2)$$

We define $x^{\text{TM}}(\theta)$ to be the solution to (2) above. Note that our assumption that $V_{x\theta} > 0$ implies that consumers with higher θ demand more care in TM. Willingness to pay for TM is then:

$$U^{\text{TM}}(\theta) = V(x^{\text{TM}}(\theta), \theta) - cx^{\text{TM}}(\theta) \quad (3)$$

²⁰ Papers that characterize rationing with a “shadow price” emphasize the responsiveness of rationing to variation in health status. See Keeler, Carter and Newhouse (1998), Frank, Glazer and McGuire (2000). Papers that characterize rationing as setting quantity emphasize that patients with different demands end up with the same use. See Baumgardner (1991), Pauly and Ramsey (1999).

This is also everywhere positive, increasing in θ at an increasing rate.²¹ Intuitively, this is because as demand shifts out, quantity increases (unlike in MA) and consumer surplus goes up approximately quadratically with quantity demanded.

Throughout the analysis we will focus on the case where x^{MA} is located in the interior; by “interior” we mean that the level of care provided under managed care is higher than the level of care consumed by the lowest type, $\underline{\theta}$, under TM and lower than the efficient level of care for the highest type, $\bar{\theta}$, individual:

Assumption 1: $x^{\text{TM}}(\underline{\theta}) < x^{\text{MA}} < x^*(\bar{\theta})$

Figure 2 depicts the two functions $U^{\text{TM}}(\theta)$ and $U^{\text{MA}}(x^{\text{MA}}, \theta)$ for an arbitrary level of x^{MA} that satisfies Assumption 1.

Efficiency involves both the level of managed care, x^{MA} , and the sorting of beneficiaries between MA and TM, given the efficient x^{MA} . We conduct our analysis in several steps. In Section 4 we take x^{MA} as given and study how to efficiently divide consumers between MA and TM. In Section 5 we keep the assumption that x^{MA} is fixed and characterize a premium policy to implement the efficient assignment. We show that, for any level of x^{MA} , a single premium cannot implement the efficient assignment whereas a premium policy with two levels of premium that depends on θ , can implement the efficient assignment. In Section 6 we allow the regulator to choose x^{MA} and analyze the first-best and second-best levels of managed care (x^{MA}) depending on whether the regulator can condition premiums on θ (the first-best case) or it can use only a single premium (the second-best case).

4. Efficient Assignment

We consider how to efficiently divide consumers between MA and TM for a given x^{MA} . We establish (Proposition 1) that the optimal division of consumers between MA and TM is characterized by consumers in a single range of θ in MA with the others in TM. We will first present our result formally and then discuss the interpretation. Maximizing

²¹ Dropping the elements inside the parentheses we have: $dU^{\text{TM}}/d\theta = V_\theta + [V_x - c][dx^{\text{TM}}/d\theta] = V_\theta > 0$
 $d^2U^{\text{TM}}/d\theta^2 = V_{\theta\theta} + V_{\theta x}(dx^{\text{TM}}/d\theta) + [V_{x\theta} + V_{xx}(dx^{\text{TM}}/d\theta)](dx^{\text{TM}}/d\theta)$

Since $V_{\theta\theta} = 0$ and $dx^{\text{TM}}/d\theta = -V_{x0}/V_{xx}$, by $V_x = c$, we get $d^2U^{\text{TM}}/d\theta^2 = -(V_{x0})^2/V_{xx} > 0$

social welfare (efficiency) requires dividing beneficiaries into the two insurance types based on social welfare defined as utility less cost.²²

Given an x^{MA} , welfare in MA as a function of θ is given by:

$$W^{MA}(x^{MA}, \theta) = V(x^{MA}, \theta) - x^{MA} \quad (4)$$

And welfare in TM is given by:

$$W^{TM}(\theta) = V(x^{TM}(\theta), \theta) - x^{TM}(\theta) \quad (5)$$

Proposition 1: For every x^{MA} there exists θ_1 and θ_2 with $\theta_1 < \theta_2$ such that $W^{TM}(\theta_i) = W^{MA}(x^{MA}, \theta_i)$ for $i = 1, 2$, $W^{TM}(\theta) < W^{MA}(x^{MA}, \theta)$ if $\theta_1 < \theta < \theta_2$ and $W^{TM}(\theta) > W^{MA}(x^{MA}, \theta)$ otherwise.

Proposition 1 says that the efficient division of beneficiaries can be expressed in terms of two cutoff values θ_1 and θ_2 . All individuals below θ_1 and all individuals above θ_2 should be TM. Beneficiaries between θ_1 and θ_2 should be in MA.²³ (Proofs of all results in this paper are contained in the Appendix.)²⁴

For the person on the boundary between TM and MA, the social net benefits (welfare) in MA and TM are equalized. Figure 3, Panel A shows demand for type θ_1 , defining the lower boundary of those who should be in MA. For this beneficiary, consumption in TM, x^{TM} just equals the given x^{MA} . Obviously, social welfare in MA and TM for this beneficiary must be the same in both MA and TM. Depicted in terms of the

²² Recently, perspectives from both “value-based insurance design” and behavioral economics have questioned the conventional welfare framework based on utility, adding to long-standing doubts about whether consumer choices should be regarded as a guide to welfare. See Chernen, Rosen and Fendrick (2007) for the proposal for the value-based approach, and Pauly and Blavin (2009) for an integration of the Pigouvian subsidy perspective into the conventional moral hazard framework. In a related point connected to behavioral economics, Newhouse (2006) illustrates how time-inconsistent preferences could lead to choices by consumers not in their best interest. As Pauly and Blavin (2009) show, however, consumer over or undervaluation can be integrated with traditional considerations of optimal health insurance. In this paper we take the conventional approach that utility is also the measure of his social welfare.

