Who Belongs in Managed Care? Using Premium Policy to Achieve an Efficient Assignment in Medicare

Jacob Glazer
Tel Aviv University
Boston University

Thomas G. McGuire
Harvard Medical School

February 26, 2009

Abstract

Abstract

Acknowledgments: The authors are grateful to the National Institute of Aging for support through P01 XXXXX. The Role of Private Plans in Medicare, J. Newhouse, PI. This paper grew out of discussions with Rhema Vaithianathan. Mike Chernew, Richard Frank and Joe Newhouse provided helpful comments on an earlier draft. The opinions and conclusions in this paper are the authors' alone.

This version is for presentation and discussion at the BU/Harvard/MIT Health Economics Workshop. Comments welcome. Please do not circulate.

1 Introduction

Beneficiaries in Medicare, the federal health insurance program for the elderly and disabled, choose between two major options: traditional Medicare and a set of private health insurance plans, including managed care plans, offered under Medicare Part C. Part C has suffered severe criticism since its inception more than twenty five years ago, and remains very controversial. Nonetheless, Part C, currently known as Medicare Advantage, is the main hope for modernizing Medicare through introduction of the cost and quality management techniques common in private health insurance. Most research on the economics of Medicare Advantage (MA) focuses on the supply prices Medicare pays to MA plans. Many papers study the rules by which Medicare pays plans, including the risk adjustment of plan payments, and their effect on plan behavior. Demand prices paid by Medicare beneficiaries receive much less attention.

At present, the MA plan chooses the premium beneficiaries pay to join an MA plan (subject to Medicare regulation), and this premium is the same for low and high-demand beneficiaries. MA plans can thus employ the classic and powerful selection device identified by Rothschild and Stiglitz (1976): low premiums attract those with a low demand for health care. We consider whether instead of allowing plans to set premiums, Medicare should set the MA plan premium. If Medicare sets beneficiary premiums for both TM and MA, it can do so conditioned on some beneficiary characteristics. Medicare now conditions its plan payments on certain demand factors, but not others. Risk adjustment of Medicare payments pays plans more for beneficiaries who use more health care because of factors related to measured health status.[1] Demand for health care is also related to nonhealth factors, such as income or wealth, which are not part of risk adjustment.[2] These demand-related factors can, however, be part of

[1] Medicare risk adjusts payments to MA plans using Diagnostic Cost Groups which measure health status based on diagnoses on hospital claims from the previous year. See Pope et al. (2004).

[2] Cross-sectional studies generally report a positive income elasticity of demand that is less than one. The

premium policy. This paper shows that pairing health status based plan payments by Medicare with beneficiary premiums adjusted for other demand factors can radically improve the performance of Medicare Part C.

After presenting institutional background on Medicare and the MA program in Section 2, Section 3 sets out our model of Medicare. Beneficiaries vary in their demand for health care according to health status and a nonhealth status factor, "income." TM and MA allocate health care spending in different ways. TM accommodates demand (from lower health status or higher income) but creates the familiar moral hazard inefficiency. MA rations care by health status but does not recognize income-related preferences. Section 4 characterizes the efficient assignment of beneficiaries to MA and TM in light of the way care is rationed in the two forms of health insurance. Our welfare framework is the conventional one in economics, but we recognize the needs-based as well as demand-based approach to welfare, and discuss in a later section how our analysis would change with a shift to a needs-based approach. An "assignment" refers to a division of beneficiaries between MA and TM. The efficient assignment looks quite different than we see at present. Under current policies, healthy, low-income beneficiaries tend towards MA while the sick and higher income groups stay in TM.[3] We argue that an efficient assignment avoids the TM-associated moral hazard costs for the high-demand groups, not the low-demand ones. The efficient assignment we describe has the high-income groups in MA and the low-income groups in TM.

Section 5 contains a series of results about the role of the premiums beneficiaries pay in implementing an efficient assignment in Medicare. First, we show that given all beneficiaries pay the same for TM, no policy in which all beneficiaries pay the same for joining MA can

Rand Health Insurance Experiment, for example, found income elasticities of between .1 and .2. (Newhouse et al., 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).

[3]Low-income beneficiaries also eligible for Medicaid (dual-eligibles) pay no premium for TM and face little TM cost sharing. These low-income beneficiaries are almost all in TM.

achieve an efficient allocation. Interestingly, risk adjustment of plan payments cannot fix this problem. Also, this result is robust to whether it is Medicare or the plan that sets the premium. Second, an efficient allocation can be attained if Medicare fixes a single premium for MA, and sets an income-related premium for joining TM.[4] Third, even if Medicare sets an income-related premium for TM if Medicare allows the MA plans to set the premium, efficiency can not be attained.

Health status risk-adjusted Medicare payments to MA plans are part of the payment system required to achieve efficiency, though risk adjustment in direct proportion to health-status based demand (conventional risk adjustment) is not part of the optimal mechanism. We show that a division of labor between Medicare plan payments conditioned on health status and beneficiary premiums conditioned on income can sort beneficiaries efficiently into MA and TM and achieve the (second-best) efficient allocation of health care, subject to the rationing rules characterizing MA and TM.

Section 6 considers some extensions of our analysis and turns to issues related to Medicare costs and alternative welfare frameworks. Section 7 briefly summarizes the paper and points to issues for further research. The main message of our paper is that Medicare should set premiums for both MA and TM, and that premium policy and plan payment policy should be considered jointly. An income-related premium policy by Medicare need not be a marginal adjustment to what happens in Part C, but could lead to a fundamental realignment of the role of MA plans in the Medicare program, a realignment we argue in the next section is much needed.

[4]By 2009, Medicare phased in a small income-related premium for Part B, effectively equivalent to an income-related premium for TM. Higher premiums are expected to affect only 5 percent of beneficiaries. Part D also has income-related premiums. See CMS (2008). Beneficiaries also eligible for Medicaid pay no premium, introducing a second connection between income and premiums under current policy.

2 Medicare and Medicare Advantage

2.1 Program Descriptions

At age 65, most Americans become eligible for Medicare.[5] If beneficiaries do not elect an MA plan, they are automatically enrolled in Part A of Medicare at no cost to them. Part A is financed largely by a payroll tax shared by employees and employers (Kaiser Family Foundation, 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. Beneficiaries may also enroll in Part B, which covers doctors visits, other ambulatory services and some drugs at federally defined benefits. The Part B premium, $96.40 per month for 2009, pays only about 25 percent of Medicare's cost of Part B, the balance being paid for by general revenues (KFF, 2008). The vast majority of beneficiaries in TM enroll in Part B. Beneficiaries are also subject to cost sharing in Part B, including an annual deductible of $135 in 2009, and 20 percent cost sharing on Medicare allowed charges. Since 2004, beneficiaries may join a Part D plan covering prescription drug costs. Part D plans receive about 25 percent of their federal revenue from general revenues, are offered by private insurers and vary in coverage. Premiums for Part D are set by the Part D insurers, and only low-income beneficiaries receive a subsidy. Beneficiaries with Part A and the optional Parts B and D, are in "traditional Medicare."

Virtually all hospitals and practicing physicians participate in traditional Medicare giving beneficiaries wide choice of providers. Medicare and its regional intermediaries make broad coverage decisions but do not interfere in 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 by provider payment policy: Medicare payments to physicians and

[5]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.

hospitals are lower than private plans on average (refs and specifics).[6]

Most beneficiaries in TM avoid cost sharing in Parts A and B via supplemental coverage.(refs) Medicaid pays cost sharing for eligible low-income beneficiaries. Some employers buy wrap-around coverage for retirees. Finally, most beneficiaries not in either of these groups buy "medigap" policies to cover some or all of the cost sharing.

