Better for Less DISCUSSION PAPER SERIES

DP No. 14/2018

The nexus between Productive Performance and Effectiveness in English General Practices under market status heterogeneity

Nikos Chatzistamoulou

School of Economics, University of Surrey, GU2 7XH, United Kingdom

n.chatzistamoulou@surrey.ac.uk, chatzistamoulou@gmail.com

https://sites.google.com/site/chatzistamoulou/

School of Economics, Faculty of Arts and Social Sciences, University of Surrey, Elizabeth Fry Building 04 AD 00, Guildford, GU2 7XH, Surrey, UK. T: +44(0)01483686623, E-mail: deliveringbetter@surrey.ac.uk, Web: https: www.surrey.ac.uk/better-for-less

Abstract
The paper explores the nexus between productive efficiency and effectiveness of the general practices in England under alternative market status types in a heterogeneity framework. For the first time consolidation patterns of eight per cent of the general practices in England from 2013 through 2016 are recorded through detailed individual matching using the address and postal code. Productive efficiency and effectiveness and brought together and compared across heterogeneous organizational structures by adopting a non-parametric metafrontier framework while effectiveness is access based on the ability to reach quality thresholds set by the regulator. Findings indicate that consolidation increases over time and thresholds rise annually making hard for general practices to reach. Practice size appears to be a big issue as healthcare is a labour intensive industry. Consolidated practices perform better in terms of efficiency-effectiveness, are bigger and manage to exploit economies of scale compared to the unconsolidated practices, probably as a result of better access to resources indicating that the two notions are interlinked. However additional investment in resources is required to reach the quality thresholds set by the regulator. Nonetheless, those are not stylized results of the general practice industry, yet this proves to be a niche for future research.

Keywords: General Practice, Efficiency, Effectiveness, Quality Thresholds, Mergers, Data Envelopment Analysis

JEL classification codes: D22, D24, I11

Acknowledgements
The author acknowledges financial support from the Leverhulme Trust, as data have been collected in the context of the project "Better for Less; Improving Productivity in the Public Services in the UK" (RL-2012-681). I thank Dr Jo Blanden and Dr Emmanouil Tyllianakis for the useful comments and suggestions on earlier versions of the paper. All remaining errors are my own.

Conflict of interest & Ethical consideration

The author declares that there is no conflict of interest. This study was reviewed using the University of Surrey's ethics procedures and was found to have no ethical concerns.

1. Introduction & Motivation

Is efficiency that comes first, is it the effectiveness or those are intertwined? The nexus between efficiency and effectiveness has been a challenging topic among scholars, practitioners, consultants as well as policy makers. However, the underlying concept is found on the grounds of the efficient resource allocation and appropriate utilisation of production means, regardless of the area of application.

In spite of being interchangeably used, the two notions of performance are distinct. The key difference is that efficiency refers to the ability of the production entity to combine resources and produce output(s) given technology or organizational structure, whereas effectiveness refers to the amount of resources required to maintain a certain level of output. The study of Szczepura et al. (1993), discerns between the two, however evidence does not suggest there is a straightforward association between the two even though a taxonomy is attempted.

In contrast to the case of efficiency where a clearer idea of its conceptual underpinnings and how to assess it has been formed via research, the case of effectiveness is considered as a hard shell due to the difficulty in determining an objective threshold. This is particularly difficult in the services sector where the output is actually part of an overall experience. That being said, setting an official threshold to discern the effective practices would enable policy makers to better design mechanisms to boost or promote the performance enhancement.

The services sector and more precisely, the primary care industry efficiency, has attracted the interest at both sides of the ocean with studies in Europe, such as in Austria (Staat, 2003), Spain (Pina & Torres, 1992), Finland (Luoma et al., 1996) and United Kingdom (Szczepura et al., 1993; Giuffrida & Gravelle, 2001; Buckell et al., 2015; Smith et al., 2016), as well as in United States (Gaynor & Pauly, 1990; Rosenman & Friesner, 2004; Collier et al., 2006a, 2006b) assessing physician efficiency. Overall, evidence suggests that primary care is a quite efficient industry (technical efficiency scores varying from .65 to 1).

As for the secondary care case, considerable research effort has emerged through the years as well. Such studies come from Finland and Norway (Linna et al., 2006), Italy (Rebba & Rizzi, 2006), Spain (Dalmau-Atarrodona & Puig-Junoy, 1998; Alonso et al., 2015), Portugal (Almeida & Fique, 2011), Greece (Aletras et al., 2007; Tsekouras et al., 2010; Halkos & Tzeremes, 2011; Kounetas & Papathanassopoulos, 2013), England (Maniadakis et al., 1999; Jacobs, 2001; Cooper et al., 2011; Takundwa et al., 2017), Germany (Tiemann et al., 2012) as well as across the globe (Hollingsworth et al., 1999; Hollingsworth, 2003; Hollingsworth, 2008; O'Neill et al., 2008) indicating that healthcare performance is not a negligible issue.

As far as the empirical tool is concerned, the Data Envelopment Analysis (DEA) technique and the Stochastic Frontier Analysis (SFA) are the methodologies that have been used to evaluate the performance of the healthcare units. Despite the choice of the method, it is highlighted that efficiency differences could be attributed at the ownership status, institutions, organization scheme, market structure, policy decisions, investment in new technology and that the absence of a solid healthcare framework leads social inequality.

However, there are limitations. A common pitfall in DEA models, is that the smaller the number of units under examination, the more will turn out to be fully efficient (Dyson et al., 2001), especially in small samples. In addition, performance evaluation models in health care have been criticised to be inadequate of assessing the actual performance of the unit as the output is not appropriately defined, among others (Newhouse, 1994).

Moreover, a direct comparison of performance between groups is problematic, as technological and organizational heterogeneity, among the units distorts the results. Industries and firms differ in resource endowments and access to resources, capabilities and knowledge stock (McGaham & Porter, 1997). The latter is what differentiates firm's performance according to Teece et al. (1997). Also, Knott (2003) mentions that strategic heterogeneity and differences in knowledge stock is what triggers an industry's growth, as firms' can learn from one another.

