

## Rationing as a Response to Supply Side Constraints

*The previous two research notes in this series highlighted the nearly unlimited demand for healthcare in South Africa, as well as the very limited supply thereof. In this note we consider various responses to the supply constraint of limited resources and examine the issue of rationing within the context of the proposed NHI. As is the case for nearly all other industries, in the health sector the fundamental question of economics also has to be solved: how can the scarce resources be allocated in the most efficient and equitable way?*

### 1. Introduction

Scarce resources in the South African healthcare system introduce the need for rationing demand and/or supply. In most sectors goods are allocated via the market price. Allocation of healthcare by price amounts to bidding based on income, and is expected to result in an overprovision of healthcare to the affluent and under-provision of healthcare to the poor. It is widely accepted that rationing the demand for healthcare via price only is not desirable – presumably because of its association with the principle of the right to life and also potentially harsh consequences of denying access to healthcare. As

Maynard eloquently described it,<sup>1</sup> the scarcity in healthcare means that “choices have to be made about who will be given the ‘right’ of access to care and who, as a result of denial, will be left in pain and discomfort, and, in the limit, to die.”

Thus, in most instances there is some degree of government intervention and free service provision in healthcare. In markets where the price is determined by the government or where there is no price (as would be the case if comprehensive healthcare benefits were available to all citizens free of charge), other ways of allocating the scarce resources have to be found. In

the healthcare sector many different words are used for the allocation of scarce resources: managed care, priority setting, cost containment, preauthorisation, gatekeeping, clinical guidelines, waiting lists, benefit design and so forth.

Although the methods whereby rationing occurs are manifold and controversial, rationing is an unavoidable part of any healthcare system, as is summarised by the extract below:

*“Priorities have to be set in all healthcare systems whatever their level of expenditure and regardless of the methods of financing and delivery that are adopted. The nature of the choices that have to be made*

1. Maynard, A., 1999. “Rationing Health Care: an Exploration,” *Health Policy*, Vol. 49, pp.5-11

This research note forms part of a series of special National Health Insurance (NHI) notes which can be accessed on the Econex website [www.econex.co.za](http://www.econex.co.za). In the interest of constructively contributing to the NHI debate, the Hospital Association of South Africa (HASA) has commissioned a comprehensive costing and human resource research project with Econex. HASA has given Econex and its partners at Stellenbosch University academic independence with respect to this project. The results of the project will be placed in the public domain in order to foster constructive debate.

and the locus of these choices do vary between systems, but the inevitability of priority setting is universal.”<sup>2</sup>

## 2. What is Rationing?

For the purpose of this note rationing can be defined as “allocating healthcare resources in the face of limited availability, [by] withholding beneficial interventions from some individuals. It is socially inevitable and prevalent. It is implicit in co-payment schedules, gatekeeper decisions, utilisation reviews and capitation contracts.”<sup>3</sup>

In theory the ideal allocation would be to direct the scarce supply of health resources and services so that one firstly fulfils the healthcare needs of those who need it most, before spending resources on lower levels of need<sup>4</sup>. This implies avoiding both errors of inclusion and errors of exclusion. Errors of inclusion imply service delivery to those who may need it less and errors of exclusion refer to cases where the system fails to meet the healthcare needs of individuals who may need it most. Of course the definition of “real need” is controversial and morally loaded – applying such a definition in

practice is empirically complex and bound to create much polarised debate.

On the demand side it is important to establish efficient mechanisms to match demand with the services required or supplied, while supply side management attempts to align reimbursement mechanisms with providers that best manage risk and performance.<sup>5</sup>

## 3. Rationing Mechanisms

From what we have learned thus far, it is clear that the question is not whether healthcare should or will be rationed, but rather one of how this will be done to ensure equity and efficiency. Since rationing occurs in all health markets, it is appropriate to consider the different forms of rationing one might encounter.

### 3.1 Explicit vs. Implicit Rationing

Friedenberg<sup>6</sup> makes the distinction between direct (explicit) and indirect (implicit) methods of rationing when he lists a few ways of rationing healthcare:

“The method of rationing can be based on rationing the physician’s time,

rationing new technologies, rationing by gatekeepers, rationing by limiting referrals, or rationing by limiting expensive procedures. Other more indirect methods of rationing include rationing by inconvenience (i.e. requiring excessive paperwork or making patients wait an exceptionally long time for an appointment), rationing by policy (i.e. declaring that a service is not covered), or rationing by contract (i.e. stating within the contract what services are covered at each level, with the patient deciding which level and amount he or she wishes to pay).”