The Medicare Modernization Act of 2003 (MMA) created MA to replace the short-lived Medicare + Choice (M+C) 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).[7] MA plans may or may not include drug coverage. Those that do are referred to as MA-PD (i.e., "Part D") plans.

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 count Health Maintenance Organizations (HMOs), the oldest and largest plan type, and the mostly tightly managed, along with 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.[8] Our model of MA plan behavior laid out below applies to managed care plans only. 13.3 percent of Medicare beneficiaries were in bona fide managed care plans as of December 2007,[9] with 3.8 percent in

[6] 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).

[7] 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.

[8] Part C also includes Special Needs Plans (SPNs) intended for beneficiaries in long-term care.

[9] Gold (2008) p.20

PFFS plans and the balance of beneficiaries in traditional Medicare.

So far, the performance of Part C has been disappointing. The original intention of Part C was to allow Medicare to share in the cost savings achieved by managed care by paying plans at 95 percent of the expected cost in TM. Favorable selection into plans thwarted Medicare savings goals from the first.[10] Attempting to address the unfavorable impact of selection, M+C introduced more sophisticated risk adjustment of plan payments and limited increases in some plan payments. Taken together, these changes drove plans from Part C. The number of contracting plans fell 50 percent under M+C, and as Gold et al. (2004) concluded, M+C was "widely viewed as a failure." In response, the MMA raised payment rates. By buying its way out of trouble, Medicare has revitalized Part C but sacrificed any cost savings potential. Analyses by the Congressional Budget Office (Orszag, 2007) and MedPAC (Miller, 2007) report that Medicare pays about 15 percent more for beneficiaries in MA plans than for beneficiaries in TM. No wonder that MA falls in Washington's cross hairs. Part C not only moves few beneficiaries into real managed care plans, it costs rather than saves Medicare money.

The recent increases in Medicare payments have expanded the supply of MA plans to beneficiaries, including those that live in rural areas. In 2008, every beneficiary could choose among multiple MA plans, though residents of many rural counties still had no offering of a bona fide managed care plan, largely because of the difficulty of assembling a network of providers in sparsely served areas.[11] The HMO form of MA plan serves the majority of beneficiaries, currently 70 percent, down from 84 percent in 1999. In absolute terms, the number of beneficiaries in HMOs is no higher than 1999, indicating that all the growth in MA enrollment since then has been in other plan types.

[10] Brown (1993) estimated that favorable selection cost Medicare 5.7% over TM costs, thus more than offsetting the efficiency discount Medicare was seeking.

[11] For recent data on MA plans, see Gold (2009).

Some of the mechanics of MA are helpful to know before proceeding with our analysis. All beneficiaries in MA enroll in Part B. HMOs set premiums below the Medicare Part B premium of $96.40 beneficiaries pay in TM. The average HMO premium was $24 per month in 2008, with 65% of HMOs setting a zero premium, the minimum possible under Medicare regulations.[12] These plans also typically have reduced cost sharing and other additional coverage so that the average beneficiary electing these plans can expect lower health care payments and can avoid buying Medicare supplemental (Medigap) coverage.

MA costs rather than saves Medicare because of the way Medicare pays MA plans. Medicare payments to plans are based on a Medicare "benchmark" rate set for each county and the plan's "bid."[13] The benchmark is based on the maximum of the CMS estimate of Medicare costs for a typical beneficiary in Parts A and B (CMS should have stopped here) and the past payment rate for the county trended forward at national average Medicare cost growth rates. These grandfathered rates pre-date the MMA and are in some cases much higher than estimates of current average cost. They are the first reason why payments to MA plans exceed expected TM costs. The second reason has to do with a "budget neutrality" adjustment. Until 2010, no plan can have its payment reduced by risk adjustment; risk adjustment only affects the magnitude of any increase. Medicare had to raise the average benchmark enough so that even the plan attracting the healthiest beneficiaries would not see a negative impact. This is hard to believe, but, "Because of the budget neutrality adjustment, all of the county's benchmarks for 2007 are above Medicare's projected FFS costs." (Merlis, 2007, page 7).

Medicare payments are also affected by the plan's bid. A plan's bid is intended to be an estimate of what regular benefits from Part A and B would cost the plan. It would be that, and, incidentally a way to rescue Medicare from its benchmark rules, if accounting principles

[12]The Part B premium is paid to Medicare whether a beneficiary joins an MA plan or elects Part B in TM. When an HMO or other MA plan reduces the premium to the beneficiary the plan in effect pays the premium to Medicare.

[13]Merlis (2007) is an excellent primer.

trumped economic incentives. The incentives push plans to "bid" at least the benchmark. If the bid is above the benchmark, the plan is paid the benchmark. If the bid falls below the benchmark, Medicare pays the bid plus 75% of the benchmark bid difference. The idea is that Medicare gets 25% of the "savings" (difference between benchmark and bid) and the plan must use the balance to reduce beneficiary premiums or provide additional coverage and benefits. The problem is that plans can increase their payment by increasing their bid up to the benchmark. In 2006, the average Plan bid was 99% of the benchmark, slightly lower (97%) for HMOs, and above the benchmark for other types of plans. In essence, plans are paid very close to the benchmark.[14] Any excess payments Medicare makes to MA plans is exacerbated by selection of low-cost beneficiaries into MA. As fare as we know, patterns of selection in MA are yet to be studied.

2.2 Previous Research

Our paper is related to a literature in public finance 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. 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 to the private system.[15] 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 Olivella, 2005). 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

[14] A bid close to the benchmark, indicating that the benchmark payment is needed to cover basic Medicare services, seems inconsistent with MA plan's ability to buy down the premium and often additional services.

[15] Normative models of local public goods trade off production efficiencies and service of heterogeneity in tastes. Oates (1972) compares inefficiencies associated with higher or lower level of government provision of public goods.

set prices. Consumers differ according to wealth and cost. Our paper shares with this literature consumers' option to leave the public sector-like rationing plan (our MA plan). In our analysis of Medicare, however, our alternative system, TM is also publicly financed and subject to regulation.

A focus on the premiums paid by beneficiaries harks back to the work of Rothschild and Stiglitz (1976) who recognized the premium as a powerful selection device. In a Medicare context, the emphasis in the literature has instead been on skimping on services as a tool for selection.[16] A recent literature identifies which services among the many provided by a health plan are subject to the strongest incentive for a plan to use as a selection device. The premium charged by a plan can fit within the same framework as service-level selection. In Ellis and McGuire's (2007) terms, services that are predictable and predictive of total health care costs would be rationed tightly. The premium is perfectly predictable (and therefore a powerful selection device) but completely uncorrelated with total health care costs (because it is a constant). The service-level selection literature implies that the "service" of a dollar payment to join (negative premium) would be overprovided by profit-oriented plans.

There are two approaches in the literature to rationing in managed care, quantity setting (Pauly and Ramsey, 1999; Baumgartner, 1991) and shadow-prices (Keeler, Carter, Newhouse, 1998; Frank, Glazer and McGuire, 2000). The quantity-setting approach sees the plan as setting the level (or maximum level) of service available to members irrespective of their demand. The shadow-price approach emanates from a distinction between the bottom-line oriented plan financial leadership and the health-minded clinicians. The plan sets a budget for health care or components of health care, and clinicians ration their resources according to the health needs of the enrollees. In papers applying the shadow price approach to date, no

[16]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.

distinction has been made between demand (benefits perceived by the enrollee) and value of services to the clinician doing the rationing. In our paper, where we explicitly introduce non health status factors affecting demand, this distinction becomes an important one.