The metafrontier framework set a new era in performance assessment (O'Donnell et al., 2008), building on the concept of metaproduction function of Hayami (1969) and Hayami and Ruttan (1970). By accounting for all the possible heterogeneity among the units under examination, facilitates comparisons and represents the knowledge stock and available technology stock where all Decision Making Units (countries, firms, industries, schools etc.) have, potentially, access to. In addition, the overall knowledge stock could be partitioned (Tsekouras et al., 2016) into distinct knowledge stocks based on a particular criterion e.g. market status type, allowing for performance comparisons. In this paper, I partition the general practice (GP) industry using the market status type of each GP to shed some light on the nexus between efficiency and effectiveness levels of English GPs.

Shifting the attention to the English National Health Services (NHS), in 2004 as part of the new General Practice Medical contract, a new scheme of performance assessment was introduced by the regulator as a means to measure how well primary care performs using an objective point system. Through the Quality of Outcomes Framework (QOF), GPs collect points on several indicators, which are subject to annual changes, which fall into different sub-domains (clinical, public health, public health additional services). The overall QOF score is comprised by those categories.

In 2013/14 NHS England increased the QOF thresholds at the 75th centile of practice achievement while from 2015/16 those will change on annually in relation to the GP QOF (NHS Employers, 2014). It becomes apparent that thresholds provide the room for effectiveness assessment as through the scheme, a proxy to GP effectiveness is provided (NHS, 2014). Although the threshold is a priori known, the actual performance, in terms of QOF score, is not. Therefore, this is akin to the carrot and stick approach to incentivize GP work harder in offering higher quality of services.

Although QOF is a voluntary point system, most practices participate as this is a means for income flow rewarding practices for the quality of services and enhanced performance,

comprising up to 15% of the practice funding. However, this scheme has been criticized for compromising practice ethos for short run reward (Rodwin, 2004) and that the pursuit of a stable income source leads to patient skimming (Whynes & Baines, 1998) while Hawkes (2014) argues that the rise of thresholds reduces practice funding from QOF payments.

The introduction of the QOF has definitely cured many quality-measurement issues, however it applied additional pressure for performance enhancement to the GPs as patients would seek for the best accessible care (Biørn & Godager, 2010, Kann et al., 2010; Santos et al., 2017) even if they have to register with a practice outside its catchment area (Mays et al., 2014). Moreover, the abolishment of the body monitoring entry in general practices in 2006 (Department of Health, 2006) created additional competitive pressures and the structure of the market is a significant determinant of a firm's ability to grow in capacity (Teruel-Carrizosa, 2010). At the same time, a shrinking trend in the primary care has been noted by The Guardian (2014) documenting a fall in the number of GPs and The Telegraph (2014) highlighting the risk of closure for small GPs especially in rural areas due to financial pressures.

As a response to the increased pressure for performance enhancement, practices come together to form large groups of practices. As practices become larger, the capacity is getting stronger and it is more likely to have a positive effect on its performance, as larger practices achieve higher quality (Kelly & Stoye, 2014). Additionally Given (1996) mentions that among Health Maintenance Organizations (HMO) there is the perception that mergers is a means to deliver higher quality. Blanden and Chatzistamoulou (2018) find that positive changes in practice size lead to higher performance by considering a variety of practice performance outcomes such as the overall QOF score and the patient satisfaction levels. Therefore, practice size has a prominent role in performance improvement as healthcare is a labour intensive industry (Szczepura et al., 1993) while size has been acknowledged as a determinant of an industry's growth (Wagner, 1992; Davidsson et al., 2002).

In conjunction to the financial pressure, GPs consolidate by joining practice groups as over the last few years, mergers have increased by 17% during the financial years 2013/14 and 2014/15 with GP leaders arguing that merging is the workhorse in GP survival (The Pulse, 2014). However evidence suggests that firms in the services sector follow a different growth pattern compared to other sectors e.g. manufacturing (Audretsch et al., 2004). Literature on the effect of mergers on practice size on firms of manufacturing and services in Switzerland has shown that the size of the initial size of the merger deal is important (Burghardt & Helm, 2015), but literature on GP performance and consolidation is quite scarce as official data on the merging activity is not maintained by the NHS England. One exception is the study by Blanden and Chatzistamoulou (2018), who find that consolidation has a positive and significant effect on practice performance, however this effect is absorbed after controlling for the size of the practice.

Since information on consolidation between practices is not easily accessible, the actual agreement scenario among the practices in the same group (federation, partnership, acquisition, merger, take over) cannot be uncovered and this a limitation of this study. Nonetheless, this is the best approximation to the GP market in relation to the market status for the time being. I use the generic term consolidation to refer to the case a practice is found to be part of a practice group without significant loss of information. Consolidation proves to be a non-negligible issue in primary care with intertemporal interest (Goddard & Ferguson, 1997; Gaynor & Haas-Wilson, 1999) for the delivery of health care (Goodwin et al., 2011)..

Considering the above, there is a number of novelties discussed in this paper. Covering most of the GPs universe in England, for the first time market status is recorded and employed as a partitioning factor of the GP industry to compare the performance of consolidated and unconsolidated practices. Moreover, this is the first study to bring together efficiency, effectiveness in the form of quality thresholds to investigate the patterns accounting for organizational heterogeneity across groups under the metafrontier framework. To the best of my

knowledge, neither the metafrontier framework has been applied to the GP performance assessment in England nor has the effect of market status been explored as a means of performance differentiation among general practices in this context.

Findings indicate that consolidation increases over the period of study and so have quality thresholds. Consolidation seems to boost performance, as consolidated practices have on average higher performance compared to the unconsolidated ones. However additional investment in resources is required to reach the quality thresholds set by the regulator. Evidence indicates that efficient practices are also effective highlighting that the two notions are intertwined and occur jointly. Consolidated practices perform better in terms of efficiency-effectiveness probably as a result of a larger capacity and access to resources and have better prospects in exploiting the economies of scale and complementarities attached to skill-mix.

The papers unfolds as follows. Section 2 presents the methods, Section 3 presents the data, Section 4 discusses the results and Section 5 concludes the paper.