In most countries rationing occurs via a mixture of explicit and implicit rationing mechanisms. In the absence of any explicit forms of rationing, one would expect to see implicit forms of rationing such as waiting lines and waiting lists – we are all familiar with waiting lists for elective surgery and organ transplants, for instance. Waiting lists and queuing may seem democratic in theory but as De Zueleta neatly explains, this rationing strategy “risks allocating resources in a piecemeal, unfair fashion with those

2. Ham, C., 1995. “Synthesis: what can we learn from international experience?” in Maxwell, R.J. (ed.), *Rationing Health Care*, Churchill Livingstone, Brit. Medical Bull, Vol. 51, No. 4, pp.819-30
3. Boscheck, R., 2004. “Healthcare Rationing and Patient Rights,” *Intereconomics*, Nov/Dec, pp.310-313
4. In health economics “need” is a relative concept and efficient allocation will usually be based on the greater value of outcome or most cost efficient uses.
5. Ruff, B., undated presentation. “‘Demand and Supply’ Side Considerations in Managing the Health Sector.”
6. Friedenber, R.M., 2003. “Rationing in Health Care: Changing the Patterns of Health Care,” *Radiology*, Vol. 227, pp.5-8. (p.6)

## About ECONEX

ECONEX is an economics consultancy that offers in-depth economic analysis covering competition economics, international trade, strategic analysis and regulatory work. The company was co-founded by Dr. Nicola Theron and Prof. Rachel Jafta in 2005. Both these economists have a wealth of consulting experience in the fields of competition- and trade economics. They also teach courses in competition economics and international trade at the University of Stellenbosch. Our newest director, Cobus Venter, who joined the company during 2008, is also a consultant economist at the Bureau for Economic Research (BER) in Stellenbosch. For more information on our services, as well as the economists and academic associates working at and with Econex, visit our website at [www.econex.co.za](http://www.econex.co.za).

'shouting the loudest', or those with the most money and/or influence gaining privileged access to the goods."<sup>7</sup>

In the absence of any explicit rationing strategy, healthcare workers such as doctors may be in the position to play a role in the rationing of healthcare by for instance deciding what type of care is allocated to whom.<sup>8</sup> This is sometimes referred to as bedside rationing. However, this implicit rationing strategy comes with problems of its own as this places doctors in a position where they "face the daunting prospect of being at once the advocate for the individual patient and the arbiter of distributive justice for the practice population."<sup>9</sup>

Explicit rationing mechanisms are often fiercely criticised though, but it is frequently for the wrong reasons. Explicit rationing is usually an effort to improve on the implicit rationing in the system, but because it draws attention to the allocation problem it tends to evoke an emotional response. In many instances explicit rationing may be a more transparent, accountable and legitimate way of allocating scarce medical services.

In many advanced healthcare systems allocation is complex and technocratic, based on advanced algorithms regarding cost to benefit ratios, with benefit often

measured as the number of healthy years a treatment or intervention would add to the patient's life. The British NHS is a good example of such an approach. The National Institute for Health and Clinical Excellence (NICE) was established to determine guidelines for healthcare provision and the treatment of specific illnesses.<sup>10</sup> In effect, NICE is responsible for allocating scarce healthcare resources in the UK, i.e. they ration medical care by determining which procedures or medication is covered by the NHS and which is not. Allocation decisions are based on an index comparing the effectiveness of any treatment in terms of the potential number of quality-adjusted life years (QALYs) gained from any suggested treatment. The institute confirmed an upper limit of £30,000 per QALY gained in January 2009. In other words, according to the NICE guidelines, the NHS will cover most treatments with an incremental cost effectiveness ratio (ICER) of £30,000 per QALY gained.

Putting an economic value on a person's life or quantifying the financial value of one extra year of life for a potentially terminally ill patient is concerned as inhumane and appalling to most of us. Prioritisation can however be established in more democratic ways (such as by consensus or through majority

voting)<sup>11</sup>, but it is important to realise that these kinds of decisions (putting a financial value on the social and other benefits gained through specific types of healthcare) will necessarily form part of any rationing mechanism.