Plan choice of benefits and premiums can serve beneficiary heterogeneity, as well as being a device for selection.[17] The idea, as expressed by Pizer, Frakt and Feldman (2003) is "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"[18] One can imagine Medicare risk-adjusted plan payments serving as a kind of voucher with premiums sorting beneficiaries by tastes for additional services. Except in very special cercumstances, however, this sorting does not lead to efficiency.[19] 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.[20]

[17]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).

[18]In their empirical work, Pizer, Frakt and Feldman (2003) study the effect of off-cycle payment increases to Part C plans (called Medicare+Choice in 2000) to counter unanticipated declines in plan supply and beneficiary participation in Part C. Some plans were already "zero-premium" plans and could only pass on increased Medicare payments on to beneficiaries in the form of extra benefits. Plans with positive premiums could reduce premiums or add benefits. Data confirmed this. If, however, the benefit-premium tradeoff were being made efficiently in the first place, Medicare contribution increase should be passed fully on to the premium, with the minor qualification of income effects, which should be ignorable. Plans did, however, change benefits, not just premiums as payments increased. Gurol and Ellis (2004) note that in the presence of selection, the effect of a Medicare payment increase is more complicated, and the prediction on premiums and benefits is not so straightforward. They find that cross-sectional differences in Medicare payment rates are related to a mix of benefit changes and premium differences. Lustig (2008) to be reviewed.

[19]Elaborate

[20]Town and Liu (2003) use the estimated elasticity and the premium price combined with profit maximization to back into an estimate of the marginal cost of a Medicare beneficiary is in an HMO. This estimated marginal

Empirical research, including work on income elasticity of demand referred to earlier, finds a role for non health status factors in demand.

After adjusting for health status, there is an independent effect of income and other indicators of socioeconomic status on 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 the National Health Interview Survey and from the Medicare Current Beneficiary Survey. They found that 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).

3 Beneficiaries and Their Health Plan Choices

3.1 Beneficiaries and Efficient Health Care

There are four types of beneficiaries, differentiated by health status and income. Beneficiaries can be healthy (h) or sick (s), poor (p) or rich (r). Type is unchanging and there is an equal number of each type.

We define $v_{ij}(x)$ to be the benefit of health spending x to a beneficiary with health status i, $i = h, s$, and income j, $j = p, r$. We assume $v'_{ij} > 0, v''_{ij} < 0$. The marginal benefit v'_{ij} can be interpreted as a demand function. We rank the demands of the healthy and sick, and poor and rich. Among those with the same income, we assume that for any price of health care, the demand by the sick exceeds the demand by the healthy; and, among those with the same

cost is substantially below the Medicare payment and the source of producer surplus in their estimates. The authors analyze the effect favorable selection into HMOs might have on their findings in some back-of-the-envelop calculations. While the results are highly sensitive to selection, particularly the impact on Medicare program costs, they argue that the research evidence does not lead to a conclusion that favorable selection could be strong enough to overturn their conclusions about net welfare gains in Part C during the 1990s.

health status, demand by the rich exceeds demand by the poor. Furthermore, we assume that at any price, the elasticity of demand for all groups is the same.[21] (We do not assume elasticity is constant for all prices.)

These assumptions can be represented as:

$$v'_{hp}(x) = v'_{sp}(\gamma x) = v'_{hr}(\beta x) = v'_{sr}(\beta\gamma x), \tag{1}$$

for some $\gamma > 1$ and $\beta > 1$. If λ is a price or a shadow price used to ration care, then $x_{ij}(\lambda)$ is given by $v'_{ij}(x_{ij}(\lambda)) = \lambda$, for $i = h, s$ $j = p, r$. Following from (1), for all λ:

$$\begin{aligned} x_{ir}(\lambda) &= \beta x_{ip}(\lambda) \quad i = h, s \\ x_{sj}(\lambda) &= \gamma x_{hj}(\lambda) \quad j = p, r \end{aligned} \tag{2}$$

Demand by the rich is magnified by a factor β over the demand by the poor. Demand by the sick is magnified by a factor γ over demand by the healthy.

Medicare health insurance may require cost sharing and/or a premium to join. We refer to any beneficiary payment as t. Utility for each beneficiary type is then simply

$$u_{ij}(x_{ij}, t_{ij}) = v_{ij}(x_{ij}) - t_{ij} \tag{3}$$

Total social welfare depends on the level of services to each type of beneficiary (not on any payments beneficiaries make). Benefits to all beneficiaries less cost is:[22]

$$w(x) = \sum_i \sum_j v_{ij}(x_{ij}) - x_{ij} \quad i = s, h \quad j = p, r \tag{4}$$

where $x = \{x_{ij}\}$, a vector.

Maximizing $w(x)$ with respect to each of the x_{ij} yields the efficient (first-best) level of spending for each type, x^*_{ij}:

$$v'_{ij}(x^*_{ij}) = 1. \tag{5}$$

[21] This assumption is stronger than we need. Our main conclusions require that the demand response to price for the rich exceeds that for the poor in absolute (not relative) terms. Equal elasticity ensures this, though it could still be true if the elasticity of demand by the poor exceeded that of the rich. For the poor to have an absolutely larger demand response, their demand elasticity would have to exceed that for the rich by a proportion approximated by how much more the rich consume than the poor at equal prices. Studies of demand response among income groups do not find differences in demand elasticity, and are thus consistent with the assumption we make. (Refs)

[22] The welfare framework (4) is based on utility. In health care an alternative approach is to consider resource allocation based on health care "need." We consider such an approach in Section 6.

3.2 Rationing and Efficiency in Traditional Medicare and Medicare Advantage

Beneficiaries have two alternatives for health insurance, Traditional Medicare (TM) and Medicare Advantage (MA). Each type of plan has a rationing rule which generates inefficiencies and which we regard to be fixed throughout this paper.

In TM we assume beneficiaries pay a fixed premium which we normalize to zero. Beneficiares do pay a fixed cost sharing c, $0 < c < 1$, for each unit of x. Thus, in TM, $t_{ij} = cx_{ij}$. In TM, beneficiaries may choose x to maximize utility. For each type ij, the utility maximizing x_{ij}^{TM} solves the following demand equation:

$$v'_{ij}(x_{ij}^{TM}) = c \quad i = h, s \;\; j = p, r \tag{6}$$

We assume that c is fixed throughout the analysis and, hence, a beneficiary's utility in TM is fixed and is equal to:

$$U_{ij}^{TM} = u_{ij}(x_{ij}^{TM}, cx_{ij}^{TM}) \quad i = h, s \;\; j = p, r \tag{7}$$

Note that since $c < 1$, $x_{ij}^{TM} > x_{ij}^{*}$; all types use too much care in TM. An expression for the welfare loss for each type is:

$$L_{ij}^{TM} = \int_{x_{ij}^{*}}^{x_{ij}^{TM}} (1 - v'_{ij}(x))dx = \int_{c}^{1} (x_{ij}^{TM} - x_{ij}(\lambda))d\lambda \;\; i = h, s \;\; j = p, r \tag{8}$$

From (2) and (8) we know:

$$L_{sr}^{TM} = \beta L_{sp}^{TM} = \gamma L_{hr}^{TM} = \gamma\beta L_{hp}^{TM} \tag{9}$$

In other words, we know the largest welfare loss is associated with the sick rich, and the lowest welfare loss with the healthy poor. Figure 1 depicts the inefficiency in TM for each group.

