NBER WORKING PAPER SERIES

IMPROVING RISK EQUALIZATION WITH CONSTRAINED REGRESSION

Richard van Kleef
Thomas McGuire
Rene van Vliet
Wynand van de Ven

Working Paper 21570
http://www.nber.org/papers/w21570

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2015

We gratefully acknowledge the Dutch Ministry of Health, the Association of Health Insurers and Statistics Netherlands for providing the administrative and survey data (see section 3). McGuire is grateful to support from the National Institute of Mental Health (R01 MH094290) and the Laura and John Arnold Foundation. The views expressed here are not necessarily those of the Foundation's officers, directors or staff. We also gratefully acknowledge the valuable comments on previous versions of this paper by Konstantin Beck, Frank Eijkenaar, Randall Ellis, Laura Hatfield, Timothy Layton, Erzo Luttmer, Joseph Newhouse, Sharon-Lise Normand, Sherri Rose, Erik Schut, Mark Shephard and Suzanne van Veen. In addition, we thank Andrew Hicks for creating the figures. The responsibility for the content of this article rests fully with the authors. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Improving Risk Equalization with Constrained Regression
Richard van Kleef, Thomas McGuire, Rene van Vliet, and Wynand van de Ven
NBER Working Paper No. 21570
September 2015
JEL No. I10,I11,I13,I18

ABSTRACT

Several countries rely on regulated health plan competition to combine affordability of health plans with incentives for cost containment and quality improvement. Typically, these policies include premium regulation supplemented with risk equalization to compensate health plans for predictable variation in medical spending. An extensive empirical literature shows, however, that even the state-of-the-art risk equalization models undercompensate some risk groups and overcompensate others, leaving systematic incentives for risk selection. A natural approach to reducing under or overcompensation for a group is to include membership in that group as an indicator in the risk equalization model. For several types of indicators, however, inclusion can be problematic or infeasible. This paper introduces and illustrates an alternative approach to reducing over or undercompensation in such cases: constraining the estimated coefficients of the risk equalization model so as to limit over or undercompensation. Our empirical illustration is based on administrative data on medical spending and risk characteristics of nearly all individuals with basic health insurance in the Netherlands. We evaluate empirically the benefits of constraints in terms of reduced under or overcompensation on indicators omitted from the Dutch risk equalization model and their costs in terms of increased under or overcompensation on indicators included in the model. Our findings imply that the benefits of introducing constraints can be worth the costs. Constrained regression adds a tool for developing risk equalization models that can improve the overall economic performance of health plan payment schemes.

Richard van Kleef
Erasus University
PO Box 1738
3000 DR Rotterdam
The Netherlands
vankleef@bmg.eur.nl

Thomas McGuire
Department of Health Care Policy
Harvard Medical School
180 Longwood Avenue
Boston, MA 02115
and NBER
mcguire@hcp.med.harvard.edu

Rene van Vliet
Erasmus University
PO Box 1738
3000 DR Rotterdam
The Netherlands
r.vanvliet@bmg.eur.nl

Wynand van de Ven
Erasmus University
PO Box 1738
3000 DR Rotterdam
The Netherlands
vandeven@bmg.eur.nl

1. Introduction

Several countries have adopted elements of Alain Enthoven's model of regulated health plan competition (Enthoven, 1993), which combines affordability of health plans with incentives for cost containment and quality improvement.[1] A crucial element of Enthoven's model is the adjustment of health plan payments to predictable variation in medical spending, also referred to as risk equalization (RE). In the absence of premium regulation, RE mitigates incentives for health plans to risk rate their premiums and thereby improves affordability of health plans for the sick. In the presence of premium regulation – as is common in practice – RE mitigates incentives for risk selection and thereby improves incentives for health plans to accept and serve the sick as well as the healthy (Newhouse, 1996).[2]

Recent research has shown that even the state-of-the-art RE models – such as those used under the Affordable Care Act in the United States or those used under the Health Insurance Act in the Netherlands – systematically undercompensate groups of consumers in relatively poor health and overcompensate the complementary groups of consumers in relatively good health (Layton et al. 2015; Van Kleef et al., 2013), exposing health plans and consumers to incentives for risk selection. Risk selection threatens the performance of (regulated) health plan markets since it may reduce 1) the quality of care (because plans may have a disincentive to meet the preferences of the sick), 2) the efficiency of care (because risk selection may be a more cost-effective strategy

[1] By 'health plan competition' we mean competition among health insurers who offer one or multiple health plans. A 'health plan' refers to a health insurance product. All consumers who have the same 'health plan' have an identical contract with the same insurer concerning benefits coverage, cost-sharing, quality, services, etc. Since objectives and strategies of insurers can differ across health plans, this paper will speak of health plans instead of insurers as decision makers.

[2] Newhouse (1996) defines risk selection as actions by consumers and health plans to exploit unpriced risk heterogeneity and break pooling arrangements. Often the term selection is also used to refer to the outcome of these actions.

for plans to reduce medical spending than improving the efficiency of care), 3) the efficient sorting of consumers among plans (when market segmentation by risk elevates premiums for particular plans), and 4) the affordability of health plans to the sick (when the same market segmentation causes the sick to face higher premiums). To contend with these potential problems, researchers and policy makers work to improve the properties of health plan payment schemes.[3] In general, three strategies can be applied to reduce incentives for risk selection in regulated health plan markets: improving RE, increasing risk sharing (e.g. via mandatory reinsurance or risk corridors) and relaxing premium-rate restrictions. This paper focusses on the first strategy.

The conventional approach to improving RE adds new/better risk indicators to the RE model, thereby improving the fit at the individual level, as measured by an R-squared, and improving the fit at the group level, as measured by the degree of under or overcompensation for particular groups. If a group of interest (for example, persons with congestive heart failure) is included with a single yes/no indicator in an ordinary least squares (OLS) RE model, the properties of least squares ensure that the payment for this group will equal the average medical spending of this group.[4] For some groups, however, it can be problematic to include an indicator of membership in the RE model. Clear examples are groups based on prior use of medical services: if indicators for these groups are included in the RE model, increased utilization in one year may increase plan payment in the following year, which reduces health plans' incentives for efficiency. In this paper we study two concrete examples of such groups: users of 'home care' in the previous year and

[3] By 'health plan payment scheme' we mean the total set of payment features, which can include risk equalization, reinsurance, risk corridors and premiums among other features.

[4] The expected value of an OLS residual conditional on a dummy variable regressor is zero. The statistical residual from an OLS model is the individual-level over or underpayment in a RE model based on the regression coefficients.

users of ‘physiotherapy’ in the previous year. While these groups are known to have relatively high future spending, inclusion of an explicit indicator for these groups in the RE model may introduce incentives to overuse these services (Van Vliet and Van Kleef, 2015). Leaving these indicators out of the model, however, confronts health plans with incentives for risk selection, e.g. by skimping on the quality of home care and physiotherapy.

In practice, RE models are estimated by OLS regression. If indicators are considered inappropriate for inclusion in the RE model (hence referred to as omitted indicators), conventional RE ignores these indicators during model estimation. This paper proposes an *indirect* use of these indicators: rather than ignoring omitted indicators, we recognize them by constraining the estimated coefficients of the RE model to reduce under or overcompensation of the groups identified by these indicators. This constrained regression can improve compensation for these groups by exploiting the empirical correlation between omitted and included indicators. We evaluate the benefits of constraints in terms of reduced under or overcompensation on indicators omitted from the RE model and their costs in terms of increased under or overcompensation on indicators included in the model. We argue and illustrate empirically that the gains from a well-chosen constraint using omitted indicators can be worth the costs in terms of reduced fit for included indicators. Constrained regressions thus expand the tools available to researchers and policy makers for modifying the fit of health plan payment schemes.

Our empirical application is the national basic health insurance for curative care in the Netherlands, a well-established example of a regulated individual health plan market based on principles of regulated competition (Van de Ven et al., 2013). In spite of a sophisticated RE model, policy researchers have identified groups that are systematically under or overcompensated (Van Kleef et al., 2013). The two groups we study in this paper are known to be undercompensated by about 1,200 euro (home care group) and 900 euro (physiotherapy group) per person per year. While individual-level indicators for membership in these groups are available in administrative data, inclusion of these indicators in the RE model is problematic due to incentives to overuse these services. Leaving the indicators out of the model, however, is also problematic due to incentives for skimping on quality of home care and physiotherapy (Van Vliet and Van Kleef, 2015). By constraining the coefficients in the RE model selection incentives against users of home care and users of physiotherapy can be mitigated without introducing incentives to overuse these services.

The paper is structured as follows. Section 2 discusses the method of constrained regression and develops measures for empirically evaluating the costs and benefits when using constrained regressions in the context of RE. In Sections 3 and 4 we apply our approach to the Dutch RE model of 2015 using data on medical spending and characteristics of nearly all individuals with basic health insurance in the Netherlands (N=16.5 million). We explore using constrained regressions to address the undercompensation of the users of home care or physiotherapy in the previous year. We apply our measures from Section 2 to show that, generally, some reduction in undercompensation for indicators omitted from the RE model can be worth the increase in under or overcompensation for indicators included in the model. Section 5 discusses our main findings and their implications, and identifies possible next steps for making use of constraints to improve overall performance of RE models.

2. Theory and concepts

Constrained regression

Least squares regression methods choose values for a set of parameters, the estimated coefficients, to minimize the residual sum of squared differences between the actual and fitted values from the regression. A researcher may place constraints on the choice of the coefficients in this minimization for various reasons. One common reason for imposing a constraint is to test a hypothesis about a set of coefficients. For example, to test the hypothesis that earned and unearned income has the same effect on household consumption, a constraint can impose the restriction that the coefficients on these two types of income are the same. The researcher can compare the model fit with and without the constraint using an F-statistic to test whether the reduction in explained variance is statistically significant; if it is, the hypothesis of constant returns is rejected.

Our motivation for introducing a constraint is different, and is akin to methods of constrained optimization. Health plan payment schemes have multiple objectives subject to tradeoffs. For example, in the design of a public health insurance program, one objective may be to reduce financial risk of the population while another objective may be to reduce public expenditures, with a tradeoff between the two. The locus of efficient policies can be found by maximizing one objective subject to a given level of attainment of the other, by, for example, maximizing financial protection for the population for a given level of public expenditures.[5] By conducting

[5] A closely related and well-known application of this approach can be found in the Appendix "On Optimal Insurance Policies" of Kenneth Arrow's "Uncertainty and the welfare economics of medical care" (Arrow, 1963).

this maximization for different levels of public expenditure, the researcher can characterize the tradeoff between spending more public money and reducing population financial risk.

Introducing constraints into a RE model serves a similar purpose. Constrained least squares regression addresses selection incentives regarding *included* and *omitted* indicators simultaneously by pursuing the "usual" objective of a RE model – minimizing squared deviations at the individual level for the *included* indicators – subject to a maximum value of under or overcompensation for the *omitted* indicators. By varying the maximum value of this second objective, researchers can trace out the tradeoffs between fit on the included indicators and over or undercompensation on the omitted ones.[6]

Our approach to constrained regressions in RE is related to some previous literature. Glazer and McGuire (2002) proposed using constrained regression to address selection problems, where the constraints were derived from first-order conditions for plan profit maximization, with one constraint for each service provided by the plan. RE weights were best fitting given a set of linear constraints that guaranteed a balanced set of incentives for plans to fund all services. This theoretical approach has never been implemented empirically, probably due to the complexity of specifying the constraints. In addition, there was no obvious way to "tighten" or "loosen" the constraints as there is here with the magnitude of undercompensation being the target of the constraint. McGuire et al. (2013) and Eijkenaar et al. (2014) have used constrained regressions in the context of RE, though for purposes other than addressing selection incentives.

[6] The form of constraints we use is written out in the Data and empirical methods section of this paper. See, in particular, equation (2').

