

Rationing of health care: is there an economic rationality to it?

Peter Zweifel^{1,2}

Published online: 7 April 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract The point of departure of this Editorial is the fact that we all are engaged in self-rationing in our everyday lives. We would like to spend more money on all sorts of nice things and devote more time to our cherished activities. Imposed rationing is characteristic of wartime governments, who seek to prevent the rich from gobbling up the resources left by the army. Since the publication in 1987 of David Callahan's *Setting Limits: Medical Goals in an Aging Society* (Callahan, *Setting limits: medical goals in an aging society*, Simon & Schuster, New York, 1987), rationing of health care has become a widely debated issue (the Internet is full of pertinent entries). While rationing has also been addressed by health economists, there are three puzzling observations. First, Callahan (Callahan, *Setting limits: medical goals in an aging society*, Simon & Schuster, New York, 1987) wrote for an American audience whereas rationing was introduced by the British National Health Service (NHS) well before 1987, with little debate. Second, the economic theory of rationing had been laid out by James Tobin [*Ectrica* 20(4): 521–533, 1952] as early as 1952—but health economists seem to have neglected his groundwork when writing about rationing. Third, they accept government-imposed rationing as inevitable in the case of health care, as though the self-rationing alternative was unavailable. An attempt is made here to provide rational explanations for these puzzles.

Demand for rationing is induced by health insurance (and even more so, NHS provision)

A well-known side effect of insurance is moral hazard. In the case of health, consider Mr X who is willing to pay €100 out of pocket for a drug, and let the rate of copayment be 33 percent. Evidently, the drug may cost as much as €300 at the pharmacy, and Mr X will still buy it. In this case, insured patients' marginal willingness to pay is inflated by a factor of three, the inverse of the rate of copayment. In Fig. 1, the 'true' demand function D_{oop} (reflecting marginal willingness to pay out of pocket) is swiveled around to become the observed demand function D_{ins}^{obs} (as seen by the pharmacist in the example). Given a positively sloped supply function S , the market equilibrium moves from E to E_{ins} . The prediction is clear: health insurance coverage causes an increase in the price of health care p_M , the quantity of healthcare services M transacted, and hence healthcare expenditure $HCE = p_M \times M$. These effects are the more pronounced, the lower the rate of copayment.

In a NHS promising access to care free of charge (as the British NHS originally did), observed demand does not react to price. Therefore, D_{NHS}^{obs} runs vertical, giving rise to equilibrium at E_{NHS} which reflects even stronger moral hazard effects than the insurance-cum-copayment alternative. In the words of NHS historian Geoffrey Rivett [19], writing about the 1960s, "Costs kept rising. The BMJ (British Medical Journal) believed that, ignoring the British capacity for muddling through, the NHS was heading in the direction of bankruptcy. The illusion that they were getting something for nothing led people to seek free supplies of household remedies for which they had previously paid, such as aspirin, laxatives, first-aid dressings and cotton

✉ Peter Zweifel
peter.zweifel@uzh.ch

¹ University of Zurich, Zurich, Switzerland

² Present Address: Kreuth 371, 9531 Bad Bleiberg, Austria

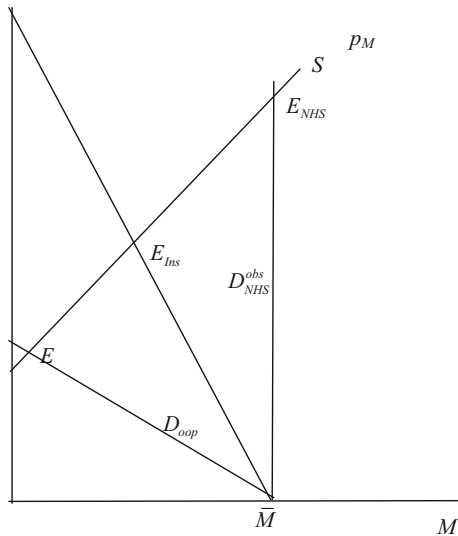


Fig. 1 Demand for rationing induced by insurance coverage

wool. Many were going round with two pairs of spectacles when one would have done”.

