

Hintermann, Beat; Minke, Matthias

Working Paper

The value of extending life at its end: Health care allocation in the presence of learning spillovers

WWZ Working Paper, No. 2018/15

Provided in Cooperation with:

Center of Business and Economics (WWZ), University of Basel

Suggested Citation: Hintermann, Beat; Minke, Matthias (2018) : The value of extending life at its end: Health care allocation in the presence of learning spillovers, WWZ Working Paper, No. 2018/15, University of Basel, Center of Business and Economics (WWZ), Basel, <https://doi.org/10.5451/unibas-ep64725>

This Version is available at:

<https://hdl.handle.net/10419/185817>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

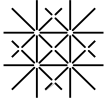
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Universität
Basel

Wirtschaftswissenschaftliche
Fakultät



June 2018

The value of extending life at its end: Health care allocation in the presence of learning spillovers

WWZ Working Paper 2018/15

Beat Hintermann, Matthias Minke

A publication of the Center of Business and Economics (WWZ), University of Basel.

© WWZ 2018 and the authors. Reproduction for other purposes than the personal use needs the permission of the authors.

Universität Basel
Peter Merian-Weg 6
4052 Basel, Switzerland
wwz.unibas.ch

Corresponding Author:
Dr. Matthias Minke
Tel: +41 (0) 61 207 33 53
Mail: matthias.minke@unibas.ch

The value of extending life at its end: Health care allocation in the presence of learning spillovers*

Matthias Minke[†] Beat Hintermann[†]

University of Basel

June 5, 2018

Abstract We investigate the social value of medical interventions at the end of life that tend to have a high cost-benefit ratio. We model the optimal allocation of health resources across a continuum of diseases that differ by severity and treatment options, and extend it to allow for learning spillovers between treatments. We calibrate our model to admissions to intensive care units in Switzerland. Cancer treatments associated with learning spillovers that decrease the mortality for non-cancer patients by 1 percentage point justify a cost-benefit ratio per additional life-year of 1.78.

Keywords: End of life; allocation of health resources; learning spillovers.

JEL Classification Numbers: I10.

*We thank Stefan Felder, Florian Kuhlmeier, Christian Rutzer, Stephanie Armbruster, Brenda Gannon and the participants of the 1st EUHEA PhD-supervisor conference in Manchester for valuable comments.

[†]University of Basel, Faculty of Business and Economics, Peter-Merian-Weg 6, 4002 Basel, Switzerland.
E-mail: matthias.minke@unibas.ch and b.hintermann@unibas.ch

1 Introduction

Expenditures for medical treatments are not evenly distributed over patients' lives, but are concentrated at the beginning and the end.¹ Given the increase in health care expenditures (HCE) towards the end of life, some authors suspect a misallocation of resources that would otherwise be available for alternative uses within or outside the health care sector [Leaf, 1977; Ginzberg, 1980; Lundberg, 1993]. This claim is based on a comparison of the costs of terminal care and the monetized benefits of a life extension implied by the value of a statistical life year (VSLY).²

In a recent article, French et al. [2017] argue against this view of wasteful spending. They confirm the high health costs at the end of life, but find that these costs accrue in the last several years of life (rather than the last) and are not due to expensive but futile efforts that prolong life by a few weeks, but the treatment of chronic diseases.

There are also theoretical arguments for why health expenditure at the end of life is not necessarily too high. For one, it is not clear whether the VSLY is a useful measure to value a life extension at the end of life, as the valuation of a life year may depend on the remaining life expectancy [Hammit, 2007]. Furthermore, the empirical literature defines end-of-life costs those that accrue during a given time period before death. One should be cautious about making normative inferences based on end-of-life HCE computed from this ex-post point of view, because a significant share of these expenditures may be associated with treatments that, on average, are quite effective.³ In contrast, if we think about end-of-life HCE from an

¹In the USA, 25-29% of Medicare expenditures occur during the patients' final year [Lubitz and Riley, 1993; Riley and Lubitz, 2010; Hogan et al., 2001] Polder et al. [2006] find that approximately 10% of health care expenditures over the whole life cycle take place in the last year of life in the Netherlands. Zweifel et al. [1999] report decedent-survivor cost ratios for all insured persons in Switzerland between 5.3 and 10.6, with an upsurge in HCE in patients' final months of life. In general, lifetime health expenditures typically follow a U-shaped curve; see, e.g., Alemayehu and Warner [2004]).

²Using data of kidney dialysis patients, Lee et al. [2009] estimate costs as ranging between USD 65,000 and USD 490,000 per quality-adjusted life year, depending on the patients' individual characteristics, with an average value of USD 129,000 for one life year in perfect health. Depending on the type of intervention, the costs of end-of-life care can be far greater.

³If a treatment has an 80%-chance of averting death, the costs associated with the remaining 20% will be labeled "end-of-life" from an ex-post-view, even though this is clearly not a treatment that would be labeled futile considering the survival rate. Alternatively, consider a two-step process that consists of diagnosis

ex-ante view, which is more appropriate in the context of a cost-benefit analysis, we would define those expenditures as “end-of-life” that are associated with treatments which do not significantly alter survival prospects or wellbeing.⁴

In this paper, we provide an additional argument for why the wasteful-spending view may not be correct. We develop a theoretical model in which a social planner allocates resources across consumption and health. To our knowledge, our model is the first to allow for heterogeneous treatment options and a continuum of health states across the population. This yields an expression for the marginal cost of saving a statistical life year (MCSLY) from an ex-ante perspective, which in the baseline model is equalized across treatment options and is equal to the VSLY.

We propose learning externalities as a rationale for why the benefits from seemingly futile medical treatments may exceed the VSLY associated with the treated patients themselves. We posit that learning externalities arise from the treatment of diseases that are currently not well understood and thus cannot be cured, but which constitute the medical frontier. By applying health care in (currently) hopeless cases, practitioners learn and eventually, via incremental improvements, can treat a condition sufficiently, thus shifting the medical frontier outwards.

Our model is motivated by the observation that life expectancy has increased significantly over time, and that technological progress has been a major cause underlying this increase [Cutler, 2008; Lakdawalla et al., 2010]. However, because we all die at some point, a decrease in the mortality rate associated with a particular disease leads to a shift in the cause of death to other diseases (or different stages of the same disease) over time. We capture this dynamic feature in a static framework by assuming that the treatment effectiveness of the end-of-life disease group, the composition of which will evolve over time, remains low regardless of the

followed by treatment. Diagnosis involves acquiring costly information about a patient’s survival prospect. Even if hopeless cases are ultimately not treated, the costs of determining that they are hopeless will count towards end-of-life HCE in an ex-post context.

⁴For a discussion on the cost of dying, the distinction between a retrospective and a prospective view on these costs and an empirical review of health care utilization in the last year of life, see Scitovsky [1994].

technical progress in the rest of the health sector, and that this technological progress is partly driven by the application of treatments in the end-of-life sector.

Our model formulation is also consistent with a situation where learning in the context of one disease may produce knowledge that turns out to be useful in the treatment of others. For example, Gelijns and Rosenberg [1994] report positive learning spillovers from using of beta blockers in cardiovascular conditions that improved the treatment effectiveness in more than twenty other diseases. Romley et al. [2011] study the effect of hospital spending on inpatient mortality in California from 1999 to 2008 and find a significant negative relationship between end-of-life expenditures and overall hospital mortality for six major diagnosis groups, thus providing indirect evidence for the existence of learning spillovers.

There are other reasons why a cost-benefit analysis of end-of-life HCE based on a simple comparison between treatment costs and VSLY may be misleading. In a paper that is conceptually related to ours, Philipson et al. [2010] study the option value inherent in a treatment that reflects the probability of surviving until a new medical treatment is innovated, which may give rise to risk-loving behavior with treatment decisions at the end of life. The total value of terminal treatment for HIV patients, including the "value of hope", was found to be well above the value implied by standard VSLY estimates. Similarly, Wessling [2013] finds experimental support for risk-loving behavior at the end of life if there is a positive exogenous chance of surviving until a better cure is developed.

We illustrate our model using health care data from Switzerland. We focus on admissions to intensive care units (ICUs) and separate diseases into cancer (which constitutes our "end-of-life" sector) vs. all other diseases. We calibrate the free model parameters using ICU admission rates, survival rates, treatment costs, overall health expenditure, aggregate income and an estimate for the VSLY from the literature. The model is solved for different magnitudes of learning spillovers that arise from the treatment of cancer cases. We find that the presence of learning spillovers that lead to a decrease in non-cancer mortality by one percentage point (relative to the situation without spillovers) implies that the optimal

MCSLY associated with cancer treatments exceeds the VSLY by 78%. This result increases as the costs of regular and end-of-life HCE diverge.

In the next section, we provide some background about progress in the medical sector, which motivates our theoretical model in Section 3. In Section 4, we calibrate the free parameters of our model using Swiss data and assess the quantitative importance of spillover effects numerically. Section 5 offers concluding remarks.

2 Medical progress and learning by doing

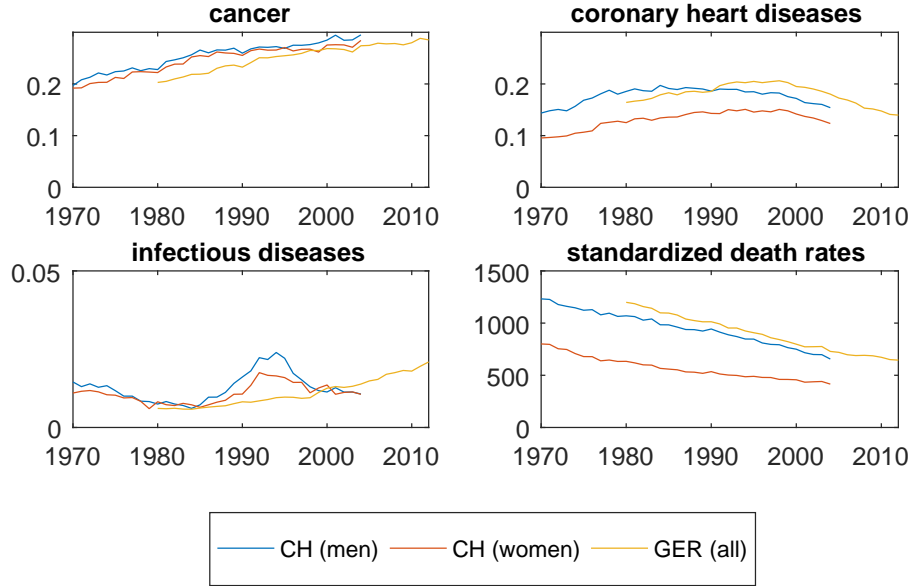
Our theoretical model is motivated by three stylized facts: First, life expectancy has increased in many countries; second, there is a continuous shift in the principal causes of death over time; and third, learning by doing is an essential component of medical progress.

Figure 1 shows the evolving pattern in the leading causes of death over the last four decades in Switzerland and Germany; similar patterns exist in other OECD countries.⁵ The lower-right panel of the Figure shows standardized mortality rates for all causes per 100,000 persons and year, which have declined significantly during the last 40 years, reflecting the increased life expectancy. For Swiss men (women), the standardized mortality rate declined from 1,230 (800) in 1970 to around 650 (420) in 2004. Similarly, the standardized mortality rate for the whole population has dropped roughly by half, from 1,200 in 1980 to less than 700 in 2012 in Germany. Mortality rates continue to decline, although the graph suggests that the speed of the reduction may be slowing down.

The remaining three panels show the contribution of the most common mortality causes as a share of total deaths over time. Although the standardized mortality rates have decreased for all of these causes, their relative importance has changed significantly since the early 1970. Figure 1 shows that cancer remains to be the leading cause of death with an increasing distance to coronary heart disease..

⁵Becker et al. [2005] show that life expectancy and disease-specific mortality rates vary significantly between industrialized and developing countries; however, an increase in life expectancy has been observed everywhere in the past 50 years.

Figure 1: Standardized mortality rates and disease shares



Source: Cause-specific deaths (ICD-10) from the Swiss Federal Statistical Office, based on deaths in 1970-2004 [Swiss Federal Statistical Office, 2014], and from Germany Federal Office of Statistics based on deaths in 1980-2012 [German Federal Statistical Office, 2013].

There is general agreement that decreased mortality rates are the result of improved living conditions, prevention, behavioral changes and enhanced medical treatments. The Figure suggests that these factors affect different types of diseases in different ways, because, although the standardized death rate as a whole is declining, its composition changes. For example, when comparing cancer and coronary heart disease (CHD), it is likely that a healthier life style and better treatment have been more successful in reducing CHD mortality than cancer mortality.

