Healthcare 7 (2019) 100356

Contents lists available at ScienceDirect

Healthcare

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

Original research

Measuring individual physician clinical productivity in an era of consolidated group practices

Neel M. Butala[a], Michael K. Hidrue[b], Arthur J. Swersey[c], Jagmeet P. Singh[a], Jeffrey B. Weilburg[b], Timothy G. Ferris[b,d], Katrina A. Armstrong[d], Jason H. Wasfy[a,b,*]

[a] Cardiology Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States
[b] Massachusetts General Physicians Organization, Boston, MA, United States
[c] Yale School of Management, New Haven, CT, United States
[d] Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

ARTICLE INFO

Keywords:
Productivity
Ambulatory care
Modeling
Risk adjustment

ABSTRACT

Background: As physician groups consolidate and value-based payment replaces traditional fee-for-service systems, physician practices have greater need to accurately measure individual physician clinical productivity within team-based systems. We compared methodologies to measure individual physician outpatient clinical productivity after adjustment for shared practice resources.
Methods: For cardiologists at our hospital between January 2015 and June 2016, we assessed productivity by examining completed patient visits per clinical session per week. Using mixed-effects models, we sequentially accounted for shared practice resources and underlying baseline characteristics. We compared mixed-effects and Generalized Estimating Equations (GEE) models using K-fold cross validation, and compared mixed-effect, GEE, and Data Envelopment Analysis (DEA) models based on ranking of physicians by productivity.
Results: A mixed-effects model adjusting for shared practice resources reduced variation in productivity among providers by 63% compared to an unadjusted model. Mixed-effects productivity rankings correlated strongly with GEE rankings (Spearman 0.99), but outperformed GEE on K-fold cross validation (root mean squared error 2.66 vs 3.02; mean absolute error 1.89 vs 2.20, respectively). Mixed-effects model rankings had moderate correlation with DEA model rankings (Spearman 0.692), though this improved upon exclusion of outliers (Spearman 0.755).
Conclusions: Mixed-effects modeling accounts for significant variation in productivity secondary to shared practice resources, outperforms GEE in predictive power, and is less vulnerable to outliers than DEA.
Implications: With mixed-effects regression analysis using otherwise easily accessible administrative data, practices can evaluate physician clinical productivity more fairly and make more informed management decisions on physician compensation and resource allocation.

1. Introduction

As physician groups nationwide consolidate and risk-based contracts replace traditional fee-for-service payment systems,[1,2] outpatient practices face new challenges.[3] Larger group size leads to greater sharing of resources, so practices must account for fixed costs in assessing physician productivity. Additionally, the push to have providers practice to the full extent of their training leads to greater team-based care,[4] and measurement of physician productivity in the context of the healthcare team requires a more sophisticated approach. In conjunction with metrics to define quality and outcomes, accurately measuring individual physician clinical productivity can allow practices to compensate physicians fairly and improve value by incentivizing physicians to utilize shared resources more efficiently.

Most academic practices currently measure clinical productivity for compensation purposes using a standardized unit, but such metrics have not yet been refined to account for shared practice resources.[5] Many shared practice-level factors are known to affect physician productivity. Physical resources, such as number of exam rooms per physician, are associated with physician productivity.[6] Incorporation of other healthcare personnel such as nurse practitioners (NPs) or medical scribes into medical teams also influence physician productivity and can improve outcomes.[7,8] However, these shared resources can also increase overall practice cost,[9] and in certain settings can actually lead

* Correspondence to: Cardiology Division, GRB 8-843, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, United States.
E-mail address: jwasfy@partners.org (J.H. Wasfy).

https://doi.org/10.1016/j.hjdsi.2019.02.001
Received 1 April 2018; Received in revised form 30 January 2019; Accepted 2 February 2019
Available online 08 February 2019

to decreased profitability.[10] Relative Value Unit (RVU)-based measures of individual physician productivity,[5,11,12] even after adjusting for administrative or research activities in an academic setting,[13] fail to capture the effect of shared practice resources and team-based care. As such, rigorous metrics to evaluate individual physician clinical productivity in a group-practice context are needed.

This study contributes to the improvement of currently used productivity measures by evaluating mixed-effects regression as a method to control for shared practice costs in assessing individual physician clinical productivity. Tactics developed in health services research that have been used for public reporting of clinical outcomes, such as risk standardization, offer the potential to refine measurements of physician productivity while still being technically and conceptually accessible to healthcare administrators and physicians responsible for changing practice operations.[14] This paper applies these risk-standardization techniques using mixed-effects regression in a novel way as a method to measure individual physician outpatient clinical productivity in order to account for shared costs within a large academic cardiology practice. This paper then quantitatively compares the novel application of mixed-effects regression in this context to other methods used in both the medical literature and the management literature to measure outpatient clinical productivity.