This paper focusses on situations where compensation for omitted variables is *desired*. This starting point is distinct from that of Schokkaert and Van de Voorde (2004) who study situations where compensation for omitted variables is *not desired*. They argue that when omitted indicators for which compensation is not desired (which they refer to as R-variables)[7] are correlated with indicators included in the RE model for which compensation is desired (C-variables), conventional RE leads to biased estimates of coefficients for C-variables since these will (partly) pick up the variation in spending due to the omitted R-variables. In Schokkaert and Van de Voorde's terminology our paper exclusively focusses on C-variables.

Evaluating incentives for selection in RE models

A major purpose of RE models is to mitigate incentives for plans to over or underserve groups among the population. Incentives to underserve enrollees with a mental illness, for example, are created if the payments a plan receives for members of this group fall below the costs they bring to the plan. A RE model that recognizes and pays more for persons with some mental illness diagnoses can reduce the gap between average costs and average payments. However, if the RE model recognizes some but not all mental conditions, a plan might seek to deter persons with mental illness from joining by limiting access to mental health services – an example of the inefficiency created by selection incentives. Incentives for a plan to "distort" its benefits away from the efficient mix to attract/deter have been studied theoretically since Rothschild and Stiglitz (1976),[8] and empirically since the beginning of the use of RE in public insurance programs (Pope et al., 2000). In the U.S. context, empirical evidence confirms that plans respond

[7] An example of an R-variable might be an indicator for smoking. Smoking may predict higher medical spending but it may be undesirable to "reward" a smoker by higher RE payments.
[8] The Rothschild-Stiglitz model was adapted to managed care health insurance by Glazer and McGuire (2000). See Breyer et al. (2012) for a recent review.

to this type of incentives in service provision.[9] In the Netherlands, several health plans have reported publicly that the imperfect RE discourages them from improving the quality of care for groups that are systematically undercompensated (Van de Ven et al., 2015).

Papers and reports concerned with incentives for selection first define the group or groups of concern and then compare average payment for members of the group to average medical spending by simulation methods. Evaluations of payment systems in Medicare and in the Marketplaces in the U.S. commonly employ "predictive ratios", a ratio with simulated RE payments for the group in the numerator and medical spending in the denominator. "Underpayment" is indicated if the predictive ratio is less than 1.0. Evaluating the RE model proposed for the Marketplaces, Kautter et al. (2014) created subgroups of individuals by predicted spending and computed predictive ratios for various subgroups ordered by these predicted spending.[10] In an evaluation of the CMS-HCC model used in Medicare, Pope et al. (2011) report predictive ratios for subgroups defined by disease, numbers of prior hospitalizations, demographic characteristics, and other factors.[11]

Other papers difference RE payments and spending to assess selection incentives, with the difference being referred to as "undercompensation" if payments are less than spending and

[9] Cao and McGuire (2003) in Medicare and Eggleston and Bir (2009) in employer-based insurance find patterns of spending on various services consistent with service-level selection among competing at-risk plans. Ellis et al. (2013) rank services according to incentives to undersupply them. Consistent with service-level selection, they show that HMO-type plans tend to underspend on services (in relation to the average) just as the selection index predicts. This pattern of spending is not observed among enrollees in unmanaged plans. See also Carey (2014).

[10] Defining subgroups of the population on the basis of predicted spending can lead to predictive ratios close to 1.0 even if the prediction model itself is weak, and therefore is not necessarily a mark in favor of the RE model. Unless the model itself does a good job at differentiating high from low cost individuals, a predictive ratio according to a ranking by predicted spending is not very informative.

[11] Other measures of individual and group fit have been proposed and applied in the literature (See Van Veen et al. 2015b for a review.)

“overcompensation” in the opposite case. Van Kleef et al. (2013) merged survey information with health claims for a subset of people in the Netherlands to calculate undercompensation for various groups of people, including those with low physical and mental self-rated health status and those reporting chronic conditions. In the current paper we track over and undercompensation, both for defining constraints and as a basis for our evaluation metrics.

In empirical research, both forms of measures, predictive ratios and monetary differences, are primarily applied to groups for whom an indicator is *not* included in the RE model. The reason is that under the ordinary least squares approach, RE models eliminate under and overcompensation for groups with indicators *included* in the RE model (see footnote 4). Under a *constrained* least squares approach, however, over and undercompensation can appear for the latter as well.

Missing from the literature is an accepted method for aggregating group-level measures of under and overcompensation to the entire population, or, in other words, there is no accepted summary measure for comparing the comprehensive performance of alternative RE models affecting multiple groups simultaneously. While we can agree that reducing undercompensation for a group of interest is an improvement for that particular group, what if a RE alternative decreases undercompensation for one group but increases it for another? Which RE model is preferred? These questions are directly relevant for this paper since the type of constraints applied here are expected to improve compensation for omitted groups but will generally worsen it for included groups. A weighted sum of under and overcompensations for all groups of concern (both omitted and included groups) is a natural basis for construction of a summary measure with the weight being the share of the population in the group of interest. In the next section, we propose a family of such measures that we apply later to examine empirically the effects of constraining estimated coefficients in order to modify under and overcompensation by the Dutch RE model.

A summary measure of potential selection inefficiency

A measure of "potential selection inefficiency" should do more than just summarize under and overcompensations for groups of concern; it should also value the potential inefficiency resulting from selection actions triggered by these under and overcompensations. Construction of such a summary measure comes with two challenges. The first is to define groups in the population for purposes of assessing under and overcompensation, and the second is to choose an appropriate efficiency weighting for under and overcompensations.

Groups can be defined according to one indicator or a combination of indicators, some of which may be included and some not included in the RE model. Some research defines groups according to a single geographic indicator under the thinking that a health plan might favor or disfavor certain regions because of systematic differences in medical spending, as was done in a study of Germany by Bauhoff (2012). Other research defines groups according to the services used, the idea being that a health plan could favor or disfavor primary versus some kinds of specialty care, for example, to encourage/discourage potential enrollees anticipating making use of those services.[12]

To explain our ideas about a summary measure, consider a mutually exclusive grouping of the population based on one or more sets of discrete indicators.[13] The indicator or indicators

[12] See Ellis and McGuire (2007) for implementation of this approach in Medicare and McGuire et al. (2014) for its application in Marketplaces.
[13] The discussion in this section draws on Layton et al. (2015) who derive a similar summary measure starting with conditions for profit maximization by a health plan.

partition the population into G mutually exclusive groups indexed by g with g = 1, …, G. We can then use data to determine:

s_g	the share of the population in group g, with $\sum_g s_g = 1$,
r_g	the average plan revenue for a person in group g,
c_g	the average plan cost for a person in group g,
$r_g - c_g$	under/overcompensation for group g.

Given these parameters, under and overcompensations can be summarized by $\sum_g s_g|r_g - c_g|$, i.e. the sum of absolute under and overcompensations weighted by the share of the affected population. We follow standard assumptions (used in calculation of both predictive ratios and over and undercompensation) by regarding *medical* spending as plan cost and figuring over and undercompensation for an *average* plan.[14] With this, in the Dutch context, over and undercompensation is solely a function of the RE payments.[15] Specifically, average plan revenues equals average predicted medical spending from the RE model and $|r_g - c_g|$ boils down to absolute residual spending for group g from the RE model. Moreover, $\sum_g s_g(r_g - c_g)$ equals zero.

As a next step, we incorporate the magnitude of the potential inefficiency associated with the selection incentives by weighting the under or overcompensation for each group. The statistical

[14] By doing so we avoid concern with selection across plans, differential premiums, and differential plan efficiency.

[15] In many health care schemes multiple payment features may coexist (e.g. risk equalization and reinsurance). Practical use of our summary measure should incorporate all these relevant features. For example, reinsurance figures into plan payments in Marketplaces in the US, and a high-cost pool, with similar effects as reinsurance was operative in the Netherlands. Simulating payments recognizing these plan payment features can be used in calculating predictive ratios or over/undercompensation (Layton et al., 2015; Van Kleef and Van Vliet, 2012).

and economic literature contains support for squaring the gap. Individual and group R-squared measures are obviously based on squared deviations.[16] In welfare economics, the efficiency loss associated with a price distortion (such as a tax) is proportional to the square of the distortion. We acknowledge, however, that the choice of the weighting function is an open issue, including the question of whether the same function should be applied to deviations for all groups. We will come back to these issues in the Discussion section. In our empirical analyses, we apply several alternative weighting functions, raising under and overcompensations to powers ranging from 1 to 2.[17]

We thus propose a summary measure of the form:

$$\mathcal{L} = \sum_g s_g \left| r_g - c_g \right|^p \quad \text{with } 1 \leq p \leq 2. \qquad (1)$$

The measure of "loss" $\mathcal{L}$ is intuitive, weighting under/overcompensation raised to a power by the share of the affected population, and is easy to compute. $\mathcal{L}$ has a minimum value of 0 and no upper bound. Comparing $\mathcal{L}$ for RE models is meaningful only for comparing models estimated on the same data with the same definitions of group membership. In our empirical analyses we estimate all RE models on the exactly the same data with exactly the same group definition. Under this procedure, RE scheme 1 will be said to be preferred to scheme 2 if $\mathcal{L}_1 < \mathcal{L}_2$.

[16] See Van Veen et al. (2015b) for discussion of the various statistical measures applied to evaluation of RE schemes.

[17] Layton et al. (2015) propose a metric similar to L with p = 2, squaring the payment-cost gap for each group, also appealing to the usual form of welfare loss in economics, in which the area of a welfare triangle is proportional to the square of a discrepancy between the first-best and the actual price. Their selection metric measures improvement in incentives over a payment system with no premium categories and no RE, and falls between 0 and 1. Lorenz (2014) considers situations in which over and undercompensation may impose different magnitude of losses, and estimators that would be appropriate for minimizing the asymmetric loss functions.

As we noted at the outset of this section the formulation of welfare loss in (1) depends on stipulation of the groups to study, which is also not a straightforward decision. In principle, definition of groups should be based on selection concerns like the extent to which groups may be vulnerable to over or underprovision of services by health plans. Taking these factors into account would ideally be based on an elicitation of the concerns of regulators and an analysis of what levers plans have to make discriminatory service decisions. We intend the current paper to be a "proof of concept" of the idea of using constraints in a RE model, and select a subset of the groups that are relevant in terms of social welfare. In the Discussion section we return to the question of how best to choose a grouping to guide design of RE schemes.

As we explain in the Data and empirical methods section, our groups are categorized by a set of indicators included in the current Dutch RE model and a set of indicators omitted from the model. The utility of constraints on regression coefficients emerges when at least one indicator is omitted from the RE model. We will use the included and the omitted indicators jointly to define mutually exclusive groups for the entire Dutch population, and compute loss $\mathcal{L}$ for this partition of the population. We also will compute the loss for the two sets of included and omitted indicators separately. These two partial classifications allow us to show the effect of constraints on groups identified by the included versus the omitted indicators. Tightening the constraint improves things for the omitted indicators but imposes a cost on fit among the included indicators.

3. Data and empirical methods

Data

The empirical analyses are based on administrative data including individual-level information on medical spending and risk indicators for almost the entire Dutch population in 2012 (N=16.5 million). These data come from various sources, including health plans, tax authorities and the registration service for social benefits.[18] The resulting merged data are those used to estimate the RE model for health plan payment in the Netherlands in 2015. As a first step in our analyses we faithfully replicate the Dutch RE model of 2015, such that our "base model" accurately indicates expected over and undercompensation for groups for 2015 (Eijkenaar et al., 2014). Our alternative RE models and simulations modify this base model and are estimated on the same data. Here we briefly describe the risk indicators included in the base model and provide some general statistics.