### 3.2. Demand Side vs. Supply Side Rationing

Rationing measures which prevent patients from freely expressing demand for healthcare, such as co-payments or user charges, provider networks and so forth, are described as demand side rationing. Supply side rationing includes measures such as regulation of the pharmaceutical market and medical technologies, controlling admissions to medical and nursing schools, only providing certain immunisation vaccines free of charge, hospital bed licensing etc.<sup>12</sup> In a national health system (such as that currently proposed in South Africa) where healthcare is provided free at the point of service, rationing measures will usually be on the supply side.

### 3.3 Managed Care

In its definition of managed care, the US National Library of Medicine<sup>13</sup> explains that this approach intends to...

*"...reduce unnecessary health care costs through a variety of mechanisms, including:*

- 
7. De Zulueta, P., 2007. "Sharing the Health: Rationing in General Practice," *The New Generalist*, Vol. 5, Nr. 3, pp.50-52
  8. Pinto, C.G. & Aragão, F., 2003. "Health Care Rationing in Portugal. A Retrospective Analysis," *Associação Portuguesa de Economia da Saude*. Available at: [http://www.apes.pt/files/dts/dt\\_012003.pdf](http://www.apes.pt/files/dts/dt_012003.pdf).
  9. Marinker M, 1990. "Changes and Developments in Primary Care," in: L'Etang, H. (editor), *Health Care Provision under Financial Constraint: A Decade of Change*, London: Royal Society of Medicine; . p.141-8.
  10. [www.nice.org.uk](http://www.nice.org.uk)
  11. Pinto, C.G. & Aragão, F., 2003. "Health Care Rationing in Portugal. A Retrospective Analysis," *Associação Portuguesa de Economia da Saude*. Available at: [http://www.apes.pt/files/dts/dt\\_012003.pdf](http://www.apes.pt/files/dts/dt_012003.pdf)
  12. Pinto, C.G. & Aragão, F., 2003. "Health Care Rationing in Portugal. A Retrospective Analysis," *Associação Portuguesa de Economia da Saude*. Available at: [http://www.apes.pt/files/dts/dt\\_012003.pdf](http://www.apes.pt/files/dts/dt_012003.pdf)
  13. <http://www.ncbi.nlm.nih.gov/sites/entrez>
-

*economic incentives for physicians and patients to select less costly forms of care; programs for reviewing the medical necessity of specific services; increased beneficiary cost sharing; controls on inpatient admissions and lengths of stay; the establishment of cost-sharing incentives for outpatient surgery; selective contracting with health care providers; and the intensive management of high-cost health care cases. The programs may be provided in a variety of settings, such as health maintenance organizations and preferred provider organizations."*

Theoretically, managed care seeks to provide the means of offering healthcare services within a defined network of service providers who are then given the responsibility of managing and providing quality and cost effective healthcare. In order for the managed care company to perform its duties, it hires or contracts with doctors, nurses and other healthcare providers. Models of managed care include the following:<sup>14</sup>

- The network gatekeeper (where a primary care physician acts as "gatekeeper" by controlling the patient's access to further specialist medical care within the scheme);
- A formulary (list of drugs that the medical scheme will cover and from which the doctor must make prescriptions);
- Capitation (a method of payment for

health services in which a provider is paid a fixed, per capita amount in advance for each enrollee – regardless of the actual number or nature of services provided to each member);

- Utilisation reviews (prior authorisation where teams of physicians within the network usually look at the medical history of patients in order to determine the appropriateness of treatment);
- Peer reviews (where data is collected and used to compare the performance of doctors relative to their peers).

## 4. Rationing in South Africa

Given the limited public health budget and the shortage of medical practitioners (detailed in NHI Note 4) as well as the preference for doctors, it is clear that there is a rationale for rationing in the South African health system.

Currently, the private sector rations through a combination of price and some managed care or gatekeeping efforts. Private healthcare is rationed largely via the medical schemes. At a first level, it is rationed by price through the monthly contributions of members and co-payments. In many instances, healthcare is then further rationed by gatekeeping and pre-authorisation in the sense that a referral is sometimes needed from the GP,

and in some cases authorisation from the particular medical scheme, before certain specialist services will be paid for by the insurer. A good example of demand side management is Discovery's Vitality programme that aims to prevent illness and injury through promoting wellness behaviour and, for instance, paying for preventative screening tests such as mammograms, blood pressure and cholesterol tests, HIV screening tests, etc.