For US data, Cutler [2008] attributes approximately 50 percent of the survival improvement for cancer to advancements in treatment and screening, and similar estimates have been reported by Lakdawalla et al. [2010]. Advances in treatment have also been instrumental in preventing deaths from cardiovascular diseases beginning with the invention of bypass surgery in the 1980s, angioplasty, and cardiac catheterization in the 1990s [see Cutler and McClellan, 2001].

Medical progress, like any technological process, is incremental. Even breakthrough discoveries typically depend on a history of painstaking and lengthy research during which many attempts at achieving an intermediate goal fail. Furthermore, medical progress is not confined to research and development: In order for effective treatments to be developed and improved, they have to be applied and refined by practitioners. Moreover, different patients may respond differently to a particular treatment, which implies that learning takes place as a treatment is applied to more patients. Gelijns and Rosenberg [1994] provide a detailed discussion of the dynamics of technological innovation and the role of learning by doing on behalf of practitioners.

These observations (increase in life expectancy, relative shift in causes of death and learning) imply that a medical condition that is considered “end of life” at one point in time, because survival rates are very low even with treatment, does not necessarily retain this status over time. Due to a combination of research and learning by doing, successful treatments can be developed that significantly improve survival rates. An important example is the treatment of cancer. The survival rate has significantly increased for many cancer types, such that some patients survive for years with a diagnosis that in the past used to lead to death within weeks or months [Jemal et al., 2017]. However, as suggested by the increasing mortality share from cancer shown in Figure 1, cancer patients tend to die later of the same (or a different) cancer, and patients that survive a different disease increasingly die of cancer. There is biomedical evidence for a positive association between life expectancy and cancer prevalence [DePinho, 2000], and the lifetime risk of being diagnosed with cancer increased significantly over time [Ahmad et al., 2015].

We build on this observation of a unbalanced manifestation of medical progress. More specifically, we argue that both the decreasing relative importance of cardiovascular diseases and the increasing importance of cancer are two sides of the same coin in the sense that improvements in the ability to deal with, for instance, cardiovascular conditions has enabled people to live long enough to develop cancer. Similar arguments apply to other diseases:

Curing one means succumbing to the next. In this sense, medicine is a victim of its own success, or as Zweifel et al. [2005] pointed out: Medical innovation is sometimes a Sisyphean task.⁶ In our theoretical model, we capture this idea by keeping the treatment effectiveness of the least curable group of diseases constant, whereas the treatment effectiveness of all other groups increases over time. Although medical progress may lead to a better treatment or even a cure for a particular disease, there will always be another disease (or disease stage) “waiting down the road” which cannot be treated effectively.

A fundamental assumption of our model is that treating these diseases, although not currently very effective in terms of life extension, is an important engine of medical progress because it generates knowledge that can be used also in non-end-of-life contexts. For the case of significant improvements of survival for diseases that previously led to death within a short time span, this assumption is in some sense tautological: It is precisely the success in treatment, partially due to treatment efforts before the breakthrough, that moves these diseases out of the end-of-life and into the “regular” health sector. However, there are many instances where a treatment success in one health context turned out to be useful in different contexts as well, such as the beta blockers originally used for cardiovascular conditions and ultimately used in a series of other diseases Gelijns and Rosenberg [1994]. In our model, we aim to capture both “intra”- and “inter”-disease spillovers and generally assume that knowledge gained in an end-of-life context can be applied in the regular health sector.

Success stories in cancer treatment include the use of chemotherapy to treat Hodgkin Lymphoma, which significantly increased survival for many different types of cancer [DeVita and Chu, 2008]. Subsequently, the use of chemotherapy was further improved using trial and error in different contexts. One particular success of a recombination of chemotherapy ingredients pertains to testicular cancer, which was associated with a 5-year survival rate of around 5 percent in the early 1970’s. A new chemotherapy regimen increased survival

⁶To describe this in more technical terms, improvements in medical technology shift the survival curve upwards, but the magnitude of the shift declines at very high ages such that the point where the expected survival probability approaches zero remains roughly constant, even as life expectation (the integral below the survival curve) increases.

rates to 64 percent [Hanna and Einhorn, 2014]. This treatment was later refined to the BEP regimen, which further increased survival and is currently also used to treat ovarian cancer.⁷

Another example is the advent of targeted drug therapies, which were initially developed in the context of treating patients with chronic myelogenous leukemia (and increased the survival rate for this cancer to almost 90 percent), but are now being used for treating other cancers and even other diseases such as stroke [Pray, 2008]. Likewise, the use of supportive care medicine such as anti-nausea drugs has increased quality of patients' lives [Cubeddu et al., 1990], and is now in widespread use for many different types of cancer and other diseases.⁸ Other diseases where drugs originally developed for cancer treatment are being used include rheumatoid arthritis and HIV [Chabner and Roberts Jr., 2005]. A different manifestation of learning by doing is indicated by the positive association between the volume and the quality of care for surgical procedures [Phillips and Luft, 1997], or between spending and inpatient mortality [Romley et al., 2011].

3 Model

In the following, we develop a model of the optimal allocation of resources across consumption and different types of health expenditures. We start with our baseline model, followed by an extension that includes learning spillovers associated with the treatment of diseases that are currently not well understood.

3.1 Baseline

There is a continuum of health conditions, indexed by their untreated survival rate $x \in [0, 1]$, which is distributed across the population according to the probability density function

⁷BEP stands for the three main ingredients bleomycin, etoposide and platinum; for different uses of BEP, see MacMillan, <http://www.macmillan.org.uk/cancerinformation/cancertreatment/treatmenttypes/chemotherapy/combinationregimen/bep.aspx>, last accessed on July 27, 2017.

⁸These and other milestones in cancer treatment are discussed on the American Society of Clinical Oncology's website, <https://www.asco.org/research-progress/cancer-progress/top-5-advances-modern-oncology>, last accessed on July 27, 2017.

$f(x)$. The untreated mortality rate is given by $m(x) = 1 - x$. There is only one period, and the timing is as follows: (1) Health types are assigned at the beginning of the period, (2) health expenditure (i.e., treatment) is determined by the planner, (3) the mortality risk materializes, and (4) survivors enjoy utility of consumption.

We separate health conditions into two groups, H and L , with shares $\pi_j, j \in H, L$ and $\pi_H + \pi_L = 1$. These groups differ with respect to the available treatment options, represented by a treatment effectiveness parameter with $\alpha_j \in (0, 1)$, where $\alpha_H > \alpha_L$.⁹ This means that conditions that are associated with the same untreated mortality rate can differ with respect to their curability. One can think of α_j as reflecting the combination of the biomedical characteristics (both of the condition and the patient) and the state of medical knowledge associated with the respective subset of diseases. Throughout the paper, we distinguish between the terms *health condition* and *disease*: Whenever we refer to the term health condition we mean the value of x which is associated with the untreated mortality rate $m(x) = 1 - x$, whereas we use the term disease to describe the pair (x, α_j) , which captures both the quality of the health condition in terms of the untreated mortality and the effectiveness of the available treatment.

We assume that for each group j there exists exactly one treatment, which is associated with a unit cost of $h > 0$.¹⁰ The mortality rate for a health condition x that belongs to group j is given by

$$\tilde{m}(x, \alpha_j) = \begin{cases} m(x), & \text{if untreated} \\ (1 - \alpha_j)m(x), & \text{if treated.} \end{cases} \quad (1)$$

⁹The model is easily generalized to $j = 1, \dots, J$, with $\sum_{j=1}^J \pi_j = 1$. In the interest of tractability and to match the theory with the calibration exercise, we focus on two types only.

¹⁰We assume that the same unit treatment cost applies to all diseases. Allowing for disease-specific treatment costs h_j would not change the results qualitatively as long as the cost associated with group L is not significantly smaller than that associated with group H (if $h_L \ll h_H$, the planner would want to treat a higher share of the type L diseases than of type H diseases, which is counter-intuitive). In our application, we use two different treatment costs for cancer and non-cancer treatments. Note also that throughout the paper, we treat h as exogenous. One could hypothesize that directed technical change is aimed at lowering the cost of particular treatments such that h_j becomes endogenous, but this is beyond the scope of our paper.

Choosing the treatment's effectiveness to be proportional to the untreated mortality ensures that the mortality rate after treatment will be between 0 and 1. It further implies that, for a given class of treatment effectiveness j , the marginal product of expanding treatment, $\frac{\partial(1-\alpha_j)m(x)}{\partial x}$, is decreasing in the untreated survival probability.¹¹

Individuals enjoy utility from a composite consumption good c according to a utility function that is increasing, continuous, concave and twice differentiable, and which depends on the state s :

$$U(c; s) = \begin{cases} u(c) \geq 0 & \text{for } s = \text{alive,} \\ 0 & \text{for } s = \text{dead.} \end{cases} \quad (2)$$

Utility when alive is strictly positive for a positive level of consumption. A fully informed utilitarian social planner maximizes social welfare W by choosing the proportion of the population that receives treatment within each group j , subject to technology and resource constraints. The planner will prioritize treatments for those individuals who have a high untreated mortality risk: the reduction in the mortality risk $\alpha_j m(x)$ associated with treatment j is decreasing in the health state x , while the marginal costs of extending expected survival is increasing. Formally, the planner chooses a per-capita consumption level c and treatment cut-offs $\mathbf{x} = \{x_L, x_H\}$, to solve

$$\max_{c, \mathbf{x}} W = u(c) \cdot S(\mathbf{x}) \quad \text{s.t.} \quad y = cS(\mathbf{x}) + hZ(\mathbf{x}) \quad (3)$$

$$S(\mathbf{x}) := 1 - \sum_{j=H,L} \pi_j \left(\int_{\underline{x}}^{x_j} (1 - \alpha_j) m(x) f(x) dx + \int_{x_j}^{\bar{x}} m(x) f(x) dx \right) \quad (4)$$

¹¹In most of the existing literature, diminishing returns in health expenditure are incorporated by assuming increasing costs [e.g., Ellis, 1998] or decreasing effectiveness as a function of the amount of treatment for a particular patient [Grossman, 1972; Hall and Jones, 2007]. Yet another approach is used in Ma [1994], where hospitals face a continuum of patients with varying costs of treatment (but the same treatment effectiveness). Our approach of holding treatment costs fixed but endogenizing the share of diseases that receives treatment (and thus the marginal effectiveness of treatment) is qualitatively similar in that it allows for a diminishing marginal productivity in the health sector overall. Allowing both treatment effectiveness and treatment costs to vary continuously across patient types would add significant complexity to the model, but with no clear gain in intuition.

$$Z(\mathbf{x}) := \sum_{j=H,L} \pi_j \int_{\underline{x}}^{x_j} f(x) dx. \quad (5)$$

Aggregate income y is exogenous and is either consumed or spent on health care. $S \in [0, 1]$ can be interpreted as the expected survival rate of the population. Throughout the paper, we denote the optimal solution by an asterisk, such that x_j^* denotes the cut-off health conditions (=survival rate) for group j : Individuals with health a condition that is weakly worse than x_j^* receive treatment, whereas those with better health do not. Z is the sum of treatments that take place in each of the groups L and H . The total treatment cost includes costs incurred for people who die despite the treatment. After substituting the households' budget constraint (and thus eliminating c as a choice variable), the optimality conditions for an interior solution $\{x_L^*, x_H^*\}$ are given by

$$u(c^*) \frac{\partial S(\mathbf{x}^*)}{\partial x_j} = -u'(c^*) \frac{\partial c^*}{\partial x_j} S(\mathbf{x}^*) \quad \text{for } j = L, H \quad (6)$$

$$c^* = \frac{y - hZ(\mathbf{x}^*)}{S(\mathbf{x}^*)}. \quad (7)$$

At the optimum, the marginal benefit of expanding treatment for a given technology (i.e., the utility gain associated with an increase in survival) is equal to the marginal decrease in consumption that is due to a shift of resources to the health sector. This decrease includes the direct health care costs h and the fact that per capita consumption declines in the number of survivors.¹² Combining (4)-(7) leads to

$$\frac{u(c^*)}{u'(c^*)} - c^* = \frac{h}{\alpha_j m(x_j^*)} \quad \text{for } j = L, H. \quad (8)$$

The left-hand side of (8) is the value of life, which is equal to the monetized utility of life net of consumption costs. This is positive whenever life is strictly preferred over death

¹²As pointed out by Meltzer [1997], the costs of HCE should not only include the direct costs of treatment, but also the indirect and future costs.

for any non-zero consumption level and when utility is concave in consumption [see Rosen, 1988]: These two assumptions are maintained throughout the paper. The right-hand side is the marginal cost of extending life using health technology j . If we think of the time frame of our model as representing one year, expression (8) states that, at the optimum, the VSLY has to be equal to the MCSLY.