²³ Notice that the θ_1 and θ_2 cutoffs depend on the x^{MA} chosen. Rather than writing the cutoff as a function of x^{MA} we stipulate that the discussion here applies for a particular value of x^{MA} .

²⁴ As can be seen in the proof of Proposition 1, the fact that there are only two cutoffs follows from the convexity of $\Delta W(x^{MA}, \theta)$, which follows from, along with other assumptions about utility, our assumption that $V_{xxx} \leq 0$. If V_{xxx} was sufficiently greater than 0, there could be some special cases where there would be more than two levels of θ at which welfare in TM was equal to the welfare in MA. We do not discuss these special cases here.

welfare loss against the first best level of care for this beneficiary, the shaded triangle welfare loss must be the same when $x^{MA} = x^{TM}$. For any θ lower than θ_1 , the beneficiary is still “over consuming” in both systems, but since x^{TM} for a θ less than θ_1 would be less than x^{MA} , the welfare loss in TM would be less than in MA. Thus, for any demand to the left of θ_1 , the beneficiary belongs in TM. For θ just greater than θ_1 , the welfare loss in MA is strictly less than the loss in TM. To see this, imagine shifting the demand shown in Panel A slightly to the right. The higher θ type is over consuming, but since $x^{MA} < x^{TM}$, the welfare loss due to over consumption is less in MA.

Panel B of Figure 3 shows demand for beneficiary θ_2 , defining the upper boundary of efficient assignment into MA. For the given x^{MA} , there is a second level of demand in TM for which the welfare loss is the same as it would be if the beneficiary were in MA, the θ for which the two shaded triangles in Panel B, the welfare loss in MA and the welfare loss in TM, are equal. The loss from “under consumption” in MA for type θ_2 just equals the loss from over consumption in TM. As θ increases beyond θ_2 , the loss from over-consumption in TM goes up by less (and, in fact, could even go down) than the increase in the welfare loss in MA. Thus, anyone with $\theta > \theta_2$ belongs in TM. A similar argument applies in the opposite case as θ decreases below θ_2 . These beneficiaries belong in MA.

Figure 4 depicts the cutoffs in terms of welfare and the distribution of demand for care. As shown in Panel A, for a fixed level of x^{MA} welfare in MA increases as a constant rate, whereas welfare in TM is concave. Panel B depicts a plausible distribution of θ among beneficiaries, with the bulk of the population falling in the “middle class” in terms of demand.

5. Implementation of Efficient Sorting

We now consider the premium policy required to implement the optimal division of beneficiaries between MA and TM, for a given x^{MA} .

Beneficiaries sort themselves between MA and TM based on their willingness to pay for each alternative and the relative premiums. The difference in willingness to pay for MA and TM, which we refer to as ΔU , is a function of θ :

$$\Delta U(x^{\text{MA}}, \theta) = V(x^{\text{MA}}, \theta) - (V(x^{\text{TM}}(\theta), \theta) - cx^{\text{TM}}(\theta)) \quad (6)$$

Beneficiaries compare this difference to the difference in premiums between the two options and choose the option with the highest net benefit.

The following Lemma regarding ΔU is central to our results about premium policies and beneficiary sorting.

Lemma 1: The beneficiary whose utility in MA exceeds TM by the *maximum* amount is θ_1 (as given by Proposition 1), the *marginal* beneficiary defining the lower bound of the types who are more efficiently served in MA.

In other words, the beneficiary for whom *social welfare is equal* in MA and TM defining the low end of the θ distribution belonging in MA is the beneficiary for whom the *difference in utility (before any premiums) is the maximum* between the two options. The advantage of TM in comparison to MA is the beneficiary's ability to choose quantity. For one type of beneficiary, that defining θ_1 , this advantage counts for nothing – the beneficiary would choose x^{MA} in any case. MA absolves the beneficiary of cost sharing, and thus for this beneficiary the utility advantage of MA is cx^{MA} . The point of Lemma 1 is that this is as good as it gets: cx^{MA} is the maximum any beneficiary gains from being in MA. Figure 5 shows why (with proof in Appendix). For a beneficiary with $\theta < \theta_1$, the gain from MA is less since some of the services are valued at less than c . For a beneficiary with $\theta > \theta_1$, there is a cost sharing gain of cx^{MA} but then a deduction because of the services lost with a value greater than c to the beneficiary. Thus, as θ moves away from θ_1 in either direction, the gain from MA is falling.