2. Methods for general practice performance assessment

2.1 Quality scores and thresholds

As mentioned above, quality thresholds raised in 2013/14 as part of the annual GMS contract changes. Increases in the thresholds are decided for all indicators in line with the 75th centile of GP achievement (NHS, 2014). Therefore as mentioned, GPs try to collect as many as possible points so as to exceed the threshold and increase practice income. Thresholds are known in advance but the GP's QOF score is not and this is the motivational characteristic of the system; to increase the effort of GPs to provide higher quality of care. At the same time, the existence of a threshold allows us to discern between those that manage to reach it and those that not.

Therefore, I define effectiveness as *the ability of the general practice to reach and exceed the quality threshold based on the 75th centile of the overall QOF achievement score's distribution on a particular year. In other words, effectiveness is about reaching targets set by the regulator.* I evaluate the effectiveness levels of the GPs based on the overall achievement score without any loss of information, as it is the sum of the points collected on the indicators of the sub-domains. Therefore, the following effectiveness types arise:

$$Achiever = \begin{cases} 1\,, if\ overall\ QOF\ score_{it} \geq Threshold_t \\ 0, \quad otherwise \end{cases}$$

2.2 Performance evaluation under heterogeneous organizational structures

Despite the fact that GP performance assessment is not newfound (Szczepura et al., 1993), direct comparisons across different production frontiers have not been attempted so far. This is attributed to the lack of an analytic framework allowing for such comparison between heterogeneous production structures. I overcome this obstacle by adopting the metafrontier framework introduced by Hayami (1969) and Hayami & Ruttan, (1970) and materialized by the work of O'Donnell et al. (2008).

The metafrontier accounts for all the possible heterogeneity among the units under examination as it envelops all the individual frontiers. It could be used, therefore, to capture any

information, micromanagement techniques, inter-practice knowledge spillover effects, leadership models, and technological advancements in the provision of services. In that sense, the metafrontier represents the *general practice industry knowledge stock,* which affects and is affected by all practices in the general practice industry allowing for comparisons, as all the units are being evaluated based on the same benchmark. Therefore, performance comparisons of individual groups become possible.

During the last few years there has been an increasing trend in consolidation activity in the general practice industry, and GPs do not operate under the same organizational structure as mentioned. The latter could be used as partition criterion to study the performance patterns of GPs, in terms of efficiency-effectiveness at an industry level. I use the market status to partition the GP industry and give rise to alternative organizational structures, the consolidated and unconsolidated GPs so as to explore whether performance is affected by the adopted organizational structure. The individual structures incorporating micromanagement techniques, practice level achievements attributed, among others, to skill-mix and leadership, convey information hard to capture about the internal operation of the practice, represent the *organizational structure knowledge stock* or market status frontier.

Formalizing the above, GPs $(i = 1,2, \dots, n)$ are multi input and output production entities using a vector of inputs $x = (x_{1i}, x_{2i}, , \dots, x_{Ni},) \in \Re_+^N$ producing a vector of outputs $y = (y_{1i}, y_{2i}, , \dots, y_{Mi},) \in \Re_+^M$ under a technology set S defined as $S \equiv \{(x, y): x\ can\ produce\ y\} \subseteq \Re_+^{N+M}$. For the output-oriented productive efficiency scores, we use the output distance function defined as $D_O(x, y) = inf\{\phi > 0: y/\phi \in P(x)\}$ while the output set is defined as $P(x) = \{y: (x, y) \in S\}$. Under alternative organizational structures, $T^{Consolidated}, T^{Unconsolidated}$, the metatechnology set, T^M, can is defined as the convex hull of the jointure of the two structures represented as $T^M = \{(x, y: x \geq 0, y \geq 0)\ x\ can\ produce\ at\ least\ one\ of\ T^{Consolidated}, T^{Unconsolidated}\}$ (for a formal definition of

production possibility and technology sets, see Batesse et al., 2004). The technology set can be defined in the same way as for the single technology.

Using DEA under variable returns to scale to account for any size effects (Halkos & Tzeremes, 2009), annually, the output-augmenting metatechnical efficiency is calculated by solving the following linear programming problem:

$$MTEff_i \equiv \hat{\phi}(x,y) = \max\{\phi | \phi > 0, y \geq \sum_{i=1}^{M} \gamma_i y_i; \phi x \leq \sum_{i=1}^{N} \gamma_i x_i \text{ for } \gamma_i \quad (1)$$

such that

$$\sum_{i=1}^{N} \gamma_i \leq 1;\ \gamma_i \geq 0, i = 1,2,..,n\Big\}$$

where ϕ corresponds to the extent that the output of the *i-th* general practice, in year *t*, could be augmented by combining all inputs and technology with the most efficient way while *y*,*x* and γ are the output and input vector and weights respectively. For the output-oriented measure, the efficiency score is given by the inverse of the above estimate.

All in all, fully efficient general practices utilize the industry's level of available technological and managerial developments and operate on the industry metafrontier achieving a productive efficiency score of one, while inefficient ones fail to do so, operating below that eventually. However, it is possible for a particular firm to exhibit inefficiency with respect to the industry frontier due to the fact that does not utilize or adopt the existing knowledge stock at an industry level but at the same time to be fully efficient utilizing the respective organizational structure knowledge stock. A graphical illustration of the analytic framework, is illustrated in Figure A1 (Appendix).

3. Data

I devise a unique dataset by combining publically available complementary sources to collect data on 7,910 general practices across England over four financial years, from 2013/14 through 2016/17, covering almost eighty per cent of the general practice universe (approximately 10,000 general practices).

In the light of no official monitoring on consolidation patterns from the NHS England, I hand-collect data on the market status of each practice using the practice address and postal code via individual searches through the NHS Choices portal, the only official source providing information on the market status of each practice. The market status variable I construct, entails three types of practices. The consolidated practices (17.62% of the sample) are the ones recorded as part of a practice group, the unconsolidated (82.33%), and the permanently closed ones (.04%).

I follow a multi-output multi-input approach. The outputs as captured by the overall QOF achievement score (in percentage points) as used in the literature (Gaynor et al., 2012; Santos et al., 2016) and the patient-reported satisfaction levels, through the GP Surveys and Report running periodically after 2007, respectively. The overall QOF score (as the sum of clinical, public health and public health additional services sub-domains points) captures many aspects of the practice performance as it includes a wide variety of health indices on the sub-domains over which each GP is evaluated. Patient satisfaction captures the percentage of the patients that reported to have a good experience with the practice which due to attrition has been weighted as if all patients had responded to the distributed questionnaires.