In MA beneficiaries pay a premium which we designate to be y_{ij}^{MA}. We assume there are no copayments in MA so $t_{ij}^{MA} = y_{ij}^{MA}$. In principle, y_{ij}^{MA} could differ among beneficiary types, and in principle, y_{ij}^{MA} could also be positive or negative. A negative y_{ij}^{MA} could be interpreted

as a subsidy to join MA. In practice plans set a single premium, making y_{ij}^{MA} the same for all types.

In MA, beneficiaries do not choose their level of spending; the MA plan rations care. Our assumption is that clinicians in the MA plan ration a budget B according to health status (but not income). Let I: $\{h, s\} \times \{p, r\} \rightarrow \{0, 1\}$ be an indicator function indicating whether type ij is in MA ($I(i, j) = 1$) or in TM ($I(i, j) = 0$). We will refer to the $I(i, j)$ as the assignment of beneficiaries to plans. Specifically, our assumption about MA plan rationing is that the care for each type, x_{ij}^{MA}, satisfies:

$$\sum_i \sum_j I(i,j) x_{ij}^{MA} = B \quad \text{s.t. } x_{sj}^{MA} = \gamma x_{hj}^{MA},\ j = p, r \tag{10}$$

The budget B should be thought of as the total amount the plan allocates to treat patients. We allow for the plan to maximize profit which means that the plan's total revenue might exceed B. Equation (10) tells us that the doctors exhaust the budget, and sicker beneficiaries are allocated more spending than the rich by factor γ. Note that the poor and the rich of the same health type get the same spending in the MA plan. If there is only one health type in the plan, say the sick poor, (10) implies that the MA plan spends equally on all enrollees. If there is both healthy and sick in the plan, the sick get a γ proportion more spending than the healthy.

From (10), we can define what the sick and healthy get in the plan, independent of income:[23] $x_s = x_{sp}^{MA} = x_{sr}^{MA}$, and $x_h = x_{hp}^{MA} = x_{hr}^{MA}$. Utility for each type in MA is then

$$u_{ij}(x_i, y_{ij}^{MA}) \quad i = s, h,\ \ j = p, r$$

We can make a number of observations about the efficiency of rationing in MA. First, if there is only a single type of beneficiary (e.g., the sick rich) in MA, the budget can be set so as to attain any x_{sr}^{MA}, including x_{sr}^{*}. Furthermore, from (1) and (10), MA rationing equates the

[23] Notice that, in principle, x_{ij}^{MA} is a function of B. We suppress the B to simplify the presentation.

marginal benefits of spending within income groups (i.e., across health status groups). An MA plan can thus be used to attain x^*_{sr} and x^*_{hr} for the sick rich and healthy rich, respectively, if only the rich were in the plan. By the same argument, an MA plan could be used to set x^*_{sp} and x^*_{hp} if only the poor were in the plan.

If two types with the same health status but different incomes are in the MA plan, the MA plan would give each income group the same spending. Since the rich value spending more than the poor, it is not, therefore, possible to efficiently serve both income groups. An expression for the welfare loss for any type and spending level in MA is:

$$L^{MA}_{ij}(x_i) = \int_{\min(x^*_{ij},x_i)}^{\max(x^*_{ij},x_i)} (|1 - v'_{ij}(x)|)dx \quad i = h, s \quad j = p, r \tag{11}$$

Figure 2 depicts the inefficiencies in MA in the case in which all four types are in the MA plan. The rationing rule (10) ensures that x_s is a certain proportion γ above x_h. Because all demands have equal elasticity, this is equivalent to assuming that rationing occurs as by a shadow price, λ, the level of which depends on the budget (and types enrolled). In Figure 2 we show the welfare loss in MA for the case where the poor get too much care in relation to their efficient level, and the rich get too little: $x^*_{sp} < x_s < x^*_{sr}$ and $x^*_{hp} < x_h < x^*_{hr}$.

3.3 Medicare Spending

Medicare pays $(1 - c)x^{TM}_{ij}$ for each type in TM. For each type choosing an MA plan, Medicare makes a risk-adjusted payment to the plan. Medicare risk adjustment is based on health status, not income, so we call the payment to MA plans for the sick r_s and the healthy r_h. We assume initially that any premiums paid by beneficiaries to join the MA plan, y_{ij}, go to Medicare. In total, therefore, Medicare spending is

$$R = \sum_{i=s,h} \sum_{j=p,r} I(i,j)((r_i - y_{ij}) + (1 - I(i,j))(1 - c)x^{TM}_{ij} \tag{12}$$

4 Efficiency

As we have seen, from the standpoint of economic efficiency, both TM and MA have plusses and minuses. TM accommodates demand differences due to income as well as health status but provides too much care to everyone. MA can control moral hazard by rationing, but ignores income-driven differences in demand. In this section we consider how to assign beneficiaries to either TM or MA to minimize the efficiency losses depicted in Figures 1 and 2. Spending for each type in TM is determined by (6) and in MA by (10). The MA plan will also need to be given a budget to allocate among its members. The budget is implicitly determined in the problem we set out below. We will refer to an allocation as an assignment (a specification of which type of beneficiary is in each type of plan) and a spending amount for each beneficiary. An efficient allocation is one that minimizes the sum of the welfare losses.

An efficient allocation (x_h^0, x_s^0, I^0) is given by the solution to the following problem:

$$\begin{aligned} &\underset{(x_h,x_s,I)}{\text{Min}} \sum_{i=s,h} \sum_{j=p,r} (I(i,j)L_{ij}^{MA}(x_i) + (1 - I(i,j))L_{ij}^{TM}) \\ &\text{s.t. } x_s = \gamma x_h \end{aligned} \tag{13}$$

Notice, that in the problem above x_i refers to the level of spending an individual with health status i will receive if he joins MA, and I refers to the assignment function.

There are only two possible solutions to the efficient allocation problem, characterized in Proposition 1.

Proposition 1: In an efficient allocation, either the rich are in MA and the poor are in TM, or, everyone is in MA.

Proof: Let (x_h^0, x_s^0, I^0) be an efficient allocation. The proposition follows from the following observations:

(a) $I^0(i,j) = 1$ iff $L_{ij}^{MA}(x_i^0) \leq L_{ij}^{TM}$,

(b) $I^0(h,j) = 1$ iff $I^0(s,j) = 1$,

(c) $I^0(h,p) = 1$ implies $I^0(h,r) = 1$ and $I^0(s,r) = 1$.

Part (a) follows trivially from the fact that the planner can always move types in and out of the MA plan, without affecting the levels of care x_h^0 and x_s^0 by adjusting the budget. If a type ij is added to (withdrawn from) the MA plan the budget is increased (decreased) by x_i^0.[24]

Part (b) follows from (a) above and the fact that

$$L_{sj}^{MA}(x_s) - L_{sj}^{TM} = \beta(L_{hj}^{MA}(x_h) - L_{hj}^{TM})$$

From (b), we know that if type hp is in MA it must also be that type sp is in MA. From (b) we also know that it cannot be the case hp and sp are in MA but only hr or only pr is in TM. Thus, to establish Part (c) it remains to be shown that it cannot be the case that only the poor (hp and sp) are in the MA plan (and the rich are in TM).