The unusual relationship between the private and social rankings of the alternatives poses a challenge for premium policy and sorting. Choice is based on utility and efficiency is governed by social welfare. The challenge is this: of the beneficiaries with the same net private utility in MA (those just to the left and right of θ_1), some (those to the left of θ_1) should be directed into TM, and some others (those to the right of θ_1) should be directed into MA. As we will show, no single premium can achieve this desired sorting. It should be emphasized that the result presented in Lemma 1 is quite general as it follows only from our assumptions about the shape of the individual's utility

function, $V(x, \theta)$, which are conventional, and on the way the level of service is determined in MA and in TM.

By Lemma 1 we know that for all $\theta > \theta_1$, $U^{\text{TM}}(\theta)$ increases at a rate faster than $U^{\text{MA}}(x^{\text{MA}}, \theta)$ and, hence, there exists a $\theta' > \theta_1$ such that U^{TM} is equal to U^{MA} at θ' . In general, θ' can be either smaller or larger than θ_2 depending on the level of the copayment c . In our analysis, hereafter, we focus on the relevant case where c is not too large and hence, U^{TM} intersects U^{MA} at a θ less than θ_2 . This situation is depicted in Figure 6 if there were no premiums for either alternative. The bolded (red) regions of the θ -axis flag the discrepancy between the beneficiaries who should be in MA and those who via utility maximization choose MA.

We consider now how the ranges of sorting can be affected by premium policy. Let $y(\theta)$ denote the premium paid by an individual of type θ if she joins TM (without loss of generality we normalize the premium in MA to zero). We allow for $y(\theta)$ to be either positive or negative. Premiums paid to TM are like a tax and do not affect services received by a beneficiary in TM. With a premium the net utility of type θ joining the MA plan is, therefore, $\Delta U(x^{\text{MA}}, \theta) + y(\theta)$.

A premium schedule $y(\theta)$ *implements the efficient sorting (assignment)* if, given x^{MA} , $\Delta U(x^{\text{MA}}, \theta) + y(\theta) \geq 0$, for all θ , $\theta_1 \leq \theta \leq \theta_2$, and $\Delta U(x^{\text{MA}}, \theta) + y(\theta) < 0$ otherwise. That is, a premium schedule implements the efficient sorting if, given that schedule, all individuals of type θ , $\theta_1 \leq \theta \leq \theta_2$ prefer MA over TM, and all other individuals prefer TM over MA.

5.1 Failure of Any Single Premium

The following proposition follows directly from Lemma 1:

Proposition 2: A single premium cannot implement the efficient sorting.

The proof is based on the following logic. According to Lemma 1, the premium beneficiary θ_1 pays for TM should make him indifferent between TM and MA so that beneficiaries with lower values of θ will not be willing to pay this premium and (efficiently) stay in TM. However, given Lemma 1, if type θ_1 is just indifferent between

MA and TM, then all other beneficiaries strictly prefer TM over MA and they will all choose TM, which is obviously not what the first best calls for.²⁵

With the premium for MA set at zero, as we raise the single premium for TM, the first beneficiary willing to join MA is at θ_1 , but then, as the premium is raised further, MA attracts beneficiaries on both sides of θ_1 ; those on the right who we want in MA, and those on the left we do not. Lemma 1 identifies the severe handicap constraining the MA program based on a single premium policy. Any (single) premium attracting enrollment to MA brings in beneficiaries who would be more efficiently served in TM.

Figure 7 shows the effect of a positive single premium for TM, shifting upward the net utility in MA. The positive premium moves some beneficiaries who should be in MA from TM (those just to the right of the θ' in Figure 6), initially improving efficiency. Beneficiaries to the left of θ_1 continue to make the wrong choice. Eventually, further increases in the single premium for TM will lead to inefficient choices in the range above θ_2 as well. There is no way to manipulate the single premium (positive or negative) to attain the efficient sorting.

5.2 Implementing Efficient Sorting Requires Two Premiums

From Proposition 2 we know that at least two premium levels are necessary to implement the efficient assignment. In Proposition 3 we describe how premiums need to depend on θ to lead to an efficient division between MA and TM. The premium policy uses two premium levels y_1 and y_2 , where $y_1(y_2)$ is the level of premium that would make type $\theta_1(\theta_2)$ just indifferent between MA and TM. The premium $y_1(y_2)$ is paid by all individuals of type $\theta < \theta_1$ ($\theta \geq \theta_1$) if they choose to be in TM.

Proposition 3: Let x^{MA} be given, let $y_1 = -cx^{\text{MA}}$, and let

$$y_2 = V(\theta_2, x^{\text{TM}}(\theta_2)) - V(\theta_2, x^{\text{MA}}) - cx^{\text{TM}}(\theta_2) \quad (7)$$

The premium policy: $y(\theta) = y_1$ if $\theta < \theta_1$, and y_2 if $\theta \geq \theta_1$, implements the efficient sorting.

²⁵ If the density function of θ did not have a full support over the interval $[\underline{\theta}, \bar{\theta}]$, then there could be a case where only the upper cutoff, θ_2 , was in the support of $f(\theta)$ and, in such a case, a single premium could implement the efficient sorting. This case is very special and we do not discuss it.