The Dutch RE model for 2015 is the product of more than twenty years of research and experience and includes the following risk indicators: 40 risk classes based on an interaction between age and gender, 25 risk classes based on the use of specified prescription drugs in the previous year referred to as pharmacy-based cost groups or PCG's (Lamers and Van Vliet, 2004), 16 risk classes based on diagnostic information from hospital treatment in the previous year referred to as diagnoses-based cost groups or DCG's (Prinsze and Van Vliet, 2007; Van Kleef et al., 2014), seven risk classes for people with high costs in multiple prior years referred to as

[18] In estimating the RE model for 2015, medical spending from 2012 has been adjusted to reflect mandatory coverage changes in the period 2013-2015. In a second stage coefficients were linearly adjusted for expected cost inflation. For reasons of simplicity we excluded the second stage from our analysis.

multiple-year high cost groups or MHCG's (Van Kleef and Van Vliet, 2012), five risk classes based on the use of durable medical equipment in the previous year referred to as durable medical equipment cost groups or DMECG's (Van Kleef and Van Vliet, 2010), four risk classes based on an interaction between two age groups and yes/no 'PCG+DCG+MHCG+DMECG>0', 12 risk classes based on an interaction between socioeconomic status and age, 10 risk classes based on regional characteristics and 19 risk classes based on an interaction between source of income and age. All risk indicators have been carefully developed in research programs sponsored by the Dutch Ministry of Health. For further details on these risk indicators, see Van Kleef et al. (2013).[19]

The RE model of 2015 was estimated by a least-squares regression with medical spending in 2012 as the dependent variable and the risk classes described above as 138 independent dummy variables.[20] Medical spending includes the expenses on primary care, pharmaceuticals, hospital inpatient and outpatient care, maternity care, obstetrics and medical devices among other categories, but excludes expenses on mental health care and home health nursing care.[21] Although the latter two categories of spending are included in the mandatory benefit package of 2015, they are omitted from the main RE model, with funds allocated for them using a separate RE model. In this paper we will be concerned with the primary RE model used to allocate more

[19] A complete description of the Dutch payment system would include some subsidiary (and less well-developed) RE models for some small cost categories, and description of some of the regional adjustments built into the main model. These details are not important for our simulations. Readers are referred to Eijkenaar et al. (2014) for full details on the RE models.

[20] Analogous to the RE methodology in the Netherlands, medical spending in 2012 (dependent variable) is annualized by dividing actual individual-level costs of 2012 by the fraction of the year an individual was enrolled in the basic health insurance in 2012. Subsequently, this 'fraction' is included as a weight in the estimation model and computation of means. For example, an individual who was enrolled for 6 months and had 1,000 Euros expenses is given annualized costs of 2,000 Euro (1,000/0.5) and a weight of 0.5.

[21] Since medical expenses for home care itself are excluded, the undercompensation of about 1,200 Euro on users of home care implies that prior use of home care has predictive value for other types of medical expenses. A possible explanation is that users of home care have a relatively high probability of dying.

than 80 percent of health care costs among health plans in the Netherlands in 2015 (Eijkenaar et al., 2014). We refer to the RE model of 2015 as the base model.

Table 1 provides some univariate information on the prevalence of risk characteristics and the distribution of medical spending in our data. For simplicity of presentation we report aggregated risk categories instead of all 138 explanatory variables separately. Average spending in the population equals 1,848 Euro per person per year. Not surprisingly, average spending is relatively high for people age 65 years or older, those who receive a disability benefit, people living at an address with more than 15 residents (which approximates being in an institution for long-term care) and those in a PCG, DCG, DMECG and/or MHCG. The latter four are the most direct indicators of morbidity; nearly 23 percent of the population is classified by at least one of these morbidity indicators.

[Table 1]

In addition to the administrative data, we use health survey information to assess how constraints on undercompensation for one omitted group affect estimates of over and undercompensation for a series of other omitted groups of interest. The survey was conducted in 2011 among a representative sample of the Dutch population and includes a broad range of questions on general health status, physical impairments, mental health problems, particular chronic diseases and prior utilization of medical care. A unique, anonymous person identifier allows merging the survey information with the administrative data. We calculate under and overcompensations for survey groups as the predicted expenses from an RE model estimated on the administrative data (N=16.5 million) minus the actual expenses. In contrast to the administrative data, the survey data are available only for a small sample (N=14,310) of the population, implying that under or overcompensation for groups identified from the survey may be vulnerable to random variation. We report on groups for which under or overcompensation by the base model is statistically significant. For the specific definition of these groups see Van Veen et al. (2015a, Appendix 2).

Selecting study indicators to include and omit from the RE model

We select a set of indicators included in the base RE model and a set omitted from this model in order to study how constrained regression methods affect fit for the two types of indicators. We select the DCG grouping as our included indicator since it is one of the most direct indicators of morbidity in the Dutch RE model and because it is already mutually exclusive (contrary to PCG's). The Dutch DCGs are a hierarchical categorization of persons based on selected diagnostic information from inpatient or outpatient hospital treatment in the previous year.[22] Persons are classified in a DCG if they received at least one of these selected treatments in the previous year. The Diagnostic Cost Group (DCG) categorization partitions the population into 16 mutually exclusive groups, from 0 to 15. As shown in Table 2, the "0" group, those with none of the selected treatments in the previous period, account for 91% of the population. Where the "0" group has below average medical expenses, the higher DCGs have above average expenses. Since all DCGs are explicitly included as dummy regressors in the Dutch RE model, average predicted spending for these groups perfectly fits average actual spending (see footnote 4), implying that for all DCGs the average under/overcompensation is zero (see Table 2).

[22] The DCG classification was developed in several steps. First, a team of medical experts carefully determined whether or not diagnoses refer to a chronic condition. Diagnoses referring to a chronic condition were categorized in 144 more or less clinically homogeneous groups, which – in a next step – were clustered into 15 DCGs based on residual cost (according to a prediction model including explanatory variables based on age, gender and PCGs) using Ward's hierarchical clustering method. If enrollees have multiple diagnoses that would fall into different DCGs, they are classified in only one DCG, i.e. the one with the highest estimated coefficient. For further details about the Dutch DCG's see Van Kleef et al. (2014).

Our omitted indicators are ‘yes/no use of home care in the year t-1’ and ‘yes/no use of physiotherapy in the year t-1’.[23] Given the zero-sum principle of the Dutch RE model (and the constrained models estimated in our empirical analyses), reductions in undercompensation for the “yes” groups imply corresponding reductions in overcompensation for the complementary “no” groups. For example, if the undercompensation for users of home care reduces by 40 percent, the overcompensation for the complementary group of non-users will reduce by 40 percent too. For simplicity of presentation we primarily focus on the two “yes” groups. As shown in Table 2, these two groups comprise 2.7% and 2.4% of the population in year t, respectively. Both groups are systematically undercompensated by the current Dutch RE model, but inclusion of these indicators in the RE model is regarded as problematic since this would introduce incentives to overuse these services (Van Vliet and Van Kleef, 2015). For some analyses we convert the two yes/no indicators into four mutually exclusive categories by crossing the indicators and classifying the population as having none, home care only, physiotherapy only, or both indicators in the previous year.

[Table 2]

<u>Constraining coefficients in the RE models</u>

We introduce a series of constraints to the base RE model that limit the under/overcompensation for one or more omitted indicators to a fixed amount. Undercompensation is limited by

[23] These indicators are based on the use of home care and physiotherapy as far as covered by the Dutch basic benefit package of 2015. This benefit package fully covers the use of home care. Physiotherapy is only covered for treatment of certain chronic conditions and above a threshold of 20 visits.

specifying in a constraint that the average payment for an omitted group is equal to certain value. For any person i, the RE payment is $Y_i = \sum_j b_j x_{ij}$ where x_{ij} is the value of the included 0/1 indicator j for person i, and b_j is the weight on the indicator in the RE formula. If the number of people in the group of interest g is n_g the average payment for a member of group g is:

$$\overline{Y}_g = \frac{1}{n_g} \sum_{i \in g} \sum_j b_j x_{ij} \qquad (2)$$

This can be rewritten as:

$$\overline{Y}_g = \sum_j b_j \bar{x}_{gj} \qquad (2')$$

Where $\bar{x}_{gj}$ is the mean value of indicator variable j for group g. This group mean must be calculated on an initial pass through the data. The constraints then take the form of setting $\overline{Y}_g$ equal to a target value which can be easily implemented with the RESTRICT statement in the PROC REG procedure in SAS. This constraint is simply an equation linear in the coefficients of the RE model, resulting in coefficient estimates that maximize the fit of the model as measured by an R-squared given that the compensation for g equals the specified value. The target value for $\overline{Y}_g$ can be chosen as any amount; here, we reduce undercompensation for the omitted group(s) by a fixed percentage in relation to the base model. Since the constraint will bind, the constrained models will always yield a lower R-squared than the unconstrained base RE model.

Empirical analyses

As shown in Table 2, the Dutch RE model of 2015 (our "base model") leads to an average per person undercompensation of 1,231 Euro for users of home care in the previous year and 922 Euro for users of physiotherapy in the previous year. Given these undercompensations, we begin by estimating a series of constrained regressions, where in each case there is just one constraint. For each of the two omitted indicators we reduce the undercompensation in series by 20%, 40%, 60%, 80% and 100%. For each model we calculate measures of overall fit (R-squared and Cummings Prediction Measure (CPM)) as well as measures for group fit (under or overcompensation for included and omitted groups).[24] To summarize overall potential selection inefficiency, we track a series of loss measures based on the expression in (1). We measure loss for the mutually exclusive set of four groups identified by the omitted indicators, the mutually exclusive set of the 16 included DCG's, and for the cross-product of the two sets (64 groups). We also consider a range of powers to apply to the payment gap for each group of interest in order to check the sensitivity of results to the form of the loss function.

In a next step, guided by the results for the single constraints, we try several combinations of constraints for the two omitted indicators to check if two constraints can produce better overall model performance than any single constraint. We find a superior two-constraint specification that performs better than any model with a single constraint.

[24] R-squared is computed as = $1-\frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$. CPM is computed as $1-\frac{\sum_{i=1}^{n}|Y_i-\hat{Y}_i|}{\sum_{i=1}^{n}|Y_i-\bar{Y}|}$.

4. Results

Table 3 shows results for the base RE model and for the same model supplemented with a series of single constraints to reduce undercompensation for the group with use of home care in year t-1. As expected, the R-squared is lower for the constrained models than for the base model. The incremental loss in R-squared goes up as the constraint is more binding, but the absolute magnitude of the reduction in R-squared is always very small: the most binding constraint in which the undercompensation for users of home care in year t-1 is completely eliminated decreases the R-squared by only 0.3 percentage points. Thus, in terms of the R-squared, the costs of the constraint appear to be very low. In terms of the CPM (not bound to fall in a "least-squares" regression) the constrained model even leads to better fit than the base model, though the actual improvement is relatively minor.

[Table 3]

To assess group-level fit, Figure 1 presents results for the sets of included and omitted groups. The capital letters H and P represent the undercompensation in year t for users of home care and physiotherapy in year t-1, respectively. By design, the undercompensation for users of home care in year t-1 is smaller for the constrained models than for the base model. More interesting is the reduction of undercompensation for users of physiotherapy in year t-1, showing that reducing undercompensation for one omitted group can also improve compensation for another omitted group. Apparently, certain risk indicators in the RE model are positively correlated with both the home care group and the physiotherapy group. When weights on these risk indicators are altered by the constraint, undercompensation for the physiotherapy group is reduced as well.

Figure 1 also shows that, as expected, the constraint for the omitted group introduces over and undercompensation for the group indicators included in the RE model. The small numbers represent the overcompensation for the 15 DCGs while the large '0' represents the undercompensation for those without a DCG. For the latter group the constrained models introduce an undercompensation up to 40 Euros per person per year; for the DCGs the models with constraints introduce overcompensation up to 1,280 Euro per person per year. The direction of these under and overcompensations can be explained by the positive correlation between the omitted groups and the DCG's: since the home care and physiotherapy groups have relatively high proportions of people in a DCG (not shown here), the constrained model overcompensates the DCG's in order to move funds to these omitted groups.[25] Like the reduction in undercompensation for users of home care, the change in under or overcompensation for the other groups in Figure 1 is also linear, a consequence of constrained least-squares estimators with linear constraints.[26]

[Figure 1]

Figure 2 reports the analogous results for the base model supplemented with a series of single constraints reducing the undercompensation for users of physiotherapy in the previous year.