When looking at problems experienced in the private sector, price remains one of the most important constraints on demand for private healthcare. Figure 1 draws on data from the General Household Surveys (GHS) for 2006 and 2007 and suggests that price is the top complaint of individuals using private sector health providers.

In contrast to the private sector where price seems to be the main complaint of users, there is no charge for primary healthcare in the public sector and the co-payments required for higher levels of care (hospitals) are minimal and often poorly enforced. In the public sector rationing occurs via a combination of explicit gatekeeping (using the clinics as the point of entry into the system) and queuing. It has been reported that gatekeeping has not been managed well and that many users are able to use higher levels of care (such as hospitals)

14. Seema Roy, March 2002. "Managed health care and preferred provider organisations." *Economic Research Report by the Competition Commission*. p.16-17

## More Information

ECONEX regularly publishes Research Notes on various relevant issues in South African competition, trade and applied economics. For access to previous editions of Research Notes, or other research reports and published articles, go to: [www.econex.co.za](http://www.econex.co.za)

If you want to add your name to our mailing list, please send an e-mail to [iris@econex.co.za](mailto:iris@econex.co.za)

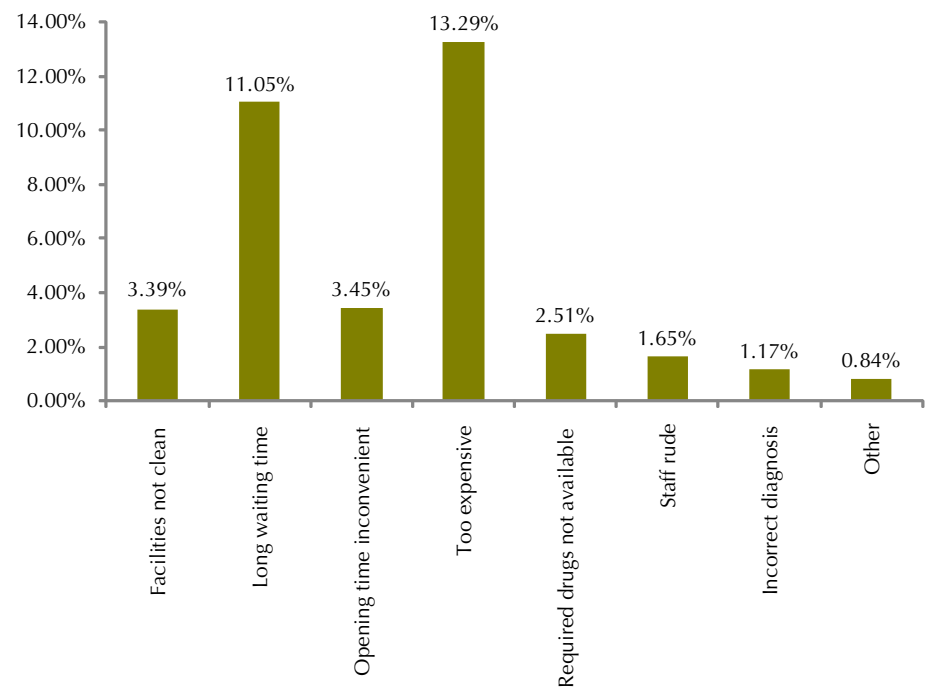
as their point of entry. Thus it appears that to a large extent demand is curtailed through rationing of the last resort: long waiting times.

In the data, waiting times are cited as the main problem by users of the public health system (see Figure 2). According to the GHS 2006 and 2007, 39.0% of users of public health services experienced long waiting times. This was the most frequent complaint amongst this group of users. In contrast only 3.2% of users of public facilities cited prices as a complaint.

