

Boundaries of Focus and Volume: An Empirical Study in Neonatal Intensive Care

Felix Miedaner

Department of Business Administration and Health Care Management, University of Cologne, Albertus-Magnus Platz, 50923 Köln, Germany, miedaner@wiso.uni-koeln.de

Sandra Sülz*

Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA, Rotterdam, The Netherlands, sulz@eshpm.eur.nl

Our study contributes to the scholarly debate whether organizational units should have a narrow focus and admit a homogeneous patient cluster or whether they should admit a pool of patient clusters. We investigate whether the benefits of increased volume through pooling patients outweigh the disadvantages of increased heterogeneity and pursue our analysis in the context of neonatal care. Our empirical studies relies on 4020 patient episodes collected in 18 German neonatal intensive care units and we distinguish between two patient clusters that differ with respect to the inherent medical risk and operational heterogeneity. Cluster 1 consists of very-low birth weight (VLBW) infants with increased risk of complications but similar service trajectories and lower operational heterogeneity. Cluster 2 contains non-VLBW infants with lower risk of complications but more diversity in disease patterns and higher operational heterogeneity. Our analysis shows that cluster volume, that is, the unit's absolute patient volume in a cluster, is positively related to process outcomes as indicated by decreasing length of stay. This relationship is found for both clusters. Regarding focus, we do not find any evidence of positive effects. In fact, we even find that cluster focus, that is, the unit's relative volume of the cluster, is detrimentally related to process outcomes for non-VLBW patients with lower risk of complications and more operational heterogeneity. This indicates that organizational units providing services for complex patients should not have a narrow focus, but should rather provide services for related patient clusters in order to achieve higher volume levels within the unit.

Key words: volume; focus; patient heterogeneity; neonatal intensive care

History: Received: March 2018; Accepted: August 2019 by Sergei Savin, after 2 revisions.

1. Introduction

There is an emerging scholarly debate with respect to redesigning hospitals and the question of whether specialized units that admit a homogeneous patient cluster are preferable or whether, instead, flexible units that admit a pool of patient clusters are better (Best et al. 2015). Specialized units, which admit one homogeneous patient cluster might benefit from a narrower range of treatment protocols, lower variability, and fewer conflicting or competing operational activities (Clark and Huckman 2012, Huckman and Zinner 2008, KC and Terwiesch 2011, McDermott and Stock 2011). The advantage of focusing solely on one cluster might, however, lead to the disadvantage of insufficiently achieving economies of scale and scope due to lower patient volume levels. Flexible units, on the other hand, might provide the benefits associated

with economies of scale and scope (Green 2012), such as higher productivity and better outcomes due to better fixed cost amortization and learning effects (Freeman et al. 2019). The drawback for flexible units lies in potential high heterogeneity and diluted focus, which might lead to a broad range of treatment protocols and conflicting operational activities. This apparent trade-off between flexible and specialized units is at the core of the scholarly debate that seeks to determine whether the benefits of increased volume through pooling patients outweigh the disadvantages of increased heterogeneity and loss of focus (Best et al. 2015).

Obviously, these trade-off decisions are not inevitably the same for all patients because patients respond differently to volume and focus. Kuntz et al. (2019), for instance, show that routine patients with a pre-planned hospital stay and no comorbidities experience substantial quality benefits from focus, yet are unaffected by volume. Complex patients are detrimentally affected by high levels of volume, but they benefit if the same types of patients are routed to the same clinical

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

department (Kuntz et al. 2019). The authors call for more research to verify their findings in the context of specific conditions while taking the peculiarities of these conditions into account. Condition-specific health service trajectories can have idiosyncratic features that can affect whether and why operational factors, such as volume and focus, are beneficial for health service quality. Consequently, zooming into the organization and conducting a setting-specific analysis allows for several theoretical mechanisms to be considered and helps to expand the evidence-base of volume and focus theories. While Kuntz et al. (2019) take a business model perspective and differentiate between routine and complex patients, we focus on the internal structure of one clinical department and the complex patients admitted to it. Our study further differentiates between these complex patients based on the inherent risk of medical complications and operational heterogeneity. We thus contribute to the body of literature by exploring the boundaries of volume and focus for complex patients admitted to neonatal intensive care units (NICUs).