Figure 8 illustrates the two-level premium policy. (See Appendix for a formal proof.) Beneficiaries below θ_1 would pay the negative premium $y_1 = -cx^{\text{MA}}$ (be subsidized) to join TM; this subsidy is large enough to keep all beneficiaries in this range in TM. Everyone else pays a positive premium for TM, calculated so as to make the beneficiary at the upper end of the efficient range just indifferent between the two options. This premium, y_2 , is high enough to keep all the beneficiaries between θ_1 and θ_2 in MA, but is not enough for the very high-demand types, those with $\theta < \theta_2$, and they go to TM (where they belong).

There are obviously other premium policies that would also sort beneficiaries efficiently (for example, the subsidy for TM need not be equal to y_1 for all the beneficiaries below θ_1 , but the (y_1, y_2) -policy we describe is the simplest – the only two-premium policy that does the job.

5.3 Second-Best Assignment Using a Single Premium

The two-premium policy described in Proposition 3 relies on the regulator using θ as a basis for premiums. Such a policy may not be feasible as the factors behind θ , e.g., education, attitudes towards health and health care, are not easily observable by the regulator and even if observable, cannot be used as a basis for a premium policy. If θ cannot be used to set premiums, the regulator must pick just one premium to sort beneficiaries, and we know from Proposition 2 that the socially efficient sorting cannot be implemented. The question then becomes what is the optimal (second-best) single premium to sort beneficiaries between MA and TM?

When the regulator chooses a single premium for TM, y , and the level of care is x^{MA} , two distinct possibilities arise. The first is when the premium is positive and, hence, there exists a single cutoff θ' such that all individuals with $\theta < \theta'$ choose MA and all individuals with $\theta \geq \theta'$ prefer TM. The second is when the premium for TM is sufficiently negative, so that there are two cutoffs θ_1' and θ_2' with $\underline{\theta} \leq \theta_1' < \theta_2' \leq \bar{\theta}$, such that all individuals with $\theta_1' \leq \theta \leq \theta_2'$ choose MA and all other individuals choose TM. In the first case the cutoff θ' is given by $-\Delta U(x^{\text{MA}}, \theta') = y$. In the second case the two cutoffs are given by $-\Delta U(x^{\text{MA}}, \theta_i') = y$, for $i = 1, 2$.

The first case is of more practical significance, since Medicare's single premium policy leads to a positive premium for TM, when the medigap and Part D premiums are

factored into the picture. This is the case we select for analysis. How should consumers be divided between MA and TM in this case? We characterize the second-best assignment in Proposition 4.

Proposition 4: Assume that x^{MA} is given and let θ' denote the second-best efficient cutoff given that a positive premium is implemented for TM (i.e., $y > 0$). Then, $\theta' = \theta_2$ where θ_2 is the first-best cutoff given by Proposition 1.

The second-best policy must accept that the low-demand beneficiaries (with the lowest willingness to pay for TM) will be in MA. With x^{MA} given, the tradeoff on the upper end of the distribution of demand is unaltered from the first-best case, the same cutoff, θ_2 , will be the target of the TM premium. Thus, the single-premium (given an x^{MA}) is y_2 from Proposition 3.

6. Efficient Level of Care

Now suppose that Medicare can set the level of x^{MA} by altering the level of its payments to plans.²⁶ How should x^{MA} be set if Medicare's objective is to maximize social welfare? We will first consider the first-best where we assume that Medicare can implement efficient sorting by using a premium policy such as the one discussed in Proposition 3. We next restrict Medicare to use a single premium and, hence, the best it can achieve is a second-best sorting as described in Proposition 4.

6.1 The First-Best Level of Managed Care

Suppose that Medicare can implement efficient sorting as given by Proposition 1. We can then express social welfare as the sum of three integrals, and the efficient outcome is given by the solution to the following problem:

²⁶ The relation between payment level (the benchmark) and services available in MA, x^{MA} , depends on the market structure of MA plan supply. Competition among MA plans implies they will make zero profit by maximizing the welfare of a representative beneficiary choosing MA. If the sector is not competitive, there will still be a positive relationship between Medicare payment and services chosen by the MA plans. In this paper we are focusing on premium policy and so do not introduce a particular model of plan behavior and MA sector equilibrium.

$$\begin{aligned} \text{Max}_{(\theta_1, \theta_2, x^{\text{MA}})} & \int_{\underline{\theta}}^{\theta_1} \left(V(x^{\text{TM}}(\theta), \theta) - x^{\text{TM}}(\theta) \right) f(\theta) d\theta + \int_{\theta_1}^{\theta_2} \left(V(x^{\text{MA}}, \theta) - x^{\text{MA}} \right) f(\theta) d\theta \\ & + \int_{\theta_2}^{\bar{\theta}} \left(V(x^{\text{TM}}(\theta), \theta) - x^{\text{TM}}(\theta) \right) f(\theta) d\theta \end{aligned} \quad (8)$$

Three first-order conditions (FOC) define the solution, $(\theta_1^*, \theta_2^*, x^*)$, to (8). Consider the FOC with respect to x^{MA} describing the x^* best for the beneficiaries in MA:

$$\int_{\theta_1^*}^{\theta_2^*} \left(V_x(x^*, \theta) - 1 \right) f(\theta) d\theta = 0 \quad (9)$$

Condition (9) is a public good-like condition stating that at x^* , the sum of the marginal valuations of x less marginal cost, weighted by the frequency of θ in MA, equals zero. x^* is the efficient care for just one value of θ (the value that solves $V_x(x^*, \theta) = 1$), but the efficient assignment puts beneficiaries with θ in a range above and below this value into MA (because of the overconsumption for all beneficiaries in TM). The compromise on x^* reflects the relative frequency among beneficiaries between the θ_1^* and θ_2^* cutoffs.