The optimal ratio of the cut-off mortality rates for treatment groups L and H is inversely proportional to the ratio of the respective treatment effectiveness:¹³

$$\frac{m(x_L^*)}{m(x_H^*)} = \frac{\alpha_H}{\alpha_L} \quad (9)$$

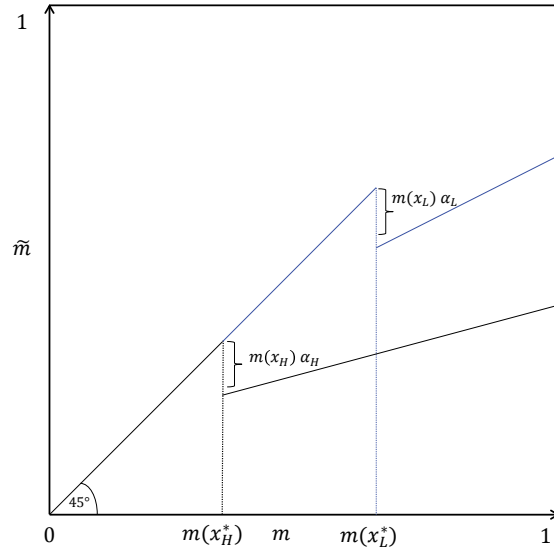


Figure 2: Optimal treatment cut-offs by group

Figure 2 plots the treated mortality rate \tilde{m} against the untreated mortality rate m for two treatment groups L and H . The treated and untreated mortality rates are equal up to the optimal cut-off mortalities $m(x_L^*)$ and $m(x_H^*)$, beyond which patients will receive treatment in each group. At the cutoffs, the function $m \rightarrow \tilde{m}$ shifts downwards discretely by $\alpha_L m(x_L^*)$ and $\alpha_H m(x_H^*)$, respectively, and the respective slopes decrease from unity to $(1 - \alpha_L)$ and $(1 - \alpha_H)$. Condition (9) states that the shift in the treated mortality must be equal for

¹³In the general case with $j = 1, \dots, J$, the condition becomes $m(x_j^*)/m(x_i^*) = \alpha_i/\alpha_j$.

both technologies. Furthermore, a greater difference in treatment effectiveness between two groups implies a greater horizontal distance between the respective cut-off mortalities.

3.2 Learning externalities

As motivated in section 2, we assume that learning by doing in the L sector affects the treatment effectiveness in “regular” health contexts, which in our model comprises the H sector. These spillovers represent either improvements in the survival rate associated with a previously untreatable disease (which moves it from L to H), or inter-disease spillovers where knowledge gained in one context is helpful in another. Assuming that the learning spillovers are effective immediately allows us to represent this process in a static framework and thus abstract from modeling the dynamics of population and income over time. We restrict learning to take place in the L sector even though learning presumably takes place throughout in the health sector. However, there is arguably more to be learned in the context of treating diseases that are currently not well understood, and which therefore are located at the technological frontier. To generate our results, the learning spillovers from applying treatments the L sector have to be larger than those originating in the H sector.

We further impose that learning in L only affects the treatment effectiveness in H , but not in L . This is motivated by observation that the composition of end-of-life diseases changes over time, as discussed in the context of Figure 1. Intuitively, technological progress pushes diseases away from the frontier over time, while at the same time new diseases or disease combinations appear, which then constitute the new frontier diseases [see Jones, 2003, for a similar rationale]. With this stationary representation of a dynamic process, we aim to capture the main interactions that govern the allocation of resources across different treatment options subject to learning-by-doing, while keeping the model as simple as possible.¹⁴ Allowing learning spillovers to increase the treatment effectiveness in L would lead to a further

¹⁴Kuhn et al. [2011] develop an overlapping generations model that incorporates both positive (learning) and negative (congestion) spillovers in health care. Their model explicitly incorporates dynamic aspects of spillovers and thus goes beyond the reduced-form treatment in our paper. However, it does not distinguish between different types of treatments, which is essential for our analysis.

extension of treatment in the L sector in the social optimum, but this effect would likely not be large, considering that the benefit from spillovers are proportional to the treatment effectiveness. Figure 3 presents our stylized model of spillovers schematically.

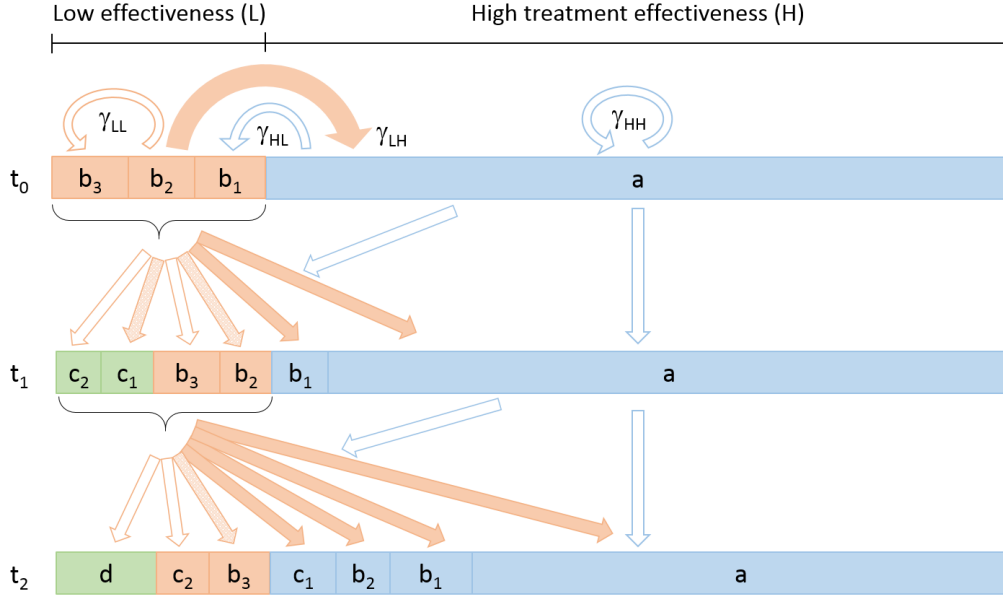


Figure 3: Static representation of dynamic learning. At time t_0 , diseases are divided into those for which an effective treatment exists (a), and those that are currently not well understood (b_1 - b_3). Medical progress renders disease group b_1 treatable in period t_1 . At the same time, new diseases appear (c_1 - c_2). In the next period, some of these new diseases have become treatable as well (c_1), along with previously existing untreatable diseases (b_2). The straight arrows represent learning spillovers in the dynamic setting. In our model, we focus on spillovers originating in L that affect diseases that are currently (solid arrows) or eventually in H (shaded arrows) but suppress all other learning (clear arrows). The static representation of this dynamic learning process is indicated by the curved arrows at the top of the figure, which contain the (dynamic) effect of the corresponding straight arrows. Among these, we focus on the shaded arrow labeled γ_{LH} .

Let the effectiveness of type H treatments be given by

$$\alpha_H = \bar{\alpha} + \gamma x_L \pi_L, \quad (10)$$

where γ captures from the low effectiveness group spilling over to the high effectiveness group (corresponding to γ_{LH} in Figure 3). Increasing the cut-off health condition x_L implies an increasing number of persons treated in group L , and also an increase in the treatment

effectiveness in group H beyond some default effectiveness $\bar{\alpha}$.

The first-order necessary conditions with respect to the cut-off survival rates x_H and x_L can be expressed as (derivation provided in the Appendix):

$$\frac{u(c^*)}{u'(c^*)} - c^* = \frac{h}{\alpha_H m(x_H^*)} = \frac{h}{\alpha_L m(x_L^*) + \gamma \pi_H \mu_H^* / x_L^*} \quad (11)$$

$$\mu_H^* \equiv \int_0^{x_H^*} (1-x)f(x)dx \quad (12)$$

For group H (first equality), this first-order condition is the same as in (6). However, for group L (second equality), there is an additional term in the denominator due to the spillovers generated by applying treatments in this group that increases the survival rate of the individuals treated in group H , and which directly depends on the strength of the spillovers parameter γ and on μ_H^* , which can be interpreted as the expected untreated mortality of the treated in the H -group.¹⁵ The higher this expected mortality, the greater is the benefit from learning spillovers.

The presence of learning spillovers implies that the MCSLY, if computed only with reference to an extension of life in group L , optimally exceeds the MCSLY in group H (which is equal to the VSLY):

$$\frac{h}{\alpha_L m(x_L^*)} > \frac{h}{\alpha_H m(x_H^*)} = \frac{u(c^*)}{u'(c^*)} - c^* \quad \text{if } \gamma > 0 \quad (13)$$

In other words, the full benefit of applying treatment in group L includes a positive spillover to group H , and neglecting this spillover by setting $\text{MCSLY}_L = \text{MCSLY}_H$ would lead to an inefficiently low use of technology L .

¹⁵To see this, note that

$$\int_0^{x_H^*} (1-x)f(x)dx = \int_0^{x_H^*} f(x)dx - \int_0^{x_H^*} xf(x)dx = F(x_H^*) - E[x|x \leq x_H^*] \cdot F(x_H^*)$$

$$\mu_H^* = F(x_H^*)E[m(x)|x \leq x_H^*]$$

The first term is the share of people within sector H that receives treatment, and the second term represents the expected (untreated) mortality within this group.

We are not able to derive a closed-form solution for the optimal treatment cut-offs without imposing very restrictive assumptions on preferences and the distribution of untreated health conditions. However, applying the implicit function theorem allows us to derive the effect of increasing the spillover parameter on the treatment cut-off x_L .

Proposition 1 *The introduction of spillovers from $\gamma = 0$ to some small $\gamma > 0$ increases the cut-off of the treatment option L if the following conditions hold with respect to the marginal number of treated $\pi_L f(x_L^*)$ at the optimum:*

$$\pi_L f(x_L^*) < \frac{\pi_H \mu_H^*}{x_L^* m(x_H^*)}, \quad \text{and} \quad (14)$$

$$\pi_L f(x_L^*) < -\frac{(u - u'c^*)u'}{u''c^*u} \frac{S^*}{x_L^* m(x_L^*) \alpha_L}. \quad (15)$$

Both conditions together are sufficient conditions, whereas either one of them must hold as a necessary condition.

Proof. By the implicit function theorem, the derivative of the optimal cut-off x_L^* with respect to the spillover parameter γ is given by

$$\frac{\partial x_L^*}{\partial \gamma} = \frac{-W_{x_L \gamma} W_{x_H x_H} + W_{x_L x_H} W_{x_H \gamma}}{W_{x_L x_L} W_{x_H x_H} - W_{x_L x_H}^2}, \quad (16)$$

where subscripts indicate partial derivatives. At the optimum, the denominator is the determinant of the Hessian matrix and is thus positive, such that it suffices to determine the sign of the numerator. In the Appendix, we show that $-W_{x_L \gamma} W_{x_H x_H} + W_{x_L x_H} W_{x_H \gamma}$ can be written as a sum of two components. The first component is positive if (14) holds, and the second component is positive if (15) holds. ■

An increase in γ increases both the marginal benefit (the left-hand side of (6), which is the marginal increase in survival) and the the marginal cost (the right-hand side, which represents the marginal decrease in consumption for the survivors) associated with treatment in group L . If the sufficient conditions in the proposition hold, the former effect dominates

the latter. To interpret the sufficient conditions, it is helpful to impose some structure on preferences. For $u(c) = c^\sigma$ with $0 < \sigma < 1$, which we employ in our numerical section below, it follows that $-\frac{(u-u'c)u'}{u''cu} = 1$, such that the second sufficient condition simplifies to $\pi_L f(x_L^*) < \frac{S^*}{x_L^* m(x_L^*) \alpha_L}$. Intuitively, it is optimal to increase the cut-off x_L in response to an increase in γ if the marginal number of patients that have to be treated to induce learning effects, $\pi_L f(x_L^*)$, is not too large relative to the share of the population that is treated with technology H and the optimal survival rate is high. Both conditions are likely to hold in advanced countries, as these countries typically have the means to treat the vast majority of health conditions quite effectively ($\pi_H \gg \pi_L$) and therefore exhibit a high life expectancy. Spillovers could therefore serve as a rationale to explain the substantial share of end-of-life expenditures in such countries.

