

Weiss Marianne (Orcid ID: 0000-0003-4217-9822)
Yakusheva Olga (Orcid ID: 0000-0001-5501-1397)

Nurse Staffing Machine Learning

1

Non-linear association of nurse staffing and readmissions uncovered in machine learning analysis

Olga Yakusheva, PhD
Department of Systems, Populations, and Leadership
School of Nursing
Department of Health Management and Policy
School of Public Health
University of Michigan,
400 North Ingalls, Suite 4343
Ann Arbor, Michigan, USA
yakush@umich.edu

James T. Bang, PhD
Department of Economics
St. Ambrose University
Davenport, Iowa, USA

Ronda G. Hughes, PhD, RN, NEA-BC, FAAN
Center for Nursing Leadership
College of Nursing
University of South Carolina,
Columbia, South Carolina, USA

Kathleen L. Bobay, PhD, RN, NEA-BC, FAAN
Marcella Niehoff School of Nursing
Loyola University Chicago
Chicago, Illinois, USA

Linda Costa, PhD, RN, NEA-BC
School of Nursing
University of Maryland,
Baltimore, Maryland, USA

Marianne E. Weiss, DNSc, RN
College of Nursing
Marquette University,
Milwaukee, Wisconsin, USA

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/1475-6773.13695](https://doi.org/10.1111/1475-6773.13695)

This article is protected by copyright. All rights reserved.

Abstract

Objective: Several studies of nurse staffing and patient outcomes found a curvilinear or U-shaped relationship, with benefits from additional nurse staffing diminishing or reversing at high staffing levels.

This study examined potential diminishing returns to nurse staffing and the existence of a “tipping point”, or the level of staffing after which higher nurse staffing no longer improves and may worsen readmissions.

Data Sources/ Study Setting: The READI study database of over 130,000 adult (18+) inpatient discharges from 62 medical, surgical, and medical-surgical (non-critical care) units from 31 US hospitals during 10/2014-3/2017.

Study Design: Observational cross-sectional study using a fully non-parametric random forest machine learning method. Primary exposure was nurse hours per patient day (HPPD) broken down by registered nurses (non-overtime and overtime) and non-licensed nursing personnel. The outcome was 30-day all-cause readmission. Partial dependence plots were used to visualize the pattern of predicted patient readmission risk along a range of unit staffing levels, holding all other patient characteristics and hospital and unit structural variables constant.

Data Collection/Extraction methods: Secondary data analysis. Missing values were imputed using the missing forest algorithm in R.

Principal Findings: Partial dependence plots were U-shaped showing the readmission risk first declining and then rising with additional staffing. The tipping points were at 6.95 and 0.21 HPPD for registered nurse staffing (non-overtime and overtime) and 2.91 HPPD of non-licensed nursing personnel. The U-shaped association was consistent with diminishing returns to nurse staffing.

Conclusions: this observational study showed that incremental gains in readmission reduction from additional nurse staffing diminish and could reverse at high staffing levels. If confirmed in future causal analyses across multiple outcomes, accompanying investments in infrastructure and human resources may be needed to maximize nursing performance outcomes at higher levels of nurse staffing.

Keywords

Readmissions, Machine Learning, Nursing, Unit Staffing, Diminishing Returns

What is known on this topic:

- Several studies have reported positive associations between hospital nurse staffing and patient outcomes that taper off or reverse at high levels of nurse staffing.
- Few explanations for these findings and no theoretical framework have yet been offered.
- No studies have examined the association of hospital nurse staffing with hospital readmissions.

What this study adds:

- The study demonstrates a novel application of machine learning methods for uncovering a non-linear association where traditional parametric regression may not be as effective.
- The study finds tapering off and reversal of the association of nurse staffing with readmissions.
- The study's findings are consistent with diminishing returns on readmissions from additional nurse staffing.

INTRODUCTION

When the United States (US) Center for Medicare and Medicaid Services implemented the Hospital Readmission Reduction Program (HRRP) in 2012, US hospitals intensified efforts to reduce readmissions. A small number of analyses have pointed to the potential benefits of monitoring and correcting low nurse staffing levels to reduce readmissions [1-3] and avoid HRRP penalties,[4] adding to a broadly accepted body of evidence that higher nurse staffing levels are associated with improved quality and safety outcomes (mortality, failure to rescue, and other patient outcomes [5-12]) and patient satisfaction with hospital nursing care.[12]

Yet, depending on the skill mix of nursing personnel and the existing level of nurse staffing in an organization, increasing the amount of nursing hours assigned to patient care may not always be uniformly beneficial. In a recent study from the United Kingdom, a uniform dose-response reduction in mortality on patient wards was associated with higher professional nurse staffing, but a U-shaped relationship with nursing assistant staffing—at low nursing assistant staffing levels, mortality decreased with additional nursing assistant staffing, but at high staffing levels, additional nursing assistant staffing was associated with higher mortality.[7] This was not the first time a U-shaped pattern was reported – nearly two decades ago, researchers reported higher rates of medication errors as registered nurse staffing increased past a certain level [13] and in a subsequent multihospital study, a robust and consistent tapering-off pattern of diminishing returns to increased nurse staffing was found across multiple outcome measures (mortality, hospital-acquired infections, pressure ulcers).[14] Two systematic reviews and meta-analyses noted a ‘curvilinear’ relationship with benefits of increased registered nurse staffing tapering off and even reversing at higher staffing levels in the US [9] and internationally

[15]. While no such evidence has ever been reported for readmissions, a curvilinear pattern was found for nurse staffing and patient satisfaction with hospital nursing care, and specifically with nurse communication, discharge information, and care transitions, in a large study of US hospitals.[12] Most recently, Needleman and Shekelle [16] proposed that the tapering-off effect may be a natural consequence of “diffusion of effort or responsibility” that may occur at high nurse staffing levels, and conjectured that this phenomenon may occur across many types of patient outcomes.

No studies to date have examined the underlying empirical patterns behind the observed associations between nurse staffing and readmissions across a range of staffing levels. Therefore, the objective of this study was to examine the pattern of the association between nurse staffing and hospital readmissions, specifically examining the possibility that additional nurse staffing has diminishing returns, with the existence of a “tipping point” after which adding more nurse staffing no longer improves (and may worsen) readmissions. Evidence of a non-linear association between nurse staffing and readmissions could inform future causal studies of the relationship between staffing and readmission, toward the ultimate goal of novel practice and policy recommendations for maximizing the value-added contribution of nursing to patient and cost outcomes and hospital performance.

THEORETICAL FRAMEWORK

Most investigations of the relationship between nurse staffing and patient outcomes are grounded in Donabedian’s structure-process-outcomes model.[17] The model views patient outcomes as the result of many organizational structural factors catalyzed during the process of

care delivery. Donabedian defines organizational structure as “the settings in which care is provided and the instrumentalities of which care delivery is a product”. Nurse staffing is but one of many structural factors that jointly comprise organizational structure; others include the availability and qualifications of other clinical staff, adequacy of facilities and equipment, and the structure and operationalization of programs that support and direct the provision of care.[17] An extension [18] of Donabedian’s model also identifies patient characteristics (specifically those directly contributing to the likelihood of an outcome, such as clinical risk factors and sociodemographic characteristics) as an additional input into the care delivery process.

In the field of economics, the theory of production [19] also views care delivery as a production process that utilizes healthcare resources as inputs to produce patient outcomes. Economic production theory categorizes structural inputs into two broad types: factors of production (“the who”) and technology (“the how”). Factors of production are the inputs that an organization employs to produce an output, and they are further categorized into labor and capital inputs. The labor input category concerns with the quantity (e.g., number of working hours, number of full-time equivalents) and the quality (e.g., education, experience, expertise) of the human resources employed by the organization (nurses, physicians, etc). The capital input category refers to the hospital’s built environment and equipment, administrative and other support systems, type and organization of the electronic health records (EHR), and other non-labor resources than enable employees to deliver care. Technology encompasses all organizational rules and norms prescribing *how* the labor and capital inputs ought to interact during the care delivery process; it refers to the established care delivery components within

the organization that form the organizational standard of care (such as protocols, programs, practice guidelines). All structural factors (labor and capital inputs, technology) are part of a joint production process. As such, the incremental productivity of any single factor is determined by the quantity of all other inputs and the technology used in the production process. As such, the contribution of the nursing input to outcomes is not static, but rather it can be enhanced by increasing the availability of other types of labor, capital and technology relevant to the delivery of nursing care.