Evidently, the equilibrium point E_{NHS} was beyond the budget allocated to the NHS early on, creating a demand for rationing. Figure 2 illustrates how a budget constituting a binding constraint induces rationing. According to the hyperbolic locus $B = p_M \cdot M$, a given HCE budget can be spent on little health care M at a high price p_M (reflected by point \tilde{F}) or much health care at a low price (point F_{NHS} ; point F_{Ins} will be discussed below). Clearly, the logic of a NHS would call for the low-price solution; however, it cannot be implemented because sufficient supply would fail to be forthcoming at that price. Therefore, the NHS has to settle for point G on the supply function, limiting the

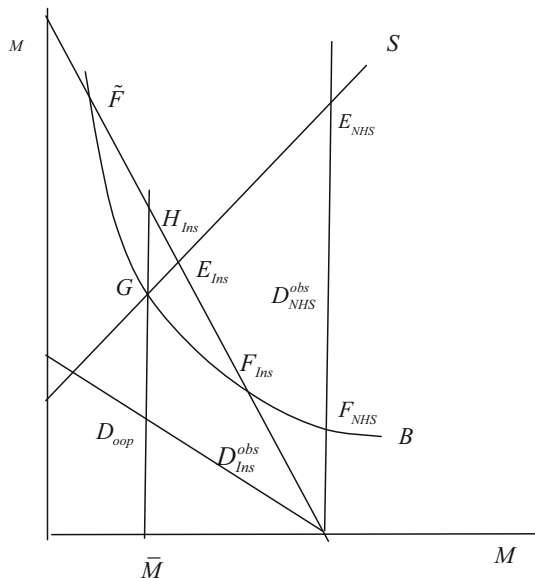


Fig. 2 Rationing induced by budgeting

volume of healthcare services to \bar{M} , well below observed demand given zero copayment (at F_{NHS}).

Since the NHS did not want to introduce charges, it had to resort to means other than price to reign in demand, e.g. waiting for access. In the 1980s, this was complemented by outright denials of costly interventions to individuals above a certain age. Currently, rationing is disguised as the NICE rule that a quality-adjusted life year should not cost more than £30,000 (which of course hits again the aged, whose number of life years are counted to begin with).

Now turn to health insurance. Social health insurance, especially, is under political pressure to stabilize (rates of) contributions, which calls for limiting HCE. This too gives rise to a fixed budget $B = p_M \cdot M$. Like a NHS, social health insurance has to settle for a point such as G on the supply function. If the budget amounts to a binding constraint, it induces rationing as well, a fact usually negated in the public debate (see [3] for Germany). However, there is a difference in that (given the same budget), pressure to ration is less than in a NHS, as can be seen by comparing the distances between E_{Ins} and G and E_{NHS} and G in Fig. 2, respectively. This is due to the fact that a positive rate of copayment serves to limit demand already.

This comparison leads to the first puzzle. If the demand for rationing is induced especially by a NHS with its zero-copayment insurance coverage as it were, why is it that the rationing debate was triggered in the United States, where substantial rates of copayment are common? A possible answer to this question is so-called dynamic moral hazard. Figure 2 can be reinterpreted as referring not to total HCE but to HCE associated with a particular medical service or pharmaceutical. As noted by [19], from the 1960s on, innovation in medical technology increasingly took place in the United States rather than the United Kingdom. Now insurance coverage encourages costly product innovation, resulting in a dynamic moral hazard effect [15]. Rationing by the NHS has the effect of limiting potential sales of new pharmaceuticals and other innovations, thus making their launch less attractive. One could say that the NHS was more successful in mitigating dynamic moral hazard than private health insurance and Medicare/Medicaid in the United States. Therefore, the situation depicted in Fig. 2 presented itself with a higher frequency in the United States than in the United Kingdom, and with it the issue of rationing, prompting Callahan [4] to trigger the debate.

The strange way health economists have been addressing the rationing issue

As early as 1952, James Tobin [20] had noted, “The question (of rationing) becomes the more compelling because in some countries rationing is evidently more than a

wartime measure ...". As reported by [10], rationing characterized the British National Health Service (NHS) early on. Yet health economists were not willing to apply Tobin's analysis to healthcare rationing by notably pointing to the efficiency loss caused by it. This efficiency loss is evident in Fig. 2. Between points G and E_{Ins} (let alone E_{NHS}), observed marginal willingness to pay (reflected by the $D_{\text{Ins}}^{\text{obs}}$ function and the $D_{\text{NHS}}^{\text{obs}}$ function, where in fact it is infinite) exceeds the marginal cost of supplying the extra service (reflected by the S function). In the case of health insurance with positive copayment, this is clearly visible as a Harberger triangle $GE_{\text{Ins}}H_{\text{Ins}}$.