---

[25] Payment weights emerging from a constrained regression hear some similarity to the finding in Glazer and McGuire (2000) that optimal risk adjusted weights should reflect the correlation between indicators part of the risk equalization and omitted factors affecting health care expenses.

[26] This can be shown by writing out the formula for the estimated coefficients in constrained least squares. Suppose we seek to estimate a vector β of regression coefficients on j variables X subject to q linear constraints of the form Q^T where Q is a jxq matrix of full rank and the superscript T indicates the transpose of a matrix. Y is a qx1 vector of constants. One of these q equations could be interpreted in our context as a constraint that the average payment for a target group is equal to an amount "y." The constrained estimator is

$\hat{\beta}^c = \hat{\beta} - (X^TX)^{-1}Q(Q^T(X^TX)^{-1}Q)^{-1}(Q^T\hat{\beta} - y)$.

From this it is clear that $d\hat{\beta}^c/dy$ is constant.

Patterns are similar to those in Figure 1, with the difference that the constraint regarding physiotherapy leads to bigger changes in under or overcompensation for other groups (both the included groups and the other omitted group).

[Figure 2]

Introduction of a constraint involves a trade-off between a reduction of undercompensation for omitted groups and an increase in under or overcompensation for the included ones. When it comes to incentives for risk selection, however, it is not only the under or overcompensation that matters, but also the size of the affected group.[27] Figure 3 combines these two aspects, showing selected results for three of the RE models from Figures 1 and 2. The height of the bars indicates the average under or overcompensation for a group and the width indicates the relative size of the group. The product of height and width, as represented by the area of the bars, indicate the total under or overcompensation for a group. The right side of Panel A shows the undercompensation for the two omitted indicators in the base model. The left side of the panels tracks the overcompensation for the larger of the included groups (DCGs with at least 1% of the population included). With no constraint, least-squares estimators eliminate over or undercompensation in the base model for the included groups. Panel B shows results for the restriction of reducing the undercompensation for the home care group by 80% compared to the base model. Undercompensation for both omitted groups falls as was reported in Figure 1, and overcompensation appears for the DCGs shown in the Figure. Panel C shows the same set of results for one of the models in Figure 2. Overall, Figure 3 illustrates that constraining

[27] Next to incentives for risk selection, another relevant aspect of the trade-off between a reduction of undercompensation for omitted groups and an increase in under or overcompensation for the included groups may be the potential effects of selection actions (that may differ across these groups). We will elaborate on this in the Discussion section.

undercompensation for one omitted group pushes funds towards that specific group as well as to the other omitted group, and to “sick people” in general, at least as indicated by a DCG. For the DCGs in Figure 3 this appears as an overcompensation for members of these groups.

[Figure 3]

Though the group results in Figures 1, 2 and 3 are informative, they cannot evaluate trade-offs between the improvement for omitted groups and the worsening for included groups. As argued in Section 2 of this paper, evaluating the trade-off can be based on a mutually exclusive grouping of individuals and a weighting of under and overcompensation, as is done by our loss measure proposed in Section 2. The essence of the loss measure in (1) is that weighted under or overcompensations are computed and aggregated for mutually exclusive groups.

Figure 4 illustrates application of our loss measure to the base model and to the same model with a single constraint for reducing the undercompensation for the home care group. The measure is computed according to formula (1) with p=2, separately for two sets of mutually exclusive groups: the four combinations of yes/no home care use and yes/no physiotherapy use in the previous year (solid line) and the 16 DCG-groups (dotted line). In the case of p = 2, we refer to our loss measure as the weighted mean squared deviation (WMSD). The results clearly show the trade-off between the improvement for the omitted groups and the deterioration for included groups.

[Figure 4]

Whereas Figure 4 illustrates application of the loss measure separately for omitted and included groups, Figure 5 integrates the losses on omitted and included groups in a single measure. As in Figure 4, the measure is calculated according to formula (1) with p=2, but this time for all 64 mutually exclusive combinations of the four omitted and 16 included groups. Up to an 80 percent reduction in undercompensation for home care the constraint reduces the overall loss, but further tightening the restriction for the omitted group increases the overall loss because of deterioration in the fit of compensation for the included groups. Based on these results we conclude that for mutually exclusive combinations of the selected included and omitted groups studied in this paper, a well-chosen constrained model can outperform the base model.

[Figure 5]

To check sensitivity of our results to different assumptions about the loss function, we calculated the summary measure (as presented in Figure 5) for values of p ranging from 1.0 to 2.0. Figure 6 displays the normalized values of weighted mean absolute deviations (WMAD, a more general term to describe our loss function for values of p other than 2.0) for the end points of 1.0 and 2.0. These are normalized so that the WMAD for each set of model comparisons is set at 100 for the base model. The pattern of loss reduction is similar for p = 1 and p = 2 (also for the intermediate values of p not shown). In Panel A, loss falls as undercompensation is reduced for users of home care, but after some point in the 60-80% reduction range, losses go up. The findings for reducing undercompensation for users of physiotherapy shown in Panel B are similar. For both weights of the over and undercompensation, although the exact minimum varies slightly, the same U-shape describes the results. At least in the application we consider here, the finding that a moderate reduction in undercompensation minimizes losses is insensitive to reasonable weights for the absolute value of the over and undercompensation.

[Figure 6]

The results presented above clearly show that for a mutually exclusive set of the selected included and omitted indicators, a single-constraint model can outperform the base model. A natural next question is whether adding a second constraint to the model can lead to further improvements. To check this in our case we started with the most effective single constraint for home care according to our loss function with p = 2: reducing the undercompensation for home care by 80%. (This is the minimum of the U-shaped dotted line corresponding to p = 2 from Figure 6 Panel A). We introduced the additional constraint that the undercompensation for physiotherapy should be reduced 20% from the base model, and then should be reduced by 40%. The value of the loss function (the WMSD) fell by slightly more than 1% of its value with the first constraint at a target undercompensation for physiotherapy of 20% less, but then went up as the second constraint was tightened to the 40% drop in undercompensation. While the improvement obtained by introducing the second constraint is considerably less when the first constraint is roughly optimized, the results show that in terms of the loss measure applied here, a two-constraint model can outperform a single-constraint model.

As a final step we examined the effects of constraints for modifying undercompensation for home care and physiotherapy users on a series of other omitted groups of interest identified by health survey information. As described in the data section, the small sample size (N=14,310) implies that under or overcompensations for groups can be vulnerable to random variation. Figure 7 shows results for the ten groups for which the initial under or overcompensation by the base model is statistically significant ($p \leq 0.05$).[28] The results are striking: a single constraint for reducing the undercompensation for users of home care (or for users of physiotherapy) in the previous year can also substantially reduce under or overcompensation for other omitted groups. Apparently, certain risk indicators in the RE model are correlated with both the home care (physiotherapy) group and the groups presented in Figure 7. When weights on these risk indicators are altered by the constraint, this reduces under or overcompensation for the groups in Figure 7 as well.

[Figure 7]

For example, consider the group on the left-hand side of Figure 7, those reporting their health status to be in the lowest three categories: bad, poor or moderate, composing 19% of the population. These people are undercompensated by an average of 331 Euros in the base model. If we impose the constraint eliminating the undercompensation for home care, the undercompensation for the bad-poor-moderates falls to 118 Euros, and if we impose instead the constraint that we eliminate the undercompensation for physiotherapy, the undercompensation disappears altogether (to only 8 Euros). Remarkably, for all eight of the undercompensated groups, imposing either constraint has a meaningful favorable impact on the undercompensation. The constraints also improve payments for the groups that were overcompensated, as shown on the right-hand side of Figure 7. The 67% of the population with no chronic illness were overcompensated by 116 Euros in the base model, and this is cut to 32 with the home care constraint imposed and 16 if the physiotherapy constraint is imposed.

[28] The test is a two-sided t-test that the difference between payment and cost for individuals in each group equals zero.

The findings in Figure 7 have two important implications. First, the observation that under and overcompensations for the groups in the figure change substantially as a result of a constraint implies that an appropriate trade-off between the benefits and costs of a constraint requires involving *all* groups of interest. This calls attention to the questions, “What are the groups of interest?” and, “Are these groups equally relevant or are some more important than others?” We discuss these issues in the next section. Second, the direction of the changes in under and overcompensation in Figure 7 implies that the benefits of constrained regression may reach far: a *single* constraint intended to improve payment fit for one relevant group can lead to an improvement for many others.

5. Discussion

This paper proposed and illustrated an innovative approach to reducing selection incentives on group indicators omitted from a RE model: constraining the estimated coefficients of the RE model such that the under or overcompensation of the group identified by an omitted indicator equals a specified amount. This approach is particularly relevant for indicators that are inappropriate for inclusion in the RE model due to perverse incentives. Two examples of such indicators are the use of home care or physiotherapy in the previous year. While these are readily available in administrative data, inclusion in the RE model creates incentives to overuse these services. Leaving the indicators out of the model, however, creates incentives for skimping on quality of these services. Constrained regression can mitigate selection incentives against users of home care or physiotherapy without introducing incentives to overuse these services.

Compared to an unconstrained model, constraints by definition reduce model fit in terms of R-squared. Our empirical application of constrained regression to the Dutch RE model of 2015, however, shows that the magnitude of this reduction may be marginal. The Cummings Prediction Measure (CPM) changes little with the introduction of the constraints considered here. On the basis of our results, an R-squared or CPM should be supplemented with other measures when evaluating constrained regression models since R-squared and CPM appear to be insensitive to changes in group-level fit induced by the introduction of the constraints considered here.

At a group level, a constraint limiting under or overcompensation for an omitted group comes at the costs of introducing under or overcompensation for included groups. To assess this trade-off, we proposed a loss measure reflecting potential selection inefficiency. Our empirical application of this measure to a set of omitted and included groups shows that the improvement for omitted groups can outweigh the deterioration for included groups. Moreover, multiple constraints can reduce potential selection inefficiency over a single constraint. Although we study a particular application of constrained regressions, we have no reason to think these findings are special to this empirical setting. If an indicator for an omitted group of interest is correlated with variables already included in the RE model, it should generally be possible to introduce at least a modest constraint that makes first-order cuts in undercompensation for the group with the omitted indicator at the cost of only "marginal" over/undercompensation for groups based on included indicators. It will be worthwhile to investigate the conditions (if they exist) under which introduction of a constraint at the margin is associated with an improvement in selection efficiency, perhaps using envelope-theorem type arguments. In any case, the practical performance of constraints in a particular plan payment application is straightforward to assess systematically for each setting.

This paper is intended to be a "proof of concept." Ultimately, to be useful in terms of plan payment redesign, application of constrained regression methods requires defining the groups that are considered relevant targets for risk selection and valuing under or overcompensation for these groups in terms of valid measures of potential selection inefficiency. Hence we discuss possible next steps to guide practical development of constrained regression methods in RE settings.

Mutually exclusive groups: which and how?

In order to optimize constrained RE models it is important to define the groups that are considered relevant targets of risk selection. It seems self-evident that there will be a social concern for some groups with indicators not included in a RE model, as well as groups with indicators already included. Once the relevant groups are defined, an appropriate method must be developed for creating mutually exclusive groups. In our empirical illustration such groups were defined by unique combinations of relevant risk indicators. It is certainly possible to apply our metric for inefficiency based on equation (1) or some variant of that to any number of groups, but for real-world applications, a group based on all combinations of indicators may lose meaning in the presence of a large number of multiple indicators. For example, twenty yes/no variables each splitting the population in two groups would result in more than a million combinations. Such a comprehensive set of mutually exclusive groups would be problematic since numbers of individuals per cell can become too small for approximating *systematic* under and overcompensation due to random variation. Moreover, not all mutually exclusive groups will represent stand-alone targets for risk selection.