Guided by these parallel theoretical frameworks, we conceptualized a patient's likelihood of a readmission as an outcome of the process of care delivery that uses an organization's labor and capital resources and the care delivery technology established in the organization. Doing so allowed us to view the relationship between readmissions and nurse staffing as a partial derivative, or partial dependence, of readmissions on nurse staffing specifically, while holding all other structural inputs (other types of labor, capital, and technology) constant as observed in the data.

Table 1 parallels the constructs and definitions of Donabedian's structure-process-outcome model, as expanded by Mitchell et al. (1998), [18] and the economic theory of production, [19] and shows the types of variables that we used as empirical measures for each construct.

METHODS

Approach

To examine the relationship between nurse staffing and readmissions, this secondary analysis applied a machine learning (ML) approach [20] to a large multi-hospital data set of adult inpatient discharges. ML methods “learn” data patterns from the data itself, without imposing restrictions on the functional form of the relationship among variables. Therefore, ML methods are capable of exposing the true data pattern behind the association of nurse staffing and readmissions, be it linear, curvilinear, a U-shape, or virtually any other potentially unknown pattern.

A unique feature of ML is that, being purely atheoretical and non-parametric, the association between the dependent variable (readmission) and an explanatory variable (e.g. staffing) is revealed as a Partial Dependence Plot (PDP). [20, 21] A PDP shows how the predicted probability of readmission varies along the observed range of nurse staffing in the sample, while holding all other variables constant as observed. In constructing a PDP, neither the direction nor the magnitude of the association between readmissions and staffing (traditionally measured by a regression coefficient) is derived as a parameter; instead the local marginal effect (the derivative of readmission likelihood with respect to staffing) is represented by the slope of the PDP at each staffing level and can vary in sign and magnitude along the range of observed staffing values.

A limitation of all current ML methods is the inability to tackle unobserved and unmeasured variables and potential reverse causality. For example, a hospital’s decisions on level of or budgeted nurse staffing are likely linked to the patient case-mix (hospitals treating

more complex patients tend to have higher nurse staffing levels), which could create a positive pattern of association between staffing and readmissions in the absence of a causal pathway. In an attempt to partially mitigate endogeneity, we used a unique proprietary data set collected by a team of economists and clinicians to examine the impact of nursing care on readmissions. As described in the Data section below, the database includes all relevant structural input variables available in the EHR; a comprehensive set of technology variables capturing organizational practices relating to discharge, and an extensive set of patient characteristics. Nevertheless, our results should not be interpreted as causal, which is one of the overt limitations of our approach.

Data

The READI (Readiness Evaluation And Discharge Interventions) study [NCT01873118; 20] was a multi-hospital cluster-randomized clinical trial, approved by [Blinded for Review] Institutional Review Board to test the impact of unit-based implementation of a discharge readiness assessment intervention on readmission and emergency department use. The study team recruited hospitals through a call for interest to Magnet-designated organizations coordinated by the American Nurse Credentialing Center of the American Nurses Association. The sample included nearly 145,000 adult (18+) inpatients discharged to home from 66 general medical, surgical, or medical-surgical units. Thirty-one US and 2 Saudi Arabia Magnet hospitals participated in the study between October 2014 and March 2017. Each hospital contributed 2 units that were randomly assigned to intervention or usual care control conditions. Data collection followed Donabedian's structure-process-outcomes model [17] and the economic theory of production.[19]

The intervention involved augmenting the existing discharge standard of care protocol with a formalized discharge readiness assessment to inform patient preparation for discharge. The study was conducted in four phases over 19-months including a baseline phase followed by three intervention phases, each testing a different version of the discharge readiness assessment protocol. The intervention had high fidelity with more than 90% of nurses trained and more than 70% of patients assessed per study protocol on intervention units. Despite some evidence of potential effectiveness in units with high baseline readmission rates, the intervention was overall not effective in changing readmissions. We treated the READI intervention as a unit-specific discharge process variable in the study along with other variables in the “technology” category.

Sample

We used 137,778 adult (18+) inpatients discharged from 62 units in 31 U.S. hospitals in the READI study. Following the scientific standards for ML methods, [20] we split the sample into a 70% learning sample of 96,444 observations and a 30% testing sample of 41,334 observations.

Study variables and measures

The outcome was 30-day all-cause same-hospital readmission, measured as a dichotomous variable (1=patient had at least one readmission within 30 days post-discharge; 0=no readmissions). The primary exposure was unit nurse staffing, measured in hours per patient day (HPPD) with three continuous range variables: direct-care registered nurse non-overtime (RN non-OT HPPD), RN overtime (RN-OT HPPD), and unlicensed nursing personnel (Non-RN HPPD). Nurse staffing was reported monthly by each study unit for a total of 1,178

unit-month observations (62 units x 19 months), and linked to patients by the discharge month.

We included three READI intervention features: a dichotomous variable for the study units' assignments (1, if randomized to intervention condition; 0, if randomized to usual standard of care control condition), a dichotomous variable for the study patients' discharge protocol (1 if discharged per study protocol on an intervention unit; 0 if discharged not per study protocol or if discharged from a control unit); and a categorical variable for the study month at discharge (1 through 19). Structural input variables, technology, and patient characteristics are included in

Table 1.

Statistical analyses

Model selection and performance: Using the learning sample and starting with the full set of structural inputs and patient variables available in the READI data set, we performed a random forest recursive feature elimination process using the caret package V6.0-84 [22] with 10-fold cross-validation, 500 trees, and a minimum node size of 6 observations. We chose a random forest model as the best tool for learning about the relationships between the outcome and individual variables, given its robust classification power and easily interpretable learning mechanism.[23-25] A random forest is a set of decision trees that each use explanatory variables as logical "if/then" splits leading down the paths, or 'branches,' ending either in a predicted readmission or in no predicted readmission. Randomly reshuffling the values of each of the explanatory variables one at a time while holding all other variables constant at the values observed for each patient in the sample, the algorithm sorts through a large number of different combinations of the variables and selects a model with the highest predictive performance. We measured predictive performance by two statistics: the random-

chance-adjusted proportion of correctly predicted outcomes, or Cohen's Kappa, and the area under the curve (AUC). Cohen's Kappa (κ) varies between 0 and 1 and has a similar interpretation to the traditional R-squared of the regression.[29] We also evaluated the importance of the three nurse staffing variables relative to other explanatory variables in our model. We measured a variable's importance by the size of reduction in the model's Cohen's Kappa ($\Delta\kappa$) when the variable's values were reshuffled at random, with larger $\Delta\kappa$'s indicating more important variables. For our final model, we reported the $\Delta\kappa$ and AUC in both the learning and the testing samples. All other results were presented for the testing sample only (results in the learning sample were similar, see the sensitivity analysis section).

Sample descriptive characteristics: Once the variables selected into the model were determined, we calculated sample descriptive statistics using counts and sample proportions for all categorical variables, means and interquartile ranges (IQRs) for continuous unit-level and hospital-level variables, and means and standard deviations for continuous patient-level variables.

Partial dependence plots: We built partial dependence plots (PDPs) for the three nurse staffing variables as predictors of readmission in the learning sample, using a 100-point evenly spaced grid from the minimum value through the observed range of each variable. We then calculated 95% confidence bounds using the standard errors of the point predictions obtained after 100 bootstrap replications for each point.

Missing data: There were no missing values in the outcome variable. The nurse staffing variables were not reported by two units during the first study month, for a total of 236 patient observations (0.17%). Among the unit and hospital variables, missing data included: unit case

mix index (12,508, 9.08% of patient observations), unit RN experience (10,180, 7.39% of observations), and unit nurse certification (6,848, 4.97% of patient observations). These missing values were imputed using the missing forest algorithm in the R statistical computing environment.[22-24] Among the patient characteristics, several categorical variables had values coded as “unknown” in the hospitals’ EHR (see **Table 2**); the unknown category was preserved in the analyses. No continuous patient variables had missing values.

RESULTS

Model selection and performance: From the full set of 141 variables in the READI data, ML selected a subset of 60, including the three nurse staffing variables (**Supplemental File: Figure A**). The predictive accuracy of the final model was $\kappa=0.24$ and $AUC=0.99$ in the learning sample, and $\kappa=0.13$, $AUC=0.71$ in the testing sample. (**Table 2**)

In variable importance analysis, patient characteristics as a group were most important in predicting readmissions (joint $\Delta\kappa=0.23$). The three nurse staffing variables as a group (joint $\Delta\kappa=0.04$) were more important for predicting readmissions than other unit-specific structure variables, but less important than hospital characteristics (joint $\Delta\kappa=0.14$). Individually, the nurse staffing variables ranked 15th (RN OT HPPD), 17th (RN Non-OT HPPD), and 19th (Non-RN HPPD) among the 60 variables in the model, and 1st, 2nd, and 7th among the 32 labor, capital, and technology variables. Among the three READI study design features, unit assignment (intervention, control) and patient treatment status (per protocol, usual care) contributed very minimally to readmissions. (**Supplemental File: Figure A**)

Sample descriptive characteristics: The readmission rate was 12.2%. Patient characteristics of the sample are presented in **Table 3**; there were no significant differences

between the learning and testing samples. The 62 study units were staffed with 10.3 hours of nursing care per day (HPPD) including 6.77 RN Non-OT HPPD, 0.20 RN OT HPPD, and 3.35 Non-RN HPPD. The proportion of BSN-prepared nurses was 68% and the average nurse experience on the unit was 6 years. (**Table 4**).

