CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Efficiency Incentives in the Nursing Home Sector

Lessons from Implementing the new Quality Framework in the Netherlands

Voeseneka, Tessa; Shestalova, Victoria; Mikkers, Misja; Agrell, Per J.; Bogetoft, Peter

Document Version
Final published version

Published in:
Decision Analytics Journal

DOI:
10.1016/j.dajour.2022.100063

Publication date:
2022

License
CC BY-NC-ND

Citation for published version (APA):
Voeseneka, T., Shestalova, V., Mikkers, M., Agrell, P. J., & Bogetoft, P. (2022). Efficiency Incentives in the Nursing Home Sector: Lessons from Implementing the new Quality Framework in the Netherlands. *Decision Analytics Journal*, *3*, Article 100063. https://doi.org/10.1016/j.dajour.2022.100063

Link to publication in CBS Research Portal

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 20. Apr. 2024

Decision Analytics Journal 3 (2022) 100063

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

Efficiency incentives in the nursing home sector: Lessons from implementing the new Quality Framework in the Netherlands

Tessa Voesenek [a,b,*], Victoria Shestalova [a], Misja Mikkers [a,b], Per J. Agrell [c], Peter Bogetoft [d]

[a] NZa Dutch Healthcare Authority, PO Box 3017, 3502 GA Utrecht, The Netherlands
[b] Tilburg University, Tilburg School of Economics and Management (TiSEM), PO Box 90153, 5000 LE Tilburg, The Netherlands
[c] UCLouvain, Center of Operations Research and Econometrics (CORE), POB L1.03.01, B-1348 Louvain-la-Neuve, Belgium
[d] CBS Copenhagen Business School, Department of Economics, Copenhagen Business School CBS, Porcelaenshaven 16 A, 2000 Frederiksberg, Denmark

ARTICLE INFO

Keywords:
Benchmarking
Incentive regulation
Quality standard
Data envelopment analysis
Nursing home sector

ABSTRACT

In many markets there is some form of regulation; and when new policy measures are introduced into a particular sector, the question of how to structure an adequate regulatory process arises. This paper contributes to the literature by describing the approach taken in the nursing home sector in the Netherlands, explaining all the steps needed for implementation and how regulatory challenges were addressed. These challenges involved making methodological choices and the integration of elements (incentives, quality, cost-efficiency, data validation, organizational commitment) into a consistent and credible regulatory framework. Drawing lessons for other sectors and countries, we particularly emphasize the usefulness of Data Envelopment Analysis as a flexible and adaptable regulatory tool for both learning and incentive provision. However, the scope and specification of the model must be consistent with costaccounting practices and quality definitions in order for DEA to be implemented effectively.

1. Introduction

In most market economies, semi-public sectors such as healthcare and utilities are characterized by a large degree of public intervention. Notably, prices in these sectors are subject to regulation in many cases. While the economic theory of regulation gives us a general guide for setting appropriate incentives for regulated companies (e.g., [1–3]), regulators still face many issues when putting these principles into practice, and numerous trade-offs are involved when the recommendations from different theories and disciplines are combined in one regulatory setting. A more unified approach to incentive regulation would therefore be helpful, and lessons from practice would provide constructive feedback for both theoretical and empirical researchers working in this field.

In this paper, we describe and analyze the essential elements of one such regulatory process, focusing on recent experiences in the Dutch nursing home sector with respect to the development of a new method of tariff regulation based on benchmarking as part of the implementation of a new Quality Standard for nursing homes [4]. The need for adequate tariff regulation arises in many other regulatory contexts in which the regulator needs to balance incentives for both quality and efficiency, such as when regulating utilities and other (semi-)public sector activities, including some healthcare services. The lessons from this case study may therefore also be of interest to other regulated sectors of the economy such as education, energy and public transport. This is especially the case, as most issues discussed in our study – for example, related to quality, input and output measurement, heterogeneity of firms – are not unique for the DEA-frontier approach taken by this study, but hold also for other methods.

The paper belongs to the strand of the theoretical and empirical literature that combines advances of the fields of DEA and incentive regulation. Our addition to this literature is both theoretical (by proposing a unified efficiency-model based regulation framework for addressing the major regulatory challenges in semi-public sectors) and empirical (by providing a case study in which we discuss the practical implementation of this unified framework to the case of tariff regulation in the nursing home care sector). In particular, the earlier literature pointed to the usefulness of frontier benchmarking methods in regulation contexts both for efficiency measurement and for providing incentives (Coelli [5],Agrell and Bogetoft [6,7]). More specifically, regulatory schemes based on Data Envelopment

* Corresponding author at: Tilburg University, Tilburg School of Economics and Management (TiSEM), PO Box 90153, 5000 LE Tilburg, The Netherlands.
E-mail addresses: tvoesenek@nza.nl (T. Voesenek), vshestalova@nza.nl (V. Shestalova), mmikkers@nza.nl (M. Mikkers), per.agrell@uclouvain.be (P.J. Agrell), pb.eco@cbs.dk (P. Bogetoft).

[1] For more on the theory of when DEA based regulation may be optimal, see (Bogetoft [3,33,34,35,36,37]) and Bogetoft and Otto [32].

[2] In particular a broader vision of elderly care and the improvement of the interpretation of the Quality Standard in the view of the labor market restrictions aggravating during the pandemics.

https://doi.org/10.1016/j.dajour.2022.100063
Received 16 February 2022; Received in revised form 22 April 2022; Accepted 28 April 2022
Available online 4 May 2022
2772-6622/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Analysis (DEA) productivity models have some considerable advantages compared to popular CPI-X revenue cap regulation Agrell et al. [8].[1] Yet, the earlier literature still provided a little guidance on how to deal with a combination of practical challenges, actually arising in nearly any real application Jamasb and Pollitt [9], Gearhart [10]. Here we argue that these challenges are interlinked and can only be solved through an evolutionary integrated dynamic regulatory process which encourages learning and improvement. While most of the scientific papers on efficiency analysis end with various sophisticated analyses of the efficiency scores obtained (see for example the research topics covered in DEA hospital applications Kohl et al. [11]), there is actually a long way to go to design a comprehensive regulatory benchmarking process to build in an incentive mechanism that encourages efficiency by creating incentives not to lag behind others. This process requires a collaborative and interdisciplinary effort from various experts from both the regulator and the sector. Most importantly for long-term application, as opposed to a one-shot academic exercise, is the credible and consistent fit between the model and the regulation framework. Our case study provides an example of initiating a benchmarking model in a real-life regulatory context in the Netherlands, thus allowing us to address both methodological and applied challenges of designing this type of model, and ways to overcome these challenges. Since the successive Dutch government places more focus on other areas regarding the organization of nursing care,[2] the further implementation steps have been put on hold. Yet, the design of such a method and the process undertaken so far provides a very valuable experience that could also be used in other sectors and contexts.

The insights obtained in this study could be of interest to various disciplines. We believe that many of these issues could well be encountered by any regulator in any sector, making this paper relevant to a broad international readership. In particular, we will address:

(1) challenges in selecting a suitable form of regulatory benchmarking method in practice;
(2) cost-accounting issues, and their consequences for the scope of the model, including the challenge related to the treatment of capital in these models;
(3) the role of quality in the regulatory benchmark-based model;
(4) model specification issues;
(5) the model integration challenge in an existing regulation framework.

In this paper we address these generic issues from the perspective of the regulatory process and show their relevance by discussing the choices considered in the context of a specific regulatory model for the Dutch nursing homes.

Before we go into more detail on the model, we will first describe the nursing home care sector in the Netherlands, the new regulatory objective and the related regulatory challenges faced by the Dutch Healthcare Regulator (NZa) in this sector. This will be followed by a discussion of the methodological approach taken by the regulator towards achieving this objective. Here, we will highlight the trade-offs related to the five major challenges outlined above, explaining the rationale behind the decisions made by the regulator on the main design issues that needed to be addressed. The data and results sections provide an overview of what was revealed in the first step of the new regulatory process, which we will refer to as the 'baseline measurement'. The final section will conclude with a discussion of the policy lessons learned.