While there are practical difficulties in defining groups of interest, we believe our approach offers an opportunity to expand the role of regulators and public policy makers. Rather than being *reactive* to problems identified in empirical study of RE models, regulators can be *proactive* and take steps to define the objectives that will be maximized by RE model estimation. Further research is clearly necessary to find a process by which a social consensus can be reached about defining groups of concern for protection against incentives for selection.

Weighting under and overcompensation: why and how?

After defining the groups that are considered relevant targets of risk selection, another crucial step for optimizing constrained RE models is to value the potential inefficiency resulting from selection actions triggered by specific under and overcompensations. Economic and statistical analysis can supply some initial guidance in this area. Literature on selection incentives provides at least four arguments why potential selection inefficiency is not necessarily constant across groups and not necessarily proportional to the size of under or overcompensation. A first argument comes from Van Barneveld et al. (2000) who contend that small predictable profits and losses are likely to be irrelevant for a health plan. Selection can be costly and the net benefits are uncertain, and small incentives may simply not induce a health plan to act. A second argument derived from standard welfare economics is made by Layton et al. (2015) who show that the welfare loss from price distortions due to under or overcompensation by a RE model is proportional to the square of these measures, implying that the inefficiency from selection goes up more than proportionally with the magnitude of the under and overcompensation. A third argument can be drawn from the work by Ellis and McGuire (2007) who argue that selection incentives do not just depend on an indicator's predictiveness (how well the indicator co-varies with total health care spending) but also on predictability of that indicator (how well the indicator can be anticipated) and demand responsiveness of individuals scoring on that indicator. For example, Ellis and McGuire find that both 'use of durable medical equipment' and 'use of anesthesia' are indicators with high predictiveness but that the first indicator is much more predictable (and therefore much more vulnerable to service level distortion) than the latter. This point implies that potential selection inefficiency of under or overcompensation for one group can be larger than that of an equal under or overcompensation for other groups. A fourth argument comes from Van de Ven et al. (2015) who contend that potential selection inefficiency depends on the specific selection actions that can occur as a consequence of under or overcompensation. They distinguish many selection actions, such as selective advertising, offering choice of deductible, making supplementary insurance (un)attractive for certain groups, offering group contracts and quality skimping. They argue that of all possible selection actions 'quality skimping' is a special threat to the functioning of regulated health plan markets because it not only reduces market efficiency, but *also* the quality of medical care. Incentives for quality skimping, however, are only present when groups with relatively strong preferences for high quality are *under*compensated.[29] After all, if these groups would be overcompensated health plans would have incentives to *improve* quality of care. Thus, for groups with strong preferences for high quality – presumably those with a chronic condition – undercompensation may be worse than overcompensation.[30]

Once under and overcompensations are valued, a loss function can readily be specified. For example, based on the idea of Van Barneveld et al. (2000) that small under or overcompensation can be ignored, the loss function could be rewritten to "count" over and undercompensation above certain absolute threshold values, as done by Beck et al. (2010). Above the threshold value, potential effects of different forms of selection can be included by using group-specific

[29] In a zero-sum RE payment scheme, undercompensation of the chronically ill implies overcompensation of the complementary group of healthy individuals, vice versa.

[30] Note that constrained regression can be a tool for changing undercompensation for groups of chronically ill into overcompensation. A related argument is made by Lorenz (2014) who also identifies empirical methods that weight over and undercompensation asymmetrically.

weights. Minimization of the loss function with respect to weights on included indicators, subject to constraints for modifying under/overcompensation on group indicators omitted from the model, implies a new way for estimating parameters of a RE model and modifying selection incentives.[31]

Although evidence is lacking to be able to definitively fine tune a loss function, the exercise of specifying a metric for economic inefficiencies caused by under or overcompensation should be a useful one for both policymakers and researchers. Without a firm consensus on a particular form of a loss function, some sensitivity analysis will be necessary to bound a reasonable range of results.

Potential applications

This paper has proposed constrained regression as an instrument for reducing under or overcompensation of groups defined by prior utilization for which it may be problematic to include an indicator of membership in the RE model due to incentives for overuse. The extent to which inclusion of such indicators may actually induce incentives for overuse, however, may differ from case to case. In general, these incentives depend on the payment weight (i.e. the regression coefficient) for the indicator in relation to the marginal cost of providing additional care to enhance higher RE payments. If the payment weight exceeds these marginal costs, incentives for overuse are present; in the opposite case they are absent. This suggests that an indicator based on prior utilization can be included in the RE model as long as its estimated coefficient does not exceed the marginal costs of enhancing higher RE payments. Although not

[31] For optimization of constrained RE models it may be interesting to consider use of inequality-type constraints (instead of equality-type constraints applied in this paper).

applied in our analyses, such a rule can be easily implemented in a constrained regression: constrain the estimated coefficients of the RE model such that 1) the *coefficient* for indicator g equals a specified value and 2) the *predicted spending* for group g equals a specified amount. An advantage of this ‘partial’ inclusion of indicator g in the RE model compared to the modality applied in this paper (no inclusion of indicator g at all) may be that the constraint is less binding.

In addition to avoiding incentives for overuse, there may be other reasons for using constrained regression instead of including a risk indicator directly in the RE model. Van de Ven and Ellis (2000) argue that for inclusion in the RE model indicators should meet criteria such as “appropriateness of incentives” and “feasibility.” The first does not only concern incentives for overuse, but also implies that RE indicators should not be “gameable”, that is, not be subject to discretionary coding by plans or providers seeking to enhance revenues. Recent research finds substantial “upcoding” in health plans paid by capitation in the US Medicare program (Geruso and Layton, 2015). Practical considerations such as these led policymakers to prune the number of disease indicators used in the RE model applied under the Affordable Care Act in the U.S. (Kautter et al., 2014). In these cases, constrained regression provides a tool for reducing under or overcompensations without inducing perverse incentives. “Feasibility” means that the information underlying an indicator should be routinely available at low cost. Information from a health survey, for example, may identify a group of interest and be a highly informative predictor of medical spending, but too expensive to collect on a regular basis for everyone in a risk pool. Also in this case, constrained regression may be an interesting tool, since data requirements are less stringent than for inclusion of an indicator in the RE model. A constraint for reducing undercompensation of omitted group g does not require that information on yes/no membership of g is available for everyone in a risk pool. Instead, it is sufficient to have a good approximation of the average per person medical spending for g, as well as the mean of all risk indicators in the

RE model conditional on g (see equation 2’). Information on these parameters obtained from a decent *sample* of the risk pool could be sufficient.

A striking result of our empirical application is that the constraints do not only reduce undercompensation for the target groups (i.e. users of home care or physiotherapy in the previous year), but also for other omitted groups (see Figure 7). It seems doubtful that such reductions can be achieved when including an indicator for the home care or physiotherapy group directly in the RE model. The reason is that direct inclusion probably moves fewer funds to the groups in Figure 7 than constrained regression. Whereas constrained regression increases compensation for almost all morbidity groups in the RE model (here: PCGs, DCGs, DMECGs and MYHCGs, together representing more than 20 percent of the population), direct inclusion of home care or physiotherapy indicators in the RE model will only increase compensation for these particular groups (representing only 2.7 and 2.5 percent of the population respectively). This raises an interesting question for further research: “Can a constrained regression model (reducing under or overcompensation for group g) outperform an unconstrained model in which an indicator for g is *included* in the RE model?” To answer this question, however, researchers should not only focus on potential selection inefficiency, but also on other dimensions of market inefficiency such as the potential for overuse, gaming and upcoding resulting from inclusion of indicator g in the model. Ideally, all relevant dimensions of market inefficiency are taken into account in the loss function used for comparing different models.

Though additional conceptual thinking and empirical testing is necessary to guide meaningful application of constrained regressions in RE settings, this paper shows the method has potential to improve the overall economic performance of health plan payment schemes.

Acknowledgements

We gratefully acknowledge the Dutch Ministry of Health, the Association of Health Insurers and Statistics Netherlands for providing the administrative and survey data (see section 3). McGuire is grateful to support from the National Institute of Mental Health (R01 MH094290) and the Laura and John Arnold Foundation. The views expressed here are not necessarily those of the Foundation's officers, directors or staff. We also gratefully acknowledge the valuable comments on previous versions of this paper by Konstantin Beck, Frank Eijkenaar, Randall Ellis, Laura Hatfield, Timothy Layton, Erzo Luttmer, Joseph Newhouse, Sharon-Lise Normand, Sherri Rose, Erik Schut, Mark Shephard and Suzanne van Veen. In addition, we thank Andrew Hicks for creating the figures. The responsibility for the content of this article rests fully with the authors.

References

Arrow KJ. Uncertainty and the welfare economics of medical care. American Economic Review 1963; 5: 941-973.

Bauhoff S. Do health plans risk select? An audit study on the Germany's social health insurance. Journal of Public Economics 2012; 96: 750-759.

Beck K, Trottmann M, Zweifel P. Risk adjustment in health insurance and its long-term effectiveness. Journal of Health Economics 2010; 29: 489-498.

Breyer F, Bundorf MK, Pauly MV 2012. Health care spending risk, health insurance, and payments to health plans. In: Pauly MV, McGuire TG, Barros PP (Eds), Handbook of health economics, vol. 2. Elsevier North Holland: Amsterdam; 2012.

Cao Z, McGuire TG. Service-level selection by HMOs in Medicare. Journal of Health Economics 2003; 22(6): 915-931.

Carey C. Government payments and insurer benefit design in Medicare part D," Working Paper, 2014.

Eggleston K, Bir A. Measuring selection incentives in managed care: Evidence from the Massachusetts state employees insurance program. Journal of Risk and Insurance 2009; 76: 159-175.

Ellis RP, Jiang S, Kuo T-C. Does service-level spending show evidence of selection across health plan types? Applied Economics 2013; 45(13): 1701-12.

Eijkenaar F, Van Kleef RC, Van Veen SHCM, Van Vliet RCJA. Research for the Dutch risk equalization model of 2015: Estimation of risk equalization weights. Research report (in Dutch), Erasmus University Rotterdam; 2014.

Ellis RP, McGuire TG. Predictability and predictiveness in health care spending. Journal of Health Economics 2007; 26(1):25-48.

Enthoven AC. The history and principles of managed competition. Health Affairs 1993; 12: 24-48.

Geruso M, Layton T. Upcoding or selection? Evidence from Medicare on squishy risk adjustment. National Bureau of Economic Research Working Paper, number W21222; 2015.

Glazer J, McGuire TG. Optimal risk adjustment in markets with adverse selection: An application to managed care. American Economic Review 2000; 90: 1055-1071.

Glazer J, McGuire TG. Setting health plan premiums to ensure efficient quality in health care: Minimum variance optimal risk adjustment. Journal of Public Economics 2002; 84: 153-173.

Kautter J, Pope GC, Ingber M, Freeman S, Patterson L, Cohen M, Keenan P. The HHS-HCC risk adjustment model for individual and small group markets under the Affordable Care Act. Medicare & Medicaid Research Review 2014; 4(3): E1-E11.

Lamers LM, Van Vliet RCJA. The Pharmacy-based cost group model: validating and adjusting the classification of medications for chronic conditions to the Dutch situation. Health Policy 2004; 68: 113-121.

Layton TJ, Ellis RP, McGuire TG. Assessing incentives for adverse selection in health plan payment systems, National Bureau of Economic Research Working Paper, number 21531; 2015.

Lorenz, N. Using quantile regression for optimal risk adjustment, University of Trier, Research Papers in Economics No 11/14. May, 2014.

McGuire TG, Glazer J, Newhouse JP, Normand SL, Shi J, Sinaiko AD, Zuvekas S. Integrating risk adjustment and enrollee premiums in health plan payment. Journal of Health Economics 2013; 32: 1263-1277.