The inputs are captured by the number of full-time equivalent (FTE, as defined in Kelly & Stoye, 2014) general practitioners in the practice (labour) while the total (in £, constant prices 2016/17) payments from the NHS to the individual practices attempt to proxy capital. This is one of the contributions in this paper, as previous studies lack a capital proxy when assessing the

productive efficiency at the general practice level. Table A1 in the Appendix presents the descriptive statistics.

Data on the QOF achievement scores, practice size and total payments was collected through the NHS Digital while data on patient satisfaction through the GP Patient Surveys and Reports. The lack of official data on consolidation does not allow us to examine any welfare effects, nor match the practices in the same practice group to evaluate the group performance. We leave that for future research, conditional on data availability.

4. Discussion and results

4.1 GP consolidation activity & industry size composition

Table 1 below presents the consolidation activity within the period of study. The cells represent row percentages.

Consolidation increases over the period, while findings indicate that one in five practices was recorded as part of a practice group over the period. More precisely, of all practices in the financial year 2013/14, 13.06% of them have been recorded as being part of a practice group, 86.91% were recorded as non being part of a practice group while the rest .03% were recorded as permanently closed. Up to 2015/16, the share of consolidated practices kept rising only to experience a slight decline in 2016/17, due to the intense activity of the precedent years most likely. Due to low representation in the sample, permanently closed practices have not been considered in the subsequent analysis.

Table 1 Consolidation patterns

Year	Consolidated	Unconsolidated	Permanently closed
2013	13.06%	86.91%	.03%
	1,018	6,774	2
2014	17.23%	82.75%	.03%
	1,288	6,187	2
2015	20.82%	79.15%	.03%
	1,334	5,071	2
2016	20.29%	79.63%	.07%
	1,383	5,427	5
Total	17.63%	82.33%	.04%
	5,023	23,459	11

Source: Own construction.

Table 2 below illustrates the composition of the English general practice industry by practice size category[1] following that of Kelly and Stoye (2014). The declining trend of number of active GP practices of the period 2004-2010, continues as the period change indicates that there was a reduction of 12.5% in the number of firms and entrepreneurial activity in the primary care, as GPs are independent contractors with the NHS. Although there was some turbulence, the reduction could be attributed to the fact that GPs decided to enrol in practice group to cope with the financial pressures by sharing the risk rather than exit the industry as Table 1 above indicates. The share of single-handed practices remains low throughout the period.

In spite of the internal changes and dynamics in the industry, it is noticeable that the share of each practice category remains almost at the same levels as time goes by. The GP industry is comprised by small and medium sized firms. The share of single-handed practices reduces and this is reasonable given the fact that GPs need increased capacity to serve the local population while only one in four every year is recorded as large and this might be attributed to the fact that those GPs have been consolidated.

Table 2 Size composition of general practice industry

Year	Single-handed	Small-Med	Med-Large	Large	All firms
2013	11.32%	31.42%	31.96%	25.3%	7,794
2014	10.41%	29.97%	32.26%	27.36%	7,477
2015	5.49%	34.63%	36.6%	23.27%	6,407
2016	6%	35.77%	36.74%	21.48%	6,815
Period	8.5%	32.8%	34.23%	24.47%	-12.52%

4.2 Heterogeneous Organizational Structures

Table 3 below showcases the profile of the general practice by organizational structure and effectiveness categories for the period of study to explore the profile of those that manage to exceed the quality threshold.

Focusing on the firm-based financial flows as captured by the QOF payments and the global sum which constitute seventy five percent of the general practice income combined, it is noticeable that consolidated practices, irrespective of the effectiveness level, are better funded

[1] The categorization is based on the number of FTE GPs and is as follows. A practice is single-handed if FTE GPs ≤1, small-med if 1<FTE GPs≤3, med-large if 3<FTE GPs≤6 and large if FTE GPs>6.

compared to the unconsolidated ones on an annual basis. We also notice that the QOF payments are being reduced as time goes by and that there is a reallocation of funds to the global sum. In relation to that, thresholds are in increasing on an annual basis and in conjunction to the fact that indicators retire without replacement, GPs earn less for the QOF (Hawkes, 2014). This adds extra pressure for performance enhancement through skill-mix diversification.

The latter brings to the forefront the importance of practice size. Kelly and Stoye (2014) mention, larger practices achieve higher QOF while Blanden and Chatzistamoulou (2018) find that increases in practice size are positively and significantly associated with changes in GP performance. Consolidated practices are bigger compared to the unconsolidated ones for both categories of effectiveness. Given (1996) mentions that HMO enrol in a merger with the perception that a higher quality could be achieved and this appears that it is confirmed in this case. The productive efficiency scores of the consolidated but non-achiever GPs (with respect to the GP industry frontier) are higher. This could be an indication that through consolidation and therefore, via a larger scale of operations GPs could learn by serving (Alcacer & Oxley, 2014) and enhance their performance. However, Bojke et al. (2001) argue that there is not just an optimal practice capacity level as operations are heterogeneous and that market structure and policy directives also affect performance levels.

Another way of drawing useful information about the firms in the GP industry is the consideration of the economies of scale via the calculation of the Minimum Efficient Scale (MES). The MES can be calculated parametrically, by specifying a Cobb-Douglas or a Translog production function (Wholey et al., 1996), however that requires information on the prices of the inputs which is now always available. It can also be calculated non-parametrically using the employed labour by calculating either the Comanor-Wilson MES index (1967, 1969), the Sutton type MES (1991) or use the median firm size (Daunfeldt & Elert, 2013). I use the Comanor-

Wilson MES index[2] as it has been proved to be adequate to capture the economies of scale (Tsekouras et al., 2009).

The MES for the consolidated is higher compared to the unconsolidated ones indicating that for the former economies of scale or the increasing returns to scale last longer. The latter means that consolidated practices have the opportunity to exploit the benefits of a larger practice capacity. This finding is present for every year of the sample. In addition, the fact that consolidated and unconsolidated practices have a different MES indicates that those are indeed distinct which provides support for the choice of market status as a partition criterion of the GP industry. Furthermore, differences in the MES between alternative organizational structures could also mirror a different extent of internal and external returns to scale. The former are firm-specific enhancing the performance of the firm while the latter are industry specific and benefit all the firms in the industry as Economic Theory dictates.