2. Problem description

The Dutch nursing home sector consists of approximately 350 providers, which together have over 2000 locations [12]. The services that they provide are multiproduct in nature. They provide inpatient care targeted at elderly persons with dementia and persons with severe physical limitations, and in addition some providers deliver home-care services or day-care treatments. This means that providers differ in terms of the product mix that they provide. In the Netherlands, almost all providers are private, non-profit organizations.

Nursing home care is almost exclusively publicly financed. All citizens of the Netherlands are covered by mandatory social insurance for long-term care (LTC) arranged by the Dutch state.[3] In addition, clients of nursing homes also make some means-tested copayments into the total budget for LTC, regardless of where and from which provider those clients receive nursing-home care.[4] The national LTC budget is allocated to regional contracting agencies, which pay for LTC in their regions. These regional agencies are responsible for funding and access to nursing-home care in their own region. The agencies face exogenous demand for care: only persons who are identified as requiring 24-hour daily care are entitled to this care; the relevant indications are provided through an assessment carried out by an independent institution acting at the national level. In 2019, around 170,000 persons (1% of the population) were eligible for nursing-home care. The regional agencies are responsible for securing access to care for all clients with indications, but bear no financial risk on their procurement activities.

Neither agencies nor providers have any strong financial incentives. The agencies act on behalf of the state, have all their care expenses reimbursed, and basically act as regional administrators. Moreover, regional budgets have traditionally been set based on previous budgets, and agencies would typically merely negotiate equal prices for all providers in a region. With a lack of any financial incentives on the buyer side, the government has to regulate provider prices as well.[5] As a result, in addition to setting regional budgets, the regulator also sets the maximum tariffs for providers, thus imposing restrictions on the prices negotiated. The actual prices paid by regional offices to providers are not allowed to exceed the maximum tariffs.

On the provider side, there is little incentive to use funds efficiently either, because non-profit providers do not seek to maximize profits, but rather focus on achieving a 'balanced budget' (i.e. spending their entire budget in some reasonable way), effort minimization or the maximization of (private) service utility. Historically, due to the constraints of budget regulation and the absence of quality standards, buyers had an incentive to put pressure on prices, regardless of quality. Because regulated maximum tariffs were set uniformly for all providers, bargaining between nursing homes and buyers usually resulted in almost uniform prices, close to the regulated maximum tariffs for providers. In view of these uniform prices and an incentive to balance their budgets, providers adjusted their quality according to the prices paid. As a result, differences in cost-enhancing circumstances such as regional differences in labor market conditions, and efficiency between providers resulted in uneven opportunities to deliver good quality care. In the absence of explicit regulatory requirements on quality, that led to excessively low prices for some providers and, eventually, to a 'race to the bottom' on quality. Constant financial pressure also undermined the trust of providers in the regulatory system.

In 2017, a paradigm shift was introduced with the publication of the new Quality Standard for nursing-home care [14]. This new standard set out legally binding quality guidelines for nursing-home care, thus stressing the importance of quality in the regulation. However, it also pinpointed a crucial problem that needed to be solved in the design of regulation. Whereas low quality could always be explained by insufficient resources, it was less evident how much reimbursement would be required to improve quality to the new standard. Facing

[3] By contrast, curative care (hospital care, GPs, curative mental healthcare etc.) is financed by mandatory insurance provided by competing private health insurers. See Croes et al. [13] for a description of the role of competition in the curative healthcare sector.

[4] Providers do not receive this money: they have no incentive to treat wealthy patients differently from patients with more modest means.

[5] Given the fact that prices of providers are regulated and demand is given, we do not see a reason to regulate the regional agencies by setting regional budgets as well.

asymmetric information on the cost of quality and a more rigorous definition of service quality, the legislator could not use the previous average cost-based regulation as a credible basis for the reimbursement model under the new quality standard. Given the social objective to ensure the cost-efficient use of public funds while improving service quality, the regulator needed to propose a new regulatory model for the sector.

The regulator faced multiple challenges in pursuing this objective. Improvements in quality was an urgent political priority, and there was political pressure to speed up progress on quality, even though the wider discussion on the specification of the Quality Standard was still on-going, and its outcome uncertain. The limited information on cost – because almost no provider could show compliance with the new standard, and thus provide a stable cost target – highlighted the need for learning and emulation in the sector. In addition, there was little trust in the system. The development of a new tariff regulation model for the nursing-home sector was therefore intertwined with the need to account for all these related issues. In other words, the new regulation also needed to align with the (unfinished) quality concretization, facilitate learning and improvements, and promote trust.

3. Methodology

Theoretically, the regulatory challenge is to establish a cost-norm for a production function with a binding quality constraint, exogenous demand with potentially heterogeneous technology (service cost). For a panel of rational profit-maximizing operators facing a homogeneous technology, the optimal cost-norm could be identified with yardstick regulation. As explained previously, the regulator could not abstract from the multi-dimensionality of the service, the potential heterogeneity in operator objectives and the absence of full information in the pursuit of the activity planning. Thus, the naïve yardstick approach had to be substituted for some type of multidimensional benchmarking, capable of dynamically handling learning, innovation and motivation.

The literature offers various methods of benchmarking, also known as *frontier methods*. These methods reveal the cost or production frontier by using data on inputs, outputs and potential additional (environmental) factors, such as quality or firm-specific exogenous circumstances. However, as explained above, in practice there are many challenges that arise when implementing such a method. The most frequent practical challenge concerns defining and measuring the relevant factors, which affects the feasible *scope* of the analysis. On the output side, the regulated activity may be not the sole activity of the firms concerned, and separate data on all activities may not be available. A typical related problem on the input side concerns the standardization of capital costs, often complicated in many sectors by an absence of data, competitive investments and observable use and state of assets. A second type challenge is accounting for *quality*, since this requires a consensus on the definition and measurement of quality, and also alignment with the existing quality regulations (if these exist). A third challenge is an adequate and collusion-proof *benchmarking model specification* that adequately captures the relationship between all the relevant factors. The outcomes of the benchmarking need to be translated into firm-specific tariffs and this requires additional choices, depending on the regulatory goals. The whole regulatory model needs to be understood and accepted by the firms being regulated, which can only be achieved by ensuring that the process is transparent, that there is openness around trade-offs and these are balanced to reach a consensus on the benchmarking model. Finally, there is the challenge of *designing the regulatory process* in a way that ensures that all parties cooperate and commit to this process. In the remainder of this section, we will discuss these issues and trade-offs in more detail, explaining the arguments behind the decisions made by the regulator during the 'baseline measurement'.

3.1. Frontier method

The regulator initiated a process of dialogue with the sector in order to develop a new method of setting tariffs. The goal was to achieve adequate tariff regulation and ensure the provision of good quality by each provider, which would also enable learning and improvements and help restore trust in the system. Frontier methods can be a useful tool for these kinds of goals, since they estimate the minimum production costs for each unit in the analysis by comparing its costs and production with other units. Since these methods can adjust for individual differences between providers, they enable individualized tariff-setting, correcting for heterogeneous operation, given quality requirements. The available frontier methods differ in terms of two dimensions and each has specific advantages and limitations. The two aspects are (i) the ability to deal with data errors and (ii) the restrictiveness of the *a priori* assumptions on the functional form of the frontier. Stochastic methods, unlike deterministic methods, use estimates and can therefore correct for possible data errors. Parametric methods use strong assumptions about the functional form of the production function, while non-parametric methods determine the frontier by using only limited assumptions on the functional form. In particular, deterministic, non-parametric methods use all data to estimate the minimum spanning frontier under very mild assumptions on the form.