Partial Dependence Plots: The PDPs for nurse staffing variables (RN non-OT HPPD, RN OT HPPD, non-RN HPPD) revealed a common quasi-parabolic data pattern with readmission likelihood first falling as staffing increased from the lowest staffing levels, reaching a minimum point, and then increasing as staffing levels increased. (**Figure 1**). The tipping points were observed at 6.95 HPPD of RN non-OT staffing, 0.21 HPPD of RN OT staffing, and 2.91 HPPD Non-RN staffing.

In sensitivity analyses, the quasi-parabolic patterns were also evident: 1) in the testing data sample, supporting the robustness of the model in out-of-sample performance (**Supplemental File: Figure B**); 2) using restricted models that eliminated various features, reducing concerns about overfitting (**Supplemental File: Figure C**); 3) in a subsample of patients discharged from control units, eliminating potential confounding from the intervention (**Supplemental File: Figure D**); and 4) after case-wise deletion, instead of imputation of missing data prior to estimation (**Supplemental File: Figure E**).

DISCUSSION

In our sample of medical, surgical, and medical-surgical units in 31 US Magnet hospitals, we found that hospital readmissions are related to nurse staffing via a robust U-shaped data pattern consistent with a tipping point past which additional staffing was associated with

increasing readmission rates. Adding to several prior studies reporting diminishing returns to nurse staffing,[7, 9, 12-15] our findings suggest that when it comes to optimizing delivery of care for maximum outcomes, “more is better” approach may not be a universally applicable principle to guide organizations in workforce planning.

From the lens of the economic theory of production, the observed U-shaped relationship between staffing and readmissions can be explained by the law of diminishing returns.[19, 26] It states that the production of additional output will decrease as more of a single factor of production (e.g., nursing labor) is incrementally added while the amounts of all other factors (other types of labor, capital) and technology stay the same. The law of diminishing returns is a direct derivative from the economic theory of production: because the contribution of any one input, such as labor, depends on the amounts of the other inputs and technology, adding more labor alone, without simultaneously increasing the levels of the other inputs or improving technology, will eventually diminish the productivity (or returns) of labor. For example, as more auto workers are hired by an automobile manufacturer, without also modifying the production floor for safety, installing additional equipment, and hiring more training and management personnel, the productivity of the expanding workforce will eventually decline.

Applied to nursing, the law of diminishing returns predicts that increasing a nursing unit’s staff will lead to initial performance gains. Eventually, however, economic theory predicts that without additional investments in all relevant structural factors that enable the delivery of high-quality care by nurses, adding more nursing hours alone will first reduce, then eliminate, and may eventually even reverse the initial productivity gains. One explanation could be what

Needleman and Shekelle called 'the diffusion of effort and responsibility'[16]—simply increasing the number of direct care nurses may lead to missed, duplicative or low-value nursing care that fails to improve and may even worsen outcomes.[27, 28] Specific to readmissions, diffusion of effort and responsibility may manifest itself as poor nurse-patient communication during discharge planning, coordination, and teaching[12] thus inhibiting a successful patient transition from hospital to home and increasing readmissions.[29] This conceptual framework and our results align with the previous studies showing tapering off and a U-shaped relationship between nurse staffing and adverse patient outcomes.[7, 12-15]

The downward-trending part of the U-shape is consistent with a large body of staffing literature reporting lower rates of readmission and other adverse outcomes with increasing RN staffing.[1-11] Higher nurse staffing and lower patient-nurse ratios can afford more direct nursing care time for assessment of patients' readiness for discharge, discharge teaching, and discharge coordination, thus reducing risk of readmission.[29] Higher nurse staffing can also contribute to reducing readmissions indirectly, by affording more time for professional development and unit-based nursing governance, which can improve nurse job satisfaction and reduce burnout and turnover.[30, 31] For units with nurse staffing levels below their tipping point, retention and recruitment of a larger unit nurse workforce can be foundational to reducing readmissions.

It is important, however, not to wrongfully infer that the observed tipping point is the optimal staffing point for benchmarking unit staffing levels; nor does our study imply that relatively well-staffed units, those on the upward sloping part of the U-shaped pattern, should be cutting back on nurse staffing. Firstly, our observational study design limits causal

interpretation of our findings. More importantly, even if our findings truly reflect a causal diminishing returns relationship, the tipping point would occur, theoretically, only when the organization fails to deploy other structural resources to support a larger nursing staff, thus effectively resulting in an inefficient substitution of nurse staffing for other hospital inputs (commonly known as ‘plugging a hole’). Therefore, to derive continuing productivity gains from nurse staffing, well-staffed organizations should be not reducing nursing staff, but instead increasing investments in other structural variables (factors of production and technology) that enable nurses to deliver high-quality discharge care. Depending on the context of each specific organization, these may include an increased supply of labor (e.g., shift managers or discharge coordinators, planners, flow coordinators, expeditors), capital (e.g., an EHR system with capacity for real-time aggregation of discharge-relevant information from multiple entry points), or technology (e.g., implementation of new organizational processes to improve communication and care coordination, such as interdisciplinary team discharge rounds). Currently, little is known about various organizational approaches to preventing and reversing diminishing returns; this area of future research presents an exciting new opportunity for informing continued outcomes improvement in already well-performing organizations.

Our study was one of the first to examine the importance of nurse staffing with other structural and patient variables interacting in a complex non-parametric model. Not surprisingly, patient characteristics were the most important predictors of patient readmission risk in our study. However, nurse hours per patient day had the highest predictive association among most other structural factors, including nurse skill mix, education, experience and expertise, and hospital- and unit-specific discharge care variables. While a patient’s risk of

readmission attributable to patient-specific factors may be difficult to identify and modify, nurse staffing strategies are well within an organization's domain of influence and should continue to be a focus of organizational outcomes improvement efforts.

Our study was the first to apply ML methods to study nurse staffing and readmissions. To date, empirical studies of the relationship between nurse staffing and patient outcomes (including readmissions) have been performed using a parametric regression analysis approach, which requires the regression equation to be chosen by the researcher prior to estimation process. Parametric regression binds the data to a pre-conceived (by the researcher) notion of the relationship, then finds the coefficients, or parameter estimates, that best support the chosen equation in a given set of data. Previous studies that reported diminishing returns to nurse staffing were designed a priori to look for particular non-linear shapes using either a piece-wise linear regression (i.e., a "V") [13] or a higher-order polynomial for staffing (i.e., a "U") [7, 12, 14]. Once chosen, the parametric methods in each of these studies were only able to confirm or reject the shape specified a priori by the researchers, and therefore unable to reveal any other data patterns. By not constraining our data to any functional form prior to estimation, our findings provide further credence to the idea that the previously reported U-shapes were likely real and not an artifact of the previously selected parametrizations. Interestingly, when we attempted to replicate the previously used parametric approaches in our data, (**Supplemental File: Figures F1-F3.**) there was little agreement regarding the shape of the pattern between methods and types of nurse staffing. One explanation is the possibility that the underlying empirical pattern in our data was not consistent with either a cubic or a linear spline model, resulting in poor fit. The other possibility lies in the different ways that

parametric regression and ML methods form predictions—while traditional regression models calculate predictions by setting all over covariates at their means, ML predictions are simulated while keeping other covariates at their observed values. [21] The relative performance of ML methods versus parametric and semi-parametric (e.g. quantile, fixed effects) regression should be examined in future studies.