The FOCs describing the optimal cutoffs between MA and TM are characterized in Proposition 1 and the two panels of Figure 3 when x^{MA} is chosen optimally at x^* . The same modification could be made to Panels A and B of Figure 4, adding stars to depict θ_1^* and θ_2^* to indicate that x^{MA} is chosen optimally. There is a welfare gain in MA for a range of beneficiaries to avoiding the over-consumption in TM. The value of x^* maximizes this gain. One consideration in this choice is the frequency of θ among the population of beneficiaries. With a plausible distribution of θ , maximizing social welfare implies setting the level of services in MA to serve the large bulk of the population.²⁷

6.2 The Second-Best Level of Care with a Single Premium

Suppose now that $y > 0$ and the premium implements a single cutoff.

Proposition 5: When a single premium is implemented, the (second-best) efficient level of care is lower than the first-best level of care.

²⁷ Another consideration is the relative welfare cost of over-consumption for beneficiaries with different values of θ . If higher θ individuals had more elastic demand, there would be more social gains to including them in MA and θ_2 would be higher. In the proof of Proposition 1, we need to rule out cases in which there are “islands” of unusual demand elasticities for some values of θ . For example, if there were a bunch of beneficiaries with perfectly inelastic demand, they could be safely put in TM with no welfare cost. The assumptions above about third derivatives of utility rule out these unusual islands.

When the single premium is positive there is a single cutoff and beneficiaries with a low θ are in MA, whereas in a first best these low-demand types would be in TM. This will lower the efficient x^{MA} , implying that the second-best high-end cutoff will fall below θ_2 . Thus, in this case, services in MA will be lower than in the first best and too few higher θ beneficiaries will be in MA.

7. Discussion

Part C of Medicare has been a disappointment for more than 25 years. Our paper identifies one fundamental reason for this. Policies governing the premiums for MA plans and TM lead to inefficient distribution of beneficiaries between the options. Although our model does not generate quantitative conclusions, the character of the optimal equilibrium is so different from what we see in practice that we are confident that the welfare losses associated with current policy are large. Enrollment in MA has vacillated between 10 and 25% in the past 15 years, and is likely to be driven down from its recent 25% highs with cuts in plan payments set in the ACA. Each MA plan sets its own “single premium.” Premiums for TM are highly subsidized and also do not vary according to demand-related factors.

We argue that MA should be directed to serve the large middle part of the distribution of demand for health care – what we refer to as the “middle class.” Instead, current premium policies encourage MA plans tend to attract beneficiaries with lower demand, both for reasons of better health status and for non-health status factors.²⁸

Increasing pressure on public budgets mean two things for Medicare Advantage (MA): first, it is even more important for Medicare to achieve the savings potential of managed care plans; second, in one form or another, beneficiary premiums (rather than tax-supported payments) will play a larger role in financing Medicare, both on the TM

²⁸ After adjusting for health status, there is an inverse relationship between income and other indicators of socioeconomic status and the likelihood that beneficiaries join an MA plan. Balsa, Cao and McGuire (2007) studied enrollment choices of 65+ Medicare beneficiaries not on Medicaid using five years of data from two sources, the National Health Interview Survey and the Medicare Current Beneficiary Survey. Lower income groups, lower education groups and members of racial/ethnic minorities were more likely to join MA, after adjustment for self-assessed health status, the presence of a series of other health conditions, age, gender, and other factors. The Kaiser Family Foundation, with more recent data, report that MA enrollees are more likely to be poor (and report poor health) than beneficiaries in TM (KFF, 2008).

and the MA side. Premium policy will inevitably be reformulated for financing reasons, yielding an opportunity to get more and more of the right beneficiaries in improved MA plans.

A comprehensive discussion of restructuring Medicare policy towards MA requires consideration of issues not addressed in this paper, including: the cost-effectiveness and quality of care in MA plans compared to TM; the desirable forms of MA plans; the level of the benchmark payment and bidding; the role medigap and employer plans; regional differences in patterns of care; the “choice architecture” for MA plans and the setup of the default option; the broader (spillover) effects of Medicare MA policy on health financing systems; and equity-distributional concerns in terms of rich-poor, healthy-sick and young-old.

While noting these concerns, we point out here two implications of our analysis of premiums and beneficiary sorting. First, our main message is that the premium for joining TM should be higher for higher-demand beneficiaries. Socially, there is more to be gained by keeping these high-demand types in MA. Currently, the Part B premium is mildly income-related, but the beneficiary pays the same premium whether they join an MA plan or stay in TM. We recommend that the income-related portion of the Part B premium be waived if the beneficiary elects MA. This simple step now would open important options down the road. In the future, as more is asked of beneficiaries in terms of premium, widening the premium differences by income will work towards the premium discrimination called for in our analysis.