The internal economies of scale bring the spotlight on the firm-specific environment which is related to the heterogeneous organizational structures introduced herein. Findings indicate that consolidated practices are in a more advantageous position compared to the unconsolidated ones and not because of an institutional reason. The stock of knowledge is what differentiates the operations and scale of the firm as larger firms need to develop leadership models and apply micromanagement techniques so as to combine their resources more efficiently. As GPs are self-employed entrepreneurs, they have devised mechanisms to enhance performance and increase their profit stream by applying more efficient micromanagement techniques and creating a culture of interpersonal relationship with their patients related to customer's loyalty argument as a longer list size is rewarded by the global sum. Moreover, consolidated firms have access to more resources and because of the risk sharing it is more possible to proceed to investments such as to adopt a new medical technology or move to bigger premises.

[2] The MES index is calculated as the average of the upper half of the observed labour distribution.

Overall, consolidated and effective (i.e. Achievers) practices appear to be better funded, larger and more efficient compared to their unconsolidated counterparts for the whole period. Moreover, they appear to be able to internalize the advantages of a larger scale as there are complementarities to exploit associated to the skill-mix. Unconsolidated practices shape the industry picture while do not seem to be able to exploit the returns to scale as those exhaust in relatively low levels overall. Moreover, given that smaller GPs have constrained resources, it is possible that their premises cannot support a larger capacity which diminishes returns to labour preventing them from exploiting economies of scale. However, this is not particularly unexpected as the literature on the US physicians provides documents that economies of scale "*exhaust in relatively small sizes*" for physicians (Given, 1996; Pope & Burge, 1996, Wholey et al., 1996) as primary care is a labour intensive industry.

These results should only be considered as indicative of a constantly changing market and by any means it is not a stylized result of the general practice industry.

Table 3 Profile of the GP by market status

		Consolidated		Unconsolidated		All practices (industry level)
Year		Non-achiever	Achiever	Non-achiever	Achiever	
2013	QOF payments	169,479	206,898	119,000	135,975	130,537
		(101,622)	(132,785)	(78,901)	(80,725)	(86,749)
	Global sum payments	325,779	368,530	242,034	276,718	261,866
		(386,762)	(398,735)	(283,482)	(294,826)	(303,627)
	Practice size	5.94	6.85	4.10	4.57	4.47
		(3.70)	(4.46)	(2.87)	(2.95)	(3.13)
	Productive Efficiency score wrt industry	.942	.993	.938	.993	.952
		(.041)	(.004)	(.049)	(.004)	(.048)
	MES	8.72	9.82	6.30	6.71	6.84
	Quality threshold			.986		
2014	QOF payments	127,461	157,288	83,279	109,393	97,586
		(73,512)	(85,923)	(54,371)	(61,068)	(63,777)
	Global sum payments	373,305	406,599	276,579	312,151	301,953
		(433,075)	(434,117)	(306,021)	(324,709)	(337,296)
	Practice size	5.97	7.18	4.06	4.95	4.63
		(3.84)	(4.44)	(2.79)	(2.98)	(3.19)
	Productive Efficiency score wrt industry	.951	.997	.945	.997	.959
		(.044)	(.003)	(.047)	(.003)	(.046)
	MES	8.83	10.14	6.21	7.24	6.94
	Quality threshold			.990		
2015	QOF payments	123,370	150,847	82,833	103,402	96,823
		(69,482)	(94,520)	(51,921)	(55,260)	(62,133)
	Global sum payments	442,325	499,101	332,931	385,433	369023
		(456,826)	(519,108)	(328,342)	(337,435)	(369,050)
	Practice size	5.49	6.22	3.82	4.46	4.34
		(3.59)	(3.86)	(2.53)	(2.63)	(2.92)
	Productive Efficiency score wrt industry	.957	.998	.955	.998	.966
		(.041)	(.002)	(.042)	(.002)	(.041)
	MES	8.07	8.95	5.70	6.51	6.49
	Quality threshold			.994		
2016	QOF payments	123,125	149,033	78079	105,015	93,870
		(78,403)	(93,325)	(50,161)	(55,021)	(63,091)
	Global sum payments	532,258	606,979	379,439	447,368	427,724
		(485,714)	(578,611)	(356,916)	(372,774)	(402,912)
	Practice size	5.30	6.32	3.62	4.48	4.18
		(3.36)	(4.14)	(2.44)	(2.62)	(2.87)
	Productive Efficiency score wrt industry	.966	.999	.962	.999	0.972
		(.036)	(.001)	(.041)	(.001)	(.038)
	MES	7.77	9.52	5.45	6.52	6.30
	Quality threshold			.996		

Note: Means and standard deviations in parentheses.

4.3 Focusing on performance

Table 4 below, presents the overall QOF scores and productive efficiency scores by organizational structure.

As regards the QOF scores, although practices appear to attaint high (on average) QOF scores annually, do not manage to reach the target, irrespective of the organizational structure, as by construction one in four reaches it. Hawkes (2014) mentions that the increase of the thresholds in conjunction to the retirement of some QOF points make GPs to be less funded by the QOF stream. Raising the thresholds could also act a carrot-and-stick approach to offer higher quality so as to get more funds. A high QOF score (on average) does not necessarily imply effectiveness, as quality scores are firm-specific (GP) whereas thresholds are industry-specific and therefore are affected by the performance of the peers. However, consolidated practices appear to perform better throughout the years and although the difference seems small, is significant.

Shifting the attention to the productive efficiency, consolidated practices, at an industry level, exhibit relatively higher performance compared to unconsolidated practices and this difference is significant for every year as indicated by the non-parametric test of Kruskal-Wallis.

On a final note, QOF scores provide a one-sided consideration of the practice performance and should the regulator aspires to promote performance, considering the whole scale of operations, from management to input complementarities, would be an appropriate approach for a spherical evaluation considering total factor productivity measures as through those input complementarities are taken into consideration (Wang et al., 2013).