Our study had a number of limitations. First, like all observational studies, our study is subject to confounding from unobserved variables and potential reverse causality. Although our study used one of the largest, richest sources of data collected specifically for analysis of readmissions in a structure-process-outcomes framework, and although the findings are robust across samples and in models with different sets of features in sensitivity analyses, our study design does not allow for causal interpretation. Second, although the READI intervention was modeled in our analysis as a unit-specific technology variable (similarly to how we accounted for all other hospital and unit discharge practices), and although our findings are robust in a subsample of patients discharged from control units in sensitivity analyses, it is difficult to know for sure to what extent our results apply to an intervention-independent sample. Third, although our sample has similar nurse staffing levels[32, 33] and readmission rates[34] to national studies of US hospitals (including Magnet hospitals [33]), in general Magnet hospitals tend to be larger, have higher levels of nurse staffing, invest more in nurse staff development and education, and deliver higher-quality care including discharge care [33, 35-37], further limiting the generalizability of our findings. Fourth, we estimated the overall association of staffing with readmissions across 31 hospitals. Even though the actual tipping points in HPPD are likely context and patient-population specific, the ML method we demonstrated can be

applied across contexts. Fifth, we only examined a single patient outcome - readmission; further research is needed to test whether these findings are equally true for other outcomes. Last, we studied 3 nurse staffing variables (RN non-OT HPPD and RN OT HPPD, and non-RN HPPD) separately, and the association of each with readmissions was obtained holding the other two constant. Future studies should examine nurse staffing variables as dynamic and interdependent to fully understand the optimal staffing strategy to produce desired patient outcomes.

CONCLUSION

While the idea of diminishing returns to labor is intuitive to most economists, it has not yet influenced the current mental models of healthcare policy makers or administrators who tend to subscribe to linear 'more is better' thinking. Our findings suggest that increasing nurse staffing alone may not always yield continuing improvements in readmissions; after a point, accompanying investments in infrastructure and human resources may be needed to support further performance improvement and outcome gains. In complex health care delivery systems where relationships among staffing variables and patient outcomes are interdependent, subject to organization-specific factors, and are not directly observed, ML methods may offer an advantage of exposing the tipping point and informing proactive organizational action to support continued high returns from nurse staffing to patient outcomes.

Funding: The READI study, the data source for this study, was commissioned by the American Nurse Credentialing Center (ANCC) following a competitive application process. Participating

hospitals paid a fee to the ANCC to participate. The ANCC had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Giuliano KK, Danesh V, Funk M. The Relationship Between Nurse Staffing and 30-Day Readmission for Adults With Heart Failure. *J Nurs Adm.* 2016;46(1):25-9.
2. McHugh M, Ma C. Hospital nursing and 30-day readmissions among Medicare patients with heart failure, acute myocardial infarction, and pneumonia. *J Nurs Adm.* 2013;43(10 Suppl):S11-8.
3. Weiss ME, Yakusheva O, Bobay KL. Quality and cost analysis of nurse staffing, discharge preparation, and postdischarge utilization. *Health Serv Res.* 2011;46(5):1473-94.
4. McHugh MD, Berez J, Small DS. Hospitals with higher nurse staffing had lower odds of readmissions penalties than hospitals with lower staffing. *Health Aff (Millwood).* 2013;32(10):1740-7.
5. Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *Jama.* 2002;288(16):1987-93.
6. Aiken LH, Sloane DM, Bruyneel L, Van den Heede K, Griffiths P, Busse R, et al. Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. *Lancet.* 2014;383(9931):1824-30.
7. Griffiths P, Maruotti A, Recio Saucedo A, Redfern OC, Ball JE, Briggs J, et al. Nurse staffing, nursing assistants and hospital mortality: retrospective longitudinal cohort study. *BMJ Qual Saf.* 2019;28(8):609-17.
8. Needleman J, Buerhaus P, Pankratz VS, Leibson CL, Stevens SR, Harris M. Nurse staffing and inpatient hospital mortality. *N Engl J Med.* 2011;364(11):1037-45.
9. Kane RL, Shamliyan TA, Mueller C, Duval S, Wilt TJ. The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. *Med Care.* 2007;45(12):1195-204.
10. Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. *N Engl J Med.* 2002;346(22):1715-22.
11. Needleman J, Buerhaus PI, Stewart M, Zelevinsky K, Mattke S. Nurse staffing in hospitals: is there a business case for quality? *Health Aff (Millwood).* 2006;25(1):204-11.
12. Opiel EM, Young GJ. Nurse Staffing Patterns and Patient Experience of Care: An Empirical Analysis of U.S. Hospitals. *Health Serv Res.* 2018;53(3):1799-818.
13. Blegen MA, Goode CJ, Reed L. Nurse staffing and patient outcomes. *Nurs Res.* 1998;47(1):43-50.
14. Mark BA, Harless DW, McCue M, Xu Y. A longitudinal examination of hospital registered nurse staffing and quality of care. *Health Serv Res.* 2004;39(2):279-300.
15. Lankshear AJ, Sheldon TA, Maynard A. Nurse staffing and healthcare outcomes: a systematic review of the international research evidence. *ANS Adv Nurs Sci.* 2005;28(2):163-74.
16. Needleman J, Shekelle PG. More ward nursing staff improves inpatient outcomes, but how much is enough? *BMJ Qual Saf.* 2019;28(8):603-5.
17. Donabedian A. Evaluating the quality of medical care. 1966. *Milbank Q.* 2005;83(4):691-729.
18. Mitchell PH, Ferketich S, Jennings BM. Quality health outcomes model. American Academy of Nursing Expert Panel on Quality Health Care. *Image J Nurs Sch.* 1998;30(1):43-6.

19. Craig C, Harris R. Total Productivity Measurement at the Firm Level. *Sloan Management Review*. 1973;Spring 1973:13–28.
20. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. New York, NY: Springer; 2013
21. Biecek P, S M. DALEX: Descriptive mACHine Learning EXplanations. 2019.
22. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, et al. *Caret: Classification and Regression Training*. R Package Version 6.0-84. 2019 [Available from: <https://CRAN.R-project.org/package=caret>.
23. Kong Y, Yu TA. Deep neural network model using random forest to extract feature representation for gene expression data classification. *Sci Rep*. 2018;8:16477.
24. Vens C, F C, editors. *Random forest based feature induction*. . IEEE 11th International Conference on Data Mining (ICDM); 2011; Vancouver, Canada.
25. Tang A, Foong J. A qualitative evaluation of random forest feature learning. In: Herawan T, Ghazali R, Deris M, editors. *Recent Advances on Soft Computing and Data Mining Advances in Intelligent Systems and Computing*: Springer Link; 2014. p. 359-68.
26. Mold JW, Hamm RM, McCarthy LH. The law of diminishing returns in clinical medicine: how much risk reduction is enough? *J Am Board Fam Med*. 2010;23(3):371-5.
27. Badgery-Parker T, Pearson SA, Dunn S, Elshaug AG. Measuring Hospital-Acquired Complications Associated With Low-Value Care. *JAMA Intern Med*. 2019;179(4):499-505.
28. Korenstein D, Chimonas S, Barrow B, Keyhani S, Troy A, Lipitz-Snyderman A. Development of a Conceptual Map of Negative Consequences for Patients of Overuse of Medical Tests and Treatments. *JAMA Internal Medicine*. 2018;178(10):1401-7.
29. Weiss ME, Bobay KL, Bahr SJ, Costa L, Hughes RG, Holland DE. A Model for Hospital Discharge Preparation: From Case Management to Care Transition. *J Nurs Adm*. 2015;45(12):606-14.
30. Chen YC, Guo YL, Chin WS, Cheng NY, Ho JJ, Shiao JS. Patient-Nurse Ratio is Related to Nurses' Intention to Leave Their Job through Mediating Factors of Burnout and Job Dissatisfaction. *Int J Environ Res Public Health*. 2019;16(23).
31. Taking action against clinician burnout: a systems approach to professional well-being: Consensus study report. . Washington, DC: National Academies Press; 2019.
32. Jiang HJ, Stocks C, Wong CJ. Disparities between two common data sources on hospital nurse staffing. *J Nurs Scholarsh*. 2006;38(2):187-93.
33. Friese CR, Xia R, Ghaferi A, Birkmeyer JD, Banerjee M. Hospitals In 'Magnet' Program Show Better Patient Outcomes On Mortality Measures Compared To Non-'Magnet' Hospitals. *Health Aff (Millwood)*. 2015;34(6):986-92.
34. Statistical Brief #248. *Healthcare Cost and Utilization Project (HCUP)*. Rockville, MD: Agency for Healthcare Research and Quality; 2019.
35. Kelly LA, McHugh MD, Aiken LH. Nurse outcomes in Magnet® and non-Magnet hospitals. *J Nurs Adm*. 2012;42(10 Suppl):S44-9.
36. Hamadi HY, Martinez D, Palenzuela J, Spaulding AC. Magnet Hospitals and 30-Day Readmission and Mortality Rates for Medicare Beneficiaries. *Med Care*. 2021;59(1):6-12.
37. Jayawardhana J, Welton JM, Lindrooth RC. Is there a business case for magnet hospitals? Estimates of the cost and revenue implications of becoming a magnet. *Med Care*. 2014;52(5):400-6.