2. Methods

2.1. Study population

We included all cardiologists at our academic medical center from January 1, 2015 to June 30, 2016. In addition to providing outpatient care, cardiologists can have other duties that are clinical (performing procedures, reviewing cardiac imaging, and inpatient service) and nonclinical (research and teaching), so the number of clinical sessions per week varies widely.

2.2. Outcomes and covariates

The primary outcome was completed patient visits per half-day clinical session per week. We chose this outcome measure for several reasons. First, including completed visits in the numerator of this measure incentivizes increasing volume of patients seen by providers, which is important for our practice given a critical need to accommodate more patients awaiting specialist appointments. Although quality is an increasingly important dimension of value-based care, given significant practice variation in accountability for quality and outcomes nationwide, we focused primarily on clinical visit volume to broaden the generalizability of the results of this study. Second, standardizing by clinic session, a discrete half-day clinical period, allowed us to compare clinicians with different proportions of effort devoted to outpatient care. For instance, a physician who sees 4 patients per clinical session and does only 1 session per week would have the same value for our outcome measure as a physician who sees 4 patients per clinical session and does 4 sessions per week. Third, examining productivity at the provider-week level (as opposed to a longer time period) allowed us to obtain multiple longitudinal observations for each physician. We identified the total number of outpatient visits completed and clinical sessions booked per week retrospectively from administrative scheduling data. We did not differentiate whether physicians actually saw patients during clinical sessions for which they were booked rooms or whether these rooms were left empty, since empty rooms represent shared group costs. Of note, nurse practitioners either jointly saw patients under the supervision of an attending cardiologist or saw patients on their own and billed incident to an attending cardiologist who could provide input, but not necessarily physically see each patient. Visits were attributed to the attending physician regardless of whether patients were seen jointly or seen incident to a nurse practitioner.

Covariates examined as possible predictor variables included NP full-time equivalents (FTEs), fellow FTEs, and secretarial FTEs, with this data obtained from hospital administrative records. We chose these covariates since they represented shared human resources factors that varied by physician and were thought to directly impact physician productivity. We also considered provider age, a dummy variable for each quarter, and patient scheduling characteristics, such number of scheduled clinic sessions per week, and proportion of new vs follow-up appointments, all summarized at the provider-week level. We additionally considered patient characteristics summarized at the provider-week level such as average patient age, average zip code median income, average Charlson score, proportion male, proportion nonwhite, proportion non-English speaking, proportion out-of-state, and proportion with Medicare, Medicaid, and private payers. Ambiguous administrative data regarding shared resources or clinical sessions were manually adjudicated in conversations with practice staff and, as such, there were no missing data at the provider level. Missing data at the patient level were minimal (limited to < 1% for all variables aside from neighborhood income).

2.3. Statistical analysis

We modeled physician productivity as a function of individual effort, shared human resources (e.g. NP, fellow, and secretary FTEs), and underlying baseline characteristics (provider, scheduling, and patient characteristics). We first examined the relationships between the outcome variable and each individual quantitative or categorical variable using the Pearson correlation coefficient and Pearson's chi-squared test, respectively. We then used mixed-effects models with a gamma distribution and log link to estimate the relationship between each covariate and physician productivity (details in Supplemental Appendix).[15] Shared resources and underlying baseline characteristics were included as fixed effects and physicians were included as a random effect in our models. The random effect in this model represents individual effort and was the primary quantity of interest for physician productivity rankings. To evaluate the impact of the different classes of covariates on physician productivity at the provider level, we estimated the model with a random effect only, and then sequentially included shared practice resources and subsequently underlying baseline characteristics as covariates. We compared successive models on their ability to explain variation in productivity using the proportional change in variance and ability to fit data using the likelihood ratio test.

We also compared our mixed-effects model with two models commonly used in the literature to evaluate physician productivity: Generalized Estimating Equations (GEE) and Data Envelopment Analysis (DEA). Since we have longitudinal data for providers, a generalized linear model with a gamma distribution and log link was estimated using a Generalized Estimation Equation (GEE) with a compound symmetry working correlation structure (details in Supplemental Appendix). GEE has been used as a comparator to mixed-effect models when comparing individual hospital quality.[16] DEA is a non-parametric method for measuring the productivity of decision-making units, which in our case are physicians. In DEA, one develops an efficient frontier, which is defined by the most productive physicians, and evaluates individual efficiencies by their deviations from the frontier.[17] The DEA model used in this paper was designated as an output-oriented, variable returns to scale model without nondiscretionary variables.