Neonatal intensive care units provide health services for patients with severe medical conditions that occur after birth. This setting has the advantages of a clearly defined patient group with a limited chance of being routed to other units. Our empirical analysis distinguishes between two clusters, with cluster 1 containing very-low birth weight (VLBW) patients and cluster 2 consisting of non-VLBW patients. We simultaneously consider cluster volume, measured as the unit's absolute patient volume in one cluster, and cluster focus, that is, the unit's relative volume in a cluster, and analyze their effects on process outcomes as indicated by length of stay. Relying on 4020 patient episodes collected in 18 German NICUs, we show that cluster volume is positively related to process outcomes for both cluster types. Regarding cluster focus, however, we do not find any evidence of positive effects. In fact, we find that cluster focus is detrimentally related to process outcomes for non-VLBW patients with lower risk of complications and more operational heterogeneity. Our results thus indicate that organizational units providing services for complex patients should not have a narrow focus, but should rather provide services for related patient segments in order to achieve higher volume levels within the unit.

2. Related Literature and Research Framework

Before reviewing the related literature and setting up our research framework, we provide an initial definition of our concepts. Our study considers NICUs, which provide health care services for a clearly distinct patient segment consisting of preterm and sick

newborns. This patient segment is composed of two medically distinct clusters, with cluster 1 containing very-low birth weight (VLBW) infants and cluster 2 consisting of non-VLBW infants (we provide more details of the clustering below). Following the recent literature on volume and focus in health care organizations (Clark and Huckman 2012, KC and Terwiesch 2011, Kuntz et al. 2019, McDermott and Stock 2011), we denote the absolute annual number of patients within a cluster who are admitted to the unit as *cluster volume*, while *cluster focus* is conceptualized as the unit's cluster volume as a proportion of the unit's overall annual volume. This conceptualizes focus as emphasis, that is, "the disproportionate emphasis on some service lines, while still maintaining others" (McDermott and Stock 2011, p. 618). Note that, following this conceptualization, cluster volume and cluster focus are inevitably related; if the NICU maintains the cluster volume in cluster 1, but increases the cluster volume in cluster 2, it automatically increases the cluster focus in cluster 2 as well. We are interested in how cluster volume and cluster focus affect health service delivery, and we will focus on process outcomes, as indicated by patient length of stay. We now proceed with reviewing the literature on volume and focus before detailing the differences between our clusters and the expected volume and focus effects therein.

The medical literature has identified a positive association between volume and outcomes for a variety of conditions and surgical procedures (Birkmeyer et al. 2002, Gaynor et al. 2005) and also within specific settings, such as neonatology (Bartels et al. 2006, Chung et al. 2010, Phibbs et al. 2007, Profit et al. 2013, 2016, UK Neonatal Staffing Study Group 2002). A positive relationship between volume and outcome has also been found in the management literature, where it is even claimed to be an "empirical regularity" (Huckman and Zinner 2008). Individuals, groups and organizations accrue experience and learn from practice, which allows them to achieve higher productivity and quality improvement as volume increases (KC and Staats 2012, Reagans et al. 2005, Theokary and Ren 2011).

From a focus perspective, the beneficial effects on productivity and outcomes are expected to accrue from less complexity due to limiting the number of routines within an organization(al unit) and from less distraction due to lower volume outside the focal activity. The focus debate dates back to Skinner's influential paper applied to the manufacturing setting (Skinner 1974); however, focus outcome effects have also recently been analyzed within the health care industry and hospitals, in particular (Clark and Huckman 2012, KC and Terwiesch 2011, McDermott and Stock 2011). Overall, this literature identifies a

positive relationship between focus and outcomes (Clark and Huckman 2012, McDermott and Stock 2011), albeit the benefits are more likely to arise within organizational units and the processes therein, as opposed to the entire organization (KC and Terwiesch 2011).