Proposition 2 *The introduction of spillovers from $\gamma = 0$ to some small $\gamma > 0$ will increase the optimal level of treatment in sector L relative to that in sector H , and thus expand the optimal share end-of-life expenditures within total HCE, if the following sufficient condition holds:*

$$\mu_H^* > x_L^* m(x_H^*) f(x_H^*) \tag{17}$$

Proof: See Appendix.

Spillovers make survival cheaper in the sense that treatments in group H become more effective at given unit treatment costs.¹⁶ Given that the planner allocates more resources to the health sector, it is thus *a priori* not clear whether it is optimal to spend relatively more in sector H or L . The sufficient conditions ensure that the benefit from engaging in treatments in L is sufficiently large, relative to the costs, in order to justify an increase of the end-of-life share in HCE. In our numerical illustration below (and in fact with all parameter

¹⁶We treat the unit cost of treatment as fixed. Alternatively, one could assume that technological progress decreases the unit cost of treatment while keeping the effectiveness constant. In both cases (and any convex combination of the two), spillovers lower the cost of reducing the expected mortality in group H .

constellations that we tried), the sufficient conditions of both propositions hold.

4 Numerical illustration

In order to assess the robustness and quantitative relevance of our analytical result, we simulate the model numerically. We stress, however, that we provide no empirical evidence in favor or against our model. The aim of the current section is to illustrate our results based on realistic assumptions about the distribution of mortality, treatment shares and overall (opportunity) costs of health expenditure.

We calibrate our model to the situation of intensive care unit (ICU) admissions in Swiss hospitals; i.e., we define the treatment decision as either being admitted to an ICU or not. The reason why we restrict our attention to ICU admissions is that it corresponds best to our model where the only aim of treatment is a reduction of the expected mortality.¹⁷

We divide diseases into two types: Cancer (type L) and all other diseases (type H), motivated by the stylized facts shown in Figure 1. We model the survival probability using a beta distribution with shape parameters a and b :

$$f(x) = \begin{cases} \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} & 0 \leq x \leq 1 \\ 0 & \text{otherwise,} \end{cases} \quad (18)$$

where $B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ is a normalizing constant which ensures that $F(1) = 1$. This distribution allows for significant flexibility. The combination $a = b = 1$ is the special case of the uniform distribution.

¹⁷In a previous version of our model, we focused on all hospital admissions. However, many inpatient treatments primarily aim to increase quality rather than quantity of life. A model based on all hospital admissions therefore falsely attributes the low mortality rate of many treatments to a high treatment effectiveness, even though the mortality would have been low even in the absence of treatment. This leads to a significant overestimation of the parameters α_j that govern the treatment effectiveness. Focusing on ICU admissions largely avoids this problem, although at the cost of excluding life-extending treatments where the patients are not transferred to the ICU. A possible alternative would involve defining a list of disease/treatment combinations where the main goal is the increasing survival. However, given the complexity and heterogeneity of diagnoses and treatments, this is not easy to do in a transparent and reproducible way.

4.1 Model parameters

In a first step, we look for suitable values for the shape parameters a and b of the density function $f(x)$. We approximate the distribution of untreated mortality rates by combining age-specific mortality rates with the age distribution of the Swiss population in 2014.¹⁸ Figure 4 shows a kernel density estimate of the age distribution (left axis) and the age-specific mortality rate (right axis) in Switzerland. The age distribution displays the typical pattern of an industrialized country, with the bulk of persons being concentrated around age 50, followed by a step decrease in the density per year of age. The mortality rate is close to zero for all ages below 70 (it exceeds the value of 1% only after age 68), and increases exponentially in old age.

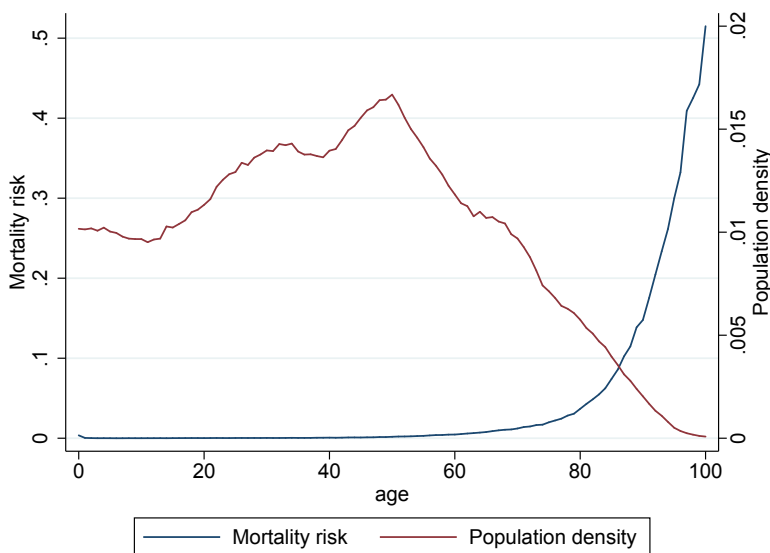


Figure 4: Age distribution and mortality rates in Switzerland

We link the age distribution with the age-specific mortality rate to obtain a distribution of (treated) mortality rates by assuming that all individuals of a given age face the average mortality rate of their age group. We then fit a beta model with shape parameters a and

¹⁸A caveat of this procedure is that observed mortality rates are the outcome of many exogenous and endogenous factors, including treatment, such that these are really the treated rather than the untreated mortality rates per age group. Nonetheless, we hypothesize that the qualitative shape of the distribution of untreated mortality rates (which is unobservable) is similar to the observed post-treatment mortality rates.

b and support $[0, 1]$. The fitted shape parameters are $\hat{a} = 41.45$ and $\hat{b} = 0.34$, so that $f(x) = \frac{x^{40.45}(1-x)^{-0.66}}{0.74167}$ for $0 \leq x \leq 1$. This describes a monotonically increasing, left-skewed distribution with an expected survival rate of $E[x] = \frac{a}{a+b} = 0.992$, which corresponds to the survival rate of all persons (including treatment for those below the cut-offs x_j^*) in the context of our model. We use this ex-post distribution of mortality as a qualitative measure for the shape of the distribution of ex-ante health conditions, namely that the vast majority of persons in the population is subject to virtually no or only a modest risk of dying, whereas a small number of persons face a substantial mortality risk. This suggests that we can focus on the case where $f'(x) > 0$ over the entire support of the distribution of health conditions.¹⁹ We fix $b = 1$ and calibrate the shape parameter a such that the share of the treated population in our model, Z , matches the empirical hospitalization rate. Figure 5 illustrates the sensitivity of the probability density functions to the value of a .

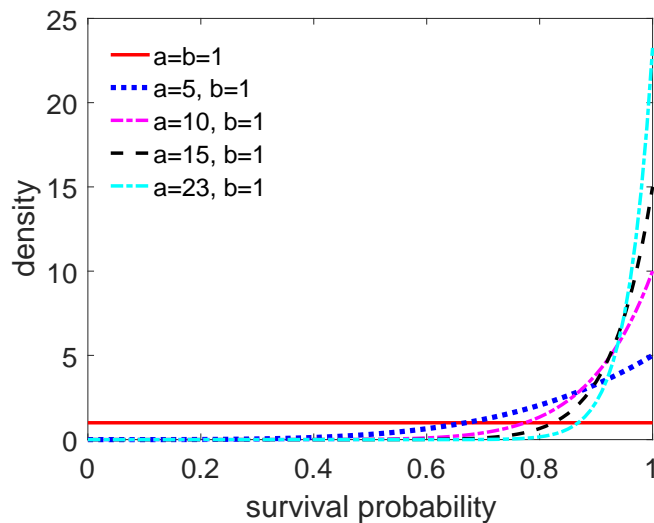


Figure 5: Untreated survival probabilities for different values of a , with $b = 1$

The remaining model parameters are chosen as follows:

Disease shares: As discussed above, we focus on cancer vs. non-cancer based on the ICD-10 main diagnosis code. According to the National Institute for Cancer Epidemiology

¹⁹For the beta distribution, $f'(x) = \frac{1}{B(a,b)} \cdot [(a-1)x^{(a-2)}(1-x)^{(b-1)} - x^{(a-1)}(b-1)(1-x)^{b-2}]$, which implies that $f'(x) \geq 0$ if $a \geq 1$ and $b \leq 1$.

and Registration (NICER), the cancer prevalence in Switzerland amounted to 3.9% in 2015.²⁰ We therefore set $\pi_L = 0.039$ and $\pi_H = 0.961$.

Health technology: To determine the values of α_L and $\bar{\alpha}$, we use their relationship with the survival rates of the treated, computed based on data from the Swiss Medical Statistics of Hospitals (MedStat) [Swiss Federal Statistical Office, 2016]. We compute the in-hospital survival rates for cancer and non-cancer patients who were assigned to the ICU in 2014. Given the endogenously determined levels of x_L and x_H , this allows us to calibrate α_L and α_H (and consequently $\bar{\alpha}$):

$$S_j(x|x \leq x_j) = \left(1 - \frac{\int_0^{x_j} (1 - \alpha_j) m(x) f(x) dx}{F(x_j)} \right) \quad \text{for } j = L, H \quad (19)$$

We calibrate the model towards the survival rates as computed from the data: $S_L(x|x \leq x_L) = 0.8601$ and $S_H(x|x \leq x_H) = 0.9103$.

Preferences: We assume a utility function of form $u(c) = c^\sigma$. We calibrate σ such that the marginal cost of saving a statistical life year (MCSLY) equals the value of a statistical life year (VSLY) in the absence of treatment, as indicated by (8). Given this calibration, our model only allows us to make statements about the divergence between the MCSLY for cancer and non-cancer treatments, but not about the level of HCE in general.

We derive the VSLY from the VSL estimate for Switzerland of CHF 10 million as reported by Baranzini and Ferro Luzzi [2001], based on an estimate of the risk wage premium for Swiss workers. An average worker faces roughly 40 years of residual life expectancy.²¹ We then calculate the value of a statistical life year (VSLY) under the assumption that it is constant over time, and given a discount rate of θ :²²

$$VSL = \sum_{t=0}^{43} \frac{VSLY}{(1 + \theta)^t} \quad (20)$$

²⁰See www.nicer.org/assets/files/statistics/prevalence/prev_counts_props_all.pdf last accessed Sep. 2016.

²¹The average age of the Swiss labor force in 2014 is approximately 41.4 years, and residual life expectancy at age 41 is about 43 years assuming, for simplicity, a female labor market participation of 50%.

²²For a discussion about the relationship between VSL and VSLY, see Hammitt [2007].

Using a discount rate of 2%, the resulting VSLY is CHF 143,879.²³ For the calibration, we keep the MCSLY_H at this amount for all levels of the spillover.

Treatment share: Of the 1,013,920 persons who were admitted at least once to a Swiss hospital in 2014, 7.12 % were admitted to an ICU [Swiss Federal Statistical Office, 2016]. Based on a population of 8.2 million in that year, this equals a treatment share of $Z = 0.9\%$.

Treatment costs: We approximate treatment costs using reimbursements categorized by their diagnosis-related group (DRG). Although in the theory section, we restrict the analysis to a single h for tractability, the treatment costs can be allowed to differ across (but not within) groups without major changes to the model. In our application, we therefore use two different treatment costs for cancer and non-cancer patients. The average reimbursement paid out for the treatment of cancer patients who were admitted to an ICU in 2014 amounted to CHF 25,173, while the respective reimbursement for non-cancer ICU patients amounted to CHF 19,130 [Swiss Federal Statistical Office, 2016].

Income: We use per-capita income of $y = \text{CHF } 78,432$ (2014 data).

Spillover parameter: The spillover γ is not observable, and it is not obvious what value would be appropriate. We therefore set $\gamma = 0$ in the baseline calibration, and recompute the model for a range of values for γ to assess its effect on the key model outcomes.

4.2 Results

We calibrate the model such that: (i) it solves the two first-order conditions and the budget constraint, (ii) it reproduces the observed treatment share (5) and the conditional survival rates given by (19), and (iii) the MCSLY matches the VSLY in the absence of spillovers. For given values of b and γ , we then have seven equations in seven unknowns: $\sigma, a, c, x_L, x_H, \bar{\alpha}$, and α_L . Finding the latter two, $\bar{\alpha}$ and α_L , involves the numerical solution of a fixed point problem: the α_j (for $j = L, H$) determine the optimal x_j , which, together

²³CHF 10 million is the lower bound reported in Baranzini and Ferro Luzzi [2001], whereas the upper bound is CHF 15 million. This would result in a VSLY of CHF 215,819. Using the upper rather than the lower bound (or an average) would change the resulting MCSLY s proportionately, but leave all results qualitatively unchanged.

with α_j , determine the survival rate S_j . The existence and uniqueness of the solution is ensured by the monotonicity of $f(x)$ and $m(x)$.