The Dutch regulator opted for a non-parametric, deterministic frontier method, Data Envelopment Analysis (DEA), initially proposed by Charnes et al. [15] later inspiring a subsequent broad academic and regulatory DEA-literature (see the introduction section). DEA was deemed the most suitable method for the current problem setting because the efficient costs determined by this method could be expressed in terms of the costs of the most efficiently operating organizations in the sector, or 'peers'. This information facilitates learning from efficient units and builds trust in the regulatory model because organizations can observe and study the individual operators (the peers) constituting their cost and service targets. An imperative feature of DEA is therefore the implication that operators –whose employees often work with a financial instead of economic background– require only limited knowledge about the model and econometric principles in order to be able to critically assess the model outcomes. The regulatory use of the model and the open identification of peers secure high data quality and an effective cross validation of data, a prerequisite for the use of DEA for due process and learning. The regulatory process was designed around the DEA to strengthen the benefits of the DEA method. So while most of the DEA literature stresses the problem of data errors in specific analyses, we believe that DEA will lead to more reliable data in a dynamic regulatory process. As a deterministic method, DEA is of course sensitive to outliers and the dataset used. Thus, attention must be paid to the detection methods for outliers and the so-called 'curse of dimensionality' issue. The latter has to do with the use of a finite dataset to reveal the frontier, which is why it is essential that the model is specified in the way that strengthen the convergence of the computed DEA-frontier to the 'perfect' frontier [16]. Newer non-parametric techniques offer some solutions to these known issues, allowing to refine the baseline insight (e.g., m-order estimator or order-α quantile estimator; see Cazals et al. [17] and Aragon et al. [18]. Given that this paper describes the first regulatory cycle, focusing on major regulatory choices, rather than on subsequent technique refinements, we will also adhere to the same focus in presenting the initial outcomes obtained during the baseline cycle, relegating all the potential DEA-refinements to further research.

3.2. Scope

The new tariff regulation was intended to facilitate the new Quality Standard for inpatient nursing care. The product scope of both the new Quality Standard and the new tariff regulation was therefore limited to inpatient nursing care. However, most nursing homes provide

both nursing home care and non-regulated services such as geriatric revalidation care but also domestic assistance beyond care, which falls outside the scope of the regulator. This raises the question of the scope of the benchmarking model.

On the one hand, there are several arguments in favor of a broader scope "— notably, the fact that providers' operational decisions on costs and efforts are taken at a higher level where some scope economies may be present, while inefficiency may vary per activity. A related cost-accounting argument in favor of a broader scope is that costs related to in-scope production and (some) out-of-scope production are entangled. This cost entanglement complicates unambiguous cost allocations at the provider level, and therefore potentially limits the reliability of the data. These arguments are particularly relevant to jointly produced care, for which the same staff members deliver nursing care services both within and beyond the scope of the regulatory problem being studied here. However, these are less relevant for the other care (i.e. care that is delivered separately, such as at a separate nursing care facility with separate cost accounting).

On the other hand, if out-of-scope production is also included in benchmarking, then the problem of cost-separation would still arise later, in the context of separating the efficient costs of the regulated activity (but that problem would be simpler to solve, because the frontier efficient cost is determined by the 'peers' spanning the frontier, so there is no need to make restrictive assumptions on the allocation of inefficiency). Taking these trade-offs into account, in the baseline measurement the regulator decided to use a benchmarking model with a broader production scope including the joint production, but which excluded other production.

Concerning the cost side of the benchmarking model, another consideration arises with respect to the handling of capital costs. From a theoretical perspective, a total expenditure model, where both operational costs and capital costs are included, would usually be preferable. In practice, however, the inclusion of capital costs may be difficult, especially in healthcare where a large portion of capital assets are health service facilities themselves. Ozcan [19, p144] notes that: "statewide databases or hospitals in their accounting books may report this variable as 'assets'; however, the value of assets depends on their recorded or acquisition time and their depreciation. Thus, using the book values of such investments does not reflect what is on the ground as a health service facility". This implies that including asset information as an input requires highly standardized information on capital costs. However, this information is hardly ever readily available and collecting it would typically be a long-term process. This problem also played a role in the case study considered here, making the inclusion of capital costs in benchmarking unfeasible. Importantly for the current case, the capital-cost-standardization problem had been recognized previously; therefore, the incumbent regulation compensated capital costs on a normative basis. Effectively, then, tariff regulation covered the operation costs only. For these reasons, the regulator decided to exclude capital costs in the first baseline measurement.

3.3. Quality

Generally, there are different options for considering quality in regulation. In practice, quality regulation typically comes either as a separate add-on regulation (added to price regulation), in the form of bonuses for good quality or penalties for bad quality, or it comes in the form of minimum standards. The latter usually means that there is an incentive to ensure that quality just reaches the standard level, because the company is not paid extra for exceeding it. However, this choice was not an issue in the current case, because the presumption in the Netherlands was that every provider should meet the minimum standard of quality. This presumption means that every nursing home client is entitled to good care, and public budgets should be allocated efficiently to safeguard that general right. As a consequence, quality indicators were not included among DEA variables, which was also a specific requirement of the Ministry of Health in its request to design the new tariff regulation. Instead, the relevant issue was to ensure that the resulting efficient costs were not too low to meet the minimum quality requirements. This could potentially be achieved by restricting the reference set for the DEA frontier to providers that were performing in accordance with the Quality Standard. The practical problem, however, was that the standard was new and hardly any provider satisfied it at that time, specifically with respect to a norm on the number of staff per client group. This implied that most providers' observed costs were too low (in comparison to efficient costs of performing in accordance with the standard). To resolve this issue, the regulator estimated the expected additional labor costs for each provider that would be needed to enable an increase in the staff number to meet the norm of the Quality Standard. By adding these additional 'prospective' costs to the observed costs, it was ensured that the initial cost level in the analysis was high enough to enable providers to meet the quality standard (more details on the derivation of prospective costs are provided in Appendix). It should be noted, however, that the estimated prospective costs are uncertain and probably exceed the costs that would actually be needed, because there is still no concrete definition of quality, and the only specific norm in the Quality Standard refers to 'the availability of two nurses per client group during the intensive care moments', thus still allowing for differing interpretations and operationalization. Since this particular norm was used to estimate the macro-budget, the individual prospective costs were computed using the same assumptions.

This normative approach for quality assurance in the face of asymmetric information can also be found in other sectors, such as in electricity distribution where the promotion of decentralized generation with unknown costs was solved by cost-plus arrangements in Sweden, in parallel with an incentive-based regulation. Usually, the definition of quality indicators and quality standards falls when there are multiple institutions with links to the client or provider-side, which are not charged with economic regulation, making the economic regulator's task very delicate in isolation. We believe that quality in the setting of nursing homes is observable but non-verifiable. That could suggest that in future the approach may involve an independent committee to judge which nursing homes are eligible to serve as organizations that could form the frontier.

3.4. Model specification

Here we explain the initial steps of the model specification, and how those were filled in the baseline measurement. Recall that the baseline measurement was only a starting point of the discussion on establishing the new regulation process, when the basic ideas and principles needed to be discussed, while only limited imperfect data were available. Thus, the emphasis of the discussion lied on the model 'carcass', capturing the major input–output relationship, while all the fine tuning (e.g. refinement for environmental effects and addressing the outlier- and finite-dataset issues mentioned in the introduction) were left for the next development cycle, which would be based on a larger and cleaner dataset. More detail on the process set up will be described in Section 3.5.

The first step in specifying a DEA model is to make assumptions regarding orientation, scale and model type. It is important that these model choices are aligned with the intended model use in regulation. As we have described, in the current case the regulatory problem is a cost-minimization problem. Because the indications for nursing care in the Netherlands are defined by an independent assessment agency (not by the provider) and, at least in the short run, the provider has much more control over the inputs than outputs, the benchmarking model is therefore naturally oriented to inputs. Based on the idea that benchmarking should encourage providers to produce at an efficient scale, the regulator has assumed constant returns to scale (making the DEA-outcomes equal to those under an output orientation). It should be noted that very small providers (which may not be able to adjust

their scale, for example, because they operate in remote rural areas) were not part of this analysis. With respect to model formulation, the choice during the first cycle was between a primal (production) and a dual (price) model. A model that encouraged learning and facilitated trust was required, and although a dual model would have allowed the regulator to set restrictions on prices, it would also have limited the possibility for sharing and explaining peer information to firms, which was one of the main reasons for using DEA in the benchmarking problem in the first place. The regulator therefore opted for a primal DEA formulation, with no restrictions on prices, and to use dual analysis only as a sensitivity check on the results.

The next step concerns the specification of inputs, outputs, and potentially other relevant factors. Ideally, environmental and complexity factors should be identified and included in a full model specification, either directly or through a second-stage analysis. However, estimations of such environmental factors require a model that is already robust based on valid data. During the baseline measurement, the regulator therefore decided to limit the discussion to input and output specification only, and to focus on understanding these results before adding more factors to the analysis.