Following the perspective of an organizational unit, we consider a patient segment that is composed of two distinct clusters (cluster 1: VLBW; cluster 2: non-VLBW), which differ based on medical and operational aspects. From a medical perspective, the clusters differ in terms of their inherent risk of complications. In cluster 1, consisting of VLBWs who—per definition—are born with a birth weight below 1500 g, neonatal complications are markedly increased and patients face a higher morbidity risk (Lee et al. 1980). From an operational perspective, a fundamental difference between the clusters is given by the variety of disease patterns, as cluster 1 is more homogeneous than cluster 2.¹ Taken together, we can define cluster 1 as a homogeneous high-risk cluster, while cluster 2 is a heterogeneous lower-risk cluster. Having outlined the differences between the cluster, we will now theorize why volume and focus effects are expected to differ between these two clusters. We are interested in assessing the total volume (and focus) effect; that is, although we rely on several theoretical arguments to derive our hypotheses, the different theoretical mechanisms are not separately tested.

With increasing levels of volume, individuals and organizations accumulate experience, which allows them to learn and, consequently, perform better. Importantly, it matters whether this experience is coming from executing the same tasks, related tasks or unrelated tasks. Concerning learning at the individual level, Boh et al. (2007) and KC and Staats (2012) find that executing the same tasks improves performance. Their findings also show that experience in related tasks and systems improves performance, albeit that the impact of same-task experiences is stronger. Staats and Gino (2012)'s findings suggest that same-task experience is beneficial in the short term but that variety is likely superior in the long run. Finding that a balance between same-task experience and variety yields the highest productivity, Narayanan et al. (2009) detect that too much variety can indeed hamper performance. If the level of variety is too high, the chance increases that the activities also cover unrelated tasks, which may cause information overload and distraction. Taken together, moderate levels of task variety may improve performance at the individual level. Equivalent results exist at the group level, where it has been shown that diverse experience gained in related tasks enhances learning at the group level and increases performance (Boh et al. 2007, Schilling et al. 2003). Translating the

learning arguments to our context, we expect the learning benefits to be higher in clusters with moderate task variety. In addition, we expect lower learning benefits in clusters with high task variety, in which a diverse set of activities need to be executed and chances are higher that individual tasks are less related to each other. Compared to cluster 1, cluster 2 is more heterogeneous in disease patterns, which results in a greater variety of tasks executed within the various service trajectories. As such, cluster 2 bears a higher likelihood of increasing distraction and unrelated activities, which translates into expecting lower learning benefits and a weaker volume outcome relationship as opposed to cluster 1. A weaker volume outcome effect for cluster 2 is also expected from a knowledge depreciation perspective. With increasing task variety, the potential to have “time gaps” between repeated executions of any one task also increases (Ramdas et al. 2018). Forgetting, in the sense of knowledge depreciation, is, therefore, more likely to occur in cluster 2, because every individual task is done less frequently as a result of the higher task variety (Ramdas et al. 2018).

Both medical clusters require the assembly and coordination of multi-disciplinary teams for service provision. The composition of this team depends on the patient's needs and we expect that, with more variety in disease patterns, there is also more variety in the multi-disciplinary team. Cluster 2, for instance, frequently has to rely on collaboration with specialists such as cardiologists, surgeons and neurosurgeons who are not part of the core care team operating at the NICU. Involving these external specialists thus reduces the likelihood that individual team members have worked with each other in the past. Common past work experience has, however, been identified to improve operational performance (Huckman et al. 2009, Reagans et al. 2005), since it facilitates and increases knowledge sharing because team members are aware of “who knows what.” This also leads to improvements in activity coordination and facilitates a learning environment. Huckman and Staats (2011), Huckman et al. (2009) and Staats (2012) also theorize that teams that are familiar with each other develop a sense of trust, thereby creating a psychologically safe environment that allows team members to speak up about mistakes (Edmondson 1999). If patient volume increases, the likelihood that individual team members have worked with each other before also increases, provided the professional group from which the teams are drawn remains the same. On the other hand, if the team needs to be made up from more different specialties as in the case of cluster 2, team familiarity rises to a lower extent. Taken together, from a team familiarity perspective, the volume outcome relationship is expected to be weaker for cluster 2.