Table I summarizes the key outcomes of the numerical solution for three levels of the spillover parameter γ . The optimal cut-off for treatment, x_H , and its implied share of treated persons, $F(x_H)$, is not sensitive to the level of spillovers we choose for the calibration. The marginal effect of learning spillovers on the optimal cut-off on treatment in group L is positive and thus consistent with Proposition 1; it is also larger than for group H .

The levels of the calibrated values of α_H and α_L reflect the simulated difference in the treatment effectiveness of the two treatment technologies, which stems from the observed difference between $S_H(x|x \leq x_H)$ and $S_L(x|x \leq x_L)$.²⁴ The numbers in the benchmark case (where $\gamma = 0$) imply that the treatment of non-cancer diseases is on average 1.22 times more effective in terms of the mortality reduction.

Table I: Model results

	$\gamma = 0$	$\gamma = 1$	$\gamma = 2$	$\gamma = 4$
x_H	0.8179	0.8179	0.8179	0.8179
x_L	0.7345	0.7672	0.7800	0.7946
α_L	0.5271	0.4709	0.4452	0.4122
$\bar{\alpha}$	0.5842	0.5542	0.5233	0.4601
α_H	0.5842	0.5842	0.5842	0.5842
$F(x_L)$	0.0008	0.0021	0.0031	0.0047
$F(x_H)$	0.0093	0.0092	0.0092	0.0091
a	23.25	23.30	23.30	23.35
$MCSLY_H$	179849	179849	179849	179849
$MCSLY_L$	179849	229626	256958	297379
σ	0.3119	0.3119	0.3119	0.3119

Parameter values: $b=1$; $\pi_L=0.039$; $\pi_H=0.961$;
 $h_L=25,173$; $h_H=19,130$; $y=78,432$; $Z=0.009$.

Table I also displays the marginal cost of saving a statistical life year (MCSLY), which is computed by dividing the unit treatment costs by the mortality reduction at the margin for each technology; i.e., $h/\alpha_H(1 - x_H)$ and $h/\alpha_L(1 - x_L)$. For $\gamma > 0$, however, the equality no

²⁴Note that here for identification of the the two α -parameters we need to have the same distribution of untreated health conditions $F(x)$ in the two groups.

longer holds, because it is optimal to expand expenditure in sector L due to the additional learning benefits that accrue in sector H .²⁵ The increase in the $MCSLY_L$ is a measure of society's willingness to forgo resources in order to induce learning effects for a given level of spillovers.

To interpret γ , recall that $\alpha_H = \bar{\alpha} + \gamma\pi_L x_L$. Increasing the spillover parameter therefore increases α_H by $\Delta\alpha_H = \Delta\gamma\pi_L x_L$. An increase from $\gamma = 0$ to some $\tilde{\gamma} > 0$ can be translated into an average increase in survival of the treated persons in group H , formulated as follows:

$$\begin{aligned} \Delta S_H &= \tilde{S}_H(x|x \leq \tilde{x}_H, \tilde{\gamma}) - S_H(x|x \leq x_H, \gamma = 0) \\ &= \left(1 - \frac{\int_0^{x_H} (1 - \bar{\alpha} - \gamma\pi_L x_L) m(x) f(x) dx}{F(x_H)}\right) - \left(1 - \frac{\int_0^{x_H} (1 - \bar{\alpha}) m(x) f(x) dx}{F(x_H)}\right). \end{aligned} \quad (21)$$

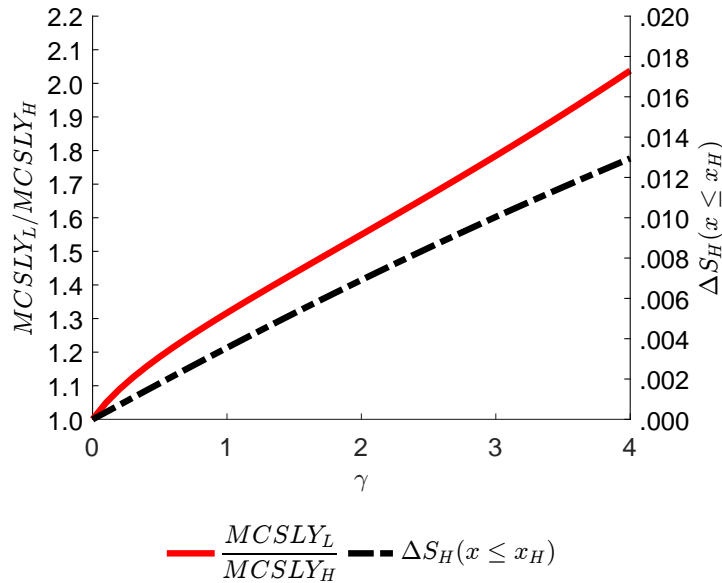


Figure 6: ICU Marginal cost of saving a statistical life year (MCSLY)

Figure 6 shows the relative marginal costs of a life extension as a function of γ (solid line,

²⁵Note that this qualitative result would hold even if we allowed for learning spillovers to originate in sector H as well, as long as they are lower than those that originate from applying treatment in sector L . The parameter γ can therefore be understood to represent the degree by which spillovers in L exceed those generated in H .

left axis), along with the resulting increase in survival probability for non-cancer patients (dashed line, right axis).²⁶ Suppose that cancer treatments are associated with learning spillovers with a magnitude of $\gamma = 3.1$, which lead to an increase in the survival rate for non-cancer patients of 1 percentage point (relative to the absence of learning spillovers). At this level of spillovers, the optimal ratio of the marginal cost of life extension is equal to 1.78, indicating that it is socially optimal to spend 78 % more on the margin for treatments in group L than in group H .

Last, we investigate the robustness of the relationship of the ratios of the MCSLYs to the choice of different parameter values. We do so by reproducing figure 6, but instead of varying the spillover parameter γ , we vary the distribution parameter a and the level of the costs for treatment L , h_L . We do so while keeping the implied increase in the average survival rate of the treated in group H constant at $\Delta S_H(x \leq x_H) = 1\%$. In each panel, the vertical line marks the result of our calibration exercise. Note that the ratio of the MCSLY is independent of the preference parameter σ , as this appears in the MCSLY of both types of diseases and drops out.

Figure 7 show the results of this exercise. The ratio of MCSLY, which governs the shape of the distribution of health states, is not sensitive to the parameter a (left panel) around the value of 23 to which we calibrated our model. For $a < 10$, the ratio of MCSLY diverge by much more than our results suggest, but such a low a is inconsistent with observed mortality rates at least in OECD countries, where most of the population is healthy.

However, the ratio of MCSLY's that is socially acceptable given that $\Delta S_H(x \leq x_H) = 1\%$ increases in the unit cost for cancer treatments (right panel). In the figure, the unit costs for noncancer treatments is held constant at CHF 19,130, and we allow the cost for cancer treatments to range from between one to five times this value. The graph shows that if end-of-life treatments are much more expensive than “regular” treatments, the optimal ratio

²⁶For this calculation, we hold all model parameters fixed at their calibrated levels based on $\gamma = 0$. We compute the counterfactual survival rate if γ were to be increased, which is qualitatively different from the calibration exercise in Table 1, where we recalibrate all values for different γ , because we do not know the level of the “true” γ .

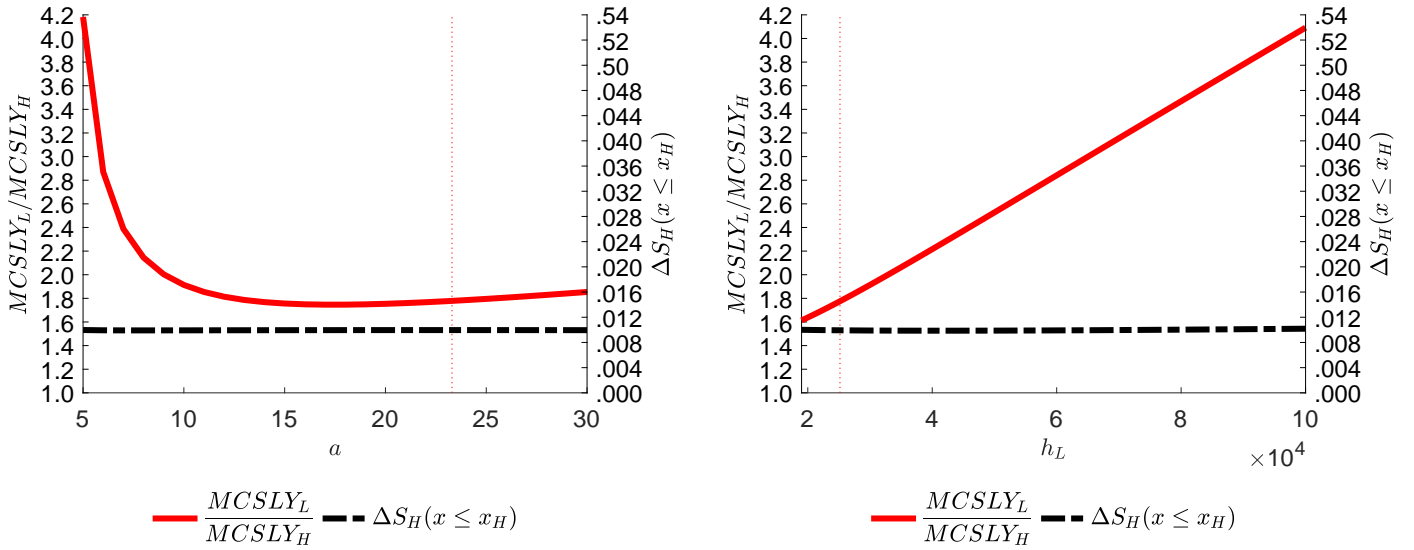


Figure 7: ICU Marginal cost of saving a statistical life year (MCSLY)

of $MCSLY_L/MCSLY_H$ increases significantly. Since the value for h_L we used in our model is close to that of h_L (because ICU admissions are costly for both cancer and noncancer patients), the optimal MCSLY-ratio associated with higher cost-ratios could significantly exceed the values in Figure 6. For example, if the treatment cost for cancer is four times the treatment cost of noncancer (i.e., around 80,000), the optimal MCSLY-ratio would be 3.4 rather than 1.78. In contrast, if medical costs are equal between cancer and noncancer, then the ratio would be 1.6.

5 Conclusions

In this paper, we develop a theory of the allocation of resources across consumption and health when the members of the population differ with respect to their health status and suitability to treatments with different productivity. We then extend the model to allow for learning spillovers accruing from the application of treatments in the sector with the low treatment effectiveness, which we label “end-of-life” because many patients die despite treatment. These spillovers improve the treatment effectiveness in the general health sector,

but not in the end-of-life sector itself. The extended model can be interpreted as a static representation of a dynamic process, in which curing one disease entails that patients subsequently die from another disease, such that the composition of the end-of-life sector changes over time. We show that under quite general conditions, the presence of learning spillovers leads to an increase in the optimal intensity of end-of-life treatments. Furthermore, we show that in the presence of learning spillovers, the MCSLY of using the ineffective technology (with reference only to the life years saved of the involved patients) exceeds the MCSLY of the technology that generates no (or fewer) spillovers.

We illustrate our model using data from the Swiss health care system and focus on admissions to intensive care units. We use cancer as the group of end-of-life diseases, whereas the remainder of the health sector comprises all other diseases. We find that the presence of learning spillovers which lead to a modest increase in survival within the group of non-cancer patients induce the optimal cost-benefit ratio of cancer treatments to significantly exceed the VSLY. Our results are conceptual in nature, and we make no claim that we can measure the strength of learning spillovers applying end-of-life treatments. Although there are indications that learning is important in the health sector, the magnitude of the spillovers is very difficult to quantify, and they will furthermore vary over different treatments and diseases within the end-of-life sector.

Our model stresses the role of technology and learning spillovers in determining the allocation of resources for different types of health care and overall consumption. We employ a social planning model and thus abstract from factors such as moral hazard or asymmetric information. The amount of resources spent on the different treatment options is determined by their relative effectiveness. Learning spillovers add to this effectiveness and therefore change the composition of resource use. The main message from our model is that if spillovers are present when treating end-of-life diseases, spending more on them than what would be expected based on the VSLY may in fact be socially optimal.