With respect to the DEA-input definition: as explained in previous two subsections, the capital cost was not included. Therefore, the input-side of the DEA only covered the total operational costs, accounting also for prospective labor costs. In other words: the sum of the incurred operational cost and the prospective labor cost was used as a single DEA-input. The prospective labor cost was included in order to cope with the new staff-availability requirement (see Section 3.3 for more detail) Thus, the remaining decisions regarding the model specification only concerned the output side.

DEA can handle multiple outputs. Ideally, each output will be produced by a multiple number of firms to enable comparison between firms. In an extreme case in which each firm provides some specific product which is not produced by other firms, all the firms would be found to be efficient because it would be impossible to find peers that delivered exactly the same product. (Some discussion of the relevant incentive issues can be found in [20,21]. Including products which can be considered close substitutes as separate outputs should also be avoided for the same reason. In essence, the outputs (the underlying product groups) in DEA are usually defined in such a way that they are made up of relatively homogeneous products, and there is heterogeneity between outputs. In practice, the DEA output specification can be conceived by grouping products that are relatively similar and for which relative prices can reasonably be determined by means of field expertise, for instance. Each output is then constructed as a weighted sum of the underlying product quantities, the weights being their relative prices. The model endogenously estimates the remaining unknown relative prices between outputs.

Table 1 provides an overview of the DEA outputs that were used in the baseline model, including their brief description. Since Dutch nursing homes provide both nursing home care (NH care, which is the main focus of the regulation) and also jointly-produced care, both types of outputs were included in the model: four outputs represented NH care and two outputs represented jointly produced care. This choice and the respective definitions were based on the following reasoning:

NH-care consists in providing inpatient care days for clients with nursing home indications. The majority of clients have medium-intensity care indications, while a much smaller group of clients have higher or lower severity levels. Higher care levels are associated with more severe patient conditions, and therefore more hours of care are delivered by more experienced and higher qualified staff. As a result, not all nursing homes serve these clients. Similarly, lower levels of care are associated with less intensive nursing requirements, fewer hours of care and lower qualification requirements for personnel. So nursing homes which provide more low-intensity care differ from other nursing homes. This implies that the difference between patients requiring different levels of care is relatively large. In addition, given the intensity level,

Table 1
Definition of DEA-outputs.

DEA-outputs	Description of the underlying production	Scope	Role in regulation
1	Less intensive residential care days, both including and excluding medical treatment	NH care	Covered by the new regulation (i.e., both the Quality Standard and the new tariff regulation model)
2	Medium intensity residential care days including medical treatment	NH care	
3	Medium intensity residential care days excluding medical treatment	NH care	
4	High intensity residential care days both including and excluding medical treatment	NH care	
5	Residential care days	Jointly produced care	Not covered by the new regulation
6	Non-residential care days	Jointly produced care	

some clients have indications for medical treatment, while others do not. This distinction is especially relevant at the medium severity level, because nearly all clients of higher levels also receive medical treatment or some additional care.[6] Based on this reasoning, four nursing care outputs were defined: one for the low severity level, one for the high severity level, and two for the medium severity level where inpatient care days with and without treatment were distinguished. Because for the NH-care –that lied in the main focus of the new regulation– the cost-based standardized regulated tariffs of inpatient care days were available. Therefore, those tariffs were used as aggregation weights to aggregate these inpatient care days into the respective output. Thus, the four DEA-outputs of nursing care were defined in terms of standardized revenues.

Jointly-produced care has more variation in type of care. However, the revenue from joint production is only 16% of total revenue on average, and thus represents a relatively small proportion of total production. The main cost-driving distinction within care produced jointly seems to be between products that represent capital-intensive care, such as nursing days in an institution (care for patients with mental disabilities, for instance), and less capital-intensive care such as nursing care provided at home, or some additional care for patients with special conditions. On these products only revenue information was available. Based on this, joint care was represented by two DEA-outputs, defined in terms of revenues: revenue from care days and revenue from other care.

Summarizing the considerations discussed above, the baseline model was specified as an input-oriented, CRS, primal DEA model with one input and six outputs. The input side of the model was represented by a single output, equal to the sum of the operational and the prospective costs of this production. The output side was represented by six outputs (Table 1), defined in terms of (standardized) revenues.

This means that the input efficiency E for nursing home o is computed for the above described single input x_o and six outputs $y_{r0}(r = 1 \ldots 6)$ using the following programming formula:

$$\min E$$

[6] In contrast, when we compare nursing care at the same intensity level for clients with dementia versus that for clients with physical problems, although the patients' conditions are different, both the nursing care and the labor required to provide this care are similar. This is reflected by the current cost-based tariffs which are equal for both groups.

$$E, \lambda^1 \dots \lambda^K$$
$$st. Ex^o \geq \sum_{k=1}^{K} \lambda^k x^k$$
$$y^o \leq \sum_{k=1}^{K} \lambda^k y^k$$
$$\lambda \in \Lambda^K$$

Where $\Lambda^K = \left\{ \lambda \in \mathbb{R}_+^K \mid \sum_{k=1}^{K} \lambda^k free \right\} = \mathbb{R}_+^K$

This baseline model set-up was meant for illustration purposes, as the model would still be further developed in next regulation cycles, for example to incorporate environmental factors or other refinements.

3.5. Process design

According to the regulator's proposal, the minimum costs estimates from the benchmarking will serve as the input for tariff setting in the future. In addition, each provider will receive a detailed, customized feedback report, which will be generated automatically and include information on the performance of the peer providers. These reports provide information on cost structure, personnel and other business related indicators, with the aim of facilitating learning and improvement among individual providers and generating insights that can help to improve tariff regulation. But benchmarking only works well if both the model and the data are adequate. The chosen regulation design therefore needs mechanisms to ensure: (i) regulatory commitment; (ii) a routinized collaboration for validating data and improving the model; (iii) data sharing. The first development steps were the following:

(i) The proposed regulatory strategy was communicated publicly from the start in a process plan. Specifically, the regulator proposed setting price bounds using (firm-specific) minimum and maximum tariffs. Here the minimum would be based on efficient costs, thus ensuring the feasibility of meeting the quality standard; the maximum would be based on the average efficiency across the sector. Capping the budget to the historical cost setting was designed as a zero-sum game, in which the reallocations create the incentives for efficiency.[7]

(ii) The regulatory process was designed as an annual cycle of data collection and validation, benchmarking analysis, a discussion of the results, and tariff setting. This would provide the possibility to further develop the baseline model in the next cycle, and to fine-tune it later, on a structural basis, working on it together with the sector. After each cycle, the feedback collected would serve as the starting point for the next cycle, thus allowing the sector to be involved at any stage by providing feedback; the idea was to enable a consensus on the model to emerge in order to restore trust in the system. The baseline measurement was simply the first cycle, designed for learning purposes and to establish a basic relationship between the cost and the output, and not for tariff setting. So the model did not adjust for any local differences that may have explained cost differences because this first cycle was meant to design the basic model and learn from the process. In future cycles, these adjustments may be incorporated through the feedback process, by means of second stage analysis for instance. The intention was that both the data and the model would be refined in subsequent cycles, and that the remaining technical issues would also be addressed — in particular, the splitting of efficient costs into nursing care costs and the cost of jointly produced care, which we include in technical appendix.

(iii) Data sharing would be necessary to facilitate the data cross validation and learning processes, which would contribute to further improvements in both efficiency and the tariff setting process.

4. Data

Here we describe how the dataset of the baseline measurement was constructed, highlighting several data issues and tradeoffs that are important from the regulatory prospective.

4.1. Data collection and validation

The baseline measurement was conducted in 2019–2020 based on data for the year 2018. As explained in the section on specification, data collection only involved nursing home care providers with revenue from nursing-home (NH) products exceeding €1 million. The major trade-offs in the data collection and validation stage concerned the fact that on the one hand, both the DEA-application and the regulation process would benefit from including more units in the dataset; and on the other hand, the quality of the results and thus their recognition by the participants of the process depend on data quality. At the start of the data collection process, all eligible nursing homes were invited to deliver data to the regulator. However, because of the fact that the baseline measurement was designed as a learning process, not for tariff-setting purposes; there was no strict enforcement of participation. Some 286 providers delivered data (i.e., 88% of all providers invited). During the validation period, in addition to prior instructions and support, nursing home received individual reports with the results of their data validation. The validation included several internal consistency checks on production and cost data. If any inconsistencies on these data were reported, providers were asked to resubmit their data. In this way, the data-cleaning process evolved an interaction between regulator and providers, whereby the providers remained responsible for and in control of their own data.