McGuire TG, Newhouse JP, Normand SL, Shi J, Zuvekas SH. Assessing incentives for service-level selection in private health insurance exchanges. Journal of Health Economics 2014; 35: 47-63.

Newhouse JP. Reimbursing health plans and health providers: Efficiency in production versus selection. Journal of Economic Literature 1996; 34: 1236-1263.

Pope GC, Ellis RP, Ash AS, et al. Principal inpatient diagnostic cost group model for Medicare risk adjustment. Health Care Financing Review 2000; 21(3): 93-118.

Pope GC, Kautter J, Ingber JJ, Freeman S, Sekar R, Newhart C. Evaluation of the CMS-HCC risk adjustment model. In final report, RTI project number 0209853.006, RTI International, March 2011.

Prinsze FJ, Van Vliet RCJA. Health-based risk adjustment: Improving the Pharmacy-based cost group model by adding diagnostic cost groups. Inquiry 2007; 44: 469-480.

Rothschild M, Stiglitz J. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. The Quarterly Journal of Economics 1976; 90: 629-649

Schokkaert E, Van de Voorde C. Risk selection and the specification of the conventional risk adjustment formula. Journal of Health Economics 2004; 23: 1237-1259.

Van Barneveld EM, Lamers LM, Van Vliet RCJA, Van de Ven WPMM. Ignoring small predictable profits and losses: A new approach for measuring incentives for cream skimming. Health Care Management Science 2000; 3: 131-140.

Van de Ven WPMM, Beck K, Buchner F, Schokkaert E, Schut FT, Shmueli A, Wasem J. Preconditions for efficiency and affordability in competitive healthcare markets: Are they fulfilled in Belgium, Germany, Israel, the Netherlands and Switzerland. Health Policy 2013; 109: 226-245.

Van de Ven WPMM, Ellis RP 2000. Risk adjustment in competitive health insurance markets. In: Culyer AJ, Newhouse JP (Eds) Handbook of Health Economics. North-Holland: Amsterdam; 2000. p. 755-845.

Van de Ven WPMM, Van Kleef RC, Van Vliet RCJA. Risk selection threatens quality of care for certain patients: Lessons from Europe's health insurance exchanges. Forthcoming in Health Affairs; 2015.

Van Kleef RC, Van Vliet RCJA, Van de Ven WPMM. Risk equalization in the Netherlands: An empirical evaluation. Expert Review of Pharmacoeconomics & Outcomes Research 2013; 13: 829-839.

Van Kleef RC, Van Vliet RCJA. Prior use of durable medical equipment as a risk adjuster for health-based capitation. Inquiry 2010; 47: 343-358.

Van Kleef RC, Van Vliet RCJA. Improving risk equalization using multiple-year high cost as a health indicator. Medical Care 2012; 50: 140-144.

Van Kleef RC, Van Vliet RCJA, Van Rooijen EM. Diagnoses-based cost groups in the Dutch risk-equalization model: The effects of including outpatient diagnoses. Health Policy 2014; 115: 52-59.

Van Veen SHCM, Van Kleef RC, Van de Ven WPMM, Van Vliet RCJA. Improving the prediction model used in risk equalization: cost and diagnostic information from multiple prior years. European Journal of Health Economics 2015a; 16: 201-218.

Van Veen SHCM, Van Kleef RC, Van de Ven WPMM, Van Vliet RCJA. Is there one measure-of-fit that fits all? A taxonomy and review of measures-of-fit for risk-equalization models. Medical Care Research and Review 2015b; 72: 220-243.

Van Vliet RCJA, Van Kleef RC. Follow-up research regarding the Dutch risk equalization of 2015: Predictive value of the prior use of district nursing for future healthcare expenses. Research report (in Dutch). Erasmus University Rotterdam; 2015.

Tables

Table 1 Population frequency and medical spending (in Euros, 2012) at aggregated levels of risk characteristics (N=16.5 million)

		Population frequency	Medical spending	
			Mean	SD
Men, <65		42%	1,207	5,893
Men, >=65		8%	4,612	11,050
Women, <65		41%	1,487	5,212
Women, >=65		9%	4,123	8,889
Region, clusters 1-5		50%	1,979	6,941
Region, clusters 6-10		50%	1,719	6,237
Source of income, reference group (age <18 or >64)		38%	2,477	8,235
Source of income, disability benefits		5%	3,817	10,570
Source of income, social security benefits		2%	2,321	7,110
Source of income, student		3%	588	2,717
Source of income, self-employment		4%	1,012	3,814
Source of income, other (including employment)		48%	1,282	4,541
Socioeconomic status, home address >15 residents		1%	4,507	10,219
Socioeconomic status, income deciles 1-3		30%	1,842	6,526
Socioeconomic status, income deciles 4-7		40%	1,869	6,527
Socioeconomic status, income deciles 8-10		30%	1,721	6,555
Pharmacy-based Cost Group (PCG)	No	82%	1,212	5,199
	Yes	18%	4,751	10,417
Diagnoses-based Cost Group (DCG)	No	91%	1,353	4,921
	Yes	9%	6,855	14,530
Durable Medical Equipment Cost Group (DMECG)	No	99%	1,772	6,382
	Yes	1%	10,933	17,099
Multiple-year High Cost Group (MHCG)	No	94%	1,378	4,957
	Yes	6%	9,536	17,056
PCG, DCG, DMECG and/or MHCG	No	77%	984	4,106
	Yes	23%	4,784	11,090
Total population		100%	1,848	6,597

Table 2 Population frequency, medical spending and under/overcompensation by the Dutch RE model of 2015 (base model) in Euros (2012) for the 4 omitted and 16 included indicators studied in our empirical analyses (N=16.5 million)

		Population frequency	Medical spending		Under/over compensation base model
			Mean	SD	
Omitted indicators:					
Use of home care in t-1	No	97.31%	1,659	5,985	34
	Yes	2.69%	8,696	16,541	-1,231
Use of physiotherapy in t-1	No	97.62%	1,737	6,313	23
	Yes	2.38%	6,422	13,124	-922
Included indicators:					
No DCG		91.00%	1,353	4,921	0
DCG1		0.67%	5,573	8,943	0
DCG2		1.49%	4,649	8,100	0
DCG3		1.11%	4,196	8,243	0
DCG4		1.80%	5,058	9,541	0
DCG5		1.16%	6,291	11,420	0
DCG6		1.26%	7,645	13,461	0
DCG7		0.55%	8,832	15,511	0
DCG8		0.12%	10,039	15,978	0
DCG9		0.30%	9,582	18,583	0
DCG10		0.33%	13,175	20,678	0
DCG11		0.04%	14,557	25,078	0
DCG12		0.07%	17,107	28,243	0
DCG13		0.04%	25,105	41,154	0
DCG14		0.04%	90,296	42,858	0
DCG15		0.01%	62,451	110,800	0
Total population		100%	1,848	6,597	0

Table 3 Results (Euros, 2012) for the base model and for ten single-constraint models

	R-squared (*100%)	CPM (*100%)
Base model	22.5%	24.8%
Base model + single constraint to limit undercompensation for users of *home care* in t-1 by:		
20%	22.5%	24.9%
40%	22.5%	24.9%
60%	22.4%	24.9%
80%	22.3%	24.9%
100%	22.2%	24.8%
Base model + single constraint to limit undercompensation for users of *physiotherapy* in t-1 by:		
20%	22.5%	24.9%
40%	22.5%	25.0%
60%	22.4%	25.0%
80%	22.2%	25.0%
100%	22.0%	24.8%

Figures

Figure 1: Results (Euros, 2012) for the base model and for the same model supplemented with a single constraint to reduce undercompensation in year t for users of home care in t-1

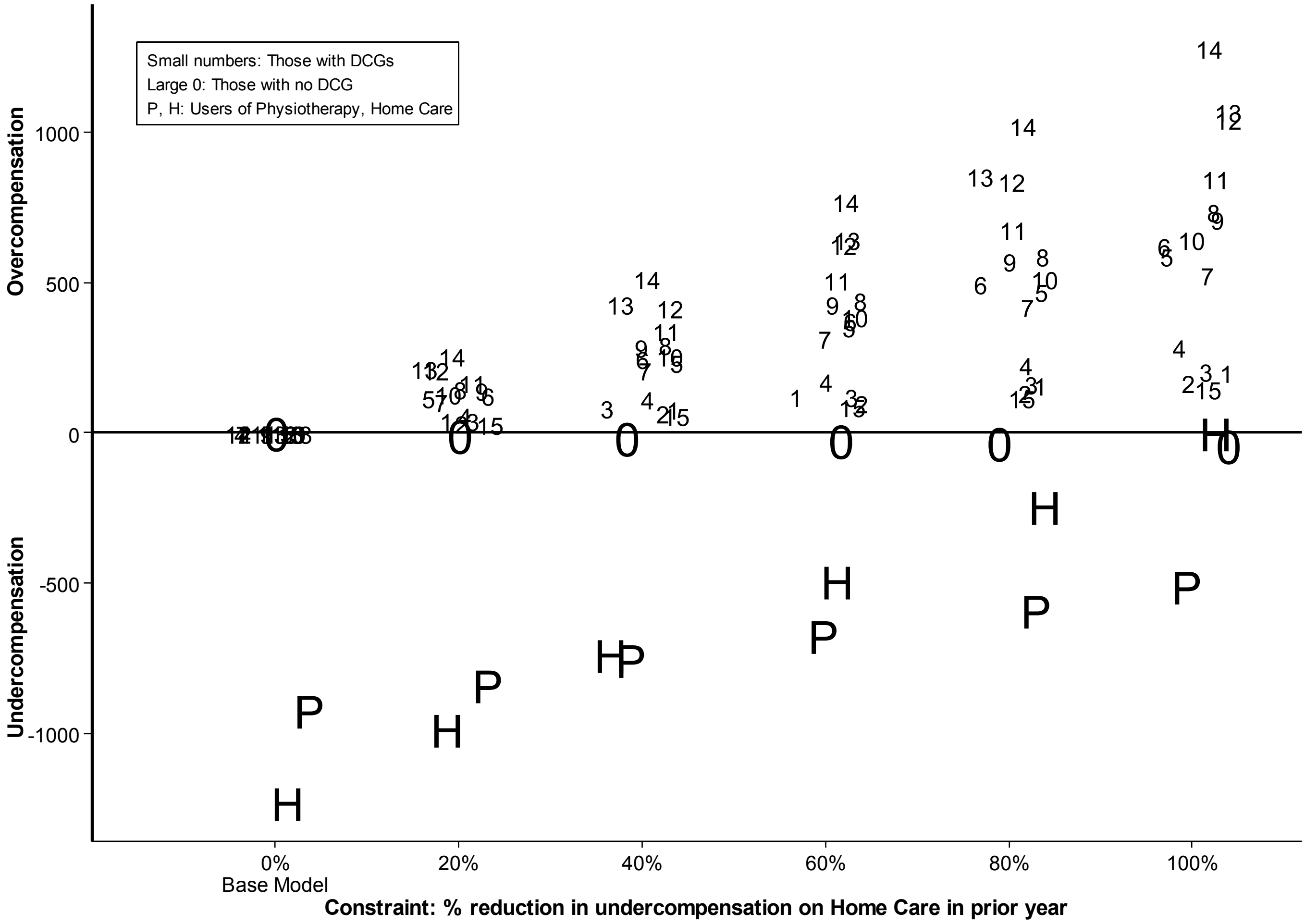


Figure 2: Results (Euros, 2012) for the base model and for the same model supplemented with a single constraint to reduce undercompensation in year t for users of physiotherapy in t-1