38. 2020 AHA Annual Survey American Hospital Association. American Hospital Association; 2020.
39. Bobay K, Bahr SJ, Weiss ME, Hughes R, Costa L. Models of Discharge Care in Magnet® Hospitals. *J Nurs Adm.* 2015;45(10):485-91.

Table 1. Theoretical Framework and Empirical Measures.

Donabedian Structure-Process-Outcome Model ^a	Structural characteristics			Patient Characteristics ^b	Process	Outcome
Economic Theory of Production ^c	Factors of production & Technology			[Not included]	Production Function	Output
Definitions	The setting in which care is provided and the instrumentalities of which the process of care is the outcome. ^a			Patient-level risk factors prior to care delivery that may guide how care is delivered and directly impact the outcome.	The process of execution of clinical care delivery	The outcome of medical care, in terms of recovery, restoration of function and of survival.
	Factors of production are inputs (labor, capital) that are used to produce an output. Technology is the organizational rules and norms prescribing <i>how</i> the labor and capital inputs ought to interact during the care delivery process ^c	Labor Inputs: the quantity (e.g., number of working hours, number of full-time equivalents) and quality (e.g., education, experience, expertise) of the human resources employed by the organization (nurses, physicians, etc.)	Capital Inputs: the hospital built environment and equipment, administrative and other support systems, type and organization of the electronic health records, and other non-labor resources than enable its employees to deliver care.			

<p>Select variables available in the READI database ^d</p>	<p><u>Unit-specific:</u></p> <p>RN hours per patient day (HPPD)</p> <p>RN overtime HPPD</p> <p>Non-RN HPPD (nurse aides, licensed practical nurses, medical assistants)</p> <p>%BSN</p> <p>% Specialty-certified RNs</p> <p>RN Experience</p> <p>% Full/Part-Time</p> <p><u>Hospital:</u></p> <p>RN full-time equivalents (FTEs)</p> <p>Non-RN FTE's^e</p>	<p><u>Unit-specific:</u></p> <p>Bed size</p> <p>Patient days</p> <p>Average Daily Census</p> <p>Case-Mix Index</p> <p>Unit Type (e.g., cardiac, medical, surgical)</p> <p><u>Hospital:</u></p> <p>Geographic region (East, West, Midwest, South)</p> <p>Bed size</p> <p>Number of med., surg., or med-surg units</p> <p>Hospital Type (community/urban; non-teaching/teaching)</p> <p>Type of the EHR system</p>	<p><u>Unit-specific:</u></p> <p>READI discharge intervention</p> <p>RN case manager/ social worker/ discharge planner/ coordinator/ expeditor involved in discharge transition care</p> <p>Prescriptions filled prior to discharge</p> <p><u>Hospital:</u></p> <p>Discharge Care Model^f</p> <p>Formal readmission risk screening imbedded in the EHR.</p> <p>Interdisciplinary care coordination rounds</p> <p>Guidelines include interdisciplinary discharge rounds</p>	<p><u>Demographic:</u></p> <p>Sex</p> <p>Age</p> <p>Race</p> <p>Ethnicity</p> <p>Marital Status</p> <p><u>Clinical:</u></p> <p>Admission type</p> <p>Admission from (source)</p> <p>Service Type (medical/surgical)</p> <p>Payment type</p> <p>Severity of Illness</p> <p>Mortality Risk</p> <p>Prior hospitalization (30 and 90 days)</p> <p>Elixhauser Comorbidity Index</p> <p>Cormorbidities</p> <p>Length of Stay</p> <p>ICU Stay</p>	<p>[Unobserved, not measured in this study]</p>	<p>Patient readmission within 30 days</p>
--	--	--	---	--	---	---

a Donabedian, 1966.[17]

b Mitchell, 1998.[18]

c Adapted from Craig, 1973.[19]

d During the random forest recursive model selection process, some of these variables were eliminated from the final model. See Digital Supplement Figure A for the full list of variables included in the final model.

e All hospital employees (physicians, pharmacists, therapists, administrative, laboratory, house-keeping, etc.) on hospital payroll, per the American Hospital Association Annual Survey definition.[\[38\]](#)

f Established programs of discharge care based on local, state, or national discharge transition improvement initiatives.[\[39\]](#)

Table 2. Predictive Accuracy of the Model in the Learning Sample (n=96,444) and the Test Sample (n=41,334)

	Learning Sample	Test Sample
Kappa	0.23	0.13
Specificity	1.00	0.99
Sensitivity	0.15	0.08
Positive predictive value	1.00	0.86
Negative predictive value	0.89	0.89
Area under Receiver Operating Characteristics curve	0.99	0.71

Table 3. Select¹ Descriptive Characteristics of Patients in the Learning Sample (n=96,444) and Test Sample (n=41,334)

Characteristic	Learning Sample n (%) or mean (SD)	Test Sample n (%) or mean (SD)
Primary Outcome: 30-Day Readmission		
No Readmission	84,585 (87.70%)	36,333 (87.90%)
Readmission	11,859 (12.30%)	5,001 (12.10%)
Patient Characteristics		
Patient Sex		
Male	47,265 (49.01%)	20,168 (48.79%)
Female	49,179 (50.99%)	21,166 (51.21%)
Patient Age	60.10 (17.38)	60.09 (17.36)
Patient Race		
American Indian Or Alaska Native	858 (0.89%)	348 (0.84%)
Asian	3,320 (3.44%)	1,416 (3.43%)
Black or African American	14,717 (15.26%)	6,288 (15.21%)
Native Hawaiian Or Other Pacific Islander	305 (0.32%)	146 (0.35%)
White	65,847 (68.27%)	28,318 (68.51%)
Unknown	11,397 (11.82%)	4,818 (11.66%)
Patient Ethnicity		
Not Hispanic	84,517 (87.63%)	36,129 (87.41%)
Hispanic	10,407 (10.79%)	4,554 (11.02%)
Unknown	1,520 (1.58%)	651 (1.57%)
Patient Marital Status		
Not Married	41,988 (43.54%)	18,197 (44.02%)
Married	43,968 (45.59%)	18,566 (44.92%)
Unknown	10,488 (10.87%)	4,571 (11.06%)

Admission Type		
Emergency	52,613 (54.55%)	22,582 (54.63%)
Urgent	17,626 (18.28%)	7,536 (18.23%)
Elective	17,451 (18.09%)	7,543 (18.25%)
Unknown	8,754 (9.08%)	3,673 (8.89%)
Admission Source		
Physician Referral	36,540 (37.89%)	15,822 (38.28%)
Clinic Referral	6,247 (6.48%)	2,662 (6.44%)
HMO Referral	185 (0.19%)	79 (0.19%)
Transfer from Hospital	3,022 (3.13%)	1,271 (3.07%)
Transfer from Another Health Care Facility	2,199 (2.28%)	966 (2.34%)
Emergency Department	19,605 (20.33%)	8,356 (20.22%)
Unknown	28,646 (29.70%)	12,177 (29.46%)
Service Type		
Medical	69,385 (71.94%)	29,770 (72.02%)
Surgical	25,748 (26.70%)	11,025 (26.67%)
Unknown	1,311 (1.36%)	539 (1.30%)
Payment Type		
Private Insurance	29,763 (30.86%)	12,730 (30.80%)
Medicare	40,820 (42.33%)	17,485 (42.30%)
Medicaid	13,993 (14.51%)	6,012 (14.54%)
Uninsured/ Unknown	11,868 (12.31%)	5,107 (12.36%)
Severity of Illness Score		
Minor	11,564 (11.99%)	5,033 (12.18%)
Moderate	27,244 (28.25%)	11,657 (28.20%)
Major	25,287 (26.22%)	10,878 (26.32%)
Extreme	4,883 (5.06%)	2,052 (4.96%)
Unknown	27,466 (28.48%)	11,714 (28.34%)
Mortality Risk Score		
Minor	23,559 (24.43%)	10,063 (24.35%)
Moderate	19,894 (20.63%)	8,672 (20.98%)
Major	14,841 (15.39%)	6,323 (15.30%)
Extreme	3,441 (3.57%)	1,464 (3.54%)
Unknown	27,466 (28.48%)	11,714 (28.34%)
Prior Hospitalization Within 30 Days	11,907 (12.35%)	5,009 (12.12%)
Elixhauser Comorbidity Index	7.04 (8.39)	7.08 (8.36)
Total Length of Stay, days	4.22 (4.22)	4.24 (4.26)
ICU Stay	17,376 (18.02%)	7,355 (17.79%)
Discharge Disposition		
Discharged to Home/Self-Care	77,120 (79.96%)	33,029 (79.91%)
Discharged to Home/Home Health Service	16,851 (17.47%)	7,197 (17.41%)
Discharged to Hospice Care	1,184 (1.23%)	535 (1.29%)

Left Against Medical Advice	1,289 (1.34%)	573 (1.39%)
Comorbidities		
Renal Failure	7,691 (7.97%)	3,344 (8.09%)
Weight Loss	3,171 (3.29%)	1,310 (3.17%)
Hypertension	21,861 (22.67%)	9,464 (22.90%)
Metastatic Tumor	2,406 (2.49%)	1,035 (2.50%)
Electrolyte Disorders	15,786 (16.37%)	6,812 (16.48%)
Congestive Heart Failure	4,796 (4.97%)	2,146 (5.19%)
READI Study Status ²		
Patient from an Intervention Unit	49,195 (51.01%)	21,224 (51.35%)
Patient Treated Per Protocol ³	25,365 (51.56%)	10,934 (51.52%)

¹ See Supplement Table A for the full set of descriptive statistics.