Second, the structure is also in place to create options attractive to low-demand beneficiaries. Low income/asset Medicare beneficiaries qualify for special premium and cost sharing subsidies through state Medicaid programs, thus encouraging them to join managed care plans. Translating some of the subsidy into income support if these beneficiaries choose economical Medicaid managed care plans would could make beneficiaries better off and reduce public health care spending.

More broadly, our analysis implies that Medicare should take a direct hand in setting premiums for MA plans. Although it would represent a big change for Medicare, there is nothing radical about a sponsor setting premiums for plan options. Virtually all large private employers do this, setting the price employees pay for any health plans

offered through employment. Employers recognize their interest in affecting incentives employees face in deciding about plans. Like private employers, Medicare should take charge of the premium paid by beneficiaries to join health plans, and set those premiums in order to forward Medicare objectives.

Pursuit of efficiency in health insurance when serving a heterogeneous population runs up against concerns for fairness. Health care policy in the U.S. leaves much more room for gradations in health insurance (e.g., the Bronze-to-Platinum range in the ACA) than in most other countries. Nonetheless, targeting MA plans to groups across the range of socioeconomic status raises fundamental positive and normative questions about serving taste heterogeneity in health insurance, and about the role and fairness of collective methods of finance.

Our analysis leaves a number of questions unanswered. Incorporation of an explicit analysis of MA plan behavior should be high on the agenda. Within a population of heterogeneous beneficiaries, many MA plans setting different levels of services will better serve diverse tastes than a single MA product.

Although we are able to make our main points in an analysis abstracting from heterogeneity in health status, issues related to risk adjustment and MA plan resource allocation according to health need also need incorporation. One could ask our question: who *should* be in MA along dimensions of health status as well as taste for health care. With new data becoming available on what goes on inside the black box of managed care plans, empirical patterns of care in MA and TM can be compared, and theories about plan behavior can be assessed.

References

1. Acemoglu, D., Finkelstein, A. and Notowidigdo, M. (2009) "Income and Health Spending: Evidence from Oil Price Stocks," NBER Working paper No. 14744.
2. Afendulis, C., Landrum, M.B. and M. Chernew, "The Impact of the Affordable Care Act on Medicare Advantage Plan Availability and Enrollment," working paper, November, 2011.
3. Balsa, A. I., Z. Cao and T. G. McGuire (2007) "Does Managed Health Care Reduce Health Care Disparities between Minorities and Whites?" *Journal of Health Economics* 26(1): 101-121.
4. Barros, P. and L. Siciliani (2012) "Public and Private Sector Interface," in Pauly, M., McGuire, T., and P. Barros (eds.) *Handbook of Health Economics, Volume 2*, Elsevier, pp. 927-1002.
5. Baumgardner, J. (1991) "The Interaction Between Forms of Insurance Contract and Technical Change in Medical Care," *RAND Journal of Economics* 22(1): 36-53.
6. Berry, S. T. (1994) "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics* 25(2): 242-262.
7. Besley, T. and S. Coate (1991) "Public Provision of Private Goods and the Redistribution of Income," *American Economic Review* 81(4): 979-984.
8. Borger, C., T. F. Rutherford and G. Y. Won (2008) "Projecting Long Term Medical Spending Growth," *Journal of Health Economics* 27(1): 69-88.
9. Breyer, F., Bundorf, M.K. and M.V. Pauly (2012) "Health Care Spending Risk, Health Insurance, and Payment to Health Plans," in Pauly, McGuire, and Barros (eds.) *Handbook of Health Economics, Volume 2*, Elsevier, pp. 691-762.
10. Bundorf, M. K., J. D. Levin and N. Mahoney. (2009) "Pricing and Welfare in Health Plan Choice," NBER Working Paper 14153, available at: <http://www.nber.org/papers/w14153>.
11. Centers for Medicare and Medicaid Services (2008) "Medicare Advantage in 2007," CMS, Baltimore, MD.