Given the lack of gold standard comparing these methods, we used a two-step approach. First, we compared mixed-effects and GEE models based on K-fold cross validation with 5 groups. In particular, we used stratified sampling, with physician as the strata, to randomly select weeks from each physician into one of the five folds. We evaluated predictive power using the Root Square Mean Error (RSME) and Mean Absolute Error (MAE). DEA was not included in this comparison since it does not provide predictive performance.

Next, we compared the three models, as well as the crude rate of patients seen per clinic session, based on ranking of physicians by productivity. Crude rate is defined as the total number of patients seen per clinic session per week without any adjustment and is akin to RVU-based productivity measures that are currently in widespread use. To allow for a fairer comparison with our DEA model, the mixed-effects and GEE models were estimated using shared resources only. Productivity was measured by the random intercept estimate in the mixed-effects model, observed to expected (i.e. predicted) ratio for the GEE model as has been done elsewhere,[16,18] and efficiency output from the DEA model. We compared rankings of providers by productivity across all models both graphically and using Spearman correlation coefficients.

Covariate selection was carried out using the backward elimination method excluding variables with a p-value greater than 0.10. Shared practice resources were included regardless of statistical significance because they were key explanatory variables of interest. Additionally, patient age, gender, and Charlson score were kept in the final model regardless of significance based on known relationships in the literature to capture patient severity.

Mixed-effects and GEE models were estimated using SAS version 9.4 (SAS Institute Inc., Cary, North Carolina) and DEA was estimated using Open Source DEA (http://opensourcedea.org/about/). Since this work was performed for administrative purposes, it was exempt from review by our hospital's Institutional Review Board (IRB) per their policies. Regardless, data security meets the common requirements of the IRB, as data was kept on a secure server behind a firewall that is used for other physician data and performance reporting and identifiable data were only accessible by three co-authors of this manuscript.

3. Results

3.1. Unadjusted statistics

The final study sample included 4212 physician-weeks from 56 cardiologists. The median provider visits per clinical session per week was 4.7 visits (range 1.0–12.5; interquartile range of 2.7). Table 1a describes the distribution of underlying baseline characteristics (patient, scheduling, and provider characteristics) as well as their correlations with visits per clinical session. Distribution of shared practice resources across cardiologists and association with visits per clinical session is detailed in Table 1b. Higher number of visits per clinical session was associated with greater NP support (3.8 for no NP, 4.8 for shared NPs, 6.7 for a full NP, and 8.3 for multiple NPs; $p < 0.001$). In contrast, the number of visits per clinical session by level of fellow support was different with any level of fellow support (3.8 for no fellow, 4.5 for a part-time fellow, 7.2 for one full-time fellow, and 5.3 for two fellows; $p < 0.001$) in univariate analysis. Having one-third secretary FTE or less was associated with fewer visits per clinic session compared to having half, or having full or more secretary FTE (4.3 vs 5.0 or 5.4, respectively; $p < 0.001$).

Table 1a
Distribution of patient covariates by provider and correlation with visits per clinical session.

Variable	Mean	Standard deviation	Correlation with outcome[a]
Provider/scheduling characteristics			
Provider age	54.9	12.1	−0.186[b]
Number of clinical sessions/week	3.7	2.0	−0.387[b]
Percentage new visits	0.2	0.2	−0.057[c]
Patient characteristics			
Average patient age	66.39	8.85	−0.002
Median income (in 000 s)	83.67	11.95	0.046[c]
Average Charlson score	1.99	1.10	−0.088[b]
Percent male	0.61	0.21	−0.015
Percent minority	0.12	0.13	−0.047[c]
Percent non-English speaking	0.06	0.09	−0.006
Percent out-of-state	0.14	0.15	−0.037[c]
Percent Medicaid	0.02	0.06	−0.015
Percent Medicare	0.51	0.22	−0.037[c]
Percent private insurance	0.47	0.22	0.042[c]

[a] Pearson correlation coefficient with clinic visits per session and its statistical significance.
[b] Statistically significant at < 0.001 significance level.
[c] Statistically significant at < 0.05 significance level.

Table 1b
Distribution of visits per clinic session by practice resources.