Notice that if only hp and sp are in MA, it must be that $x_i^0 = x_{ip}^*$ for $i = h, s$. In such a case the total welfare loss will be $L_{hr}^{TM} + L_{sr}^{TM}$. In such a case, however, welfare could be improved by assigning only types hr and sr to MA (and assigning hp and sp to TM) and allocating them the level of services $x_i^0 = x_{ir}^*$. In such a case the welfare loss will only be $L_{hp}^{TM} + L_{sp}^{TM}$, which must be smaller. ■

Corollary 1: There exists a $\beta^0 > 1$ such that if $\beta > \beta^0$, then in an efficient assignment the rich are in MA and the poor are in TM. If $\beta < \beta^0$, in an efficient assignment, everyone is in MA.

Proof: If the rich are in MA and the poor are in TM, the rich will get the efficient level of care and welfare loss will be:

[24] For convenience, we assume that if $L_{ij}^{MA}(x_i^0) = L_{ij}^{TM}$ the individual is assigned to MA.

$$L^1 = L^{TM}_{hp} + L^{TM}_{sp} > 0$$

If everyone is in MA welfare loss will be

$$L^2 = \sum_{i=s,h} \sum_{j=p,r} L^{MA}_{ij}(x_i)$$

where x_i $i = s, h$ is the level of care that minimizes the welfare loss given that everyone is in MA.

Notice that L^1 is independent of β whereas $L^2 = 0$ when $\beta = 0$ and it increases with β. Thus for β small enough $L^2 < L^1$ and it is socially efficient to have everyone in MA whereas for β large enough $L^2 > L^1$ and it is socially efficient to have only the rich in MA. ■

Proposition 1 implies that an efficient assignment takes one of two forms: all types are in MA, or, only the rich are in MA with the poor in TM. Corollary 1 indicates that the form of efficient assignment depends on β, the parameter capturing how much different demand by the rich is from demand by the poor. If the demands of the rich and the poor are not very different (β is near one), there is little welfare loss from combining the rich and the poor in MA, and the efficient assignment will be everyone in MA. If the demands are very different (β is much greater than one), the rich and the poor demand very different health care spending for a given health status. An MA plan providing the same health care to both will create large welfare losses. In this case, the poor should not be in MA.

The interesting form of efficient assignment is when the one-size-fits-all resource allocation in MA does create a significant welfare loss, and only the rich should be in MA. Why the rich? The rich demand more health care than the poor and their welfare loss in TM is greater. Using MA to ration health care avoids more efficiency losses if the rationing is oriented to the rich. The healthy and sick (rich) can both be efficiently served in MA because the MA plan

respects health status differences in demand (though not income differences). Thus, the efficient allocation with only the rich in MA will have the rich sick and healthy getting their efficient health care spending, and the poor being served inefficiently in TM.

As we noted in our review of the performance of MA, it is in fact the poor that choose MA, not the rich. This is because Medicare's current policy sorts the low, not the high-demand groups, into MA.

5 Implementation

In this section we consider allocations that can be implemented with Medicare policy. For an allocation to be implementable it must be consistent with beneficiary and MA plan behavior. Beneficiaries choose between MA and TM to maximize utility. The MA plan must accept all beneficiaries who seek to join, but may choose the level of services to offer. We adopt the conventional assumption that the MA plan seeks to maximize profits. However, many of our results hold under other plan objectives, such as maximizing the number of enrollees subject to breaking even.

Our specific objective in this section is to identify the allocations that are implementable with various Medicare policies. We define Medicare policy to consist of a pair (r_{ij}, y_{ij}) $i = s, h$, $j = p, r$, indicating what Medicare pays to the MA plan and what premiums beneficiaries pay to join MA. Note that we are supposing that Medicare chooses MA plan premiums. In principle, Medicare payment and beneficiary premium could depend on both health status and income. We regard the premium beneficiaries pay as going directly to Medicare. If the premium is negative (a subsidy), this would be paid by Medicare to beneficiaries. We say that an allocation (x_h, x_s, I) is implementable by (r_{ij}, y_{ij}) if, given (r_{ij}, y_{ij}), the plan chooses (x_h, x_s) and beneficiaries choose plans according to I.

The first policy we consider is current policy: Medicare pays plans according to health status, and all beneficiaries pay the same premium to join the MA plan. As we show this policy cannot in general achieve an efficient allocation as defined in the previous section. Next, we consider an alteration of current policy in which Medicare continues to pay plans on the basis of health status, but charges beneficiaries a premium that can differ by income. This second policy enables Medicare to implement any efficient allocation.

There are obviously other forms of policy that could be considered. Conceivably, for example, Medicare could make a flat payment to plans for every beneficiary and then set a premium for each type that depends on both health status and income. Such a voucher-complete risk adjustment-subsidy scheme could achieve efficiency. We do not believe, however, that Medicare premium policy requiring beneficiaries to pay their own risk adjustment is interesting. We consider policies with a division of labor between Medicare plan payments, which can be risk adjusted on health status but not income, and Medicare premium policy that can be adjusted on income but not health status.

5.1 The Fundamental Problem with a Single Premium

Current Medicare policy involves two elements. First, Medicare makes a health status (but not income) related risk adjusted payment to the plan for each enrollee, r_h for each healthy enrollee and r_s for each sick enrollee. Second, MA plans set the premium to be the same for all beneficiaries. In other words, $y_{ij} = \bar{y}$, $i = h, s$, $j = p, r$. This premium could be positive or negative. Our first significant observation about implementable allocations under current policy follows directly from beneficiary utility maximization.

In what follows we make the following assumption, which we will refer to as an assumption that "coverage in TM is sufficiently good:"

$$U_{hp}^{TM} - v_{hp}(x_{hp}^*) = \int_c^1 (x_{hp}(\lambda) - x_{hp}^*)d\lambda - cx_{hp}^* > 0$$

Under this assumption, if the premium the individual has to pay to join the MA plan is set at zero and the level of care at MA is set at the efficient level for type hp, the individual of this type will prefer TM over MA. One can easily see that the condition above holds for $c = 0$ and that it is vilotated at $c = 1$. Thus, we assume that the level of copayment is sufficiently low.

We can now make the following proposition:

Proposition 2: Suppose coverage in TM is sufficiently good, $x_i^{MA} = x_{ir}^*$ and $y_{ij} = \bar{y}$ for $i = h$, s and $j = p, r$. If type sr chooses MA, everyone chooses MA.

Proof: If type sr chooses MA it must be that

$$U_{sr}^{TM} - v_{sr}(x_{sr}^*) + \bar{y} < 0.$$

Notice that:

$$U_{sr}^{TM} - v_{sr}(x_{sr}^*) = \gamma\beta(U_{hp}^{TM} - v_{hp}(x_{hp}^*)) > U_{hp}^{TM} - v_{hp}(x_{hp}^*) > U_{hp}^{TM} - v_{hp}(x_{hr}^*).$$

Hence, if $U_{sr}^{TM} - v_{sr}(x_{sr}^*) + \bar{y} < 0$ it must be that $U_{hp}^{TM} - v_{hp}(x_{hr}^*) + \bar{y} < 0$, and type hp also chooses MA. In the same way it can be shown that the other types also choose MA. ■

Corollary 2 follows directly from Propositions 1 and 2:

Corollary 2: Suppose that coverage in TM is sufficiently good, then the only efficient allocation that can be achieved with a single premium policy is if everyone is in MA.