12. Centers for Medicare and Medicaid Services (2010), "Fact Sheet: Premiums and Deductibles for 2011," November 4, 2010.
13. Centers for Medicare and Medicaid Services (2011), Medicare & You.
14. Chernew, M.E., Rosen, A.B. and Fendrick, A.M. (2007) "Value-Based Insurance Design," *Health Affairs*, 26(2): w195-203.
15. Ellis, R. P. and T. G. McGuire (2007) "Predictability and Predictiveness in Health Care Spending," *Journal of Health Economics* 26(1): 25-48.
16. Frank, R. G., J. Glazer and T. G. McGuire (2000) "Adverse Selection in Managed Health Care," *Journal of Health Economics* 19(6): 829-854.
17. Fogel, R.W. (2008) "Forecasting the Cost of U.S. Health Care in 2040," National Bureau of Economic Research Working Paper 14361.
18. Glazer, J. and T. G. McGuire (2011) "Gold and Silver Health Plans: Accommodating Demand Heterogeneity in Managed Competition," *Journal of Health Economics*, 30(5): 1011-1020.
19. Gold, M. (2008) "Medicare Advantage in 2008," Issue brief 7775, Henry Kaiser Family Foundation, June 2008.
20. Gold, M. (2009) "Medicare's Private Plans: A Report Card on Medicare Advantage," *Health Affairs* 23(1): w41-w54.
21. Gold, M., L. Achman, J. Mittler, et al. (2004) "Monitoring Medicare+Choice: What Have We Learned? Findings and Operational Lessons for Medicare Advantage," Mathematica Policy Research: Washington, DC.
22. Government Accountability Office, "Medicare Advantage Plan Bids," GAO-11-247R, February 4, 2011.
23. Grassi, S. and C-T A. Ma (2008) "Public Sector Rationing and Private Sector Selection," Department of Economics, Boston University.
24. Kaiser Family Foundation (2008) "Medicare Fact Sheet: Medicare Spending and Financing," Publication # 7305-03 at www.kff.org.

25. Kaiser Family Foundation (2009) "Medicare: A Primer," available at <http://www.kff.org/medicare/upload/7615-02.pdf>.
26. Kaiser Family Foundation (2010a), "Medicare Chartbook," 4th edition.
27. Kaiser Family Foundation (2010b), "Medicare Advantage 2011 Data Spotlight: Plan Availability and Premiums," October, 2010.
28. Keeler, E. B., G. Carter and J. Newhouse (1998) "A Model of the Impact of Reimbursement Schemes on Health Plan Choice," *Journal of Health Economics* 17(3): 297-320.
29. McGuire, T., Newhouse, J. and A. Sinaiko (2011), "An Economic History of Medicare Part C," *Milbank Quarterly* 89(2):289-332.
30. McWilliams, J. M., Afendulis, C. C., McGuire, T. G., and B. E. Landon, (2011), "Complex Medicare Advantage Choices May Overwhelm Seniors – Especially those with Impaired Decision Making," *Health Affairs*,
31. MedPAC (2008), Report to Congress; Medicare and the Health Care Delivery System, June.
32. MedPAC (2011a), Report to Congress; Medicare and the Health Care Delivery System, June.
33. MedPAC (2011b) A Data Book: Health Care Spending and the Medicare Program, June, 2011.
34. Miller, N. (2005) "Pricing Health Benefits: A Cost Minimizing Approach," *Journal of Health Economics* 24(5): 931-949.
35. Newhouse, J.P. (1993) *Free for All: Lessons from the RAND Health Insurance Experiment*, Harvard University Press.
36. Newhouse, J. P. (2002) *Pricing the Priceless: A Health Care Conundrum*, MIT Press.
37. Newhouse, J.P. (2006). "Reconsidering the Moral Hazard-Risk Avoidance Tradeoff," *Journal of Health Economics*, 25(5): 1005-1014.
38. Pauly, M. and Blavin, F. (2009). "Moral Hazard in Insurance, Value-based Cost Sharing, and the Benefits of Blissful Ignorance." *Journal of Health Economics*, 27(6):1407-17.

39. Pauly, M. and S. Ramsey (1999) "Would You Like Suspenders to Go with that Belt? An Analysis of Optimal Combinations of Cost Sharing and Managed Care," *Journal of Health Economics* 18(4): 443-458.
40. Pizer, S., A. Frakt and R. Feldman (2003) "Payment Policy and Inefficient Benefits in the Medicare+Choice Program," *International Journal of Health Care Finance and Economics* 3(2): 79-93.
41. Pope, G. C., J. Kautter, R. P. Ellis, A. S. Ash, J. Z. Ayanian, L. I. Iezzoni, M. J. Ingber, J. M. Levy and J. Robst (2004) "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model," *Health Care Financing Review* 25(4): 119-141.
42. Rothschild, M. and J. Stiglitz (1976) "Equilibrium in Competitive Insurance Markets: An Essay in the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90(4): 629-649.
43. Town, R. and S. Liu (2003) "The Welfare Impact of Medicare HMOs," *RAND Journal of Economics*, 34(4): 719-736.
44. Van de Ven, W.P.M.M., and R. P. Ellis, (2000) "Risk Adjustment in Competitive Health Plan Markets," in A. Culyer and J. Newhouse (eds.), *Handbook of Health Economics, Volume 1*, Elsevier, pp. 755-846.

Appendix

Proof of Proposition 1:

We define the difference between social welfare in MA and social welfare in TM to be:

$$\Delta W(x^{MA}, \theta) = W^{MA}(x^{MA}, \theta) - W^{TM}(\theta)$$

The change in ΔW as θ changes is given by (recall that $V_{\theta\theta} = 0$):

$$\begin{aligned} d\Delta W(x^{MA}, \theta)/d\theta &= V_{\theta}(x^{MA}, \theta) - V_{\theta}(x^{TM}(\theta), \theta) - [V_x(x^{TM}(\theta), \theta) - 1]dx^{TM}/d\theta \\ &= V_{\theta}(x^{MA}, \theta) - V_{\theta}(x^{TM}(\theta), \theta) + [V_x(x^{TM}(\theta), \theta) - 1]\left[\frac{V_{\theta x}(x^{TM}(\theta), \theta)}{V_{xx}(x^{TM}(\theta), \theta)}\right] \end{aligned}$$

The second equality follows from our definition of $x^{TM}(\theta)$.