4.2. Data coverage

There is a tradeoff between the number of units included in the analysis and the quality requirements for the data, especially relevant for a non-parametric method such as DEA. While the regulatory process benefits from the inclusion of as many providers as possible, so that all participants of the efficiency measurement get engaged by receiving their results, the regulator needs to ensure the reliability of the results by only allowing high-quality data to be used as the basis for tariff setting. While the providers' production data was relatively easy to collect and validate with external data sources, the submitted cost data were sometimes unreliable. In the baseline measurement the regulator resolved this issue by using the concept of a 'reference set' which only included those providers with high-quality cost data. The reference set included 109 providers with reliable data, while the complete set included 286 providers in total, covering about 80% of the total nursing home production.

Only the reference set was used further in the DEA efficiency analysis to define the DEA frontier. The constructed DEA-frontier could afterwards be used to define efficient cost levels for any output combination, thus also for providers with inconsistent data on costs. Therefore, if any inconsistencies were encountered in the submitted cost data remained after final submission, the data were excluded from the reference set. As a result the reference set only contained operators eligible to become a peer-unit for others, based on the quality of their cost data. This way, all the providers were included in the analysis and could thus be engaged in the process, and at the same time it was possible to guarantee that the outcomes (scores) were based only on high-quality data.

4.3. Descriptive statistics on the reference set

Table 2 shows descriptive statistics on DEA-inputs and outputs of the firms in the reference set. The input variable represents the total

[7] The capping of budgets to historic costs or by pegging to the average firm is also found in other regulatory applications, such as in electricity distribution in Norway.

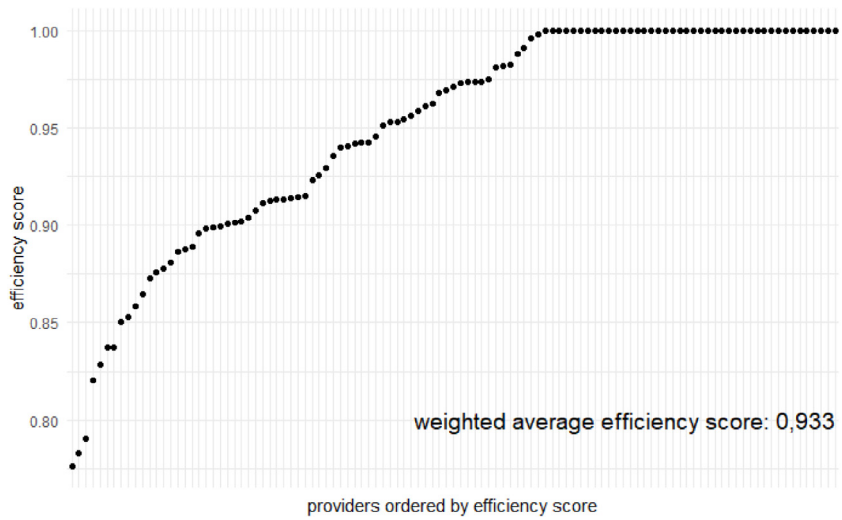

Fig. 1. Distribution of efficiency scores for providers in the reference set.

Table 2
Descriptive statistics for inputs and outputs in the reference set.

	Mean (st.dev)	Min	Max	Total
Input	53.66 (51.1)	4.03	237.17	5849.48
Output1	6.19 (5.95)	0.05	23.58	674.33
Output2	24.53 (23.33)	0	87.28	2673.61
Output3	4.98 (6.08)	0	29.82	542.92
Output4	5.83 (6.89)	0	42.44	635.6
Output5	6.26 (7.76)	0.01	37.27	682.3
Output6	2.78 (5.93)	0	44.04	303.33

Note: 109 observations, all numbers are expressed in M €

operational cost, computed as a sum of observed and prospective costs. Prospective costs account on average for 20% of the total. With respect to outputs, it can observed that not all companies produce all outputs. Output2 (medium severity care days with treatment) is typically the largest, while Output6 (one of the jointly-produced outputs) is the smallest.

4.4. Refining the reference set for DEA

As explained in Section 3, the baseline measurement was the first regulatory cycle, mainly intended to communicate the base principles, with the intention to refine it in next cycles. During this cycle, only basic outlier-detection criteria were applied to exclude from the initial referenceset those providers which would not be a reasonable benchmark for others. In particular, the regulator identified which units in the reference set could be considered outliers, based on their 'superefficiency'. [See e.g. 22, for more detail on this]. Superefficiency is an intuitively clear concept, determined by running a unit-specific DEA model, and removing this unit from the DEA reference set. A unit which is extremely far away from the frontier that would result from an analysis without that particular unit, could possibly set the minimum cost standard for other units at an unrealistically low level (possibly due to currently unknown specific circumstances of the outlier unit). The refined reference set was therefore determined after excluding six outliers, in order to prevent those from being peers for other units. These six 'superefficient' units were considered as fully efficient by definition, and assigned a score of 1 in the analysis. (see appendix for technical details on the exclusion of outliers).

5. Results

In this section, we report on the quantitative results from the baseline measurement, followed by a summary of the feedback received from providers. As said, these first results were not meant for the use in regulation, but were made to give the sector the idea about how the model works and what sort of results and feedback it generates in order to enable potential refinements on both the model and the data by bringing the discussion to the next level.

5.1. Distribution of DEA efficiency scores

Fig. 1 presents the distribution of efficiency scores for providers in the reference set. A total of 40 providers (36%) were fully efficient with a score of 1, 34 of which were peers for others and 6 were superefficient. The lowest score was 0.786 and the mean was 0.933. The unweighted mean efficiency in the reference set was above the weighted efficiency (0.949 > 0.933), pointing to a marginally lower efficiency of larger providers.

Further inspection of the providers with efficiency score 1 revealed that some had relatively rare output combinations, pointing to the possibility that these might look efficient because of the lack of comparators. This suggests that the baseline measurement might overstate actual efficiency.

5.2. Regional coverage and efficiency allocation

Table 3 shows the geographic coverage of the dataset, based on a division into three regions. Each of these regions is well represented in both the reference set and the residual set, and there is no statistically significant difference between the mean scores for the three geographic regions.

5.3. Relationship between prospective costs and efficiency scores

Fig. 2 shows a scatterplot of DEA-scores against prospective costs per care day, leading to two observations. First, the negative correlation that we see in this figure stresses the trade-off between two social objectives: improving service quality by increasing the number of nursing home personnel and cost-efficiency; and thus a need to account for this trade-off in tariff regulation. Secondly, a large variation in prospective cost estimates between providers and the presence of some extremely high values suggested that the definition might needs reevaluation. In particular, this second observation raised the issue of accounting for contribution of (some types of) stagers to the care process (see Section 5.6).

Table 3
Regional coverage and efficiency allocation in the reference set.

Region	Share of providers in the dataset across regions[a]			Efficiency score[b]
	Complete set	Reference set	Residual set	Reference set
Central	37%	15%	23%	0.945 (0.010)
North	33%	12%	21%	0.954 (0.008)
South	29%	12%	17%	0.950 (0.010)
Unknown	1%	0%	1%	

Notes:
[a]Each provider is allocated to a single region, where the most clients receive care.
[b]Unweighted mean with standard deviation in parentheses.