Our assumption about c ensures that TM is attractive enough to be a viable option to MA. Obviously, if c were very close to 1, TM would be a bad option for beneficiaries, even if it were free, and the question of dividing beneficiaries between TM and MA would not be interesting. In practice, c is low, near zero at the margin for Part A, and 20% for Part B. With a single premium for MA (which could be positive or negative), all types pay the same price to join MA. Proposition 2 states that if the highest demand type, the sick rich, prefers MA at this price, then every other type will prefer MA as well. This is because

higher-demand types give up relatively more leaving TM; if the single price makes it worthwhile for the high-demand type to give up TM, it must also be worthwhile for a lower demand type.

Taken together, Proposition 1 and Proposition 2 expose a fundamental problem of current policy. A single MA premium can implement an efficient assignment only in the uninteresting case in which all beneficiaries should be in MA. (Recall there is never an efficient assignment in which no beneficiary is in MA.) This would be easy to do: If we want everyone in MA, set a very large negative premium (subsidy) to join MA, large enough so that all types would rather be in MA. In the interesting case in which beneficiaries should be divided between MA and TM, a single premium can never achieve this assignment. Note that we have not specified in Proposition 2 how plans behave. This flaw in current policy therefore holds for any plan objective or any means of paying plans. It holds whether Medicare or plans decide on the single premium. Said another way, the flaw cannot be overcome by the way Medicare pays plans.

Proposition 2 may not hold if coverage in TM is not sufficiently good. For example, suppose c is very close to 1 and TM confers virtually no benefits to beneficiaries. Then one can readily see that if the poor and rich have quite different demands, a high-premium high benefit MA plan might attract the rich only, and thus might achieve an efficient allocation.

5.2 An Income-Based Premium Can Implement Any Efficient Allocation

We now consider Medicare payment policies that can implement an efficient allocation. Since implementation involves beneficiary and plan behavior, we must specify both the premiums beneficiaries pay, and the payments Medicare makes to the plans.

Proposition 3: Medicare policy of risk-adjustmented payment to MA plans on the basis of health status paired with an income-adjusted premium to beneficiaries can implement any efficient allocation.

Implementing the efficient allocation when everyone is in MA is straightforward. We are concerned here with the case in which only the rich are in MA.

Proof: Let y_p denote the premium that types hp and sp pay and set

$$y_p \geq \text{ Max}\{v_{hp}(x^*_{hr}) - U^{TM}_{hp}, v_{sp}(x^*_{sr}) - U^{TM}_{sp}\} \tag{14}$$

Let y_r denote the premium that both types hr and sr pay and set

$$y_r = v_{sr}(x^*_{sr}) - U^{TM}_{sr} < v_{hr}(x^*_{hr}) - U^{TM}_{hr} < 0 \tag{15}$$

The two inequalities above follow from the assumption that coverage in TM is sufficiently good.

Let r_h denote the risk adjusted payment the plan receives for each individual of either type hr or hp that joins the plan and set $r_h \leq \underline{x}$ where $\underline{x}$ solves:

$$v_{hr}(\underline{x}) - y_r = U^{TM}_{hr}. \tag{16}$$

Notice that if $x^{MA}_h < \underline{x}$ type hr will choose TM when the premium he has to pay is y_r.

Let r_s denote the risk adjusted payment the plan receives for each individual of either type sr or sp that joins the plane and set

$$r_s = x^*_{hr} + x^*_{sr} - r_h. \tag{17}$$

(A special case of the risk adjustment payment scheme above is $r_h = 0$ and $r_s = x^*_{hr} + x^*_{sr}$).

Notice that under the premiums above if the plan provides the levels of services $x_h = x^*_{hr}$ and $x_s = x^*_{sr}$ both poor types hp and sp will prefer TM over the MA plan, whereas the rich types hr and sr will prefer to be in MA.

Furthermore, the profit maximizing plan will choose $x_h = x^*_{hr}$ and $x_s = x^*_{sr}$. If the plan chooses $x_h > x^*_{hr}$ (and, hence, $x_s > x^*_{sr}$), it will attract the rich types (and possibly also the poor) and lose money. If, on the other hand, the plan chooses $x_h < x^*_{hr}$ (and, hence, $x_s < x^*_{sr}$) it will attract (at most) the healthy types for which the risk adjusted payment is insufficient to cover the cost unless $x < r_h$. However, if $x \leq r_h \leq \underline{x}$ the individual will choose TM. ■

Premiums to beneficiaries must be set so that at the efficient allocation with only the rich in MA, the poor do not want to join MA but the rich do. Given the assumption that coverage in TM is sufficiently good, we know that $y_r < 0$. Furthermore using (14) [we can show that using transitivity of preferences for MA and TM across groups] $y_r < y_p$. In other words, the rich must be subsidized to join MA; the subsidy available to the poor must be less. The poor may possibly be charged a positive premium to keep them out of MA.

Proposition 3 also describes the risk adjusted payments that would implement the efficient allocation with just the rich in MA. Equation (17) ensures that the plan will break even if both the health and sick rich join. Payment for the healthy must be sufficiently low that the plan cannot cover cost for a level of services that will attract the rich health only. Equation (16) characterizes the level of spending on the rich healthy at which they are just indifferent between TM and MA at the premium they are charged for MA. The risk-adjusted payment for the rich can be no greater than this amount.

Some risk-adjustment is necessary to implement the efficient allocation. Paying the same for healthy and sick rich would allow the plan to reduce spending below the optimal, attract only the healthy and make profits. Notably, however, conventional risk adjustment,[25] with payment to the sick directly in proportion to their greater use, would not implement the efficient

[25]The term "conventional" risk adjustment refers to risk adjustment in proportion to expected spending, usually based on an empirical predictive model of health care costs. See Glazer and McGuire (2002) for discussion and contrast with "optimal" risk adjustment which is derived as a solution to a resource allocation problem. The analyses in the current paper are in the spirit of the optimal risk adjustment literature.

allocation. More specifically, consider a Medicare risk adjustment policy where x^*_{hr} is paid for the healthy and x^*_{sr} is paid for the sick. This would obviously allow the plan to break even at the efficient level of service for the two types we want in the plan. The clinical spending rules at the plan would mean that if the plan spent its entire budget, the healthy and sick would get the efficient care. But would the plan choose to spend the whole budget? From (15), the rich sick are just indifferent to joining MA, but the rich healthy get a surplus at the premium for the rich. The MA plan being paid conventional risk adjustment could therefore spend a little less on health care for both types, attract only the rich healthy, and make a profit. This is why the risk adjustment formula must be set to make the healthy rich unprofitable if they alone join the plan. This result parallels the literature on optimal risk adjustment which calls for "overpaying for the sick" (in order to encourage spending on services used by the sick).

5.3 Reinterpretation: An Income-Related Premium to Join TM

While the idea of charging the poor more than the rich to join MA seems odd, the income-related premium called for in Proposition 3 can be thought of in a different and more appealing way. Since the premium from (15) is likely to be negative, it can be thought of as a positive premium to join TM. In other words, suppose the MA plan were free (normalize the MA plan premium to zero, and charge beneficiaries to join TM). Set the premium for TM to the rich to be

$$y^{TM}_{ir} = U^{TM}_{sr} - v_{sr}(x^*_{sr}) \quad i = s, h. \tag{18}$$

Set the premium for TM for the poor

$$y^{TM}_{p} = \text{Min}\{U^{TM}_{hp} - v_{hp}(x^*_{hr}), U^{TM}_{sp} - v_{sp}(x^*_{sr})\}. \tag{19}$$

At these premiums, the poor will be in TM and the rich will not pay the premium at TM and go to MA.

5.4 MA Plan Choosing A Premium Interferes with Efficiency

So far we have regarded premiums as being set and collected by Medicare. In what follows, we allow the MA plan to choose a premium. We show that in such a case, the plan will not choose the efficient level of care.