Variable	# Providers	Visits per session mean (STD)	P-Value[a]
Full Sample	56	4.9 (3.5)	
NP FTE[b]			< 0.001
No NP	28	3.8 (2.4)	
Shared NP	13	4.8 (2.9)	
One NP	11	6.7 (4.0)	
> one NP	4	8.3 (4.1)	
Fellow FTE[c]			< 0.001
No fellow	30	3.8 (2.5)	
Shared fellow	7	4.5 (2.2)	
One fellow	15	7.2 (4.4)	
Two fellow	4	5.3 (3.0)	
Secretary FTE[d]			< 0.001
One-third or less	13	4.3 (2.7)	
Half	32	5.0 (3.4)	
Full or more	11	5.4 (4.2)	
Season	# Visits		0.042
Q1 2015	9351	4.5 (3.3)	
Q2 2015	11,888	5.2 (3.6)	
Q3 2015	10,627	5.0 (3.5)	
Q4 2015	11,105	5.0 (3.5)	
Q1 2016	10,069	4.8 (3.2)	
Q2 2016	10,634	4.9 (3.5)	

[a] Pearson's chi-squared test of equality of mean of the visits per clinic session across categories.
[b] Among those with shared NP, 10 providers have 0.5 FTE and 3 have 0.33 FTE. Among those with more than one NP, 3 have 2 NPs and one has 1.5 NP.
[c] Among those with shared fellow, 4 have 0.55 FTE and 3 have 0.67 FTE.
[d] One provider who has 1.5 secretary FTE is grouped as full FTE and one provider who have no secretary FTE and one who has 0.1 FTE are grouped into the one third FTE. NP = nurse practitioner, FTE = full time equivalent.

3.2. Mixed-effects model

We estimated three models sequentially. Adding shared resources to a random effect-only model significantly improved model fitness (χ^2 (9) = 54, $P < 0.001$) and drastically reduced variation in productivity among providers (proportional change in variance = 63%). Adding underlying baseline characteristics (scheduling, patient, and provider level variables) to the shared resource model slightly improved model fitness (χ^2 (12) = 89, $P < 0.001$) and only marginally increased variation in productivity (proportional change in variance = 3%).

The results of the complete mixed-effects model are presented in Table 2. Having < 1 NP FTE was associated with a 34% increase in productivity (p = 0.018), 1 NP FTE was associated with a 31% increase in productivity (p = 0.020) and greater than one NP FTE was associated with an 67% increase in productivity compared to not having an NP (p = 0.001). Having 1 fellow FTE was associated with a 28% increase in productivity compared to not having a fellow (P = 0.03). Secretary FTE was not statistically significant.

3.3. Comparison to other models

In comparing the mixed-effects and GEE models using K-fold cross validation, the mixed-effects model performed better than the GEE

Table 2
Mixed effect analysis predicting number of visits per clinical session.

Predictors	Rate ratio			P-value
	Mean	[95% CI]		
Intercept	4.40	1.96	9.88	0.0008
Quarter (ref = Q2_2016)				
1st Q 2015	0.90	0.85	0.96	0.0006
2nd Q 2015	1.06	1.01	1.10	0.019
3rd Q 2015	1.02	0.97	1.07	0.473
4th Q 2015	1.00	0.95	1.06	0.8925
1st Q 2016	0.97	0.92	1.02	0.2412
NP FTE (rel = No NP)				
< 1 FTE	1.34	1.05	1.71	0.0177
1 FTE	1.31	1.04	1.65	0.0199
> 1 FTE	1.67	1.25	2.23	0.0005
Fellow FTE (rel = No Fellow)				
< 1 FTE	1.13	0.88	1.45	0.340
1 FTE	1.28	1.02	1.49	0.0304
> 1 FTE	1.04	0.70	1.55	0.8375
Secretary FTE (rel ≤ 0.33)				
0.5 FTE	1.12	0.84	1.49	0.4489
≥1 FTE	1.06	0.71	1.57	0.7875
Payor (rel = Per Medicare)				
Percent Medicaid	0.99	0.57	1.70	0.9698
Percent private	1.03	0.77	1.38	0.8368
Other Variables				
Clinical session	0.90	0.86	0.95	< 0.0001
Average patient age	1.01	1.00	1.02	0.2074
Percent male	1.03	0.85	1.26	0.7338
Average charlson score	0.94	0.90	0.98	0.004
Percent out-of-state	0.59	0.45	0.76	< 0.001
Median income (in 000 s)	1.00	0.99	1.00	0.2306

NP = nurse practitioner, FTE = full time equivalent.

Table 3
Comparing actual number of visits per clinical session and predicted number of visits per clinical session using K-fold cross validation.

Model	GEE		Mixed-effects	
	RMSE	MAE	RMSE	MAE
K-fold_1	3.43	2.45	2.69	1.92
K-fold_2	2.84	2.07	2.54	1.79
K-fold_3	2.95	2.19	2.70	1.92
K-fold_4	3.06	2.23	2.81	1.96
K-fold_5	2.82	2.06	2.56	1.86
Average	**3.02**	**2.20**	**2.66**	**1.89**

GEE = Generalized estimating equation, RMSE = root mean squared error, MAE = mean absolute error.

model (RMSE 2.66 vs 3.02; MAE 1.89 vs 2.20, respectively; Table 3).