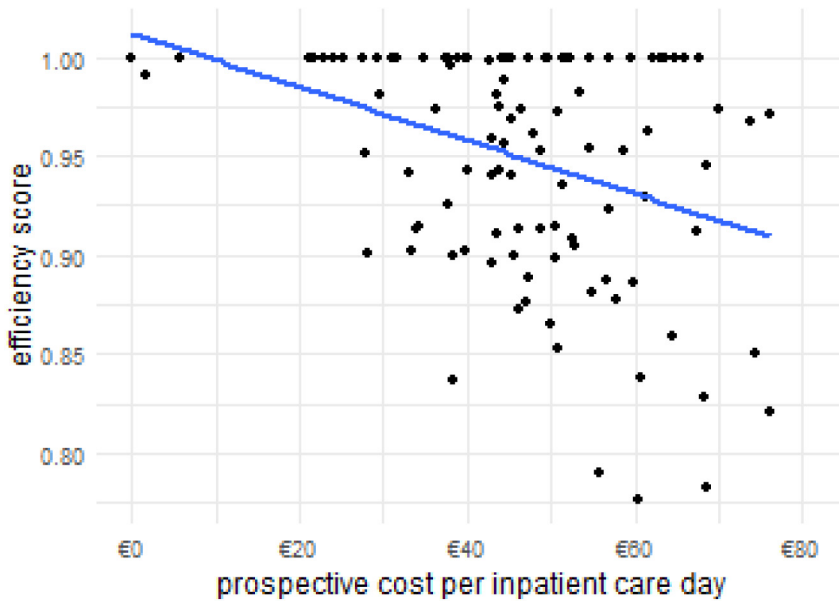

Fig. 2. Relationship between the prospective costs and DEA-scores from the base model.

5.4. Peer identification

As already noted above, some companies with rare output compositions appeared own peers, possibly because of the lack of comparators. This points to the need of increasing the dataset in future cycles, as this increases the chance of more fair comparison. Besides, this suggests the need of model development to tackle this (in line with methodological advances mentioned at the end of Section 3.1).

In addition, it may be impossible to find "perfect" peers for all organizations. In some cases, the efficient cost also includes some 'slack': the excess output produced by peers relative to their non-peer organizations. This is an effect of the radial efficiency measure, exceeding or equal to the output of the compared unit. In the current analysis, cases with slack (e.g. >5% of output value) were relatively frequent. This problem could be reduced as the quality of data increases in future analysis and thus units with a more diverse output mix can be included in the reference set. However, some level of slack is likely to remain and from the regulator's perspective, the question of how to deal with slack levels in tariff setting remains. Although the translation of the model outcomes into tariff setting was beyond the scope of the baseline measurement, a theoretical exploration of how to deal with slack levels is included in the appendix.

5.5. Feedback reports

All providers were sent individual reports with a comparison to their peers (in terms of several indicators) and the resulting feasible minimum costs from the DEA model. The providers in the reference set were also sent the efficiency scores. After the release of the results, the regulator organized a number of sessions in groups to discuss the results and to ask for feedback on the process. On the positive side, the reports were taken seriously by providers, and some participants expressed their appreciation. However, the meetings also showed that despite all the effort on communication, the approach is often misunderstood or mistrusted. This clearly suggests that the process needs time to bear fruit, because on the one hand expertise from the sector is needed to improve the benchmarking model and make it fit for purpose, transparent and robust in application; and on the other hand, the sector needs to understand the method and to trust the process before it can provide such support.

5.6. Follow up on the results of the baseline DEA

The results of the baseline measurement had direct implications for the subsequent model development. First, the limited amount of comparators in the reference set highlighted the necessity of continuous focus on data collection in future model cycles. With more providers submitting reliable cost data, the reference set could be expanded and form a better representation of the variety of output combinations among providers. This would make the estimated DEA frontier lie closer to the actual (unobserved) production frontier, thus reducing the slack problem. In addition, a broader reference set would partly tackle the 'slack' problem by enabling the model to find proper peers for more providers, thus contributing to the recognition and the support to this model by the sector. Second, it was recognized that the further methodological development of the efficiency model must go hands-in-hands with the improvement of data definitions for the consistent integration of this model into regulation framework. As a first suggested action, the provider feedback led the regulator to revise two basic assumptions behind the construction of model inputs and outputs in the

next step of the model development, the integration of the DEA model into tariff regulation. Specifically, the product aggregation weights used for constructing DEA-outputs were revised[8] and the computation of prospective costs was adjusted.[9] These first follow-up revisions were motivated by the need to streamline the definitions of the variables in the DEA model to allow for the subsequent translation of the model results into tariff regulation (as explained in the footnotes). In addition, it was recognized that later follow-up steps might need to involve more advanced efficiency modeling methods (as discussed in Section 3.1) as well as addressing more general issues (discussed in Section 6).

6. Discussion

In theory, regulatory economics provides us with a powerful framework for creating proper incentives when there are information asymmetries regarding cost and quality. However, as always, the devil is in the details. Putting the theory into practice requires a careful analysis of elements, their integration into a coherent framework, and acceptance from the relevant sector regarding their development and configuration. In this paper we have described and analyzed the essential elements of such a regulatory process by focusing on the current experience in the Dutch nursing home sector, but there are several insights that we believe can be useful across sectors and countries.

First of all, with respect to benchmarking methodology, although in theory frontier methods are optimal for tariff setting, the case shows that choosing a particular method should take the regulatory context into account. The current study has highlighted the particularly useful features of DEA as a regulatory tool in such contexts. In particular, the possibility of comparison with other units is useful, as this makes it possible to increase learning by sharing best practices (as also pointed out in earlier literature on both healthcare and other sectors: e.g., A. et al. [23] argues on the usefulness of DEA within a regulatory system for energy distribution and Dai and Kuosmanen [24] illustrate how the combined use of DEA and clustering methods may optimize the selection of best practice targets, given similarities in operating environments of units). Other practical considerations such as the level of trust vis-à-vis regulation in the sector and the proposed involvement of firms in tariff setting, are also likely to shift the emphasis to non-technical aspects, particularly the possibility of explaining the method to a non-technical audience, data availability and data quality. Similarly, the specification of the model requires sector expertise from the employees working at the regulated companies. Their knowledge can only be used when participating firms understand the topics adequately, have confidence in the process and contribute to efficiency analysis by improving the quality of the data through an iterative process. In the current case, the involvement of 286 providers limited the options for one-on-one physical guidance and communication about the model. Therefore, the importance of building in customized automated feedback on both the data and the model results should not be underestimated, in order to facilitate the communication process throughout the entire regulatory cycle.

Second, cost-accounting issues affect the feasible scope of the regulatory benchmarking model, particularly the position of capital and non-regulated production which is produced jointly with regulated production. If capital is included, it must be standardized, because the standardization of capital costs ensures adequate accounting for the capital stock of the firm in financial regulation, which secures financing possibilities for reinvestment. However, capital cost-standardization is an important hurdle. In the regulatory DEA applications, capital cost proxies are often used instead of capital costs (such as nursing home beds[see e.g. 19, for healthcare applications]. An exception is found in utility regulation, where capital costs could in principle be constructed on the basis of physical assets (some discussion can be found in [23]. While some data on capital asset proxies such as beds are available to the regulator and could therefore be included in the model, the bulk of capital costs typically relate to the facility itself. Including proxies in the model would not therefore solve the issue of capital cost remuneration, which would still need to be addressed through financial regulation. This may be a decisive consideration in favor of using an operational cost benchmarking model instead of a total expenditure model.

Third, quality clearly plays an important role in public service provision and there are multiple studies on the relationship between quality and efficiency in the nursing home market Blank and Eggink [25], Weech-Maldonado et al. [26], Laine et al. [27], Schnelle et al. [28], Dulal [29], Ni Luasa et al. [30]. However, this literature is unrelated to the regulatory context of financial remuneration, while both the design and use of quality measures in benchmarking for tariff-setting purposes pose additional technical challenges. Technically there are various ways of handling quality measures (see Varabyova and Schreyögg [31] for a theoretical overview of techniques for incorporating quality in DEA models), but their use for regulatory purposes requires consensus among – and the involvement of – stakeholders. Information asymmetry and the credibility of the regulation are often at stake when defining such technical indicators, which makes the regulator's task a very delicate one. Service quality in most sectors (healthcare, transport, utilities, education) is handled by multiple institutions with links to the client or provider-side, which are not charged with economic regulation. This requires the regulator not only to focus on technical aspects, but also to keep a close eye on these external or political developments and manage any changes or uncertainties that may affect the regulatory benchmarking model and its outcomes. While the theory provides ways of handling the (prospective) costs of operating under higher quality standards on personnel availability, it should also be recognized that such a standard remains a very incomplete quality measure and so the process should also prepare the ground for better measures to be developed.