Let $y^{TM} = \left\{y_p^{TM}, y_r^{TM}\right\}$ denote the premium that Medicare charges for TM to beneficiaries with income $i = p, r$, and let $y^{MA} = \left\{y_p^{MA}, y_r^{MA}\right\}$ denote the premium that the MA plan charges a beneficiary with income i. Let $r = \{r_h, r_s\}$ denote the payment Medicare pays the plan for each individual with health status $j = h, s$ that joins the plan. The order of moves is as follows: First, Medicare announces y^{TM} and r. Next, the MA plan chooses $x = \{x_h, x_s\}$ (where, $x_s = \gamma x_h$) and announces y^{MA}. Finally, individuals observe y^{TM}, y^{MA} and x and decide whether to join MA or TM.

Proposition 4: If the MA plan chooses y^{MA}, it will not choose the socially efficient level of care.

Proof: Suppose that the socially efficient allocation is such that only the rich are in MA. In such a case, in order for the socially efficient allocation to be implemented, it must be that the MA plan attracts only the rich types and provides each type with the socially efficient level of care. If only the rich types are in MA, the plan's profit is $r_h + r_s + 2y_r^{MA} - x(1+\gamma)$. The plan's choice of y_r^{MA} and x must satisfy:

$$\underset{\{y_r^{MA},\ x\}}{\text{Max}}\ r_h + r_s + 2y_r^{MA} - x(1+\gamma)$$

s.t.

$$v_{hr}(x) - y_r^{MA} \geq U_{hr}^{TM} + y_r^{TM}$$

$$v_{sr}(\gamma x) - y_r^{MA} \geq U_{sr}^{TM} + y_r^{TM}$$

By (15) we know that (at $\{x_{hr}^*, x_{sr}^*\}$) only the second constraint is binding and, hence, the first-order conditions imply:

$$2 + \lambda = 0$$

$$-(1+\gamma) - \lambda\gamma v'_{sr}(\gamma x) = 0$$

where λ is the multiple for the second constraint. Hence,

$$v'_{sr}(\gamma x) = \frac{1+\gamma}{2\gamma} < 1 = v'_{sr}(x^*_{sr})$$

which also implies

$$v'_{hr}(x) = \frac{1+\gamma}{2\gamma} < 1 = v'_{hr}(x^*_{hr})$$

Hence, the plan chooses a level of care higher than the socially efficient one both for the healthy and for the sick rich types. ■

The intuition for the result above is as follows: Suppose that the plan attracts only the rich and provides the socially efficient levels of care, x^*_{hr} and x^*_{sr}, and charges a premium y^{MA}_r. Notice, first, that setting y^{MA}_r to maximize profits implies that if the plan attracts both rich types, it must be that sr (weakly) prefers MA over TM and type hr strictly prefers MA over TM. Next observe that if the plan increases y^{MA}_r and x_s by ϵ and x_h by $\frac{\epsilon}{\gamma}$, and both the healthy and the sick rich types stay with the plan, the plan's profit will be increased by $\epsilon(2-(1+\frac{1}{\gamma}))$. Next observe that if at x^*_{sr} and y^{MA}_r, the individual of type sr prefers MA over TM, then, since $v'(x^*_{sr}) = 1$, if the plan increases both x_s and y^{MA}_r by ϵ,the individual of type sr will still prefer MA over TM. Furthermore, since at x^*_{hr}, type hr strictly prefers MA over TM, there is an ϵ small enough such that even if the premium is increased by ϵ and x_h is increased by $\frac{\epsilon}{\gamma}$ the individual of type hr will still prefer MA over TM. Thus, the plan can increase its profit by increasing the level of services beyond their socially efficient level and increasing the premium accordingly.

It should be mentioned here that even in the case where it is socially efficient for all types to be in MA, letting the plan charge a premium will result in the plan providing a level of

services beyond the socially desired one. The proof of this result is similar to the one provided above for the case where it is socially efficient for only the rich to be in MA.

6 Extensions

The connection between premiums and risk adjustment.

Results are affected by the distribution of types.

One extension to consider is more heterogeneity in demand (e.g., continuous income or health status groups). We believe most of our results go through with modifications in the case of one MA plan.

More than one MA plan raises a number of issues. Sticking with existing model of beneficiaries:

- Efficiency is straightforward: need two types of plan, one for poor, one for rich.
- Implementation can be accomplished with general set of premiums and risk adjustment.
- Interesting questions revolve around constricted Medicare policy and sorting which we haven't worked out. What are reasonable restrictions on what Medicare could do with an MA Basic and MA Gold?
- Another normative, empirical issue is the degree to which non-healthstatus heterogeneity should be served in Medicare. This is partly a value question of whether demand or need is the right goal for policy, and partly an empirical question of how much different care do high and low income (and other dimensions) demand.

Welfare Based on Need

The efficiency of Medicare can also be examined from a more narrow but critical perspective of the Medicare Program itself. Medicare program spending is composed of part of health care costs in TM and risk-adjusted per capita payments to private plans in MA. We study how Medicare policy sorting beneficiaries to TM and MA affects Medicare spending. Retaining TM as an option for beneficiaries, we identify Medicare policies, based on current rules and modifications we recommend, that minimize program spending.

7 Conclusion

Note so far. Please ignore!

The social benefits of switching a beneficiary from TM to an managed care plan are roughly proportional to magnitude of health care spending brought under "management." It is thus the beneficiaries with *high demand*, because of health status or other factors such as income or education, that yield more social benefits (and save Medicare more money) by moving to MA. By contrast, MA plans actually attract the *low-demand* beneficiaries by lowering premiums below what beneficiaries pay for TM. While this empirical pattern is known, the usual conclusion is that inadequate risk adjustment is responsible for favorable selection into private plans. We argue that the design flaw in Medicare has to do with beneficiary premiums, not payments to plans.

Health plans incorporating some features of managed care are the only options for employees getting insurance through work, in individual health insurance markets, and in the state-run Medicaid programs for the poor.

The federal Medicare program will spend approximately $xxx billion in 2009 for health care for the elderly and disabled beneficiaries. (ref) Projected Medicare spending accounts for almost all of growth in budget deficits in the next several decades. (ref)

Some income-related premiums already in Medicare. Limited in Part B. Part D has an income-related premium.

Relation to 1998 National Bipartisan Commission on the Future of Medicare.

References

[1] 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.

[2] Barros, P. and P. Olivella (2005) "Waiting Lists and Patient Selection," *Journal of Economics and Management Strategy* 14(3): 623-646.

[3] Baumgardner, J. (1991) "The Interaction Between Forms of Insurance Contract and Technical Change in Medical Care," *RAND Journal of Economics* 22(1): 36-53.

[4] Berenson, R. A. (2008) "From Politics to Policy: A New Payment Approach to Medicare Advantage," *Health Affairs* web exclusive, March 4, 2008, w156-w164.

[5] Berry, S. T. (1994) "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics* 25(2): 242-262.

[6] Besley, T. and S. Coate (1991) "Public Provision of Private Goods and the Redistribution of Income," *American Economic Review* 81(4): 979-984.

[7] Brown, R., et al. (1993) "Do Health Maintenance Organizations Work for Medicare?" *Health Care Financing Review* 15(1): 7-23.

[8] Cao, Z. and T. G. McGuire (2003) "Service-level Selection by HMOs in Medicare," *Journal of Health Economics* 22(6): 915-931.