Health economists have consistently failed to emphasize this efficiency loss (see e.g. [14, 17], and [3]). A possible reason is the profession's widespread skepticism regarding the concept of a demand function, citing the malleability of patient preferences due to the asymmetry of information between them and healthcare providers (see e.g. [18]). Yet health economists willingly espouse e.g. the concept of generic substitution. However, if patients are to consent to it in response to price differentials between branded and generic drugs, they must exhibit a nonzero elasticity of substitution. This in turn implies a nonzero own-price elasticity of demand because the two elasticities are intimately related (see [16], Sect. 3.1 for the case of CES preferences).

Thus, health economists typically accept government-imposed rationing of health care as a necessity. For instance, Uwe Reinhardt [17] titled an article, "Rationing health care: what it is, what it is not, and why we cannot avoid it"; the title of a book by Alan Maynard and Karen Bloor [14] reads, "Our Certain Fate: Rationing in Health Care", both referring to government-imposed rationing. The lead in the fight against rationing was left to philosophers (e.g. [12] and ethicists (e.g. [1])). Economists even missed out on the fact, cited by Binstock [1]), that only 3.5 percent of U.S. Medicare HCE (as of 1999) was spent on high-cost, high-tech medical interventions for persons aged 65 and older. They also failed to notice that countries with a high share of old-aged persons are not characterized by a high GDP health share [2]), supporting the 'red herring' hypothesis according to which closeness to death rather than age is the main driver of HCE [6].

At the other extreme, the rationing issue is not addressed at all in important works. For instance, the Oxford Handbook of Health Economics (with a British co-editor) [5] does not even cite 'rationing' (or 'prioritization', a euphemism for rationing, as noted by [3]) in its index (the Elgar Companion to Health Economics [9] does).

Self-rationing as the rational alternative

Increasingly, access to health care is seen as a human right. However, this is also true of food and shelter, where individuals nevertheless continue to perform self-rationing. Indeed, there is scope for self-rationing in the case of healthcare services as well. Here are a few possibilities.

- Higher deductibles in return for lower contributions. This is an option Swiss social health insurers are allowed to offer. Admittedly, plans with deductibles in excess of the legal minimum attract favorable risks. Yet, when controlling for these risk-selection effects, [21] still find deductibles to have a 'true' limiting effect on moral hazard. Note that individuals of higher age who prefer to eschew this type of self-rationing at least contribute to their increased HCE through higher contributions. This alternative can also be implemented in a NHS-type system; the government would have to offer a tax reduction (e.g. in the guise of a fixed pay-back of income or sales tax) to citizens willing to accept a deductible.
- Higher rates of copayment in return for lower contributions. Based on the RAND Health Insurance Study, Manning and Marquis [13] estimated the optimal rate of copayment (i.e. where the difference between the risk premium and the efficiency loss due to moral hazard is maximum) to be 50 %. While European citizens may be more risk-averse than their U.S. counterparts, permitting them to select plans with higher than the near-zero rates of copayment presently prevailing in social health insurance would enable them to perform self-rationing. Again, they would have to be promised a tax reduction in a NHS context.
- Bonus options for no claims. The alternatives cited up to this point share the downside of exposing individuals to a cumulation of risks. When ill, they suffer both a health loss and a financial loss, having to bear part of the medical expense. In countries where income replacement in the event of sickness is limited (as in the Netherlands), there is even the third risk of an income loss. Small wonder that the Dutch would have to be hugely compensated for accepting an annual deductible of just €500 [11]. At least the cumulation of health loss and financial loss due to copayment can be relieved by a bonus option for no claims. The insured can always sacrifice a bonus in return for full insurance coverage if they deem the out-of-pocket payment excessive. In this way, they are able to shift the financial loss to later periods by paying a higher future contribution. A NHS could offer reduced taxation during a period of illness, to be made up by higher tax payments later.

- Medical saving accounts (MSAs). By not claiming health insurance benefits, the insured can build up credit. In the case of the United States, this credit can only be used for financing medical care later in life, in contradistinction to the bonus option. MSAs therefore shift moral hazard to higher age, potentially to the most costly 2 years before death in life where reigning it in might make a great difference [6]. By way of contrast, in Singapore the balance of a MSA can be bequeathed, in particular to members of the family who are free to use the money as they see fit—a powerful motive not to squander it on medical care that holds little promise of prolonging healthy life (Central Providence Act [5]). In her critical review of MSAs, [8] neglects this crucial difference.
- Managed Care-type plans in return for lower contributions. This option characterizes notably employer-contracted U.S. health insurance and Swiss social health insurance, with [21] again finding these plans to lower HCE after controlling for risk selection. Individuals signing up for Managed Care explicitly permit the treating physician to act as the rationing agent—in return for a reduced contribution. Note the difference between these options and legally imposing gatekeeping by primary care physicians as in the Netherlands (with no commitment on the part of politicians or health insurers to reduce contributions). This amounts to government-imposed rationing, which is resisted by the Dutch since experimental evidence suggests that they exhibit positive willingness to pay for returning to free physician choice [11].