All in all, the lessons from this case study stress the need for a diligent and evolving regulatory process in which all aspects and actors are addressed and allowed to contribute by means of collaborative interdisciplinary efforts by various experts from both the regulator and the industry. Since the development of the model may take several years, during which circumstances may change, the regulator should be flexible and maintain clear and unambiguous communication about the model and its results. Only when the data issues have been solved and the benchmarking model has been adequately developed can the regulator base tariff regulation on the results of the model. The benchmark-based individualized tariff-setting requires additional decisions, such as how to deal with missing or unreliable data on some regulated organizations. Reflecting beyond these technical tariff-setting aspects, the regulator will often be even more concerned about other general issues in the regulated sector. For example, regulation theory typically assumes firms that will seek to maximize their profits, but many regulated sectors are dominated by not-for-profit firms, which have other objectives. It is therefore no trivial matter to anticipate their actual response to incentive regulation when there are non-standard firm preferences. A relevant question relating to the design of the broader regulatory process is how to deal with a situation in which providers do not meet efficiency targets. The regulator should have a strategy for dealing with situations in which quality deteriorates, or bankruptcy threatens.

[8] Since capital costs were not included on the cost side DEA, it was decided that also the product aggregation weights used to construct DEA-outputs should be adjusted to exclude capital component.

[9] Initially, providers took a view that this computation must disregard stagers. However, providers pointed out that it would be fair to account for the presence of stagers at least to some extent. Indeed, those companies where there were a lot of stagers were having both additional costs on providing those stage places and supervision, but they also could to some extent rely on extra available hands that could be added should the situation requires this.

7. Conclusions

Regulation is ultimately an information game, in which a proxy for the political principal tries to solicit production and cost information from a set of agents in order to limit their information rents. Considering the three objectives — motivation (incentives), coordination (allocation) and information (learning) [32]: if the agents are providing a homogeneous and observable service under equal conditions in the pursuit of a purely economic goal, a simple incentive structure may work wonders. In a setting in which the agents are entrusted only with easily controllable tasks, and the principal is able to replace an agent at any time, a coordination-oriented, command-and-control method will be sufficient. However, in situations where assumptions on the agent's objectives, the observability of the quality of the service by the regulator and the homogeneity of conditions are not met, the regulatory approach must strike a fair balance between the benefits and costs of different regulatory choices. An information-oriented regulatory approach, such as frontier-based analysis, may be the only feasible option. On the one hand, this provides an endogenous cost norm based on real observations with minimal dependency on a priori assumptions. This avoids the conventional conflictual bargaining position that is inherent in price regulation. On the other hand, the firms may receive informative input on feasible targets and where to look for practical tips on how to achieve them. In short, both sides gain from the project, where each party plays a well-defined role as long as the information game is understood and endorsed by all. The application in this paper documents one attempt to move in this direction. These open questions may inspire analysts and practitioners to think creatively on how to improve the game.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Per Agrell reports a relationship with SUMICSID that includes: consulting or advisory. Peter Bogetoft reports a relationship with SUMICSID that includes: consulting or advisory. All the authors were involved in the development of the regulation, in particular, in the 'baseline measurement' project described in this paper. Misja Mikkers, Tessa Voesenek and Victoria Shestalova are employees at the Dutch Healthcare Authority (NZa); and Per Agrell and Peter Bogetoft were consulting the NZa during the implementation of this project. The views and opinions expressed in this paper are those of the authors and do not necessarily reflect the official policy or position of the NZa. The regulatory choices described in the paper are not binding for future policy in the Dutch nursing care sector.

Per Agrell serves as an associate editor for DAJ.

Technical Appendix

Derivation of prospective costs

Prospective costs were defined by taking the difference between the required and the available nursing staff, and multiplying this by the average wage level of a standard nursing staff mix (in which qualification levels 2,3,4 were used in proportions 2:2:1). This methodology followed the same principles that were adopted to define the total nation-wide additional budget for implementing the Quality Standard. The most important assumptions underlying the total budget computation were the assumption of the group size of 8 clients, and an assumed roster with 2 nurses (of qualification levels 2,3,4 in standard proportions) per 8 clients available during the daytime. Together with the data on the number of client care days provided, this determined the required number of nursing staff that would be needed to provide this care under Quality Standard. The available nursing staff was computed based on company FTE data, corrected for generic non-productive time due to vacations and non-nursing activities (such as education on the job) and company-specific sick leave. The estimate of average nursing wages for new personnel that would be needed to satisfy the Quality Standard was based on company-specific wages for the standard qualification level mix of the levels 2,3,4. For providers for which the available nursing staff exceeded the required staff, the prospective cost was set to zero. In the case of missing data, the country-average nursing requirement was assumed per care day produced, to obtain a prospective cost estimate. After applying these assumptions to the dataset analyzed, the prospective cost per client care day at a company level ranged between €0 and €76, with an average of €46. As explained, the total earmarked nation-wide budget accounted for nearly 20% of total sector budget. Note, however, the temporal nature of this computation: as soon as sufficiently many companies satisfy the requirements, it should be possible to base the reference set on those companies.

Outliers and super-efficiency

Some units turned out to be outliers as defined by the concept of super-efficiency. Outliers could have low costs due to certain specific operating conditions, but also due to an (undetected, and thus not corrected) application of atypical cost-allocation rules. The exclusion of these units from the reference set, prevents unrealistic outcomes for other units in the analysis. Super-efficiency for each unit from the reference set was determined by running a DEA model for the reference set excluding this particular unit. The threshold value for super-efficiency was determined based on the 'German' criterion, which does not allow super-efficiency scores in the DEA reference set to exceed a critical value of $q(0.75) + 1.5(q(0.75)-q(0.25))$, where q represents the quartiles of the distribution of super-efficiency. [22].)

Slack

In some cases it is impossible to find "perfect" peers for all organizations. In those cases, the efficient cost also covers some slack. Slack refers to the excess output that peers produce relative to their non-peer organizations. From the regulatory perspective, it is undesirable for the regulated tariffs to include slack. This issue could be tackled by subtracting the estimated slack-related costs from the DEA efficient costs. Note also, that a dual DEA would in theory be a more consistent choice with respect to the treatment of prices; however, as already discussed main text, the dual model may not be feasible given the requirements with respect to communication and learning and improvements in the industry. The issue of accounting for slack in regulation lies beyond the baseline measurement; but is an important subject for future work.

Fig. A.1 illustrates the amount of slack in the current DEA run for the reference set. We use the respective sums of the DEA products to highlight where this slack is attributed, and the figure shows that it could be attributed both to NH and to jointly produced care. It also can be seen that nearly a half of the dataset has some (sometimes very limited amount of) slack, and in roughly 10% of the cases slack is above 5%.

Splitting the joint-production cost

The DEA model defines the total efficient cost of nursing home (NH) care and jointly-produced care, while the regulatory focus is only on NH care. Thus, in order to translate efficient costs into tariffs, the regulator would need to isolate the costs related to NH care from the joint-production cost. This cost split was not a part of the baseline measurement and thus needs to be done as part of future work, along with the slack-correction discussed above. Under the assumption of the equal profit margins on both NH and jointly produced products for an efficient nursing home, the efficient cost split for this nursing home could be derived as follows:

*efficient opex of NH products = prospective cost + (opex component of NH revenue/ opex component of total revenue) * reported total opex*

Here revenue opex components are defined by subtracting the (normative) capital component from the revenue. The data on normative capital component is available for the most nursing home services, because this component is set by the regulator as a part of the regulated

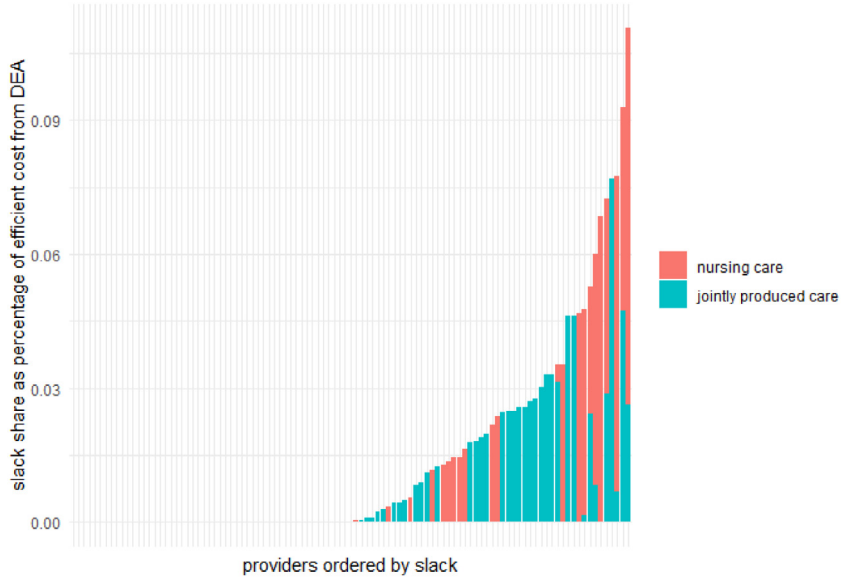

Fig. A.1. Slack as share of DEA efficient costs in de reference set (excluding outliers).

maximum tariff. For a small share of jointly produced products, which are not under regulation, the capital component can be estimated based on the capital component for similar products from the regulated segment.