Table 4 DEA VRS efficiency scores and thresholds

		Year			
		2013	**2014**	**2015**	**2016**
	Quality threshold	.986	.990	.994	.996
QOF score	All practices (industry level)	.937 (.075)	.949 (.068)	.959 (.060)	.966 (.059)
	Unconsolidated	.936 (.076)	.947 (.070)	.958 (.061)	.964 (.060)
	Consolidated	.946 (.060)	.958 (.058)	.965 (.056)	.972 (.051)
	Difference (T-test)	-.010*** (.002)	-.011*** (.002)	-.007*** (.002)	-.008*** (.002)
Productive Efficiency w.r.t. industry					
All practices (industry level)		.952 (.048)	.959 (.046)	.966 (.041)	.972 (.038)
Unconsolidated		.951 (.490)	.958 (.046)	.966 (.041)	.971 (.039)
Consolidated		.956 (.042)	.964 (.043)	.969 (.039)	.976 (.034)
Kruskal-Wallis p-value		.058	.000	.000	.000

Note: Parentheses correspond to standard deviations. Stars indicate significance at 1% ***, 5% **, 10% *.

4.4 Efficiency and Effectiveness patterns

Table 5 below presents the allocation of general practices between efficiency and effectiveness under heterogeneous organizational structures over the period of study. Cells represent row percentages.

The percentage of consolidated practices that are fully efficient and effective is higher than that of the unconsolidated ones for almost all the years of the sample. Consolidated practices appear to perform quite well indicating that organizational structure they operate under could be a source of performance boost due to the fact that consolidated practices are bigger and perform better (Given, 1996; Kelly & Stoye, 2014, Blanden & Chatzistamoulou, 2018).

However, the following clarification is necessary. After a GP has been classified as efficient, does not imply that it cannot improve its performance any further. On the contrary, it means that the particular practice utilizes its resources i.e. inputs, technology management techniques, in the most efficient way possible avoiding waste. In order to continue attaining high scores of efficiency, or in other words improve quality, more investment is required. This could imply either more resources or better management as firms mix resources under human

supervision conditional on technological regimes, resources and production environment (Dimara et al., 2008).

Therefore, it is most likely that consolidated practices i.e. larger practices become more competent in reaching quality targets in the long run, probably because they need time to exploit the benefits of the merger such as better management, skill-mix or learning by providing to the local population (Alcacer & Oxley, 2014) or this might be attributed to the learning by doing (Fudenberg & Tirole, 1983) activity of the practice. As primary care is a labour-intensive industry general practices might have learnt to offer services tailored to the needs of the patients due to the fact that GPs built an interpersonal relationship with the patients. Even if primary care is a mature industry, the exploitation of some sort of inter-practices or even clinical commissioning group spillover effects cannot be ruled out (Irwin & Klenow, 1994; Teece et al., 1997; Knott, 2003; Baicker & Robbins, 2015).

Concluding, there is a strong correlation between the two notions of performance evaluation. Also it appears that market status matters in the provision of care as it provides the opportunity to have access to better resources and medical technology due to more funds. On a final note, although the differences in the performance between market status types are significant, those are not too large indicating that unconsolidated practices also offer high quality of services to the patients, even with less resources.

Table 5 Efficiency and Effectiveness by market status

		Non-achiever	Achiever
2013			
Unconsolidated	Non Fully Efficient	76.94%	23.06%
	Fully Efficient	19.47%	80.53%
Consolidated	Non Fully Efficient	74.22%	25.78%
	Fully Efficient	16%	84%
2014			
Unconsolidated	Non Fully Efficient	80.32%	19.68%
	Fully Efficient	7.52%	92.48%
Consolidated	Non Fully Efficient	76.56%	23.44%
	Fully Efficient	8.99%	91.01%
2015			
Unconsolidated	Non Fully Efficient	83.29%	16.71%
	Fully Efficient	6.11%	93.89%
Consolidated	Non Fully Efficient	79.44%	20.56%
	Fully Efficient	2.04%	97.96%
2016			
Unconsolidated	Non Fully Efficient	86.12%	13.88%
	Fully Efficient	5.60%	94.40%
Consolidated	Non Fully Efficient	81.85%	18.15%
	Fully Efficient	1.55%	98.45%

5. Conclusions

The importance of primary care performance assessment as a mechanism in promoting wellbeing is not newfound. Many studies have provided useful insights in measuring the performance of general practices. However due to limited information, aspects that appear to affect performance have been unintentionally overlooked.

This papers is the first paper to assess the performance patters of general practices in England bringing together productive efficiency and effectiveness under heterogeneous organizational structures. The latter correspond to alternative market status types which are recorded, using hand-collected data though the NHS Choices site. Since official data on consolidation is not maintained at a GP level, the market status was uncovered via detailed matching using the address and postal code of each practice in the sample.

Using official data on from the QOF scheme and considering the thresholds set by the regulator, I calculate DEA efficiency scores under a metafrontier framework which allows for direct comparisons between different market status types.

However, lack of official data on consolidation does not allow us to explore its effect on practice performance before and after being part of a practice group as we account for consolidation only using a snapshot of the industry. As a result, there could be some bias in identifying the market status unintentionally. The analysis could also be benefited by data on the staff composition of the practice such as nurses, trainees, pharmacists and by a wider time span so as to explore the effect of consolidation in the long run.

I find that although most of the practices in the industry perform quite well, do not manage to meet quality targets on average. Consolidated practices have higher performance compared to unconsolidated ones, receive more funding (on average) and operate in a larger scale by exploiting the economies of scale via the internal ones attributed to micromanagement techniques applied to larger groups as well as the exploitation of complementarities in the

practice capacity. These results are only indicative of the GP industry, do not constitute a stylized result of the healthcare market and should be interpreted with caution.

References

Alcacer, J., & Oxley, J. (2014). Learning by supplying. *Strategic Management Journal*, *35*(2), 204-223.

Aletras, V., Kontodimopoulos, N., Zagouldoudis, A., & Niakas, D. (2007). The short-term effect on technical and scale efficiency of establishing regional health systems and general management in Greek NHS hospitals. *Health Policy*, *83*(2), 236-245.

Alonso, J. M., Clifton, J., & Díaz-Fuentes, D. (2015). The impact of New Public Management on efficiency: An analysis of Madrid's hospitals. *Health Policy*, *119*(3), 333-340.