Fig. 1 and Table 4 show how the three models rank physician productivity. The mixed-effects model was more strongly correlated with the GEE model (Spearman 0.991) than with the DEA model (Spearman 0.692) or the crude rate (Spearman 0.545). Of note, the DEA model correlated less strongly with the crude rate (Spearman 0.637) than with the GEE model (Spearman 0.726).

In examining the results of the DEA model, we found that efficiency results were extremely sensitive to the inclusion of one provider who had 12.5 visits per session (the highest among providers). The efficiency score rankings in the DEA model changed significantly when this provider was excluded. An additional provider only saw an average of 1.9 patients per session but had an efficiency score of 1 (indicating the provider is on the frontier), which was misleading. Upon excluding these outlier observations, the correlation between the mixed-effects model and the DEA model rankings increased significantly from Spearman 0.692 to Spearman 0.755 (Table 4). Upon removing these providers from the mixed-effects model, provider rankings stayed largely similar. Taken together, these results suggest extreme sensitivity of the DEA model rankings to potential outliers.

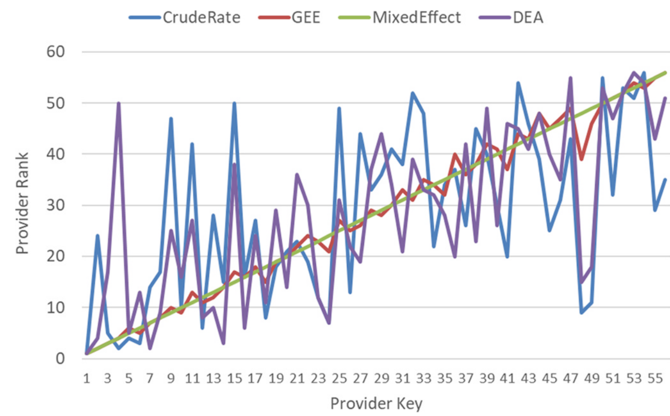

Fig. 1. Comparing provider ranking among alternative models. Pairwise Spearman coefficients among different models are 0.991 for mixed-effects vs GEE, 0.545 for mixed-effects vs crude rate, 0.692 for mixed-effects vs DEA, 0.596 for GEE vs crude rate, 0.726 for GEE vs DEA and 0.637 for crude rate vs DEA (all $p < 0.001$). Rankings go from lowest productivity to highest. (e.g. 1 refers to the provider with lowest productivity and 56 with highest productivity). GEE = generalized estimating equation, DEA = data envelopment analysis.

Table 4
Spearman correlation coefficient between different provider ranking methods with and without excluding outliers.

Models compared	All providers	Excluding outlier 1	Excluding outlier 2
Crude Rate vs DEA	0.637	0.624	0.684
Crude Rate vs GEE	0.596	0.609	0.602
Crude Rate vs Mixed-effects	0.545	0.553	0.552
GEE vs DEA	0.726	0.720	0.778
Mixed-effects vs DEA	0.692	0.687	0.755
GEE vs Mixed-effects	0.991	0.995	0.992

DEA = data envelopment analysis, GEE = generalized estimating equation.

4. Discussion

In this paper, we apply a mixed-effects methodology for evaluating individual physician clinical productivity in a novel way by incorporating aspects of practice-level resource utilization and compare it to other models for measuring physician productivity. We found that adjusting for shared practice resources reduced variation among providers by more than half, but subsequently adjusting for additional underlying baseline characteristics had only marginal impact on variation. Furthermore, we found that a mixed-effects model outperformed GEE on K-fold cross validation, and that the mixed-effects model was much less vulnerable to outliers than DEA in ranking providers according to productivity. Variation in physician productivity as measured by the random effect in our mixed-effects model represents individual productivity that is not accounted for by differences in shared group resources. As such, in conjunction with quality metrics, this can be used as a basis for adjusting compensation with the intent of enhancing value in healthcare delivery.

Our model estimates for the effects of shared resources add to the literature on shared resources and productivity. The estimated productivity increase from adding an NP varies widely from 20% to 90%.[7,19,20] The 31% increase from one additional NP in our mixed-effects model falls within this range. The ambulatory clinical productivity change from adding an additional trainee varies widely as well,[21] ranging from improvements of up to 3% to losses of up to 8% for internal medicine residents in aggregate.[22,23] We found that an additional fellow was associated with an additional 28% increase in productivity in multivariate analyses, but having less than 1 fellow FTE or greater than 1 fellow FTE was not statistically significant. Understanding whether an optimum clinician-trainee ratio balancing assistance from a trainee clinician with inefficiencies resulting from teaching time exists would be a rich area for future inquiry. These insights are important for understanding optimal allocation of resources among providers, especially as practices grow and expand.