A final argument relates to the differences between the clusters in terms of process uncertainty, which has recently been theorized to moderate volume outcome relationships (Kuntz et al. 2019). Process uncertainty is thereby defined as the level of incompleteness of a hospital's information at the start of the service episode about the exact service configuration, that is, what needs to be done, when, where and by whom (Kuntz et al. 2019). With more uncertainty in the differential diagnostic and less information being present at the start of the service trajectory, cluster 2 is characterized by higher process uncertainty. Higher process uncertainty poses more challenges for care coordination, which diminishes the volume effects (Kuntz et al. 2019). Therefore, we expect the volume outcome effect to be weaker for cluster 2.

Overall, we are interested in the aggregated effect of the theoretical mechanisms listed above. Based upon the arguments discussed above, we hypothesize the following:

HYPOTHESIS 1A. *For cluster 1, an increase in cluster volume is associated with decreasing length of stay.*

HYPOTHESIS 1B. *For cluster 2, an increase in cluster volume is associated with decreasing length of stay.*

HYPOTHESIS 2. *The volume length of stay association is weaker for cluster 2 than for cluster 1.*

One of the theoretical arguments regarding why focus is supposed to be beneficial for performance is that there is less organizational complexity due to limiting the number of routines (Huckman and Zinner 2008). Homogeneous clusters have, c.p., fewer different work routines than heterogeneous clusters. Consequently, focusing on a heterogeneous cluster does not reduce the number of different routines as effectively as focusing on a homogeneous cluster, that is, the reducing routine effect is expected to be weaker for cluster 2.

Additionally, an important factor moderating focus effects is the availability of related services outside the focal activity; e.g., a hospital focusing on cardiac care can benefit from providing services in areas related to cardiac care (Clark and Huckman 2012). These related services can improve the hospital's performance in its cardiac care activity either as a direct spillover from the related area or indirectly through complementing the hospital's cardiac focus (Clark and Huckman 2012). In our context, this means that the NICU's performance in one cluster can be directly or indirectly affected by the level of activity and the services provided in the other cluster. The reason is as follows: If the NICU increases its cluster focus for cluster 1 patients and devotes more technological and personnel resources to provide services for these

high-risk patients, this can have two distinct consequences for the performance in cluster 2: Firstly, it can negatively affect the performance in cluster 2 because an increase in cluster 1 focus directly implies a decreasing cluster 2 focus. Cluster 2 performance is thus negatively affected as a direct result of lower cluster focus. Secondly, increasing the cluster focus in cluster 1 can positively affect the performance in cluster 2 because patients could benefit from the technological and medical expertise gained in cluster 1. This resembles an indirect gain in cluster performance as a result of spillovers from the other cluster. We have already argued before that the homogeneous cluster 1 can benefit more from focus than the heterogeneous cluster 2. Cluster 2 is expected to benefit more from spillovers, since providing services for the high-risk cluster 1 requires by law a substantial number of health care professionals with subsequent training in neonatal care. Taken together, the positive effect of increasing cluster focus for cluster 2, which is already expected to be weaker than for cluster 1, is further mitigated due to the decreasing spillover potential from cluster 1. Therefore, we posit the following:

HYPOTHESIS 3A. *For cluster 1, an increase in cluster focus is associated with decreasing length of stay.*

HYPOTHESIS 3B. *For cluster 2, an increase in cluster focus is associated with decreasing length of stay.*

HYPOTHESIS 4. *The focus length of stay association is weaker for cluster 2 than for cluster 1.*

While we presented multiple arguments as to why the cluster difference in heterogeneity is expected to moderate the volume/focus outcome relationship, we did not list arguments concerning the medical risk difference. A higher medical risk is likely to go along with poorer outcomes and increasing length of stay. However, this is not an argument why volume (or focus) effects are supposed to differ between the two clusters. It rather captures the direct influence of cluster affiliation on length of stay and will empirically be taken into account with cluster fixed effects.