Álvaro Almeida & José Pedro Fique, 2011. "Evaluating Hospital Efficiency Adjusting for Quality Indicators: an Application to Portuguese NHS Hospitals,"FEP Working Papers 435, Universidade do Porto, Faculdade de Economia do Porto.

Audretsch, D. B., Klomp, L., Santarelli, E., & Thurik, A. R. (2004). Gibrat's Law: Are the services different?. *Review of Industrial Organization*, *24*(3), 301-324.

Baicker, K., & Robbins, J. A. (2015). Medicare payments and system-level health-care use: the spillover effects of Medicare managed care. American journal of health economics, 1(4), 399-431.

Battese, G. E., Rao, D. P., & O'donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis*, *21*(1), 91-103.

Biørn, E., & Godager, G. (2010). Does quality influence choice of general practitioner? An analysis of matched doctor–patient panel data. *Economic Modelling*, *27*(4), 842-853.

Bojke, C., Gravelle, H., & Wilkin, D. (2001). Is bigger better for primary care groups and trusts?. *British Medical Journal*, *322*(7286), 599-602.

Blanden, J., & Chatzistamoulou, N., (2018). Strength in numbers; Does Horizontal Consolidation affect General Practice Performance in England? Better for Less Project Discussion Paper (2018).

Buckell, J., Smith, A., Longo, R., & Holland, D. (2015). Efficiency, heterogeneity and cost function analysis: empirical evidence from pathology services in the National Health Service in England. *Applied Economics*, *47*(31), 3311-3331.

Burghardt, D., & Helm, M. (2015). Firm growth in the course of mergers and acquisitions. *Small Business Economics*, *44*(4), 889-904.

Comanor, W. S. (1967). Market structure, product differentiation, and industrial research. *The Quarterly Journal of Economics*, *81*(4), 639-657.

Comanor, W. S., & Wilson, T. A. (1969). Advertising and the Advantages of Size. *The American Economic Review*, *59*(2), 87-98.

Collier, D. A., Collier, C. E., & Kelly, T. M. (2006a). Benchmarking physician performance, Part 1. *The Journal of medical practice management: MPM*, *21*(4), 185-189.

Collier, D. A., Collier, C. E., & Kelly, T. M. (2006b). Benchmarking physician performance, Part 1. *The Journal of medical practice management: MPM*, *21*(4), 185-189.

Cooper, Z., Gibbons, S., Jones, S., & McGuire, A. (2011). Does hospital competition save lives? Evidence from the English NHS patient choice reforms. *The Economic Journal*, *121*(554).

Dalmau-Atarrodona, E., & Puig-Junoy, J. (1998). Market structure and hospital efficiency: evaluating potential effects of deregulation in a national health service. *Review of Industrial Organization*, *13*(4), 447-466.

Davidsson, P., Kirchhoff, B., Hatemi–J, A., & Gustavsson, H. (2002). Empirical analysis of business growth factors using Swedish data. *Journal of small business management*, *40*(4), 332-349.

Daunfeldt, S. O., & Elert, N. (2013). When is Gibrat's law a law?. *Small Business Economics*, *41*(1), 133-147.

Department of Health. (2006). Our Health, Our Care, Our Say: A New Direction for Community Services. Cm 6737. London, Stationery Office.

Dimara, E., Skuras, D., Tsekouras, K., & Tzelepis, D. (2008). Productive efficiency and firm exit in the food sector. *Food Policy*, *33*(2), 185-196.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of operational research*, *132*(2), 245-259.

Fudenberg, D., & Tirole, J. (1983). Learning-by-doing and market performance. The Bell Journal of Economics, 522-530.

Gaynor, M., & Pauly, M. V. (1990). Compensation and productive efficiency in partnerships: evidence from medical groups practice. *Journal of Political Economy*, *98*(3), 544-573.

Gaynor, M., & Haas-Wilson, D. (1999). Change, consolidation, and competition in health care markets. *Journal of economic perspectives*, *13*(1), 141-164.

Gaynor, M., Laudicella, M., & Propper, C. (2012). Can governments do it better? Merger mania and hospital outcomes in the English NHS. *Journal of health economics*, *31*(3), 528-543.

Giuffrida, A., & Gravelle, H. (2001). Measuring performance in primary care: econometric analysis and DEA. *Applied Economics*, *33*(2), 163-175.

Given, R. S. (1996). Economies of scale and scope as an explanation of merger and output diversification activities in the health maintenance organization industry. *Journal of health economics*, *15*(6), 685-713.

Goddard, M. and Ferguson, B. (1997). Mergers in the NHS: Made in heaven or marriages of convenience? Nuffield Trust.

Halkos, G. E., & Tzeremes, N. G. (2009). Exploring the existence of Kuznets curve in countries' environmental efficiency using DEA window analysis. *Ecological Economics*, *68*(7), 2168-2176.

Halkos, G. E., & Tzeremes, N. G. (2011). A conditional nonparametric analysis for measuring the efficiency of regional public healthcare delivery: An application to Greek prefectures. *Health policy*, *103*(1), 73-82.

Hayami, Y. (1969). Sources of agricultural productivity gap among selected countries. *American Journal of Agricultural Economics*, *51*(3), 564-575.

Hayami, Y., & Ruttan, V. W. (1970). Agricultural productivity differences among countries. *The American economic review*, *60*(5), 895-911.

Hawkes, N. (2014). GPs earn less from QOF points as thresholds for some indicators rise.

Jacobs, R. (2001). Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis. *Health care management science*, *4*(2), 103-115.

Irwin, D. A., & Klenow, P. J. (1994). Learning-by-doing spillovers in the semiconductor industry. Journal of political Economy, 102(6), 1200-1227.

Kann, I. C., Biørn, E., & Lurås, H. (2010). Competition in general practice: prescriptions to the elderly in a list patient system. *Journal of health economics*, *29*(5), 751-764.

Kelly, E. & Stoye, G. (2014). Does GP Practice Size Matter? The relationship between GP practice size and the quality of health care. Institute for Fiscal Studies. Available at https://www.ifs.org.uk/publications/7445

Knott, A. M. (2003). Persistent heterogeneity and sustainable innovation. *Strategic Management Journal*, *24*(8), 687-705.