Our methodology for measuring outpatient individual physician clinical productivity extends the existing published models of measuring physician productivity by employing mixed-effects modeling to account for shared practice resources, which is now increasingly important given recent healthcare trends. Most academic departments of medicine measure clinical productivity for compensation purposes, and 84% of these use a standardized unit, such as an RVU generated for billing purposes, which is akin to measuring productivity using a crude rate.[5] In this study, crude rate had only moderate correlation with the mixed-effects model (Spearman 0.545), suggesting that our mixed-effects model would result in substantially different physician productivity rankings compared to rankings from currently used RVU-based productivity measures. While some have augmented RVU-based metrics to account for research and teaching or to normalize physician data to account for complexities in billing codes,[13,24–26] such metrics have not yet been refined to account for shared practice resources in the medical literature. In examining physician productivity with repeated observations, several clinical studies have used GEE, though none have controlled for shared practice characteristics.[27–29] From a statistical standpoint, mixed-effects models are better suited for assessing individual physician clinical productivity than GEE models by the inclusion of a random-effect to account for individual effort. Advanced analytic models to calculate productivity such as DEA or stochastic frontier analysis developed in operations and economics literature have been applied to healthcare,[30–38] though most focus on inpatient efficiency or compare variation at aggregated levels, and few control for shared practice characteristics such as ours to measure individual physician productivity. In a landmark review of 317 papers on measuring efficiency and productivity in healthcare,[30] Hollingsworth finds that most studies use nonparametric DEA and concludes by suggesting a need to translate this work to end users such that it can be implemented in routine clinical practice. Furthermore, a review of 39 studies applying DEA in the outpatient primary care setting found that much of the literature targeted researchers and health economists and only 6 studies discussed how to put findings into practice.[31] In contrast to DEA, from a statistical standpoint, the mixed-effects model in our study enables practices to define and quantify the nature of the relationships between shared resources and output, which can provide further actionable insights on resource allocation, otherwise not be available in DEA, which only generates efficiency scores for each provider.

Our comparison of different models has practical implications for administrators wanting to measure physician outpatient productivity in the context of shared practice resources. Several studies have compared methods to measure productivity, though few have made this comparison in healthcare, much less in looking at individual physician productivity in the clinical outpatient setting.[30,39,40] One study comparing hospital productivity using different methods found a correlation between a single output constant returns to scale DEA model and a mixed-effects model ranging from 0.75 to 0.91, depending on mixed-effects model specification.[41] In contrast, our study found a lower correlation of 0.69 between the ranking of physicians by productivity calculated using the mixed-effects vs DEA models, but this improved to 0.76 after 2 outliers were removed. As has been noted elsewhere,[41–45] we found that DEA is very vulnerable to outliers, and hence care must be exercised in interpreting DEA results since outliers can be difficult to identify. Furthermore, in our study, there is relatively little difference in provider ranking between DEA and crude rate, reflecting the marginal change in clinical practice from using the DEA model over currently used RVU-based metrics for physician productivity. Additionally, we found strong correlation in ranking of providers by productivity between the mixed-effects and GEE models. In comparing GEE and mixed-effects methods based on K-score cross validation, we found that the mixed-effects model had better predictive power, arguing for its use in clinical practice over GEE. Finally, in comparing different specifications for the mixed-effects model, we found that including shared resource variables had higher impact on model fitness than subsequent inclusion of underlying baseline characteristics, and the addition of underlying baseline characteristics did not improve provider-level variance substantially. This has implications for practices where administrative data regarding shared resources may easily be available, but underlying patient-level baseline characteristic data may not. In such cases, the marginal benefit gained from adding these underlying patient-level baseline characteristics may not be worth the additional cost of obtaining this information. In our practice, in response to the results of this analyses, standard expectations for NP-physician ratios as well as utilization of clinic rooms have been established and the outpatient clinics have been centralized.

The findings of our study should be interpreted in the context of its limitations. First, our study was conducted at a single academic cardiology practice so the extent to which point estimates in the mixed-effects model are generalizable to other settings is uncertain. Nevertheless, the application of the underlying mixed-effects methodology is reproducible, so a similar approach could be replicated for other clinical settings. Second, our measure of productivity does not consider quality, which is often subjective and difficult to assess, but remains an important dimension of value-based care. Any attempt to use our method to adjust compensation should also account for context-specific quality metrics. Third, the associations between shared practice resources and physician productivity do not necessarily imply causation. As such, providing more shared practice resources to less productive physicians may not necessarily lead to the same impact on productivity. Nevertheless, the use of shared practice resources still imposes a practice-wide cost and accounting for them in physician productivity rankings may still be helpful in determining individual physician compensation. Further refining this model is a rich area for further research, especially as digital administrative data becomes more easily accessible.