We do not claim that the Swiss health system is in fact optimal nor that informational

problems are negligible. The main purpose of the calibration to the Swiss system using a social planning model is to narrow the reasonable range within which the model parameters can be expected to be located, which is particularly important in the light of the fact that the effect of adding spillovers is ambiguous when allowing for the entire parameter space. We furthermore make no claim that the treatment effectiveness is the only determinant of the allocation of resources, but argue that technology is a particularly important aspect. Although other factors such as aging of the society and longevity, increased income, and insurance-induced moral hazard are important, they are unlikely to explain the increase in health spending across almost all countries despite substantial institutional differences [Newhouse, 1992]. In contrast, technological change affects countries in a similar way.

In terms of policy implications, our results show that in the absence of learning, the marginal cost of extending life by a statistical life year (MCSLY) should equal the value of a statistical life year (VSly) for all types of treatments. Simply being closer to death does not justify an increase in health expenditure. This suggests that the recent concern about an increase in costly yet seemingly futile treatments at the end of life is warranted. Basic health insurance should therefore cover health expenditures at the end of life (and in general) to the point where the benefits justify the costs, whereas excess coverage should be contingent on supplemental insurance. However, our calibrated model implies that even moderate levels of learning spillovers accruing from applying treatments in seemingly hopeless cases may lead to a substantial increase of the optimal MCSLY, relative to the VSly. This provides a caveat to the rationing argument: When limiting the expenditure for certain treatments, regulators should not only take into account the benefits that accrue to the individuals that are treated, but also any benefits accruing to the population in the long run due to learning. If learning effects are likely to be important in a particular health context, then it will be optimal to allow for treatment costs to exceed the VSly by a significant margin, and this margin increases with the difference in treatment costs across the two groups.

To operationalize the implications of our model, one would have to find a method for

an unbiased assessment of particular treatments with respect to their potential to generate learning benefits. This will likely require the collaboration of health professionals, representatives of health insurance firms, and the government. Even though it is clearly impractical to carry out such an analysis for all diseases and treatments, this could be done for the subset of diseases for which medical coverage is contentious from a cost-benefit point of view, and where learning effects are to be expected.

Possible extensions of this model include an explicit treatment of the dynamics, which would relax the assumption that the share of end-of-life diseases is constant, but which in turn would require that the growth dynamics of both the population and the economy have to be modeled. Moreover, it would be interesting to depart from the social planner model and instead investigate different existing health care systems and the resulting incentives for an under- or over-provision of health interventions at the end of life.

Compliance with ethical standards

This Study was funded by the Swiss National Science Foundation under the National Research Program 67 ("End of life") with grant number NWW1513. Since finishing his Ph.D. at the University of Basel in July of 2017, Matthias Minke has been employed by the Swiss Federal Agency of Health (BAG). Beat Hintermann is a full-time professor at the University of Basel. Both authors declare that they have no conflict of interest.

References

- Ahmad, AS, N Ormiston-Smith and PD Sasieni (2015). "Trends in the lifetime risk of developing cancer in Great Britain: comparison of risk for those born from 1930 to 1960." *British journal of cancer* 112(5), p. 943.
- Alemayehu, Berhanu and Kenneth E. Warner (2004). "The Lifetime Distribution of Health Care Costs." *Health Services Research* 39(3): 627–642.
- Baranzini, Andrea and Giovanni Ferro Luzzi (2001). "The economic value of risks to life:

- Evidence from the Swiss labour market.” *Swiss Journal of Economics and Statistics* 137(2): 149–170.
- Becker, Gary S., Thomas J. Philipson and Rodrigo R. Soares (2005). “The Quantity and Quality of Life and the Evolution of World Inequality.” *American Economic Review* 95(1): 277–291, February.
- Chabner, Bruce A. and Thomas G. Roberts Jr. (2005). “Chemotherapy and the war on cancer.” *Nature Reviews* 5: 65–72.
- Cubeddu, Luigi X., Irene S. Hoffmann, Nery T. Fuenmayor and Andrew L. Finn (1990). “Efficacy of Ondansetron (Gr 38032F) and the Role of Serotonin in Cisplatin-Induced Nausea and Vomiting.” *New England Journal of Medicine* 322(12): 810–816. PMID: 1689807.
- Cutler, David M. (2008). “Are We Finally Winning the War on Cancer?” *Journal of Economic Perspectives* 22(4): 3–26.
- Cutler, David and Mark McClellan (2001). “Is Technological Change in Medicine Worth It?” *Health Affairs* 20: 11–29.
- DePinho, Ronald A. (2000). “The age of cancer.” *Nature* 408(6809): 248–254.
- DeVita, Vincent T and Edward Chu (2008). “A history of cancer chemotherapy.” *Cancer research* 68(21): 8643–8653.
- Ellis, Randall P. (1998). “Creaming, skimping and dumping: Provider competition on the intensive and extensive margins.” *Journal of Health Economics* 17(5): 537–555.
- French, Eric B, Jeremy McCauley, Maria Aragon, Pieter Bakx, Martin Chalkley, Stacey H Chen, Bent J Christensen, Hongwei Chuang, Aurelie Côté-Sergent, Mariacristina De Nardi et al. (2017). “End-of-life medical spending in last twelve months of life is lower than previously reported.” *Health Affairs* 36(7): 1211–1217.
- Gelijns, Annetine and Nathan Rosenberg (1994). “The dynamics of technological change in medicine.” *Health Affairs* 13(3): 28–46.
- German Federal Statistical Office (2013). “Gesundheitsberichterstattung des Bundes.”
- Ginzberg, Eli (1980). “The high costs of dying.” *Inquiry* 17(4): 293–295.
- Grossman, Michael (1972). “On the concept of health capital and the demand for health.” *Journal of Political Economy* 80(2): 223–255.

- Hall, Robert E. and Charles I. Jones (2007). “The Value of Life and the Rise in Health Spending.” *The Quarterly Journal of Economics*: 39–72.
- Hammitt, James K. (2007). “Valuing Changes in Mortality Risk: Lives Saved Versus Life Years Saved.” *Review of Environmental Economics and Policy* 1(2): 228–240.
- Hanna, Nasser and Lawrence H. Einhorn (2014). “Testicular Cancer: A Reflection on 50 Years of Discovery.” *Journal of Clinical Oncology* 32(28): 3085–3092. PMID: 25024068.
- Hogan, C., J. Lunney, J. Gabel and J. Lynn (2001). “Medicare Beneficiaries’ Costs of Care in The Last Year of Life.” *Health Affairs* 20(4): 188–195.
- Jemal, Ahmedin, Elizabeth M. Ward, Christopher J. Johnson, Kathleen A. Cronin, Jiemin Ma, A. Blythe Ryerson, Angela Mariotto, Andrew J. Lake, Reda Wilson, Recinda L. Sherman, Robert N. Anderson, S. Jane Henley, Betsy A. Kohler, Lynne Penberthy, Eric J. Feuer and Hannah K. Weir (2017). “Annual Report to the Nation on the Status of Cancer, 1975–2014, Featuring Survival.” *JNCI: Journal of the National Cancer Institute* 109(9), p. djx030.
- Jones, Charles I. (2003). “Why Have Health Expenditures as a Share of GDP Risen So Much?”
- Kuhn, Michael, Stefan Wrzaczek, Alexia Prskawetz and Gustav Feichtinger (2011). “Externalities in a life cycle model with endogenous survival.” *Journal of Mathematical Economics* 47(4): 627–641.
- Lakdawalla, Darius N., Eric C. Sun, Anupam B. Jena, Carolina M. Reyes, Dana P. Goldman and Tomas J. Philipson (2010). “An economic evaluation of the war on cancer.” *Journal of Health Economics* 29(3): 333–346.
- Leaf, A. (1977). “Medicine and the Aged.” *New England Journal of Medicine* 297(16): 887–890.
- Lee, Chris P., Glenn M. Chertow and Stefanos A. Zenios (2009). “An Empiric Estimate of the Value of Life: Updating the Renal Dialysis Cost-Effectiveness Standard.” *Value in Health* 12(1): 80–87.
- Lubitz, James D. and Gerald F. Riley (1993). “Trends in Medicare Payments in the Last Year of Life.” *New England Journal of Medicine*.
- Lundberg, George D. (1993). “American health care system management objectives.” *JAMA: The Journal of the American Medical Association* 269(19): 2554–2555.

- Ma, Ching-to Albert (1994). "Health care payment systems: Cost and quality incentives." *Journal of Economics & Management Strategy* 3(1): 93–112.
- Meltzer, David (1997). "Accounting for future costs in medical cost-effectiveness analysis." *Journal of Health Economics* 16(1): 33–64.
- Newhouse, Joseph P (1992). "Medical care costs: how much welfare loss?" *The Journal of Economic Perspectives* 6(3): 3–21.
- Philipson, Tomas J., Gary Becker, Dana Goldman and Kevin M. Murphy (2010). "Terminal Care and the Value of Life Near Its End." Working Paper 15649, National Bureau of Economic Research.
- Phillips, Kathryn A. and Harold S. Luft (1997). "The policy implications of using hospital and physician volumes as indicators of quality of care in a changing health care environment." *International Journal for Quality in Health Care* 9(5): 341–348.
- Polder, Johan J., Jan J. Barendregt and Hans van Oers (2006). "Health care costs in the last year of life: The Dutch experience." *Social Science & Medicine* 63(7): 1720–1731.
- Pray, Leslie A (2008). "Gleevec: The breakthrough in cancer treatment." *Nature Education* 1(1), p. 37.
- Riley, Gerald F. and James D. Lubitz (2010). "Long-Term Trends in Medicare Payments in the Last Year of Life." *Health Services Research* 45(2): 565–576.
- Romley, John A., Anupam B. Jena and Dana P. Goldman (2011). "Hospital Spending and Inpatient Mortality: Evidence From California." *Annals of Internal Medicine* 154(3): 160–167.
- Rosen, Sherwin (1988). "The value of changes in life expectancy." *Journal of Risk and Uncertainty* 1(3): 285–304.
- Scitovsky, Anne A. (1994). "The High Cost of Dying: Revisited." *The Milbank Quarterly* 72(4), p. 561.
- Swiss Federal Statistical Office (2014). *Todesursachenstatistik*. Bundesamt für Statistik.
- (2016). "Medizinische Statistik der Krankenhuser." http://www.bfs.admin.ch/bfs/portal/de/index/infothek/erhebungen__quellen/blank/blank/mkh/01.html.
- Wessling, Jens (2013). "Risk loving in the end: The role of hope and risk preferences in end-of-life situations." Technical report, CINCH Health Economics Research Center.

Zweifel, Peter, Stefan Felder and Markus Meier (1999). “Ageing of population and health care expenditure: A red herring?” *Health Economics* 8(6): 485–496.

Zweifel, Peter, Lukas Steinmann and Patrick Eugster (2005). “The Sisyphus syndrome in health revisited.” *International Journal of Health Care Finance and Economics* 5(2): 127–145.

A Appendix

Optimality condition with spillovers

Substituting S and α_H into (3), the welfare function becomes

$$W(c, x_L, x_H) = u(c) \cdot S = u(c) \left[1 - \pi_L \left(\int_0^{x_L} (1 - \alpha_L)(1 - x)f(x) dx + \int_{x_L}^1 (1 - x)f(x) dx \right) - \pi_H \left(\int_0^{x_H} (1 - (\bar{\alpha} + \gamma x_L \pi_L))(1 - x)f(x) dx + \int_{x_H}^1 (1 - x)f(x) dx \right) \right],$$

subject to the budget constraint and the definition for Z as in (5). The first-order necessary conditions are given by

$$W_{x_H} = u(c^*)\pi_H\alpha_H^*m(x_H^*)f(x_H^*) - u'(c^*)(c^*\pi_H\alpha_H^*m(x_H^*)f(x_H^*) + h\pi_Hf(x_H^*)) = 0$$

$$W_{x_L} = u(c^*)\pi_L\alpha_L^*m(x_L^*)f(x_L^*) - u'(c^*)(c^*\pi_L\alpha_L^*m(x_L^*)f(x_L^*) + h\pi_Lf(x_L^*)) + \gamma\pi_L\pi_H \int_0^{x_H^*} (1 - x)f(x) dx \cdot [u(c^*) - u'(c^*)c^*] = 0.$$

Substituting $\mu_H \equiv \int_0^{x_H^*} (1 - x)f(x) dx$ and combining leads to eq. (11).