In sum, there is a wide variety of options for permitting individuals to self-ration health care. While open also to NHS-type systems, they tend to be offered by health insurers under the pressure of competition. To retain their clientele, they need to develop policies featuring a favorable ratio of expected utility to contribution, which rarely involves outright denial of access to life-saving therapies. Admittedly, this implies that HCE will continue to grow. But then, what is wrong with that if it is the reflection of citizens' preferences for self-rationing, expressed through their choice of health insurance policy?

References

1. Binstock, R.H.: Aged-based rationing of health care. In: Encyclopedia of aging (2002) (website)
2. Binstock, R.H.: Healthcare costs around the world: is aging a fiscal 'black hole'. *Generations* **27**(4), 37–42 (1993)
3. Breyer, F.: Implicit versus explicit rationing of health services. *CESifo DICE Rep* **11**(1), 7–15 (2013)
4. Callahan, D.: *Setting limits: medical goals in an aging society*. Simon & Schuster, New York (1987)
5. Central Providence Act of Singapore (Chapter 36), <http://statutes.agc.gov.sg/aol/search/display/view.w3p;ident=49476963-b5dd-47e3-ad86-abf8ab544f7d;page=0;query=DocId%3A%22f1b7803c-71dd-4732-b7f0-231fef8142ca%22%20Status%3Ainforce%20Depth%3A0;rec=0#pr25-he->. (courtesy Dr Kelvin Tan, Ministry of Health). Accessed 25 Feb 2015
6. Felder, S., Werblow, A., Zweifel, P.: Do red herrings swim in circles? Controlling for the endogeneity of time to death. *J. Health Econ.* **29**(2), 205–212 (2010)
7. Glied, S., Smith, P.C. (eds.): *The Oxford handbook of health economics*. Oxford University Press, Oxford (2011)
8. Hsu, J.: Medical savings accounts: what is at risk? World health report background paper, p. 17. World Health Organization, Geneva (2010)
9. Jones, A.M. (ed.): *The Elgar companion to health economics*, 2nd edn. Edward Elgar, Cheltenham (2012)
10. Kapp, M.B.: De facto health-care rationing by age: the law has no remedy. *J. Legal Med.* **19**, 232–249 (1998)
11. Leukert-Becker, K., Zweifel, P.: Preferences for health insurance in Germany and the Netherlands—a tale of two countries. *Health Econ. Rev.* **4**(1), 1–11 (2014)
12. Malone, N., Rycroft-Malone, J.: Equity and rationing in the NHS: past to present. *J. Nurs. Manag.* **6**(6), 325–332 (1998)
13. Manning, W.G., Marquis, M.S.: Health insurance: the tradeoff between risk pooling and moral hazard. *J. Health Econ.* **15**(5), 609–639 (1996)
14. Maynard, A., Bloor, J.: *Our certain fate: rationing in health care*. Office of Health Economics, London (1998)
15. Newhouse, J.P.: Medical care costs: how much welfare loss? *J. Econ. Perspect.* **6**(3), 3–21 (1992)
16. Ramskov, J., Munksgaard, J.: *Elasticities—a theoretical introduction*. The Balmorel Project, University of Copenhagen (2001) (Website)
17. Reinhardt, U.E.: Rationing health care: what it is, what it is not, and why we cannot avoid it. *Baxter Health Pol Rev* **2**, 66–69 (1996)
18. Rice, T.: *The economics of health reconsidered*, 3rd edn. Health Administration Press, Chicago (2009)
19. Rivett, G.: *NHS history, Ch 1: 1948–1957, Establishing the NHS/emerging problems (1948)* (Website)
20. Tobin, J.: Survey of the theory of rationing. *Ectrica* **20**(4), 521–533 (1952)
21. Trottmann, M., Beck, K., Zweifel, P.: Supply-side and demand-side cost sharing in de-regulated social health insurance: which is more effective? *J. Health Econ.* **29**(4), 489–498 (2010)