² Study month not shown, see Supplement Table A

³ Patients from intervention units only; the percentage includes the baseline period, see Supplement Table B.

Table 4. Select¹ Descriptive Structural Input Characteristics of Units (n=62) and Hospitals (n=31)

Input variable	n (%) or mean (IQR)
Labor inputs:	
<u>Unit-specific nursing labor inputs:</u>	
Quantity: nursing hours per patient day: ¹	
RN non-overtime	6.77 (5.98, 7.42)
RN overtime	0.20 (0.13, 0.25)
Non-RN total HPPD	3.35 (2.47, 3.68)
Quality: education, expertise, experience, flexibility:	
% BSN-Prepared RNs	67.65 (55.66, 80.78)
% Certified RNs	29.35 (10.98, 40.98)
Average RN Experience (on the unit)	6.18 (4.50, 8.00)
% Full-Time RNs ²	78.28 (71.25, 93.18)
<u>Hospital-wide labor inputs:</u>	
RN full-time equivalents	1,101.84 (486.91, 1,271.35)
Non-RN full-time equivalents (all combined) ³	2,774.88 (570.91, 3,259.22)
Capital inputs:	
<u>Hospital facility and resources:</u>	
Hospital bed size	540.87 (315.00, 607.50)
Hospital Type:	
Rural, non-teaching	1 (3.23%)
Community, non-teaching	11 (35.48%)
Community, teaching	7 (22.58%)
Urban, non-teaching	4 (12.90%)

Urban, teaching	3 (9.68%)
Academic medical center	5 (16.13%)
<u>Unit-specific facility and resources:</u>	
Unit type:	
Cardiac Care	11 (17.74%)
Medical	8 (12.90%)
Surgical	2 (3.23%)
Medical/Surgical	17 (27.42%)
Medical with Telemetry	17 (27.42%)
Surgical with Telemetry	3 (4.84%)
Neurology/Neurosurgery	2 (3.23%)
Orthopedics	1 (1.61%)
Respiratory	1 (1.61%)
Average daily patient census ²	26.49 (22.19, 28.70)
Patient case-mix index ²	1.76 (1.42, 1.80)
Technology for readmission avoidance:	
<u>Discharge Care Models⁴</u>	
Transitional Care Model	1 (3.23%)
Care Transitions Model	6 (19.35%)
Re-Engineered Discharge	2 (6.45%)
Institute for Healthcare Improvement/ State Action on Avoidable Readmissions	2 (6.45%)
Centers for Medicare/Medicaid Services	3 (9.68%)
State hospital association initiative	3 (9.68%)
Local/regional collaborative initiative	4 (12.90%)
None of the above/other	18 (58.06%)

¹ See Supplement Figure A and Table A for the full set of structural input variables.

² Measured monthly (19 data points); otherwise measured annually (2 data points).

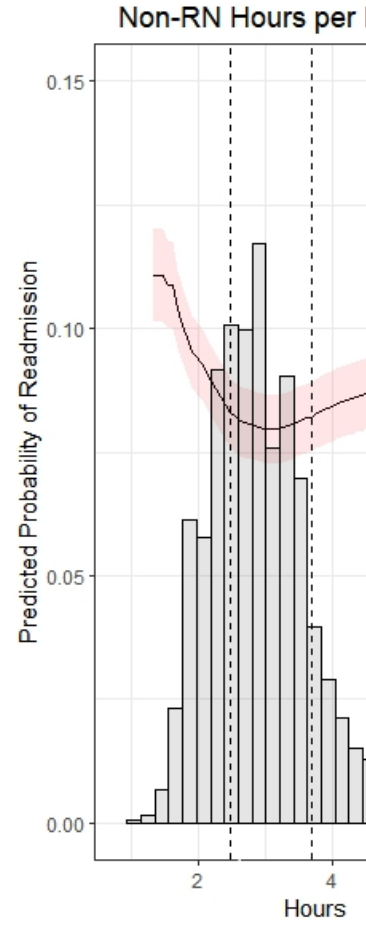
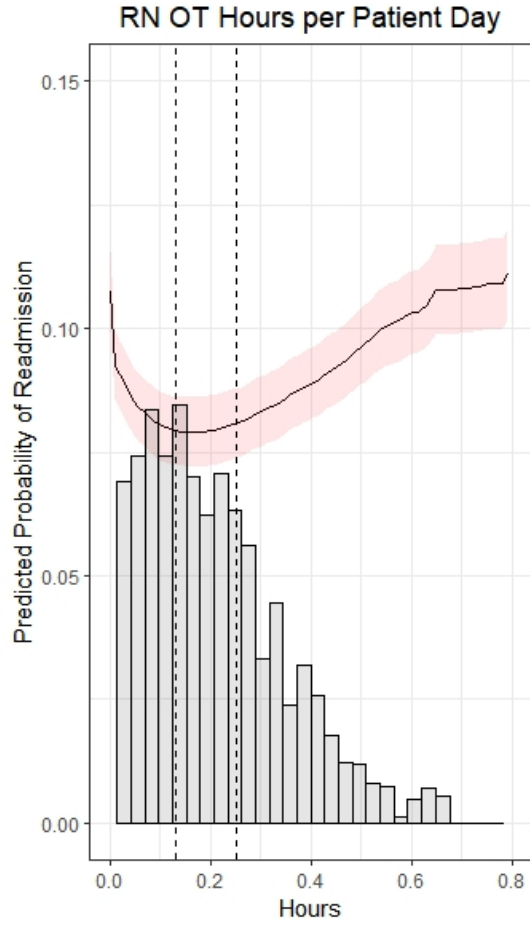
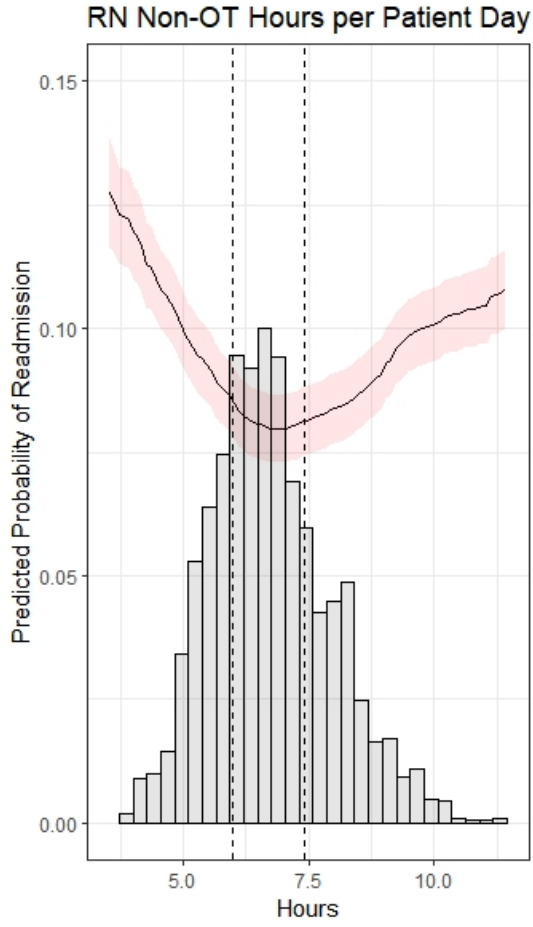
³ All hospital employees (physicians, pharmacists, therapists, administrative, laboratory, house-keeping, etc.) on hospital payroll, per the American Hospital Association Annual Survey definition.[38]

⁴ Categories are not mutually exclusive.