Proof of Proposition 1

We are interested in the signs of partial derivatives of the optimal treatment cut-off x_L with respect to the spillover parameter γ . Using the implicit function theorem, we can express both derivatives in terms of the second derivatives of the objective function $W(x_L, x_H)$:

$$\frac{\partial x_L}{\partial \gamma} = \frac{-W_{x_L\gamma}W_{x_Hx_H} + W_{x_Lx_H}W_{x_H\gamma}}{W_{x_Lx_L}W_{x_Hx_H} - W_{x_Lx_H}^2}. \quad (\text{A.1})$$

The denominator in (A.1) is the determinant of the Hessian matrix of the two-dimensional optimization problem, which is positive by assumption. It therefore suffices to determine the

sign of the numerator. Suppressing the asterisks for convenience, the first-order conditions for the planner's problem are

$$W_{x_H} = u'(c) \frac{\partial c}{\partial x_H} S + u(c) \frac{\partial S}{\partial x_H} = 0 \quad (\text{A.2})$$

$$W_{x_L} = u'(c) \frac{\partial c}{\partial x_L} S + u(c) \frac{\partial S}{\partial x_L} = 0. \quad (\text{A.3})$$

Taking the derivatives w.r.t. x_H , x_L and γ , the terms in (A.1) are given by

$$\begin{aligned} W_{x_H x_H} &= u'' \left(\frac{\partial c}{\partial x_H} \right)^2 S + u \frac{\partial^2 S}{\partial x_H^2} + u' \left(\frac{\partial^2 c}{\partial x_H^2} S + 2 \frac{\partial c}{\partial x_H} \frac{\partial S}{\partial x_H} \right) \\ &= u'' \left(\frac{\partial c}{\partial x_H} \right)^2 S + u \frac{\partial^2 S}{\partial x_H^2} + u' \left(-h \frac{\partial^2 Z}{\partial x_H^2} - \frac{\partial c}{\partial x_H} \frac{\partial S}{\partial x_H} - c \frac{\partial^2 S}{\partial x_H^2} - \frac{\partial c}{\partial x_H} \frac{\partial S}{\partial x_H} + 2 \frac{\partial c}{\partial x_H} \frac{\partial S}{\partial x_H} \right) \\ &= u'' \left(\frac{\partial c}{\partial x_H} \right)^2 S + \frac{\partial^2 S}{\partial x_H^2} (u - u'c) - u'h \frac{\partial^2 Z}{\partial x_H^2} \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned} W_{x_H x_L} &= u'' \frac{\partial c}{\partial x_L} \frac{\partial c}{\partial x_H} S + u \frac{\partial^2 S}{\partial x_H \partial x_L} \\ &+ u' \left(-\frac{\partial c}{\partial x_H} \frac{\partial S}{\partial x_L} - c \frac{\partial^2 S}{\partial x_H \partial x_L} - \frac{\partial c}{\partial x_L} \frac{\partial S}{\partial x_H} + \frac{\partial c}{\partial x_H} \frac{\partial S}{\partial x_L} + \frac{\partial c}{\partial x_L} \frac{\partial S}{\partial x_H} \right) \\ &= u'' \frac{\partial c}{\partial x_L} \frac{\partial c}{\partial x_H} S + \frac{\partial^2 S}{\partial x_H \partial x_L} (u - u'c) \end{aligned} \quad (\text{A.5})$$

$$\begin{aligned} W_{x_H \gamma} &= u'' \frac{\partial c}{\partial \gamma} \frac{\partial c}{\partial x_H} S + u \frac{\partial^2 S}{\partial x_H \partial \gamma} + u' \left(-\frac{\partial c}{\partial \gamma} \frac{\partial S}{\partial x_H} - c \frac{\partial^2 S}{\partial x_H \partial \gamma} - \frac{\partial c}{\partial x_H} \frac{\partial S}{\partial \gamma} + \frac{\partial c}{\partial x_H} \frac{\partial S}{\partial \gamma} + \frac{\partial c}{\partial \gamma} \frac{\partial S}{\partial x_H} \right) \\ &= u'' \frac{\partial c}{\partial \gamma} \frac{\partial c}{\partial x_H} S + \frac{\partial^2 S}{\partial x_H \partial \gamma} (u - u'c) \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned} W_{x_L \gamma} &= u'' \frac{\partial c}{\partial \gamma} \frac{\partial c}{\partial x_L} S + u \frac{\partial^2 S}{\partial x_L \partial \gamma} \\ &+ u' \left(\frac{c}{S} \frac{\partial S}{\partial \gamma} \frac{\partial S}{\partial x_L} - c \frac{\partial^2 S}{\partial x_L \partial \gamma} - \frac{\partial c}{\partial x_L} \frac{\partial S}{\partial \gamma} + \frac{\partial c}{\partial x_L} \frac{\partial S}{\partial \gamma} - \frac{c}{S} \frac{\partial S}{\partial \gamma} \frac{\partial S}{\partial x_L} \right) \\ &= u'' \frac{\partial c}{\partial \gamma} \frac{\partial c}{\partial x_L} S + \frac{\partial^2 S}{\partial x_L \partial \gamma} (u - u'c). \end{aligned} \quad (\text{A.7})$$

Substituting (A.4)-(A.7) into (A.1) and rearranging, we get

$$\begin{aligned}
-W_{x_L\gamma}W_{x_Hx_H}+W_{x_Lx_H}W_{x_H\gamma} &= -u''(u-u'c)S \left(\frac{\partial^2 S}{\partial x_L\partial\gamma} \left(\frac{\partial c}{\partial x_H} \right)^2 - \frac{\partial^2 S}{\partial x_H\partial\gamma} \frac{\partial c}{\partial x_H} \frac{\partial c}{\partial x_L} \right) \\
&\quad - \left(\frac{\partial^2 S}{\partial x_H^2} (u-u'c) - u'h \frac{\partial^2 Z}{\partial x_H^2} \right) \left(u''S \frac{\partial c}{\partial\gamma} \frac{\partial c}{\partial x_L} + (u-u'c) \frac{\partial^2 S}{\partial x_L\partial\gamma} \right). \quad (\text{A.8})
\end{aligned}$$

We start by determining the sign of the first line in (A.8). The term $(u-u'c) = u'(u/u'-c)$ is positive whenever utility is concave and life is strictly preferred over death (Rosen, 1988), such that the sign of the first line is equal to the sign of the parenthesis. Substituting for the partial derivatives (listed below for convenience) and $\mu_H \equiv \int_0^{x_H^*} (1-x)f(x)dx$, we obtain

$$\begin{aligned}
&-u''(u-u'c)S \left(\frac{\partial^2 S}{\partial x_L\partial\gamma} \left(\frac{\partial c}{\partial x_H} \right)^2 - \frac{\partial^2 S}{\partial x_H\partial\gamma} \frac{\partial c}{\partial x_H} \frac{\partial c}{\partial x_L} \right) \\
&= -u''(u-u'c) \left[\pi_H\pi_L\mu_H \left(-\frac{\pi_H f(x_H)}{S} (h + c\alpha_H(1-x_H)) \right)^2 \right. \\
&\quad \left. - x_L\pi_H\pi_L(1-x_H)f(x_H) \left(-\frac{\pi_H f(x_H)}{S} (h + c\alpha_H(1-x_H)) \right) \left(-\frac{\pi_L f(x_L)}{S} (h + c\alpha_L(1-x_L)) \right) \right] \\
&= -u''(u-u'c) \left(\frac{\pi_H f(x_H)}{S} \right)^2 (h + c\alpha_H(1-x_H))^2 \left[\pi_H\pi_L\mu_H - x_L(1-x_H)\pi_L^2 f(x_L) \right],
\end{aligned}$$

where we use the equality $\alpha_H(1-x_H) = \alpha_L(1-x_L)$ from the FONC (evaluated at $\gamma = 0$).

The sign of this term depends on the content of the square brackets and is positive if

$$\pi_L f(x_L) < \frac{\pi_H\mu_H}{x_L(1-x_H)}. \quad (\text{A.9})$$

Conditional on A.9 being true, a sufficient condition for proposition 1 to hold is that the second line in A.8 is non-negative. This is the case if the parentheses in the second line have a different sign (such that together with the minus sign, the expression is positive). We start

with the first parenthesis in the second line of (A.8):

$$\begin{aligned} \frac{\partial^2 S}{\partial x_H^2}(u - u'c) - u'h \frac{\partial^2 Z}{\partial x_H^2} &= \pi_H \alpha_H ((1 - x_H) f'(x_H) - f(x_H))(u - u'c) - u'h \pi_H f'(x_H) \\ &= -\pi_H \alpha_H (u - u'c) f(x_H) < 0 \end{aligned} \quad (\text{A.10})$$

where we have substituted the first-order condition $u'h = (u - u'c)\alpha_H(1 - x_H)$.

Last, we turn to the second parenthesis in the second line in A.8:

$$\begin{aligned} &\left(u'' S \frac{\partial c}{\partial \gamma} \frac{\partial c}{\partial x_L} + (u - u'c) \frac{\partial^2 S}{\partial x_L \partial \gamma} \right) \\ &= \left(\frac{u''}{u - u'c} S \frac{\partial c}{\partial \gamma} \frac{\partial c}{\partial x_L} + \frac{\partial^2 S}{\partial x_L \partial \gamma} \right) (u - u'c) \\ &= \left(\frac{u''c}{u - u'c} x_L \pi_H \pi_L \mu_H \frac{\pi_L f(x_L)}{S} (h + c\alpha_L(1 - x_L)) + \pi_H \pi_L \mu_H \right) (u - u'c) \\ &= \pi_H \pi_L \mu_H \left[\frac{u''cu}{(u - u'c)u'} \frac{x_L \pi_L \alpha_L (1 - x_L) f(x_L)}{S} + 1 \right] (u - u'c), \end{aligned}$$

where we used the equality $\frac{\pi_L f(x_L)}{S} (h + c\alpha_L(1 - x_L)) = \frac{u}{u'} \frac{\pi_L \alpha_L (1 - x_L) f(x_L)}{S}$ from the first-order condition. The sign of this expression is determined by the term within the square bracket.

It follows that the sufficient conditions for $\frac{\partial x_L}{\partial \gamma} > 0$ are given by

$$\pi_L f(x_L) < \frac{\pi_H \mu_H}{(1 - x_H) x_L} \quad \text{and} \quad (\text{A.11})$$

$$\pi_L f(x_L) < -\frac{(u - u'c)u'}{u''cu} \frac{S}{x_L(1 - x_L)\alpha_L} \quad \blacksquare \quad (\text{A.12})$$

Proof of Proposition 2

In order for the extension in L to exceed that in H , and thus for the health care sector to expand in relative terms, it must be that $-W_{x_L \gamma} W_{x_H x_H} + W_{x_L x_H} W_{x_H \gamma} > -W_{x_H \gamma} W_{x_L x_L} +$

$W_{x_L x_H} W_{x_L \gamma}$, or

$$\begin{aligned}
& -u''(u-u'c)S \left[\frac{\partial^2 S}{\partial x_L \partial \gamma} \left(\frac{\partial c}{\partial x_H} \right)^2 - \frac{\partial^2 S}{\partial x_H \partial \gamma} \frac{\partial c}{\partial x_H} \frac{\partial c}{\partial x_L} \right] \\
& - \left(\frac{\partial^2 S}{\partial x_H^2} (u-u'c) - u'h \frac{\partial^2 Z}{\partial x_H^2} \right) \left(u''S \frac{\partial c}{\partial \gamma} \frac{\partial c}{\partial x_L} + (u-u'c) \frac{\partial^2 S}{\partial x_L \partial \gamma} \right) > \\
& -u''(u-u'c)S \left[\frac{\partial^2 S}{\partial x_H \partial \gamma} \left(\frac{\partial c}{\partial x_L} \right)^2 - \frac{\partial^2 S}{\partial x_L \partial \gamma} \frac{\partial c}{\partial x_L} \frac{\partial c}{\partial x_H} \right] \\
& - \left(\frac{\partial^2 S}{\partial x_L^2} (u-u'c) - u'h \frac{\partial^2 Z}{\partial x_L^2} \right) \left(u''S \frac{\partial c}{\partial \gamma} \frac{\partial c}{\partial x_H} + (u-u'c) \frac{\partial^2 S}{\partial x_H \partial \gamma} \right)
\end{aligned}$$