The 1998 and 2003 Demographic and Health Surveys tell the same story (Figure 3). It is also important to note the increase in people complaining of long waiting times (26% in 1998 and 42% in 2003). However, the question regarding reasons for dissatisfaction is structured in a different way: it first asks if the respondent was dissatisfied and then probes what the reason for the dissatisfaction was. It can be argued that this will lower the prevalence of complaints regarding waiting times – at least compared to the General Household Survey question where users were simply asked whether they experienced long waiting times. This is also in line with the qualitative research employing focus group interviews and exit interviews.<sup>15</sup>

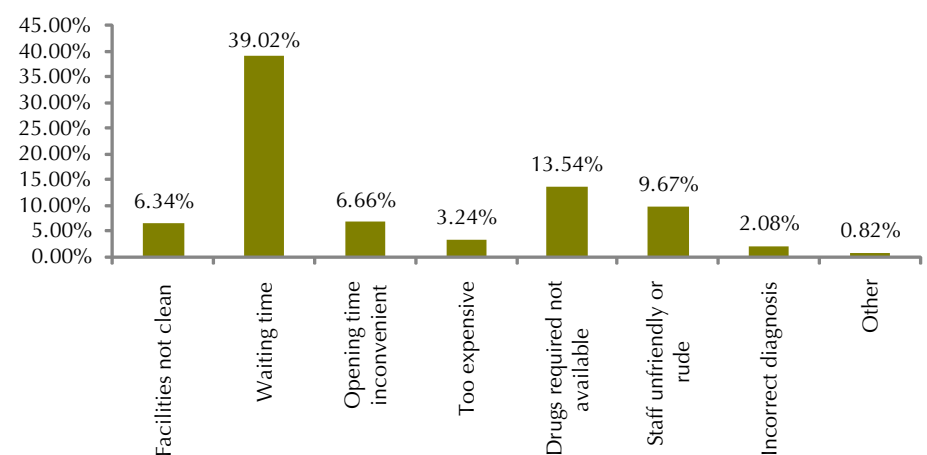
Although it may effectively curtail demand, queues can waste a considerable amount of otherwise productive time. A 2004 study found that the average waiting time in private clinics was between 10 and 40 minutes, while waiting times ranged from 50 minutes to 3 hours in public clinics.

Figure 1: Problems when Consulting a Health Worker in a Private Sector Facility, 2006 & 2007



Source: GHS, 2006 & 2007

Figure 2: Problems when Consulting a Health Worker in a Public Sector Facility, 2006 & 2007



Source: GHS, 2006 & 2007

Individuals experience these waiting times as costly: as reported in NHI Note 3, long queues have been shown to be one of the most important reasons

motivating poor families to consult private clinics<sup>16</sup> or GPs at their own cost although primary healthcare is available free of charge.

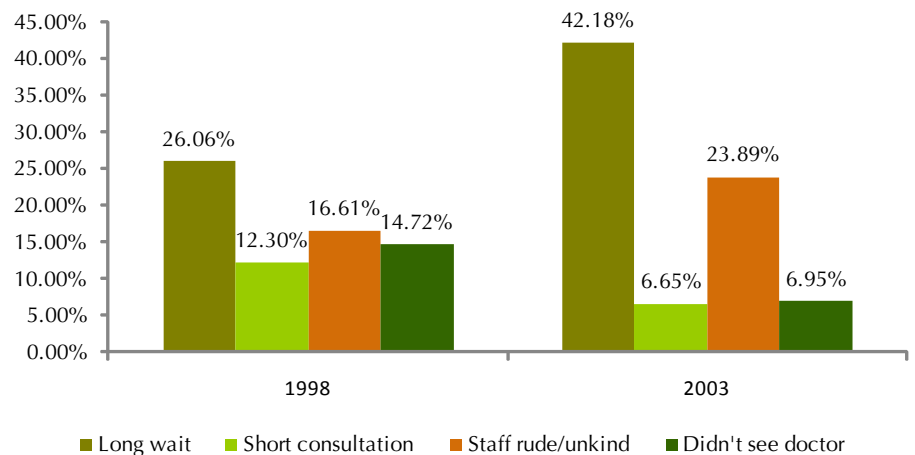
15. See for instance Palmer, N. 1999. "Patient choice of primary health care provider," *South African Health Review*. Durban: Health Systems Trust  
 16. Mills, A., Palmer, N, Gilson, L, McIntyre, D. Schneider, H, Sinanovi, E. 2004. "The performance of different models of primary care provision in Southern Africa," *Social Science Medicine* Vol 59, pp. 931-43.