3. Methods

3.1. Setting and Cluster

Our study setting focuses on NICUs in Germany. NICUs are highly specialized and focus on a very particular group of patients with severe medical conditions that occur after birth. Within neonatal care, an important criterion to differentiate between the patient cluster and the corresponding health service processes is the infant's birth weight. Newborns with a birth weight below 1500 g are referred to as very

low birth weight (VLBW) infants. This threshold was found to be a major breakpoint for higher medical risk and increased neonatal complications (Lee et al. 1980), and the medical literature subsequently distinguishes VLBW from non-VLBW infants using a cut-off threshold of 1500 g. We follow this line of literature and differentiate between these two clusters as follows: Cluster 1 comprises VLBW infants born with a birth weight below 1500 g and cluster 2 contains non-VLBW infants born with a birth weight of at least 1500 g.

Notably, these clusters do not only differ with respect to their medical risk but also form an operational perspective in terms of service trajectories. Cluster 1 (VLBW) consists of a fairly homogeneous group of patients whose trajectories typically involve measures for developmentally supportive care, nutrition and respiratory support. These infants are at risk of developing similar complications associated with preterm delivery, such as intraventricular hemorrhage, cystic periventricular leukomalacia, bronchopulmonary dysplasia, necrotizing enterocolitis or retinopathy of prematurity. Cluster 2 (non-VLBW) consists of a more heterogeneous group of patients involving late-preterm infants with initial support and newborns suffering from various problems, such as newborn infections, newborn jaundice, and meconium aspiration syndrome, and newborns with surgical problems, such as cardiac defects and esophagus obstruction. We refer to Appendix S1, chapter 2, for more descriptive details on the operational and medical differences between these two clusters.

In Germany, NICUs are typically divided into either high (Level 1) or lower levels of care (Level 2). Both levels have a public mandate to provide services for both clusters (at least initial care), yet the levels differ in a broad range of structural characteristics, e.g., the required number of physicians and nurses, as well as the required ratio of health care professionals with subsequent training in neonatal care. In 2018, Germany had 165 level 1 NICUs and 46 level 2 NICUs (Institut für Qualitätssicherung und Transparenz im Gesundheitswesen 2018), which are nationally dispersed, and the average distance between any two NICUs is equal to 19.18 km (SD = 17.56 km).

3.2. Data Source

This project was part of a prospective multicenter study (Health Services Research in Neonatal Intensive Care Units – HSR-NICU), conducted in German NICUs in 2013. The study is registered in the German Clinical Trial Register (DRKS00004589) and was approved by the corresponding Ethics Commission, Faculty of Medicine, University of Cologne (#12-228). Out of the 229 identified and approached NICUs in 2013, 66 NICUs agreed to participate. For the purpose

of this project, data were collected from two different sources: (i) a self-administered survey, which was completed by the medical director of each NICU to ascertain characteristics at the NICU level, and (ii) administrative data to gather information about all treated infants within the respective NICU in 2013. In line with the study protocol and to meet the ethical guidelines, patient and hospital data were collected such that the research team only retrieved data, which did not include patient names but only pre-defined pseudonyms – for patients as well as hospitals. With this procedure, it was ensured that data from different sources could be matched in the data analyses through these pseudonyms. Out of the 66 participating NICUs, 24 provided data from both data sources, yielding a total of 7576 patient episodes.

3.3. Dependent Variable: Length of Stay

We consider as our process outcome the length of stay at the NICU, because it has previously been shown to be an important process measure for severe outcomes in neonatal care (Profit et al. 2013, 2016). Reducing a patient's length of stay in an NICU to the extent deemed possible for medical reasons is a desirable objective because patients in NICUs are increasingly affected by hospital-acquired infections or other diseases, which are often preventable and associated with a prolonged hospital stay (Payne et al. 2004).