FIGURES

Figure 1: Partial Dependence Plots of the Relationship between Nurse Staffing and Readmissions, the Learning Sample (n=96,444)

Notes: For each of the staffing variables, the figure shows: 1) the partial dependence plot for the relationship between the corresponding staffing hours variable and the predicted probability of readmission (PDP, solid black curve), and the 95% confidence interval (pink sleeve); 2) the relative frequency distribution of the patient sample by nursing hours per patient day in 25 increments (grey shaded bars); and 3) the 1st and 3rd quartile of the unit-level distribution of nursing hours per patient day (dashed vertical lines).



Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit <http://www.adobe.com/go/acrreader>.

Windows is either a registered trademark or a trademark of Microsoft Corporation in the United States and/or other countries. Mac is a trademark of Apple Inc., registered in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.



REQUIRED INFORMATION AND FORM FOR ACCEPTED MANUSCRIPTS AT HSR [SEPT., 2009]

HSR Authorship Responsibility, Disclosures, and Acknowledgments

General instructions:

The corresponding author is responsible for coordinating with each author to obtain all information disclosed on this form. This information is required to be submitted in the section on Manuscript Central labeled for the editors only, i.e., as supplementary files not for review, at the time of the first revision and should be updated as necessary for subsequent revisions or in response to queries from the editor-in-chief.

Each author must read and provide the information to the corresponding author on:

- (1) Authorship Responsibility, Criteria, and Contributions,
- (2) Financial Disclosures, and
- (3) Other Disclosures,

The **corresponding author** must complete and submit on behalf of all authors the summary form, affirming the information provided is accurate. The submitted form should contain the answers for the corresponding author on items 1-3 and summarize all authors' answers in the (4) Author and Other Contributors Section and (5) Acknowledgment/Disclosure Statement.

This form is available online at HSR.org *and* <http://mc.manuscriptcentral.com/hsr>. If necessary, it may be photocopied and distributed to coauthors or other contributors.

NOTE: We require this form to be completed before publication. The joint acknowledgement/disclosure statement and author matrix will be published electronically; the joint acknowledgement/disclosure statement [including any contributors to be acknowledged] will also appear in the print version. Each form will be approved by the editors to make sure that all necessary and relevant information that readers need to know to evaluate the work published in HSR has been reflected in the acknowledgement/disclosure paragraph; detailed disclosures will not be published.

The corresponding author should collect the disclosure information from each co-author and provide that information in the appropriate box below. The corresponding author should then prepare a draft joint acknowledgement/disclosure statement along with the author matrix. Each author should review and approve of the final joint statement and authorship information.

The wording of the acknowledgement/disclosure statement and author matrix may undergo some negotiation with the editors, either to eliminate some material that is not needed or to include some that should be disclosed. As only the corresponding author can upload information to Manuscript Central, he or she will coordinate the process. All authors should submit drafts of their disclosures electronically to the corresponding author. The corresponding author may submit them and a draft of the joint statement electronically at any time during our processing of the manuscript. This material will be reviewed for consistency and completeness when the manuscript is close to acceptance.

**Health Services Research
Certification and Disclosures by Authors**

Manuscript Number: HSR-

[After submitting your manuscript online, your manuscript is assigned a number that must be entered in the space above].

Corresponding Author:

Please fill in the following table: AUTHOR ORDER AND CONTACT INFORMATION

Authors, in preferred order for publication	Telephone Number	Fax Number	e-mail address

INSTRUCTIONS: EACH AUTHOR MUST READ SECTIONS 1-3 AND PROVIDE FULL DISCLOSURE TO THE CORRESPONDING AUTHOR. THE SUBMITTED FORM SHOULD PROVIDE INDIVIDUAL INFORMATION FOR THE CORRESPONDING AUTHOR AND SUMMARIZE ALL AUTHORS' INFORMATION.

1. Authorship Responsibility, Criteria, and Contributions. Each author should meet all criteria below (A, B, C, and D) and should affirm his or her general and specific contributions to the corresponding author. The corresponding author should then complete the information for him/herself by checking the appropriate boxes below and summarize the information for all authors in section 4 and 5. This summary statement should be consistent with and reflect all significant disclosures of each of the individual authors. Research and manuscripts may depend critically on various individuals who do not meet all the criteria for authorship; they should be acknowledged as contributors in section 4 below.

A. I certify that

- the manuscript represents valid work and that neither this manuscript nor one with substantially similar content in which I am an author has been published or is being considered for publication elsewhere, except as described in an attachment, and that copies of closely related manuscripts have been provided to the editors or HSR; and
- [for papers with more than 1 author], I agree to allow the corresponding author to serve as the primary correspondent with the editorial office, to review the edited typescript and proof, and to make decisions regarding release of information in the manuscript to the media, federal agencies, or both; or,
- [for papers for which I am the only author] I will be the corresponding author and agree to serve in the roles described above.

B. I have given final approval of the submitted manuscript and each of its revisions, if any.

C. I have participated sufficiently in the work to take public responsibility for (check 1 or 2 below)

1. the whole content.

2. part of the content: (fill in as appropriate)
limited to

all but

D. To qualify for authorship, you must check at least 1 box for *each* of the 2 categories of contributions listed below. For examples of contributions that qualify for acknowledgment but not for authorship, see section 4 below.

I have made substantial contributions to the intellectual content of the paper as described below.

1. (check at least 1 of the 3 below)

- a. conception and design
- b. acquisition of data
- c. analysis and interpretation of data

2. (check at least 1 of 2 below)

- a. drafting of the manuscript
- b. critical revision of the manuscript for important intellectual content

2. Financial Disclosure. For the purposes of this disclosure and the one below, “conflicts of interest” include not only situations in which decisions and judgment have been influenced, but also those that may, if subsequently disclosed or uncovered, lead readers to question whether decisions or judgment may have been influenced by such situations. We anticipate that the vast majority of “conflicts,” if disclosed in advance, will have little bearing on a reader’s assessment of the research and no bearing on our decision to publish. Except in cases in which all the authors certify they have nothing to disclose, the corresponding author will submit a brief acknowledgement/disclosure of the conflicts of all the authors in Section 4.

Please check the appropriate boxes below (applies to the past five years and foreseeable future):

- a. I have no relevant financial interests pertaining to this manuscript.
- b. I certify that all my conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (e.g., employment / affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are disclosed in an attachment.
- c. I certify that all financial and material support (including those providing data or access to patients, interviewees, etc.) for this research and work are clearly identified in an acknowledgement/disclosure section to be published with the manuscript.
- d. I certify that I agree with the description in the Acknowledgements/Disclosures statement in Section 5 of any support for the project or conflicts of interest as they pertain to myself.

3. Other Disclosures. Aside from financial interests, there may be other real or potential conflicts that need to be disclosed. In particular, we are concerned about situations in which an external sponsor, provider of data, or other entity may be able to influence either an author's ability to publish or to shape what he or she seeks to publish. We are also concerned that an author, or organizations with which he or she is affiliated, may have taken certain positions relevant to the research that should be disclosed. We may require that such information be publicly acknowledged. For the purposes of this disclosure, the term "contractual right" includes not only formal contracts and memoranda of understanding with outside parties, but also the rights of your employer to review and or approve your publications. Please check the appropriate boxes below and provide the information requested:

- a. Sponsors and/or supporters of this research, e.g., organizations providing data, or supervisors within my own organization have:
 - i. a contractual right to review and approve the manuscript before submission or publication.
 - ii. a contractual right to review and comment on the manuscript within days, after which it can be submitted without constraint.
 - iii. no contractual rights to review the manuscript before submission, but there is a requirement that the sponsor/supporter be given a copy of the accepted manuscript prior to publication.
 - iv. no requirement for prior approval or notification, but I solicited feedback and/or plan to provide an advance copy as a courtesy.
 - v. the manuscript has not been reviewed or commented upon by the sponsor(s) and I have no plans to provide advance notification.
 - vi. this research did not have a sponsor as defined above.

If you checked any of the boxes i through iv, please indicate the organizations involved and the circumstances. (For example, some organizations require that manuscripts undergo internal peer review for comment, but an author may publish any findings he or she wishes.)

IN THE SUBMITTED FORM, THE CORRESPONDING AUTHOR SHOULD PROVIDE DETAILS FOR EACH AUTHOR [NAMED] WHO HAS CHECKED ANY BOX i THROUGH iv.

Please describe:

A contractual right by a sponsor, supporter, or employer to review and approve (and hence censor) a manuscript makes it appear to be a “work for hire” rather than research. On occasion such work may warrant publication in HSR, e.g., because it illustrates novel methods, but we will be reluctant to accept a manuscript with findings or conclusions that might be influenced by the ability of the sponsor to withhold permission to publish.

b. I have taken public stands (e.g., in print, media, expert witness, legislative testimony or other venues, with or without compensation) that are identified with a particular advocacy position relevant to the manuscript.