It can first be noted, that the LHS of the above inequality is positive by Proposition 1 ($\frac{\partial x_L}{\partial \gamma} > 0$). We can therefore proceed by splitting this problem into two parts, that is, we compare the terms in brackets and the terms in parentheses on each side of the inequality. Then the above inequality is necessarily true if both of the following conditions hold:

$$\begin{aligned}
& -u''(u-u'c)S \left[\frac{\partial^2 S}{\partial x_L \partial \gamma} \left(\frac{\partial c}{\partial x_H} \right)^2 - \frac{\partial^2 S}{\partial x_H \partial \gamma} \frac{\partial c}{\partial x_H} \frac{\partial c}{\partial x_L} \right] > \\
& -u''(u-u'c)S \left[\frac{\partial^2 S}{\partial x_H \partial \gamma} \left(\frac{\partial c}{\partial x_L} \right)^2 - \frac{\partial^2 S}{\partial x_L \partial \gamma} \frac{\partial c}{\partial x_L} \frac{\partial c}{\partial x_H} \right] \tag{A.13}
\end{aligned}$$

$$\begin{aligned}
& - \left(\frac{\partial^2 S}{\partial x_H^2} (u-u'c) - u'h \frac{\partial^2 Z}{\partial x_H^2} \right) \left(u''S \frac{\partial c}{\partial \gamma} \frac{\partial c}{\partial x_L} + (u-u'c) \frac{\partial^2 S}{\partial x_L \partial \gamma} \right) > \\
& - \left(\frac{\partial^2 S}{\partial x_L^2} (u-u'c) - u'h \frac{\partial^2 Z}{\partial x_L^2} \right) \left(u''S \frac{\partial c}{\partial \gamma} \frac{\partial c}{\partial x_H} + (u-u'c) \frac{\partial^2 S}{\partial x_H \partial \gamma} \right) \tag{A.14}
\end{aligned}$$

Because $-u''(u-u'c)S > 0$, it is sufficient to consider the terms within the brackets, such that the inequality in A.13 holds if

$$\frac{\partial^2 S}{\partial x_L \partial \gamma} \left(\frac{\partial c}{\partial x_H} \right)^2 - \frac{\partial^2 S}{\partial x_H \partial \gamma} \left(\frac{\partial c}{\partial x_L} \right)^2 > \frac{\partial c}{\partial x_L} \frac{\partial c}{\partial x_H} \left(\frac{\partial^2 S}{\partial x_H \partial \gamma} - \frac{\partial^2 S}{\partial x_L \partial \gamma} \right)$$

Substituting the partial derivatives, using $\alpha_H(1 - x_H) = \alpha_L(1 - x_L)$, setting $\gamma = 0$ and simplifying gives

$$\begin{aligned}
& \pi_H \pi_L \mu_H \left(\frac{\pi_H f(x_H)}{S} (h + c\alpha_H m(x_H)) \right)^2 - \pi_H \pi_L x_L m(x_H) f(x_H) \left(\frac{\pi_L f(x_L)}{S} (h + c\alpha_L - m(x_L)) \right)^2 > \\
& \frac{\pi_H f(x_H)}{S} (h + c\alpha_H m(x_H)) \frac{\pi_L f(x_L)}{S} (h + c\alpha_L m(x_L)) \pi_H \pi_L (x_L m(x_H) f(x_H) - \mu_H) \\
& \iff \mu_H (\pi_H f(x_H))^2 - x_L m(x_H) f(x_H) (\pi_L f(x_L))^2 > \pi_H f(x_H) \pi_L f(x_L) (x_L m(x_H) f(x_H) - \mu_H) \\
& \iff \mu_H \left((\pi_H f(x_H))^2 + \pi_H f(x_H) \pi_L f(x_L) \right) > x_L m(x_H) f(x_H) \left((\pi_L f(x_L))^2 + \pi_H f(x_H) \pi_L f(x_L) \right) \\
& \iff \mu_H \cdot \pi_H f(x_H) > x_L m(x_H) f(x_H) \cdot \pi_L f(x_L) \\
& \iff \pi_H \cdot \mu_H > \pi_L f(x_L) \cdot x_L m(x_H)
\end{aligned}$$

Since the last line is identical to the first sufficient condition derived for Proposition 1, it follows that A.13 holds if the sufficient conditions for Proposition 1 are met.

We now derive the sufficient condition under which the left-hand side in A.14 exceeds the right-hand side (recall that if Proposition 1 holds, the LHS is positive). Substituting the partial derivatives and setting $\gamma = 0$ leads to the following condition for Proposition 2 to hold:

$$\begin{aligned}
& - \left(\pi_H \alpha_H \left((1 - x_H) f'(x_H) - f(x_H) \right) (u - u'c) - u' h \pi_H f'(x_H) \right) \cdot \\
& \quad \left[u'' c \pi_H \pi_L x_L \mu_H \cdot \pi_L f(x_L) A + (u - u'c) \pi_H \pi_L \mu_H \right] > \\
& - \left(\pi_L \alpha_L \left((1 - x_L) f'(x_L) - f(x_L) \right) (u - u'c) - u' h \pi_L f'(x_L) \right) \cdot \\
& \quad \left[u'' c \pi_H \pi_L x_L \mu_H \cdot \pi_H f(x_H) A + (u - u'c) \pi_H \pi_L x_L (1 - x_H) f(x_H) \right] \quad (\text{A.15})
\end{aligned}$$

where we substituted $A \equiv \frac{h + c\alpha_L(1 - x_L)}{S} = \frac{h + c\alpha_H(1 - x_H)}{S} > 0$ from (8).

The parenthesis on the LHS in (A.15) corresponds to A.10 in Proposition 1 and is therefore negative. Using equivalent operations, the parenthesis on the RHS in (A.15) becomes

$\pi_L \alpha_L (u - u'c)f(x_L)$. Simplifying leads to

$$\begin{aligned} \pi_H \alpha_H f(x_H) [u'' c x_L \mu_H \cdot \pi_L f(x_L) A + (u - u'c) \mu_H] > \\ \pi_L \alpha_L f(x_L) [u'' c x_L \mu_H \cdot \pi_H f(x_H) A + (u - u'c) x_L (1 - x_H) f(x_H)] \end{aligned} \quad (\text{A.16})$$

The term that pre-multiplies the bracket on the LHS is unambiguously larger than the corresponding term on the RHS. Furthermore, the first term in each bracket is negative, because $u'' < 0$. Because $\pi_L f(x_L) < \pi_H f(x_H)$, it follows that the first term in the brackets on the LHS is smaller in absolute magnitude (i.e., closer to zero) than the corresponding term on the RHS. For the inequality to hold, a sufficient condition is therefore that the second term in brackets is larger on the LHS than on the RHS, or that

$$\mu_H > x_L (1 - x_H) f(x_H) \quad (\text{A.17})$$

which corresponds to condition (17) in Proposition 2. ■

Partial derivatives used for proofs

$$\begin{aligned} \frac{\partial c}{\partial x_H} &= -\frac{h \frac{\partial Z}{\partial x_H} + c \frac{\partial S}{\partial x_H}}{S} = -\frac{\pi_H f(x_H)}{S} (h + c \alpha_H m(x_H)) < 0 \\ \frac{\partial^2 c}{\partial x_H^2} &= (h + c \alpha_H m(x_H)) \pi_H \left(\frac{2 \pi_H f(x_H)^2 \alpha_H m(x_H)}{S^2} + \frac{f(x_H) c \alpha_H}{S (h + c \alpha_H m(x_H))} - \frac{f'(x_H)}{S} \right) \\ \frac{\partial c}{\partial \alpha_H} &= -\frac{c}{S} \frac{\partial S}{\partial \alpha_H} < 0; \quad \frac{\partial^2 c}{\partial \alpha_H^2} = \frac{2c}{S^2} \left(\frac{\partial S}{\partial \alpha_H} \right)^2 > 0 \\ \frac{\partial c}{\partial x_L} &= -\frac{\frac{\partial Z}{\partial x_L} h + c \left(\frac{\partial S}{\partial x_L} \right)}{S} = -\frac{\pi_L f(x_L)}{S} (h + c \alpha_L m(x_L)) - \frac{\pi_H \pi_L \gamma \mu_H}{S} < 0 \\ \frac{\partial^2 c}{\partial x_L^2} &= -\frac{\left(h \frac{\partial^2 Z}{\partial x_L^2} + \frac{\partial c}{\partial x_L} \frac{\partial S}{\partial x_L} + c \frac{\partial^2 S}{\partial x_L^2} \right) S - \left(h \frac{\partial Z}{\partial x_L} + c \frac{\partial S}{\partial x_L} \right) \frac{\partial S}{\partial x_L}}{S^2} \end{aligned}$$

$$\frac{\partial^2 c}{\partial \alpha_H \partial x_L} = -\frac{\frac{\partial c}{\partial x_L} S - \frac{\partial S}{\partial x_L} c}{S^2} \frac{\partial S}{\partial \alpha_H} > 0$$

$$\frac{\partial c}{\partial \gamma} = -\frac{c}{S} \frac{\partial S}{\partial \gamma}$$

$$\frac{\partial^2 c}{\partial x_H \partial \gamma} = -\frac{\left(\frac{\partial c}{\partial \gamma} \frac{\partial S}{\partial x_H} + c \frac{\partial^2 S}{\partial x_H \partial \gamma}\right) S - \left(\frac{\partial Z}{\partial x_H} h + c \frac{\partial S}{\partial x_H}\right) \pi_H \pi_L x_L \mu_H}{S^2}$$

$$\frac{\partial^2 c}{\partial x_L \partial \gamma} = -\frac{\left(\frac{\partial c}{\partial \gamma} \frac{\partial S}{\partial x_L} + c \frac{\partial^2 S}{\partial x_L \partial \gamma}\right) S - \left(\frac{\partial Z}{\partial x_L} h + c \frac{\partial S}{\partial x_L}\right) \frac{\partial S}{\partial \gamma}}{S^2}$$

$$\frac{\partial^2 c}{\partial x_L \partial x_H} = -\left(\frac{\left(\frac{\partial c}{\partial x_H} \frac{\partial S}{\partial x_L} + c \frac{\partial^2 S}{\partial x_L \partial x_H}\right) S - \left(\frac{\partial Z}{\partial x_L} h + c \frac{\partial S}{\partial x_L}\right) \frac{\partial S}{\partial x_H}}{S^2}\right)$$

$$\frac{\partial Z}{\partial x_H} = \pi_H f(x_H); \quad \frac{\partial^2 Z}{\partial x_H^2} = \pi_H f'(x_H)$$

$$\frac{\partial Z}{\partial x_L} = \pi_L f(x_L); \quad \frac{\partial^2 Z}{\partial x_L^2} = \pi_L f'(x_L)$$

$$\frac{\partial S}{\partial x_H} = \pi_H \alpha_H m(x_H) f(x_H) > 0; \quad \frac{\partial^2 S}{\partial x_H^2} = \pi_H \alpha_H (m(x_H) f'(x_H) - f(x_H))$$

$$\frac{\partial S}{\partial \alpha_H} = \pi_H \mu_H > 0; \quad \frac{\partial^2 S}{\partial \alpha_H^2} = 0$$

$$\frac{\partial^2 S}{\partial \alpha_H \partial x_H} = \pi_H f(x_H) m(x_H); \quad \frac{\partial^2 S}{\partial \alpha_H \partial x_L} = 0$$

$$\frac{\partial S}{\partial x_L} = \alpha_L \pi_L m(x_L) f(x_L) + \pi_H \pi_L \gamma \mu_H > 0; \quad \frac{\partial^2 S}{\partial x_L^2} = \alpha_L \pi_L (m(x_L) f'(x_L) - f(x_L))$$

$$\frac{\partial^2 S}{\partial x_L \partial x_H} = f(x_H) \pi_H (1 - x_H) \frac{\partial \alpha_H}{\partial x_L}$$

$$\frac{\partial S}{\partial \gamma} = \pi_H \pi_L x_L \mu_H; \quad \frac{\partial^2 S}{\partial \alpha_H \partial \gamma} = 0$$

$$\frac{\partial^2 S}{\partial x_H \partial \gamma} = \pi_H \frac{\partial \alpha_H}{\partial \gamma} m(x_H) f(x_H); \quad \frac{\partial^2 S}{\partial x_L \partial \gamma} = \pi_H \pi_L \mu_H$$

$$\frac{\partial \alpha_H}{\partial \gamma} = \pi_L x_L; \quad \frac{\partial^2 \alpha_H}{\partial x_L \partial \gamma} = \pi_L$$

$$\frac{\partial \alpha_H}{\partial x_L} = \gamma \pi_L; \quad \frac{\partial^2 \alpha_H}{\partial x_L^2} = 0$$