3.4. Independent and Moderating Variables: Cluster Volume, Cluster Focus, and Cluster Fixed Effects

For both of our clusters $c = \{1, 2\}$, we calculate the cluster volume in the NICU $n = \{1, \dots, N\}$ as the annual number of patients in cluster c admitted in the study year 2013. Since the number of admissions varies considerably between clusters, we standardize the cluster volume by calculating the z-scores for both clusters across NICUs. In line with the literature on focus in hospitals (Clark and Huckman 2012, KC and Terwiesch 2011, McDermott and Stock 2011), we conceptualize focus as emphasis and measure the NICU's cluster focus as the annual number of admissions in cluster c as a proportion of the NICU's total number of annual admissions. Since we only have two clusters under consideration, the distribution of the cluster focus variable is bi-modal and we mitigate that by computing z-scores for both clusters across NICUs. To capture substantial differences between the two clusters, we incorporate a dummy variable C_{icn} which is equal to 1 if patient i belongs to cluster 2 and is admitted to NICU n . Effect modifications are incorporated via interaction terms between the dummy variable and the standardized cluster volume/focus variable, that is, $C_{icn} \times Vol_{cn}$ and $C_{icn} \times Foc_{cn}$, respectively.

Following this operationalization, our analyses incorporate cluster volume and cluster focus but we neglect the unit's total patient volume. Since our analysis only considers two clusters, an increase in total volume can be achieved (i) via an increase in the patient's own cluster volume and (ii) via an increase in the other cluster volume, which is captured via reduced cluster focus levels. As such, the total volume of both clusters is incorporated via the denominator of the cluster focus variable and does not need to be included separately.

3.5. Control Variables

Several variables were used in this study to control for potential confounders at the individual and NICU level. To account for differences in individual patient characteristics, we control for the admission month, risk of illness and comorbidity and complexity level. To capture the risk of illness and risk differences within the cluster, we incorporate the birth weight information (in its continuous form but centered within the cluster). In addition, we control for the patient's comorbidity and complexity level (PCCL). This information is extracted from a patient's diagnosis-related group (DRG), which classifies patients by conditions and procedures. The DRG complexity measure is based on all actual secondary diagnoses in the discharge records, whereby every secondary diagnosis obtains a CC score (CCL Wert). The German DRG system then calculates a patient-level complexity score (PCCL Wert) based upon the aggregated CC scores for each patient. The different PCCL levels can be inferred from the letters in the fourth digit of the DRG code, whereby A indicates the highest category, B the second-highest category, etc. We capture these differences in PCCL scores using a categorical variable distinguishing between four categories of PCCL.

In addition, we control for the occupancy level a patient was exposed to on his or her day of discharge to ensure that a potential beneficial volume effect on a reduced length of stay might not be a reflection of early discharge due to congestion. Therefore, we calculate the occupancy level (midnight census) a patient i experienced on the day of discharge d in NICU n as the number of patients treated in n on day d relative to the n 's capacity. In line with the literature (Berry Jaeker and Tucker 2016, Kuntz et al. 2015), the capacity of each unit n is given by the maximum number of patients treated in n on any given day t during the observation period $t = 1, \dots, T$.

At the NICU level, we control for unit characteristics via the NICU level of care (level 1 or level 2) and staff-mix differences using the number of neonatologists as a proportion of all NICU physicians.

Descriptive statistics of all model covariates are provided in Appendix S1, chapter 1.

3.6. Data Sample and Exclusions

Of the 7576 eligible infants, patients were excluded because no information about their length of stay or severity of illness was provided ($n = 637$) or because patients died during their NICU stay ($n = 59$). To avoid censoring when calculating the occupancy level a patient experienced on the discharge day, we excluded patients discharged in January or February 2013 ($n = 1433$), because the average length of stay of VLBW infants exceeded 1 month. Lastly, patients were excluded for which no sufficient DRG information was provided to extract the PCCL scores ($n = 1427$ patients). This resulted in an overall sample of 4020 patients from 18 NICUs.