IN THE SUBMITTED FORM, THE CORRESPONDING AUTHOR SHOULD PROVIDE DETAILS FOR EACH AUTHOR [NAMED] WHO HAS CHECKED BOX b.

Please describe:

c. The organization with which I am currently (or was, at the time of the manuscript’s preparation) affiliated (as a spokesperson, board member, or similar prominent position) is identified with a particular advocacy position relevant to the manuscript.

IN THE SUBMITTED FORM, THE CORRESPONDING AUTHOR SHOULD PROVIDE DETAILS FOR EACH AUTHOR [NAMED] WHO HAS CHECKED BOX c.

Please describe:

d. I certify I agree with the description in the Acknowledgements/Disclosures statement in Section 5 of disclosures as they pertain to myself.

Author Manuscript

4.A. Author matrix. Please complete the authorship matrix affirming which roles each author has fulfilled. To be listed as an author, persons should have made contributions in both of the shaded areas; all others should be acknowledged as contributors. In the box in 4.B., we invite you to name other contributors. In addition, you may use the box in 4.B. if you wish to name other kinds of contributions by authors.

Section 4 A. and B. will be made available in the electronic version of accepted manuscripts.

*You may wish to copy the author's names from the box on page 2.

Manuscript Number:	Each author must check at least one blue box showing important contributions to:			& check contributions in at least one orange box:	
	Conception and design	Acquisition of data (arranging for or obtaining)	Analysis and interpretation of data	Drafting the manuscript	Critical revision for important intellectual content
Name of Each Author*					

Section 4.B. Other Contributions

Some research projects involve very extensive teams with many people who do not meet the criteria for authorship but who have contributed importantly to the work. The box below is designed to acknowledge other types of contributions. As appropriate, use this box to name other people who have made important contributions to the work reported in this manuscript. Please also name people who have played a significant role in preparing the manuscript but who are not listed as authors. NOTE: If you wish to acknowledge any of these contributors in print, please include their names and roles in the joint acknowledgment section below.

Please name contributors to be acknowledged for:

Statistical analysis

Data collection (under supervision)

Interviewing (under supervision)

Programming and data management

Editing for presentation or style

Obtaining funding

Administrative, technical, or material support

Supervision of research staff

Other (specify)

INSTRUCTIONS: THE CORRESPONDING AUTHOR MUST FILL OUT SECTION 5.

5. Joint Acknowledgment/Disclosure Statement. Please enter a 1-3 sentence statement that acknowledges all forms of financial and material support for the project, the roles of key individuals who should be recognized; (this may or may not extend to everyone listed as contributors). Please also include summary statements about what you feel are the conflicts of interests and disclosures without which some readers may feel that important relevant information is being withheld. The editors will determine whether these disclosures are sufficient or excessive and may return a revised version. All authors will need to approve the final version. As appropriate, please add at the end of your statement: 'Disclosures: None' or 'No Other Disclosures'.

Section 5 will be made available in both the electronic and print version of accepted manuscripts.

Joint Acknowledgement/Disclosure Statement:

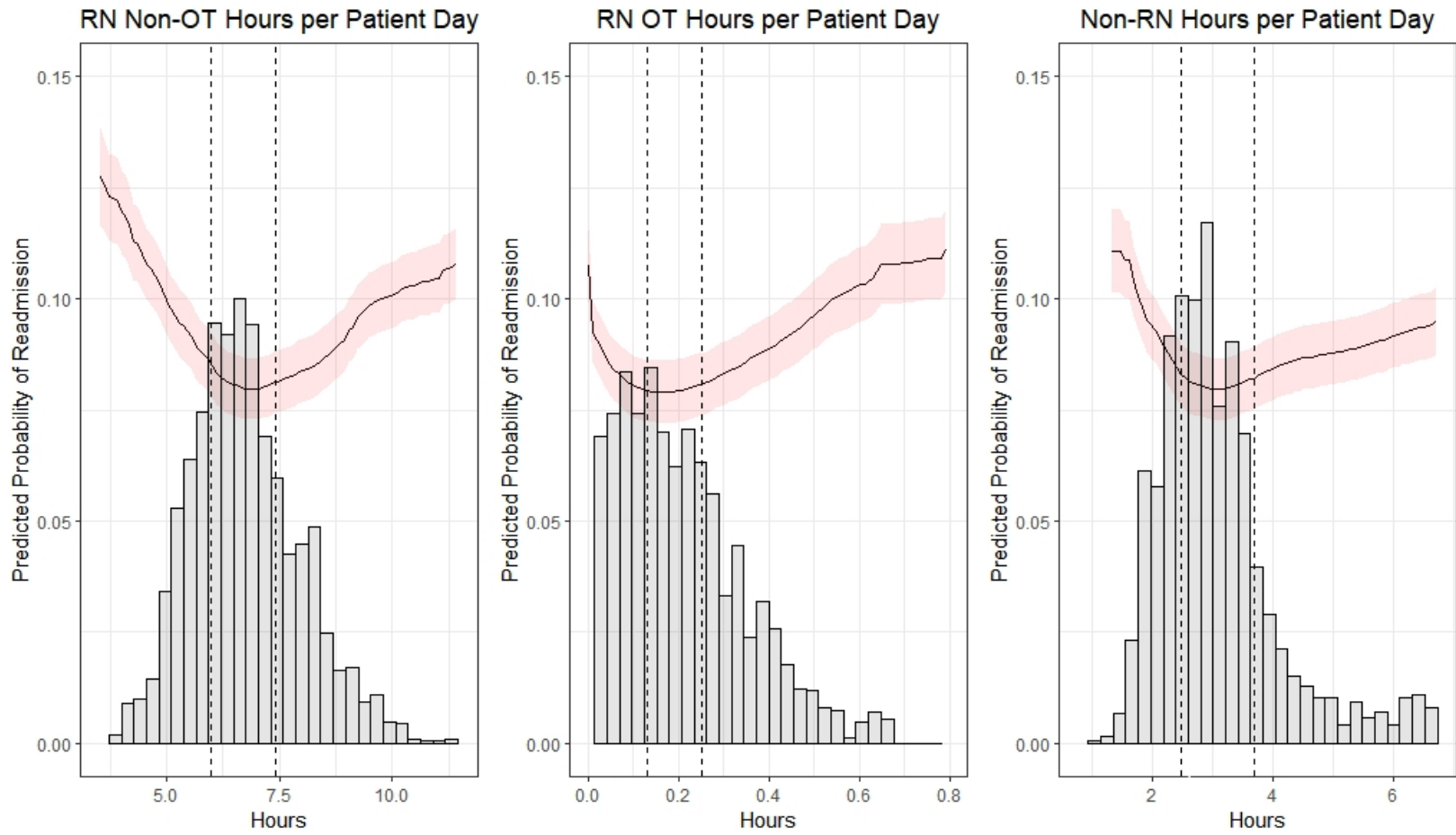
The corresponding author should obtain permission to name all individuals named in an Acknowledgment or the Contributorship Section because readers may infer their endorsement of data and conclusions.

The corresponding author must check the box below to affirm his/her certification that:

- all persons who have made substantial contributions to the work reported in this manuscript (e.g., data collection, analysis, or writing or editing assistance) but who do not fulfill the authorship criteria are named with their specific contributions in the Contributorship Box associated with the manuscript.
- all persons named in the Contributorship box have provided me with permission to be named.
- no other persons have made substantial contributions to this manuscript.
- all authors have approved the Joint Acknowledgement/Disclosure Statement intended for publication and the author matrix.

I, the corresponding author, certify that the above statements are true.

(Adapted, with permission, from the Journal of the American Medical Association, 2006)



Non-linear association of nurse staffing and readmissions uncovered in machine learning analysis

Olga Yakusheva, PhD
Department of Systems, Populations, and Leadership
School of Nursing
Department of Health Management and Policy
School of Public Health
University of Michigan,
400 North Ingalls, Suite 4343
Ann Arbor, Michigan, USA
yakush@umich.edu

James T. Bang, PhD
Department of Economics
St. Ambrose University
Davenport, Iowa, USA

Ronda G. Hughes, PhD, RN, NEA-BC, FAAN
Center for Nursing Leadership
College of Nursing
University of South Carolina,
Columbia, South Carolina, USA

Kathleen L. Bobay, PhD, RN, NEA-BC, FAAN
Marcella Niehoff School of Nursing
Loyola University Chicago
Chicago, Illinois, USA

Linda Costa, PhD, RN, NEA-BC
School of Nursing
University of Maryland,
Baltimore, Maryland, USA

Marianne E. Weiss, DNSc, RN
College of Nursing
Marquette University,
Milwaukee, Wisconsin, USA

Word Count: 4483