In this work, we apply a risk-standardization methodology used in public reporting to account for shared resources in measuring individual physician clinical productivity and compare it to other methods for measuring physician productivity. Mixed-effects modeling appears to be superior in terms of predictive performance than GEE and far less vulnerable to outliers than DEA. By employing mixed-effects regression analysis of shared resources using otherwise easily accessible administrative data, practices can evaluate physician clinical productivity more fairly and make more informed management decisions on physician compensation and resource allocation, which can improve practice efficiency. In conjunction with context-specific quality metrics, this pragmatic methodology for measuring productivity can be used more broadly to empower practices to deliver higher-value care.

Sources of funding

None for this manuscript; Jason H. Wasfy received significant research grant support from a KL2 career development award funded by the National Institutes of Health through Harvard Catalyst (KL2 TR001100).

Disclosures

Michael K. Hidrue, Jeffrey B. Weilburg, Timothy G. Ferris, and Jason H. Wasfy received significant salary support from the Massachusetts General Physicians Organization.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.hjdsi.2019.02.001.

References

1. Welch WP, Cuellar AE, Stearns SC, Bindman AB. Proportion of physicians in large group practices continued to grow in 2009–11. *Health Aff.* 2013;32:1659–1666.
2. Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *Jama.* 2007;297:740–744.
3. Kaplan KC, Elizondo I, Schaff S. Value-based physician compensation tackling the complexities. *Healthc Financ Manag: J Healthc Financ Manag Assoc.* 2013;67:60–69.
4. Brush JE, Handberg EM, Biga C, et al. 2015 ACC health policy statement on cardiovascular team-based care and the role of advanced practice providers. *J Am Coll Cardiol.* 2015;65:2118–2136.
5. Kairouz VF, Raad D, Fudyma J, Curtis AB, Schünemann HJ, Akl EA. Assessment of faculty productivity in academic departments of medicine in the United States: a national survey. *BMC Med Educ.* 2014;14:1.
6. Duck E, DeLia D, Cantor JC. Primary care productivity and the health care safety net in New York City. *J Ambul Care Manag.* 2001;24:1–14.
7. Hurdle S, Pope GC. Improving physician productivity. *J Ambul Care Manag.* 1989;12:11–26.
8. Bank AJ, Gage RM. Annual impact of scribes on physician productivity and revenue in a cardiology clinic. *Clin Outcomes Res: CEOR.* 2015;7:489.
9. Litaker D, Mion LC, Planavsky L, Kippes C, Mehta N, Frolkis J. Physician-nurse practitioner teams in chronic disease management: the impact on costs, clinical effectiveness, and patients' perception of care. *J Interprof Care.* 2003;17:223–237.
10. Thornton J. Do physicians employ aides efficiently?: some new evidence on solo practitioners. *J Econ Financ.* 1998;22:85–96.
11. Albritton TA, Miller MD, Johnson MH, Rahn DW. Using relative value units to measure faculty clinical productivity. *J Gen Intern Med.* 1997;12:715–717.
12. Hsiao WC, Braun P, Dunn DL, et al. An overview of the development and refinement of the resource-based relative value scale. The foundation for reform of U.S. physician payment. *Med Care.* 1992;30:NS1–NS12.
13. Coleman DL, Moran E, Serfilippi D, et al. Measuring physicians' productivity in a Veterans Affairs medical center. *Acad Med.* 2003;78:682–689.
14. Wasfy JH, Borden WB, Secemsky EA, McCabe JM, Yeh RW. Public reporting in cardiovascular medicine accountability, unintended consequences, and promise for improvement. *Circulation.* 2015;131:1518–1527.
15. Hox J. *Multilevel Analysis: Techniques and Applications*. 2nd ed. New York: Routledge; 2010.
16. AHRQ. *AHRQ Quality Indicators: Risk Adjustment and Hierarchical Modeling Approaches*. 2007; 2007 (Accessed 30 January 2019). 〈https://www.qualityindicators.ahrq.gov/Downloads/Resources/Publications/2012/AHRQ_RAHM_Workgroup_Technical_Report_Final.pdf〉.
17. Cooper WW, Seiford LM, Tone K. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* Second Edition US: Springer; 2007https://doi.org/10.1007/978-0-387-45283-8.
18. Shwartz M, Ash A, Pekoz E. Risk adjustment and risk-adjusted provider profiles. *Int J Healthc Technol Manag.* 2006;7:15–42.