For inefficient nursing homes, the efficient opex of NH products can be expressed as a linear combination of that of peers, weights being the DEA-weights, corrected for slack:

efficient opex of NH products =

*sum over peers (peer weight * efficient opex of NH products of the peer) – slack value*

References

[33] P. Bogetoft, Incentive efficient production frontiers: An agency perspective on DEA, Manage. Sci. 40 (1994a) 959–968.

[34] P. Bogetoft, Non-Cooperative Planning Theory, Springer-Verlag, Berlin, 1994b.

[35] P. Bogetoft, Incentives and productivity measurements, Int. J. Prod. Econ. 39 (1995) 67–81.

[3] P. Bogetoft, DEA-based yardstick competition: The optimality of best practice regulation, Ann. Oper. Res. 73 (1997) 277–298.

[36] P. Bogetoft, DEA and activity planning under asymmetric information, J. Prod. Anal. 13 (2000) 7–48.

[37] P. Bogetoft, Performance Benchmarking: Measuring and Managing Performance, Springer, New York, 2012.

[32] P. Bogetoft, L. Otto, Benchmarking with DEA, SFA, and R, Springer, New York, 2011.

[1] A. Shleifer, A theory of yardstick competition, Rand J. Econ. 16 (3) (1985) 319–327.

[2] J.-J. Laffont, J. Tirole, A Theory of Incentives in Procurement and Regulation, MIT Press, Cambridge, Massachusetts, The USA, 1993.

[4] NZa – Dutch Healthcare Authority, Advies Integrale Vergelijking Verpleeghuiszorg, NZa, The Netherlands, 2021, published on 21-01-2021.

[5] Tim Coelli, in: Tim Coelli (Ed.), A primer on efficiency measurement for utilities and transport regulators, 953, World Bank Publications, 2003.

[6] P. Agrell, P. Bogetoft, Regulatory benchmarking: Models, analyses and applications, Data Envel. Anal. J. 3 (2017) 49–91.

[7] P. Agrell, P. Bogetoft, Theory, techniques and applications of regulatory benchmarking, in: E. Grifell-Tatjé, C.A. Knox Lovell, R.C. Sickles (Eds.), The Oxford Handbook of Productivity Analysis, Oxford University Press, 2018, pp. 523–555, Chapter 16.

[8] P.J. Agrell, P. Bogetoft, J. Tind, DEA and dynamic Yardstick competition in scandinavian electricity distribution, J. Prod. Anal. 23 (2005) 173–201.

[9] T. Jamasb, M.G. Pollitt, Benchmarking and Regulation of Electricity Transmission and Distribution Utilities: Lessons from International Experience, 2004, http://dx.doi.org/10.17863/CAM.5398.

[10] R. Gearhart, No theory: An explanation of the lack of consistency in cross-country health care comparisons using non-parametric estimators, Health Econ. Rev. 6 (40) (2016).

[11] S. Kohl, J. Schoenfelder, A. Fügener, et al., The use of Data Envelopment Analysis (DEA) in healthcare with a focus on hospitals, Health Care Manag. Sci. 22 (2019) 245–286, http://dx.doi.org/10.1007/s10729-018-9436-8.

[12] NZorgkaart Nederland, 2021. Retrieved on 26-01-2021 from https://www.zorgkaartnederland.nl/verpleeghuis-en-verzorgingshuis.

[13] R.R. Croes, Y.J.F.M. Krabbe-Alkemade, M.C. Mikkers, Competition and quality indicators in the health care sector: empirical evidence from the dutch hospital sector, Eur. J. Health Econ. 19 (1) (2018) 5–19.

[14] KwaliteitskaderVerpleeghuiszorg, 2017. https://www.zorginstituutnederland.nl/publicaties/publicatie/2017/01/13/kwaliteitskader-verpleeghuiszorg. (Accessed 22 April 2021).

[15] A. Charnes, W.W. Cooper, E. Rhodes, Measuring the efficiency of decision making units, European J. Oper. Res. 2 (6) (1978) 429–444.

[16] A. Kneip, B.U. Park, L. Simar, A note on the convergence of non-parametric DEA efficiency measures, Econom. Theory 14 (1998) 783–793.

[17] C. Cazals, J.P. Florens, L. Simar, Nonparametric frontier estimation: a robust approach, J. Econometrics 106 (2002) 1–25.

[18] Y. Aragon, A. Daouia, C. Thomas-Agnan, Nonparametric frontier estimation: A conditional quantile-based approach, Econom. Theory 21 (2005) 358–389.

[19] Y.A. Ozcan, Evaluation of performance in health care, in: Health Care Benchmarking and Performance Evaluation, in: International Series in Operations Research & Management Science, vol. 210, Springer, Boston, MA, 2014, http://dx.doi.org/10.1007/978-1-4899-7472-3_.

[20] R. Antle, P. Bogetoft, Mix stickiness under asymmetric cost information, Manage. Sci. 65 (6) (2019) 2787–2812.

[21] P. Bogetoft, A. Eskesen, Balancing Rent Extraction and Service Differentiation, Working Paper, CBS, Denmark, 2021.

[22] P.J. Agrell, P. Niknazar, Structural and behavioral robustness in applied best-practice regulation, Socio-Econ. Plan. Sci. 48 (1) (2014) 89–103.

[23] Arcos-Vargas. A., F. Núñez-Hernández, G. Villa-Caro, A DEA analysis of electricity distribution in Spain: An industrial policy recommendation, Energy Policy 102 (2017) 583–592, http://dx.doi.org/10.1016/j.enpol.2017.01.004.

[24] X. Dai, T. Kuosmanen, Best-practice benchmarking using clustering methods: Application to energy regulation, Omega 42 (1) (2014) 179–188.

[25] J.L. Blank, E.A. Eggink, Quality-adjusted cost function in a regulated industry: The case of Dutch nursing homes, Health Care Manag. Sci. 4 (2001) 201–211.

[26] R. Weech-Maldonado, G. Neff, V. Mor, Does quality of care lead to better financial performance? The case of the nursing home industry, Health Care Manag. Rev. 28 (3) (2003) 201–216.

[27] J. Laine, M. Linna, U. Häkkinen, A. Noro, Measuring the productive efficiency and clinical quality of institutional long-term care for the elderly, Health Econ. 14 (3) (2004) 245–256.

[28] J.F. Schnelle, S.F. Simmons, C. Harrington, M. Cadogan, E. Garcia, B.M. Bates-Jensen, Relationship of nursing home staffing to quality of care, Health Serv. Res. 39 (2) (2004) 225–250.

[29] R. Dulal, Technical efficiency of nursing homes: do five-star quality ratings matter? Health Care Manag. Sci. 21 (2018) 393–400.

[30] S. Ni Luasa, D. Dineen, M. Zieba, Technical and scale efficiency in public and private Irish nursing homes – A bootstrap DEA approach, Health Care Manag. Sci. 21 (2018) 326–347.

[31] Y. Varabyova, J. Schreyögg, Integrating quality into the nonparametric analysis of efficiency: A simulation comparison of popular methods, Ann. Oper. Res. 261 (1) (2018) 365–392.