[9] Centers for Medicare and Medicaid Services (2007) "Medicare Advantage in 2007," CMS, Baltimore, MD.

[10] Ellis, R. P. and T. G. McGuire (2007) "Predictability and Predictiveness in Health Care Spending," *Journal of Health Economics* 26(1):25-48.

[11] Frank, R. G., J. Glazer and T. G. McGuire (2000) "Adverse Selection in Managed Health Care," *Journal of Health Economics* 19(6): 829-854.

[12] Fogel, R.W. (2008) "Forecasting the Cost of U.S. Health Care in 2040," *National Bureau of Economic Research* Working Paper 14361.

[13] Glazer, J. and T. G. McGuire (2000) "Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Health Care," *American Economic Review* 90(4): 1055-1071.

[14] Glazer, J. and T. G. McGuire (2002) "Setting Health Plan Premiums to Ensure Efficient Quality in Health Care: Minimum Variance Optimal Risk Adjustment," *Journal of Public Economics* 84(2): 153-173.

[15] Gold, Marsha (2008) "Medicare Advantage in 2008," Issue brief 7775, Henry Kaiser Family Foundation, June 2008.

[16] Gold, Marsha (2009) "Medicare's Private Plans: A Report Card on Medicare Advantage," *Health Affairs* 23(1): w41-w54.

[17] 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.

[18] Grassi, S. and C-T A. Ma (2008) "Public Sector Rationing and Private Sector Selection," Department of Economics, Boston University.

[19] Gurol, I. and R. P. Ellis (2004) "Medicare Managed Care Choice of Benefits and Premiums: Competition and Selection Incentives Really Matter," *unpublished*, Boston University, Department of Economics.

[20] Kaiser Family Foundation (2008) "Medicare Fact Sheet: Medicare Spending and Financing," Publication # 7305-03 at www.kff.org.

[21] 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.

[22] Lustig, J. (2008) "The Welfare Effects of Adverse Selection in Privatized Medicare," Department of Economics, Boston University.

[23] Mayes, R. and R. Berenson (2006) *Medicare Prospective Payment and the Shaping of U.S. Health Care*, The Johns Hopkins University Press.

[24] MedPAC Report to Congress, Chapter 3, March 2008.

[25] Merlis, M. (2007) "Medicare Advantage Payment Policy," *Background Paper*, George Washington University: Washington, DC.

[26] Miller, M. E., Executive Director, Medicare Payment Advisory Commission, (June 28, 2007), "Medicare Advantage Program and MedPAC Recommendations," Committee on the Budget, U.S. House of Representatives: Washington, DC.

[27] Miller, N. (2005) "Pricing Health Benefits: A Cost Minimizing Approach," *Journal of Health Economics* 24(5): 931-949.

[28] Newhouse, J.P. (1993) *Free for All: Lessons from the RAND Health Insurance Experiment*, Harvard University Press.

[29] Newhouse, J. P. (2002) *Pricing the Priceless: A Health Care Conundrum*, MIT Press.

[30] Oates, W. (1972) Fiscal Federalism, Harcourt Brace: New York, NY.

[31] Orzag, P. R. (June 28, 2007) "CBO Testimony: The Medicare Advantage Program," before the Committee on the Budget, U.S. House of Representatives.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] Skinner, J., E. Fisher and J. Wennberg "The Efficiency of Medicare," *National Bureau of Economic Research* Working Paper 8395.

[37] Town, R. and S. Liu (2003) "The Welfare Impact of Medicare HMOs," *RAND Journal of Economics* 34(4): 719-736.

Figure 1: Spending and Inefficiencies in Traditional Medicare

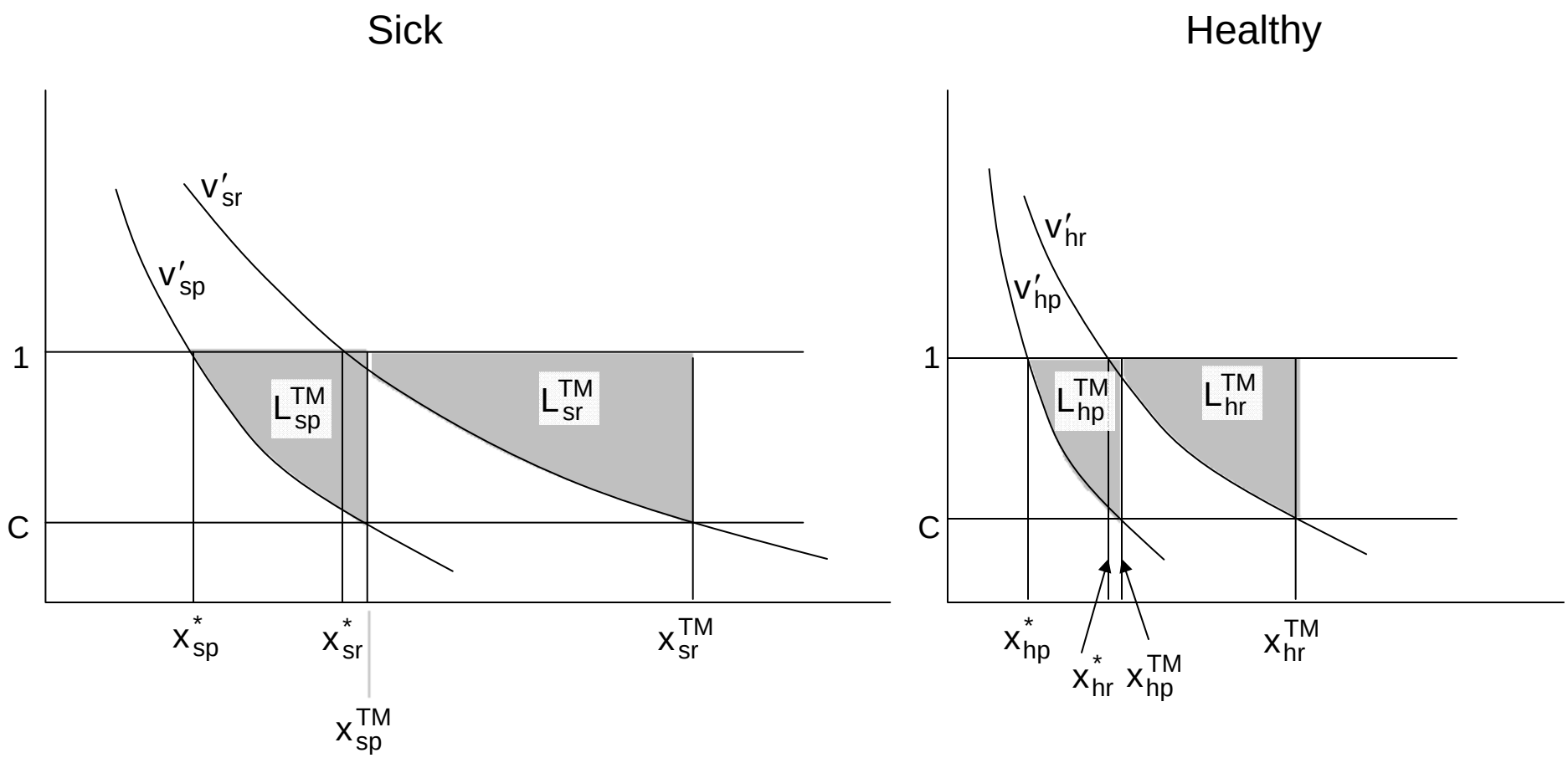

Figure 2: Spending and Inefficiencies in Medicare Advantage