19. Roblin DW, Howard DH, Becker ER, Kathleen Adams E, Roberts MH. Use of midlevel practitioners to achieve labor cost savings in the primary care practice of an MCO. *Health Serv Res.* 2004;39:607–626.
20. Larson EH, Palazzo L, Berkowitz B, Pirani MJ, Hart LG. The contribution of nurse practitioners and physician assistants to generalist care in Washington State. *Health Serv Res.* 2003;38:1033–1050.
21. Wynn BO, Smalley R, Cordasco KM. Does it cost more to train residents or to replace them?: a look at the costs and benefits of operating graduate medical education programs. *RAND Health Q.* 2013;3(3).
22. DeLia D, Cantor JC, Duck E. Productivity vs. training in primary care: analysis of hospitals and health centers in New York City. *Inq: J Health Care Organ Provis Financ.* 2002;39:314–326.
23. Johnson T, Shah M, Rechner J, King G. Evaluating the effect of resident involvement on physician productivity in an academic general internal medicine practice. *Acad Med.* 2008;83:670–674.
24. Mezrich R, Nagy PG. The academic RVU: a system for measuring academic productivity. *J Am Coll Radiol.* 2007;4:471–478.
25. Abouleish AE, Zornow MH, Levy RS, Abate J, Prough DS. Measurement of individual clinical productivity in an academic anesthesiology department. *J Am Soc Anesthesiol.* 2000;93:1509–1516.
26. Arenson RL, Lu Y, Elliott SC, Jovais C, Avrin DE. Measuring the academic radiologist's clinical productivity: applying RVU adjustment factors. *Acad Radiol.* 2001;8:533–540.
27. Joseph JW, Davis S, Wilker EH, et al. Modelling attending physician productivity in the emergency department: a multicentre study. *Emerg Med J.* 2018;35:317–322.
28. Fairchild DG, McLoughlin KS, Gharib S, et al. Productivity, quality, and patient satisfaction: comparison of part-time and full-time primary care physicians. *J Gen Intern Med.* 2001;16:663–667.
29. Conrad D, Fishman P, Grembowski D, et al. Access intervention in an integrated, prepaid group practice: effects on primary care physician productivity. *Health Serv Res.* 2008;43:1888–1905.
30. Hollingsworth B. The measurement of efficiency and productivity of health care delivery. *Health Econ.* 2008;17:1107–1128.
31. Pelone F, Kringos DS, Romaniello A, Archibugi M, Salsiri C, Ricciardi W. Primary care efficiency measurement using data envelopment analysis: a systematic review. *J Med Syst.* 2015;39:1–14.
32. Zuckerman S, Hadley J, Iezzoni L. Measuring hospital efficiency with frontier cost functions. *J Health Econ.* 1994;13:255–280.
33. Kuntz L, Scholtes S, Vera A. Incorporating efficiency in hospital-capacity planning in Germany. *Eur J Health Econ.* 2007;8:213–223.
34. Wagner JM, Shimshak DG, Novak MA. Advances in physician profiling: the use of DEA. *Socio-Econ Plan Sci.* 2003;37:141–163.
35. Chilingerian JA. Evaluating physician efficiency in hospitals: a multivariate analysis of best practices. *Eur J Oper Res.* 1995;80:548–574.
36. Rosenman R, Friesner D. Scope and scale inefficiencies in physician practices. *Health Econ.* 2004;13:1091–1116.
37. Andes S, Metzger LM, Kralewski J, Gans D. Measuring efficiency of physician practices using data envelopment analysis. *Manag Care.* 2002;11:48–54.
38. Giuffrida A. Productivity and efficiency changes in primary care: a Malmquist index approach. *Health Care Manag Sci.* 1999;2:11–26.
39. Gannon B. Testing for variation in technical efficiency of hospitals in Ireland. *Econ Social Rev.* 2005;36:273.
40. Cubbin J, Tzanidakis G. Regression versus data envelopment analysis for efficiency measurement: an application to the England and Wales regulated water industry. *Uti Policy.* 1998;7:75–85.
41. Siciliani L. Estimating technical efficiency in the hospital sector with panel data. *Appl Health Econ Health Policy.* 2006;5:99–116.
42. Sexton TR, Silkman RH, Hogan AJ. Data envelopment analysis: critique and extensions. *New Dir Program Eval.* 1986;1986:73–105.
43. Banker RD, Chang H. The super-efficiency procedure for outlier identification, not for ranking efficient units. *Eur J Oper Res.* 2006;175:1311–1320.
44. Chen W-C, Johnson AL. A unified model for detecting efficient and inefficient outliers in data envelopment analysis. *Comput Oper Res.* 2010;37:417–425.
45. Pastor JT, Ruiz JL, Sirvent I. A statistical test for detecting influential observations in DEA. *Eur J Oper Res.* 1999;115:542–554.