## 5. Improving Rationing and Allocation in South Africa

In the design of any national or social health insurance system, policy makers have to take into account many different country-specific factors, in order to develop a system that offers sufficient benefits while still being affordable, and most importantly, sustainable over time. Scarce healthcare resources will have to be allocated in the most efficient way, taking into account various constraints (financial, political, staffing, facilities, training etc.) while trying to maximise social welfare. Designing an appropriate health insurance system for this country will necessarily also entail difficult decisions with far reaching consequences – such as the trade-off between providing a comprehensive benefits package to a smaller portion of the population, or providing basic cover to a larger portion or the entire population.<sup>17</sup> The large proportion of individuals that are unemployed, the prevalence of poverty, the shortage of doctors and the relatively small tax base<sup>18</sup> would be important considerations when implementing a NHI in South Africa.

The WHO also emphasize five areas that need specific attention when designing a health financing system. These are administrative efficiency and transparency, stability of funding, equity, pooling, and purchasing. The literature suggests that management is a substantial obstacle to improving the delivery of public healthcare and thus the first and the last factors listed here, may be of particular importance in the case of South Africa. In terms of design there is clearly no “one-size-fits-all” solution and the concept of

Figure 3: Reasons for Dissatisfaction for Users of Public Clinics and Hospitals, 1998 & 2003



Source: DHS, 1998 & 2003 as quoted in Pelzer, K. 2009, "Patient experiences and health system responsiveness in South Africa," BMC Health Service Research, Vol. 9

optimal design is defined in the context of the country's own needs, institutions and the size of its health budget.

In terms of the proposed National Health Insurance plan, one of the key concerns is regarding the effectiveness, adequacy and fairness of the outlined rationing strategies. The proposed examples of ceilings on utilisation (3 GP visits per person per year) are somewhat higher than current levels of utilisation in the public sector<sup>19</sup> and would require resource increases that are unlikely to be feasible in the short to the medium term unless there was evidence of spare capacity. There is, unfortunately, no evidence of spare capacity, and it was shown in NHI Note 4 that the country is already experiencing a severe shortage of doctors and nurses.

The NHI plan also proposes capitation payments for GPs. The motivation is to remove any perverse incentives to overmedicate or overcharge patients. However, while there is an analytical

argument for suspecting that GPs may have an incentive to overmedicate and overcharge, the extent of such behaviour will differ in practice and is difficult to measure. Given the scarcity of doctors in South Africa, there is reason to argue that there is not a problem with utilisation, and thus overmedication and overcharging may be limited. If this is true, the strategy will not have a large impact. Capitation may however also provide an incentive for under provision or the provision of poor quality services as the doctor receives a fixed fee per patient in advance. The doctor will have an incentive to refer patients to specialists, rather than treating them him/herself. With the limited amount of specialists locally (see NHI Note 4) this could be even more problematic

## 6. Conclusion

It is vital to consider rationing options and strategies carefully, seeing that the

17. Hsiao, W.C. & Shaw, R.P., 2007. "Social Health Insurance for Developing Nations," WBI Development Studies, The World Bank, Washington D.C.

18. Only 5.2 million individuals pay personal income taxes and 25% of them cover 75% of personal tax. See Chait, G., 2009. "The Risk of Upending the Tax Base," 29 April. Available at: <http://whythawk.com/analysis/the-risk-of-upending-the-tax-base.html>

19. In NHI Note 4 it was pointed out that public sector utilisation is currently between 2.1 and 2.3 visits per person per year.

NHI plan envisions that there will be minimal rationing by price and that there are apparently no plans to implement the private sector demand management practices that are currently in place (e.g. savings accounts, pre-authorisation and selective contracting). Looking at the experience of other countries who have implemented successful NHI programmes, it can be seen that some form of rationing will be required under a NHI.

If the provision of services is not strictly managed under a NHI, then other

rationing mechanisms will evolve over time. If the efforts to ration in an explicit and purposeful way are not sufficient to contain demand at levels that are feasible given the limited supply of healthcare resources, there will necessarily be implicit rationing. With the anticipated increase in demand under the NHI and the lack of a clear and effective rationing strategy, this could result in even longer queues. These rationing strategies are wasteful (time of patients) and often unfair. Another likely consequence of inadequate rationing is more rationing via price – and given the

expected increase in demand, prices will only increase without the appropriate rationing mechanisms in place. In addition, rationing helps to contain costs, and if this is not used effectively, healthcare costs are expected to escalate. These prospects are concerning as they threaten the central vision of the NHI, namely of providing equitable access to high quality care.