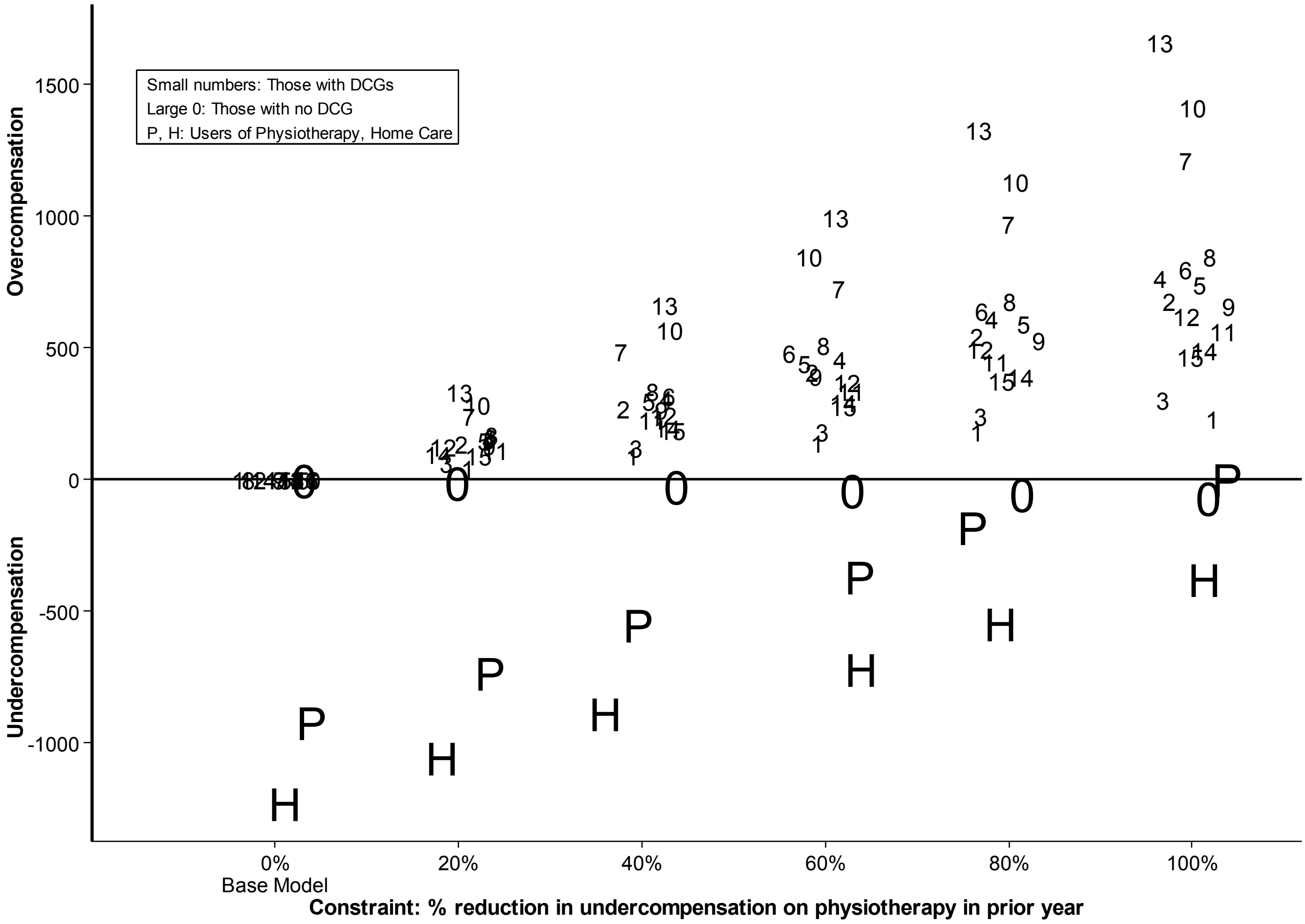


Figure 3

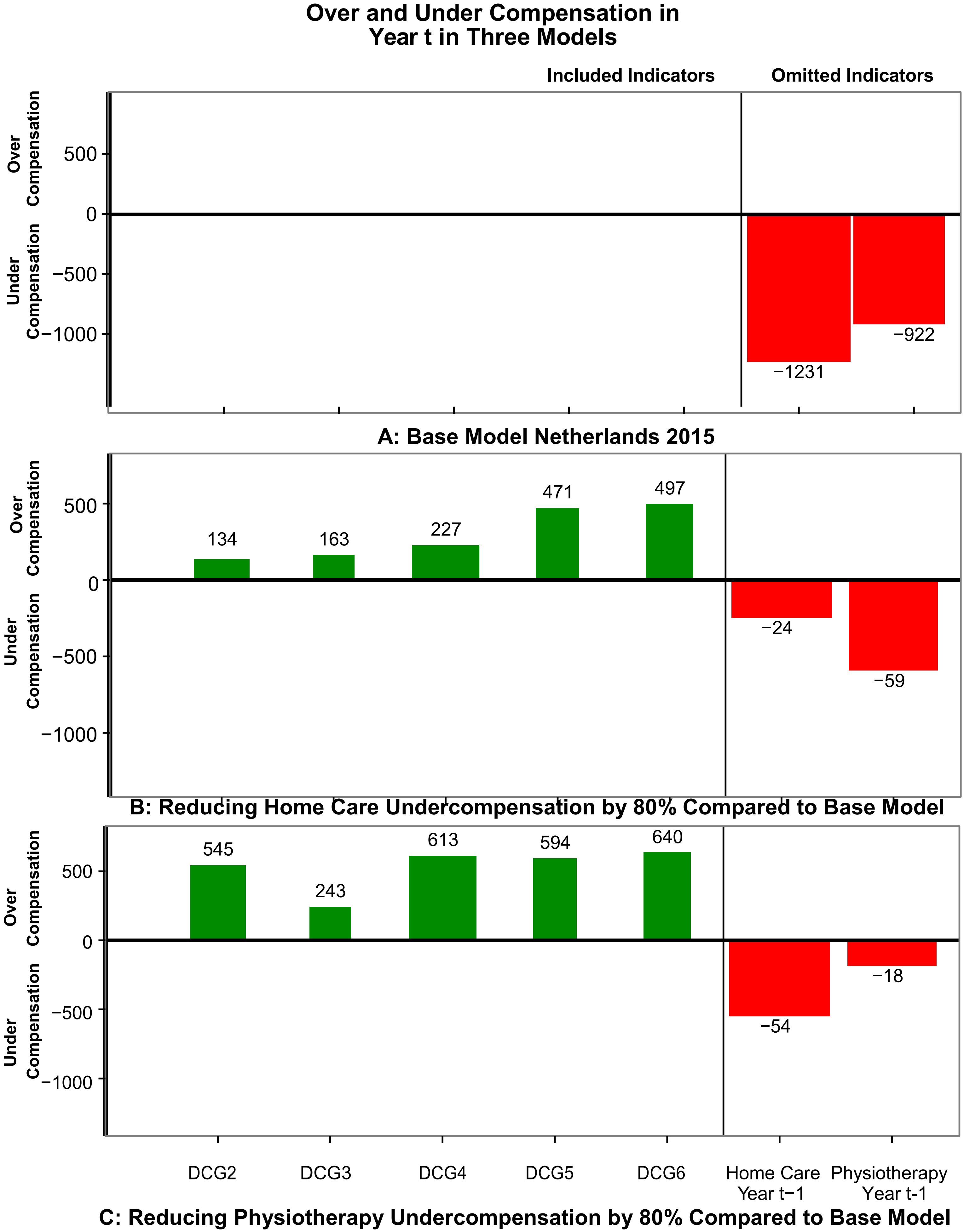

Over and Under Compensation in
Year t in Three Models
Included Indicators
Omitted Indicators
Over Compensation
Under Compensation
500
0
−500
−1000
−1231
−922
A: Base Model Netherlands 2015
134
163
227
471
497
−24
−59
B: Reducing Home Care Undercompensation by 80% Compared to Base Model
545
243
613
594
640
−54
−18
DCG2
DCG3
DCG4
DCG5
DCG6
Home Care Year t−1
Physiotherapy Year t-1
C: Reducing Physiotherapy Undercompensation by 80% Compared to Base Model

Figure 4: Base model with constraints on undercompensation for users of Home Care in prior year: Weighted Mean Squared Deviation (WMSD) for two sets of mutually exclusive groups based on included and omitted indicators

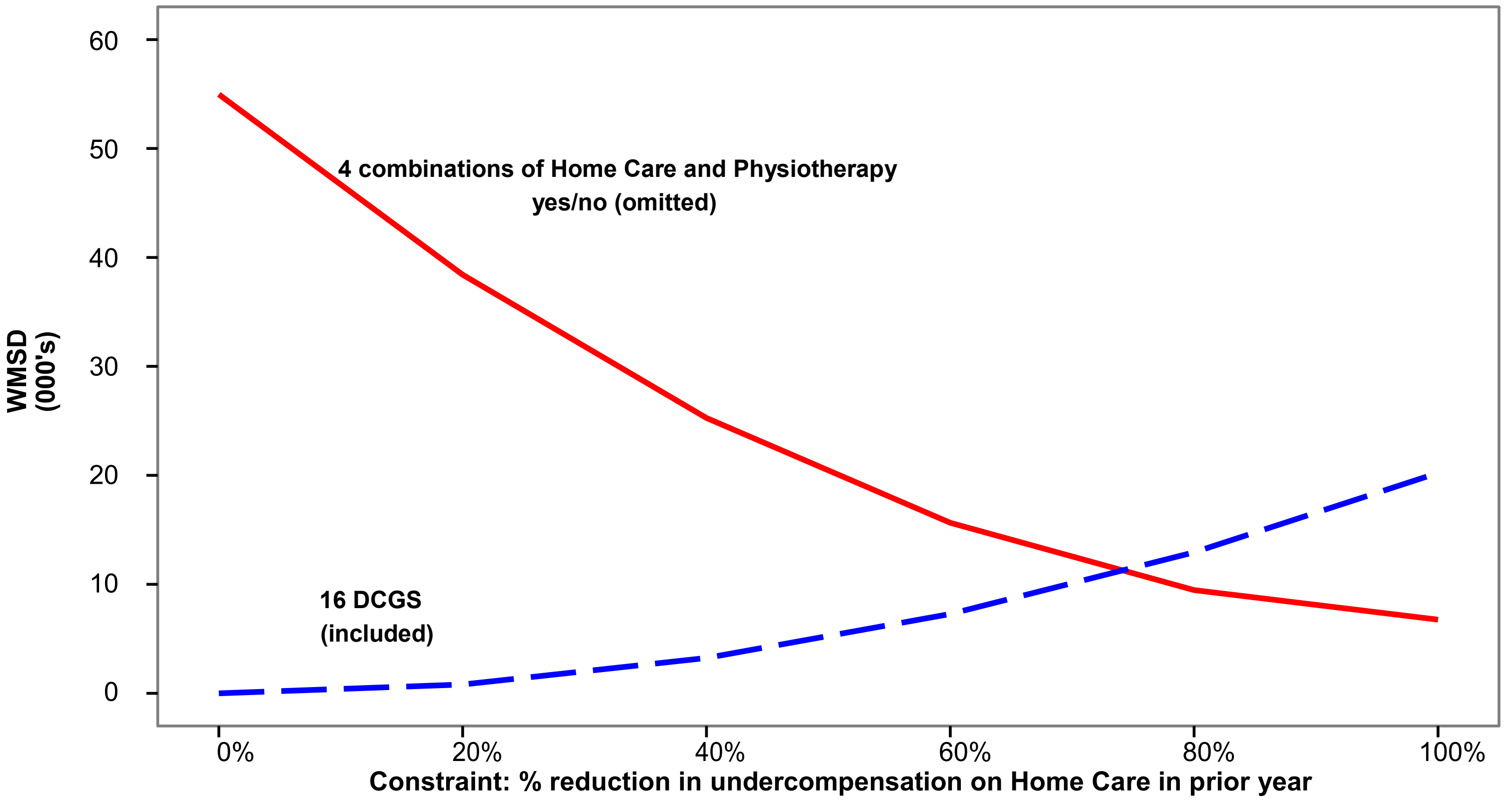


Figure 5: Base model with constraints on undercompensation for users of Home Care in prior year: Weighted Mean Squared Deviation (WMSD) for one set of mutually exclusive groups based on included and omitted indicators