To assess the representativeness of our sample, we use the nationwide quality report of the Institute for Applied Quality Improvement and Research in Health Care (2013), which contains birth weight information for all newborn infants admitted to German NICUs in 2013. This report publishes birth weight information in seven categories, and a comparative analysis between the proportions yields the following results (included infants in our sample vs. all NICU newborns in the population): <500 g: 0.9% vs. 0.5%; 500–1499 g: 18.2% vs. 8.9%; 1500 g–2499 g: 33.4% vs. 30.2%; >2499 g: 47.5% vs. 60.5%). We test the equality of proportions in these seven categories and observe significant differences in three categories, while four categories do not show significant differences in the proportions (we account for multiple testing). Based on these birth weight categories, we see that our sample does not deviate substantially from the nationwide birth weight distribution.

3.7. Statistical Analysis

To account for the hierarchy in our data where patients are nested within NICU clusters, we rely on multilevel regression models. These models are increasingly used in Operations Management (see, e.g., Ang et al. 2002, DeHoratius and Raman 2008, McDermott and Stock 2011) and are appropriate if observations are not independent from each other due to sharing group characteristics. Multilevel models take individual and group level variation into account while estimating group level regression coefficients (Gelman and Hill 2006), which is an important consideration in our context, where the individual length of stay is supposed to be explained by a standardized cluster volume (and standardized cluster focus) that only varies at the NICU cluster level.

We estimate the NICU length of stay of patient i in cluster c in NICU n as follows:

$$\begin{aligned} \ln(LOS_{icn}) = & \beta_0 + \beta_1 Vol_{cn} + \beta_2 Foc_{cn} + \beta_3 Vol_{cn} \times C_{icn} \\ & + \beta_4 Foc_{cn} \times C_{icn} + \beta_5 C_{icn} + \beta_6 X_{icn} + u_{cn} \\ & + \epsilon_{icn}, \end{aligned}$$

where Vol_{cn} denotes the standardized cluster volume of cluster c in NICU n , Foc_{cn} denotes the standardized cluster focus of cluster c in NICU n , C_{icn} is equal to 1 if patient i belongs to cluster 2 and is admitted to NICU n , X_{icn} denotes the vector of control variables, $u_{cn} \sim N(0, \tau^2)$ denotes the random error at the NICU cluster level, $\epsilon_{icn} \sim N(0, \sigma^2)$ the idiosyncratic error, and u_{cn} and ϵ_{icn} are assumed to be orthogonal. We estimate our models with the *mixed* command, Stata Version 14.2. We allow the idiosyncratic errors to correlate within groups and cluster standard errors at the NICU cluster level.

4. Results

The descriptive statistics are shown in Table 1. At the individual patient level, we observe that cluster 1 infants have a substantially longer average NICU length of stay than cluster 2 patients (37.7 vs. 9.2 days), yet the coefficient of variation is smaller for cluster 1 ($29.1/37.7 = 0.772$) than cluster 2 patients ($8.9/9.2 = 0.967$), indicating more heterogeneity in NICU length of stay for the latter. At the organizational level, we observe substantial variations in the patients treated in each unit. While NICUs, on average, treated more cluster 2 (211.6) than cluster 1 patients (51.6), the coefficient of variation is smaller for cluster 2 ($86.2/211.6 = 0.407$) than for cluster 1 patients ($35.9/51.6 = 0.696$), indicating a larger dispersion for the latter. These substantial differences in distributions support our decision to standardize these variables.

The results of the multilevel model are shown in Table 2, which lists our main variables in the first panel, notes the control variables in the second panel and provides basic model statistics in the bottom panel. Within all models, there is substantial variation at the group level (clusters in NICUs), as indicated by the intraclass correlation (ICC). This supports our choice of a multi-level model. We will base the inference on the full model (4) and present the other models for completeness and to allow the reader to assess the differences in coefficients between the models.