Kounetas, K., & Papathanassopoulos, F. (2013). How efficient are Greek hospitals? A case study using a double bootstrap DEA approach. *The European Journal of Health Economics*, *14*(6), 979-994.

Linna, M., Häkkinen, U., & Magnussen, J. (2006). Comparing hospital cost efficiency between Norway and Finland. *Health policy*, *77*(3), 268-278.

Luoma, K., Järviö, M. L., Suoniemi, I., & Hjerppe, R. T. (1996). Financial incentives and productive efficiency in Finnish health centres. *Health Economics*, *5*(5), 435-445.

McGahan, A. M., & Porter, M. E. (1997). How much does industry matter, really?. *Strategic management journal*, 15-30.

Maniadakis, N., Hollingsworth, B., & Thanassoulis, E. (1999). The impact of the internal market on hospital efficiency, productivity and service quality. *Health Care Management Science*, *2*(2), 75-85.

Mays, N., Tan, S., Eastmure, E., Erens, B., Lagarde, M., & Wright, M. (2014). Potential impact of removing general practice boundaries in England: a policy analysis. *Health policy*, *118*(3), 273-278.

Newhouse, J. P. (1994). Frontier estimation: How useful a tool for health economics?. *Journal of health economics*, *13*(3), 317-322.

National Health Services Employers (2014), via http://www.nhsemployers.org/your-workforce/primary-care-contacts/general-medical-services/quality-and-outcomes-framework/changes-to-qof-2013-14

O'Donnell, C. J., Rao, D. P., & Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical economics*, *34*(2), 231-255.

O'Neill, L., Rauner, M., Heidenberger, K., & Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. Socio-Economic Planning Sciences, 42(3), 158-189.

Pina, V., & Torres, L. (1992). Evaluating the efficiency of nonprofit organizations: an application of data envelopment analysis to the public health service. *Financial Accountability & Management*, *8*(3), 213-224.

Rebba, Vincenzo and Rizzi, Dino, Measuring Hospital Efficiency Through Data Envelopment Analysis When Policy-Makers' Preferences Matter: An Application to a Sample of Italian NHS Hospitals (March 2006). University Ca' Foscari of Venice, Dept. of Economics Research Paper Series No. 13/06. Available at SSRN: https://ssrn.com/abstract=946885 or http://dx.doi.org/10.2139/ssrn.946885

Rodwin, M. A. (2004). Financial incentives for doctors: Have their place but need to be evaluated and used to promote appropriate goals. *BMJ: British Medical Journal*, *328*(7452), 1328.

Rosenman, R., & Friesner, D. (2004). Scope and scale inefficiencies in physician practices. *Health Economics*, *13*(11), 1091-1116.

Smith, A. S. J., Buckell, J., Wheat, P., & Longo, R. (2016). Hierarchical performance and unobservable heterogeneity in health: A dual-level efficiency approach applied to NHS pathology in England. In *Productivity and Efficiency Analysis* (pp. 119-143). Springer, Cham. Staat, M. (2003). The efficiency of treatment strategies of general practitioners. *The European Journal of Health Economics, formerly: HEPAC*, *4*(3), 232-238.

Santos, R., Gravelle, H., & Propper, C. (2017). Does quality affect patients' choice of doctor? Evidence from England. *The Economic Journal*, *127*(600), 445-494.

Szczepura, A., Davies, C., Fletcher, J., & Boussofiane, A. (1993). Efficiency and effectiveness in general practice. Journal of Management in Medicine, 7(5), 36-47.

Sutton, J. (1991). *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration*. MIT press.

Takundwa, R., Jowett, S., McLeod, H., & Peñaloza-Ramos, M. C. (2017). The Effects of Environmental Factors on the Efficiency of Clinical Commissioning Groups in England: A Data Envelopment Analysis. *Journal of medical systems*, *41*(6), 97.

Teruel-Carrizosa, M. (2010). Gibrat's law and the learning process. *Small Business Economics*, *34*(4), 355-373.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, *18*(7), 509-533.

The Guardian (2014), "GP numbers tumble in England as recruitment crisis bites" via https://www.theguardian.com/society/2014/jun/14/gp-numbers-fall-recruitment-crisis-bites

The Telegraph (2014), "Decline of the traditional family doctor revealed" via http://www.telegraph.co.uk/news/health/news/10778519/Decline-of-the-traditional-family-doctor-revealed.html

The Pulse (2014), "Sharp rise in GP mergers as smaller practices struggle to stay above water" via http://www.pulsetoday.co.uk/hot-topics/stop-practice-closures/sharp-rise-in-gp-mergers-as-smaller-practices-struggle-to-stay-above-water/20007879.article

Tiemann, O., Schreyögg, J., & Busse, R. (2012). Hospital ownership and efficiency: a review of studies with particular focus on Germany. *Health Policy*, *104*(2), 163-171.

Tsekouras, K., Dimara, E., Skuras, D., & Tzelepis, D. (2009). Back to basics: The Comanor–Wilson MES index revisited. *Small Business Economics*, *32*(1), 111-120.

Tsekouras, K., Papathanassopoulos, F., Kounetas, K., & Pappous, G. (2010). Does the adoption of new technology boost productive efficiency in the public sector? The case of ICUs system. *International Journal of Production Economics*, *128*(1), 427-433.

Tsekouras, K., Chatzistamoulou, N., Kounetas, K., & Broadstock, D. C. (2016). Spillovers, path dependence and the productive performance of European transportation sectors in the presence of technology heterogeneity. *Technological Forecasting and Social Change*, *102*, 261-274.

Wagner, J. (1992). Firm size, firm growth, and persistence of chance: Testing GIBRAT's law with establishment data from Lower Saxony, 1978–1989. *Small Business Economics*, *4*(2), 125-131.

Wang, Q., Zhao, Z., Zhou, P., & Zhou, D. (2013). Energy efficiency and production technology heterogeneity in China: a meta-frontier DEA approach. *Economic Modelling*, *35*, 283-289.

Wholey, D., Feldman, R., Christianson, J. B., & Engberg, J. (1996). Scale and scope economies among health maintenance organizations. *Journal of health economics*, *15*(6), 657-684.

Whynes, D. K., & Baines, D. L. (1998). Income-based incentives in UK general practice. *Health policy*, *43*(1), 15-31.