The curvature of the ΔW is given by:

$$\frac{d^2\Delta W(x^{MA}, \theta)}{d\theta^2} = \frac{dx^{TM}}{d\theta} \left[-V_{\theta x} + (V_x - 1) \left(2 \frac{V_{\theta xx}}{V_{xx}} - \frac{V_{xxx} V_{\theta x}}{(V_{xx})^2} \right) \right] < 0$$

The inequality follows from the fact that

$$V_{\theta\theta} = 0, V_{\theta x} > 0, \frac{dx^{TM}}{d\theta} > 0, V_{xx} < 0, V_{xxx} \leq 0, V_x(x^{TM}(\theta), \theta) < 1,$$

$V_{\theta\theta x} = 0$ and $V_{\theta xx} = V_{xx}\theta$ by our assumption that $V_{\theta\theta} = 0$.

Given X^{MA} there exists a θ , call it θ^{MA} , for which, x^{MA} is socially efficient, that is $V_x(x^{MA}, \theta^{MA}) = 1$ and, hence, $\Delta W(x^{MA}, \theta^{MA}) > 0$. Furthermore, there exists another $\theta < \theta^{MA}$, call it θ_1 for which $x^{MA} = x^{TM}(\theta_1)$ and, hence, $\Delta W(x^{MA}, \theta_1) = 0$. By the concavity of ΔW , we know, therefore, that there exists at most one more θ , call it θ_2 , such that $\theta_2 > \theta^{MA}$ and $\Delta W(x^{MA}, \theta_2) = 0$.

Proof of Lemma 1:

Using the envelope theorem one can see that:

$$\frac{d\Delta U}{d\theta} = V_{\theta}(x^{MA}, \theta) - V_{\theta}(x^{TM}(\theta), \theta)$$

We know that at θ_1 , $x^{TM}(\theta_1) = x^{MA}$, and thus at θ_1 , $\Delta U/d\theta = 0$. Furthermore, by the linearity of U^{MA} and the convexity of U^{TM} we know that ΔU is convex with respect to θ . Thus, ΔU is maximized at θ_1 .

Proof of Proposition 2:

Assume a single premium y for TM. For the premium to implement the efficient assignment it must be that $-y + \Delta U(x^{MA}, \theta) \leq 0$ for all $\theta < \theta_1$ and $-y + \Delta U(x^{MA}, \theta) \geq 0$ for $\theta > \theta_1$. However, since with $\underline{\theta} < \theta_1 < \bar{\theta}$ and $\Delta U(x^{MA}, \theta)$ is maximized at θ_1 , the two conditions above cannot be met simultaneously.

Proof of Proposition 3:

Since $y_1 = -cx^{MA} = \Delta U(x^{MA}, \theta_1) > \Delta U(x^{MA}, \theta)$ for all $\theta < \theta_1$, all individuals with $\theta < \theta_1$ will prefer TM over MA. Since $\Delta U(x^{MA}, \theta)$ is decreasing with θ , for all $\theta > \theta_1$ and since $y_2 = -\Delta U(x^{MA}, \theta_2)$, it must be that $\Delta U(x^{MA}, \theta) + y_2 > 0$ for all $\theta, \theta_1 < \theta < \theta_1$ and $\Delta U(x^{MA}, \theta) + y_2 < 0$ for all $\theta > \theta_1$.

Proof of Proposition 4:

When $y = 0$ there is a single cutoff to the left of θ_2 . By increasing the premium the single cutoff moves towards θ_2 and welfare is increased. Once the cutoff reaches θ_2 , welfare will be decreased if the premium is further increased. (See Figure 7.)

Proof of Proposition 5:

When $y > 0$, and there is a single cutoff between MA and TM, the second-best outcome is given by:

$$\text{Max}_{(\theta', x)} \int_{\underline{\theta}}^{\theta'} (V(x, \theta) - x) f(\theta) d\theta + \int_{\theta'}^{\bar{\theta}} (V(x^{TM}(\theta), \theta) - x^{TM}(\theta)) f(\theta) d\theta \quad (10)$$

The first-order condition implies:

$$V(x', \theta') - x' = V(x^{TM}(\theta'), \theta') - x^{TM}(\theta'), \quad (11)$$

and:

$$\int_{\underline{\theta}}^{\theta'} (V_x(x', \theta) - 1) f(\theta) d\theta = 0 \quad (12)$$

Condition (11) is similar to the result obtained in Proposition 4, namely that $\theta' = \theta_2$.

Using the fact that $\theta' = \theta_2$ and $\underline{\theta} < \theta_1$, where θ_1 is the first-best lower cutoff, (12) implies:

$$\int_{\theta_1}^{\theta_2} (V_x(x', \theta) - 1) f(\theta) d\theta > 0$$

Thus, when the cutoffs can be chosen efficiently (i.e., at the first-best), welfare at x' is increasing in x . Thus, $x^* > x'$.