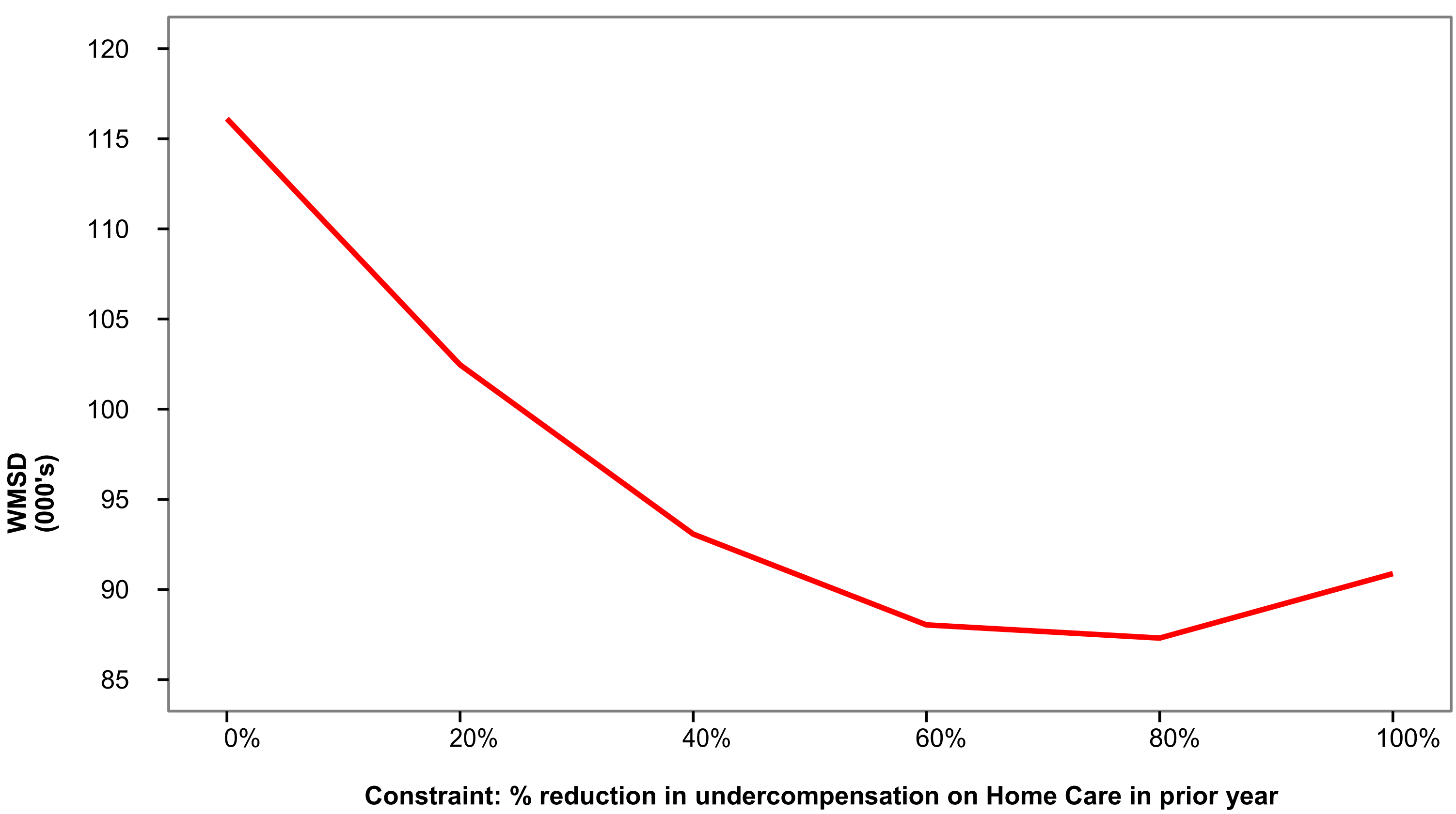


Figure 6
Normalized Weighted Mean Absolute Deviations (Euros, 2012)
for 64 Mutually Exclusive Groups for Weights P=1 and P=2

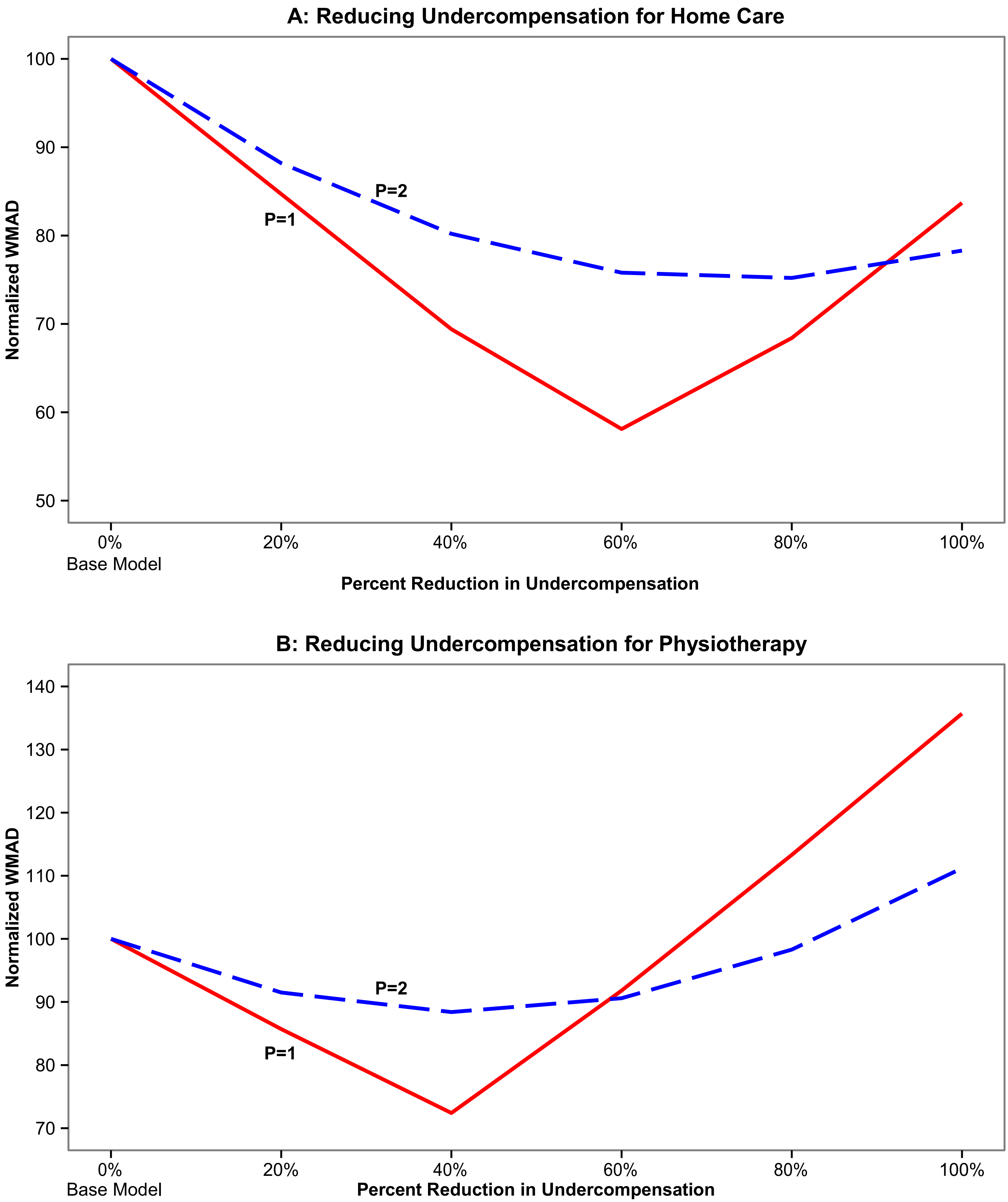
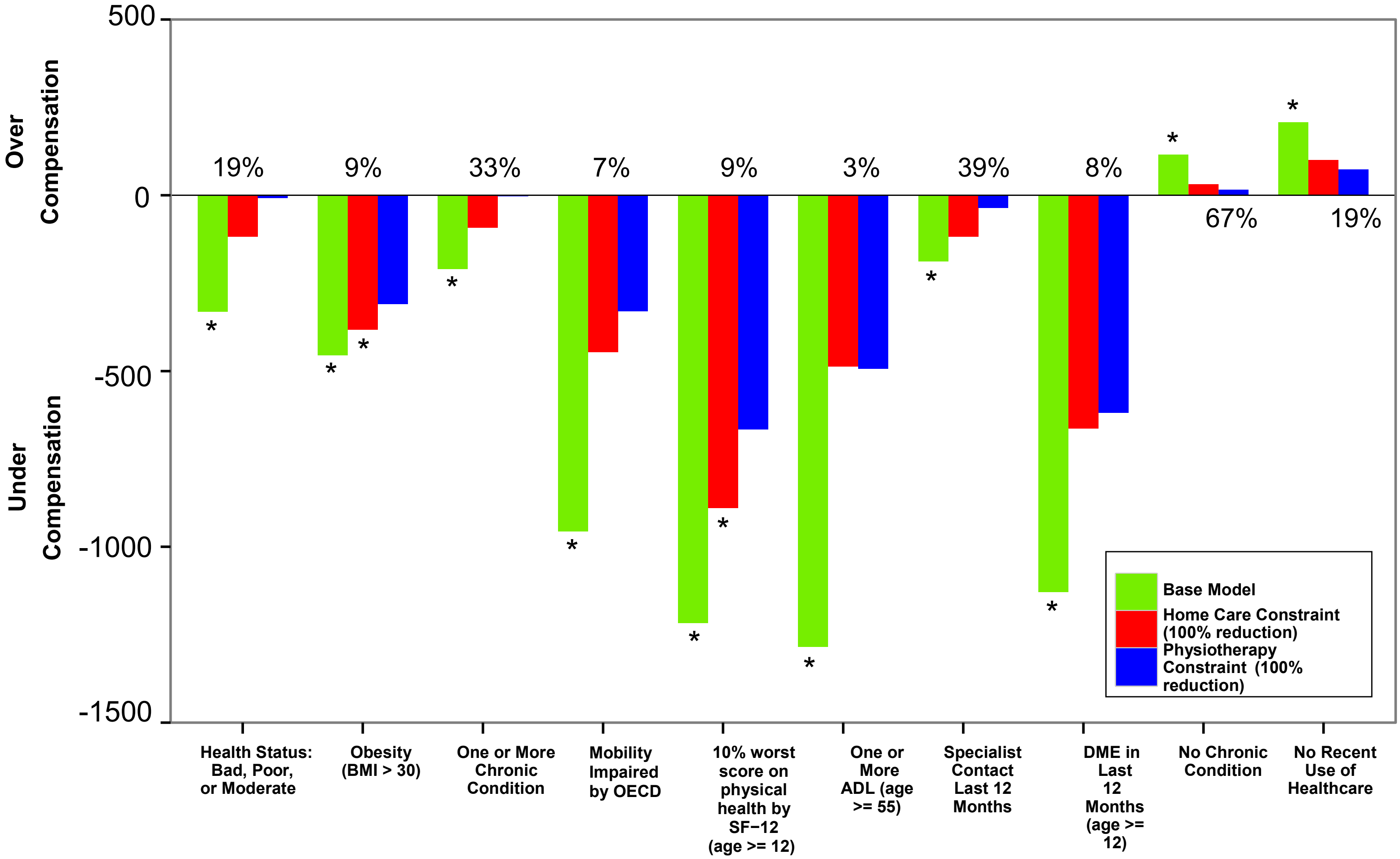


Figure 7
Population Frequency (%) and Average Under or Overcompensation (Euros, 2012) for Prior-year Survey-based Indicators Not Included in the Base Model and Not Included in Constraints



* Statistically significantly different from zero (p<0.05).