Our first set of hypotheses relates to the volume effects. Hypotheses 1a and 1b, which state that an increase in cluster volume is associated with decreasing length of stay, is supported for cluster 1 ($\beta_1 = -0.546, p < 0.001$) and cluster 2 ($\beta_1 + \beta_3 =$

Table 1 Descriptive Statistics of Individual and Organizational Characteristics

	Cluster 1 (VLBW)	Cluster 2 (non-VLBW)
Individual characteristics	$N = 768$	$N = 3252$
Length of stay in days, mean (SD)	37.7 (29.1)	9.2 (8.9)
Birth weight in g, mean (SD)	1074.7 (303.7)	2787.8 (764.2)
PCCL, highest level	12.76%	10.85%
PCCL, second-highest level	27.47%	36.35%
PCCL, third-highest level	30.47%	43.97%
PCCL, residual levels	29.30%	8.83%
Occupancy level on discharge day, mean (SD)	64.4% (19.2%)	65.8% (17.7%)
Organizational characteristics	$N = 18$	$N = 18$
Cluster volume, mean (SD)	51.6 (35.9)	211.6 (86.2)
Cluster focus, mean (SD)	19.6% (14.3%)	80.4% (14.3%)
Level 1 NICU	83.30%	83.30%
Proportion of neonatologists, mean (SD)	27.1% (10.7%)	27.1% (10.7%)

Table 2 Effect of Cluster Volume and Cluster Focus on Log. Length of NICU Stay

	Model (1)	Model (2)	Model (3)	Model (4)
Volume	-0.429*** (0.113)		-0.694*** (0.090)	-0.546*** (0.099)
Focus		-0.025 (0.103)	0.261** (0.111)	-0.096 (0.059)
Volume × Cluster 2				-0.183 (0.339)
Focus × Cluster 2				0.612*** (0.120)
Cluster 2 (non-VLBW)	-1.765*** (0.143)	-1.665*** (0.164)	-1.768*** (0.126)	-1.796*** (0.151)
Occupancy	Yes	Yes	Yes	Yes
Birth weight	Yes	Yes	Yes	Yes
CC Score	Yes	Yes	Yes	Yes
% Neonatologists	Yes	Yes	Yes	Yes
NICU level	Yes	Yes	Yes	Yes
Admission month	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes
Observations	4020	4020	4020	4020
Number of groups	36	36	36	36
ICC: 95% CI	[0.185; 0.369]	[0.247; 0.439]	[0.150; 0.320]	[0.087; 0.236]

Note: Standard errors clustered on group level (clusters in NICUs). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

$-0.546 - 0.183 = -0.729, p < 0.05$). Hypothesis 2, stipulating that the volume length of stay association is weaker for cluster 2 than for cluster 1, is not supported as the interaction term is insignificant ($\beta_3 = -0.183, p = 0.589$).

Our second set of hypotheses relates to the focus effects and we expected to find an increase in cluster focus associated with decreasing length of stay for cluster 1 (Hypothesis 3a) and cluster 2 (Hypothesis 3b). We find neither support for Hypothesis 3a ($\beta_2 = -0.096, p = 0.105$) nor for Hypothesis 3b ($\beta_2 + \beta_4 = -0.096 + 0.612 = 0.516, p < 0.001$). In fact,

we even find the reverse for Hypothesis 3b, that is, an increase in cluster focus is associated with an increase in length of stay. The reverse result is explained by finding strong support for Hypothesis 4, which argues that the focus length of stay association is weaker for cluster 2 than for cluster 1 ($\beta_4 = 0.612$, $p < 0.001$).

In order to assess the effect sizes, we predict length of stay for varying levels of cluster volume and cluster focus, leaving all other variables as observed.² Figure 1 outlines these counterfactual predictions averaged across patients. If cluster volume increases by one standard deviation from the mean, length of stay decreases from 50.6 days to 29.3 days for VLBW infants in cluster 1 and declines from 9.1 days to 4.4 days for non-VLBW infants in cluster 2. This corresponds to a decrease of 42.1% for VLBW infants and 51.6% for non-VLBW infants. If cluster focus increases by one standard deviation from the mean, length of stay falls from 43.0 days to 39.1 days for VLBW infants in cluster 1 and increases from 9.7 days to 16.2 days for non-VLBW infants in cluster 2. This corresponds to a decrease of 9.1% for VLBW infants, albeit it is insignificant, but to a 67% increase for non-VLBW infants.

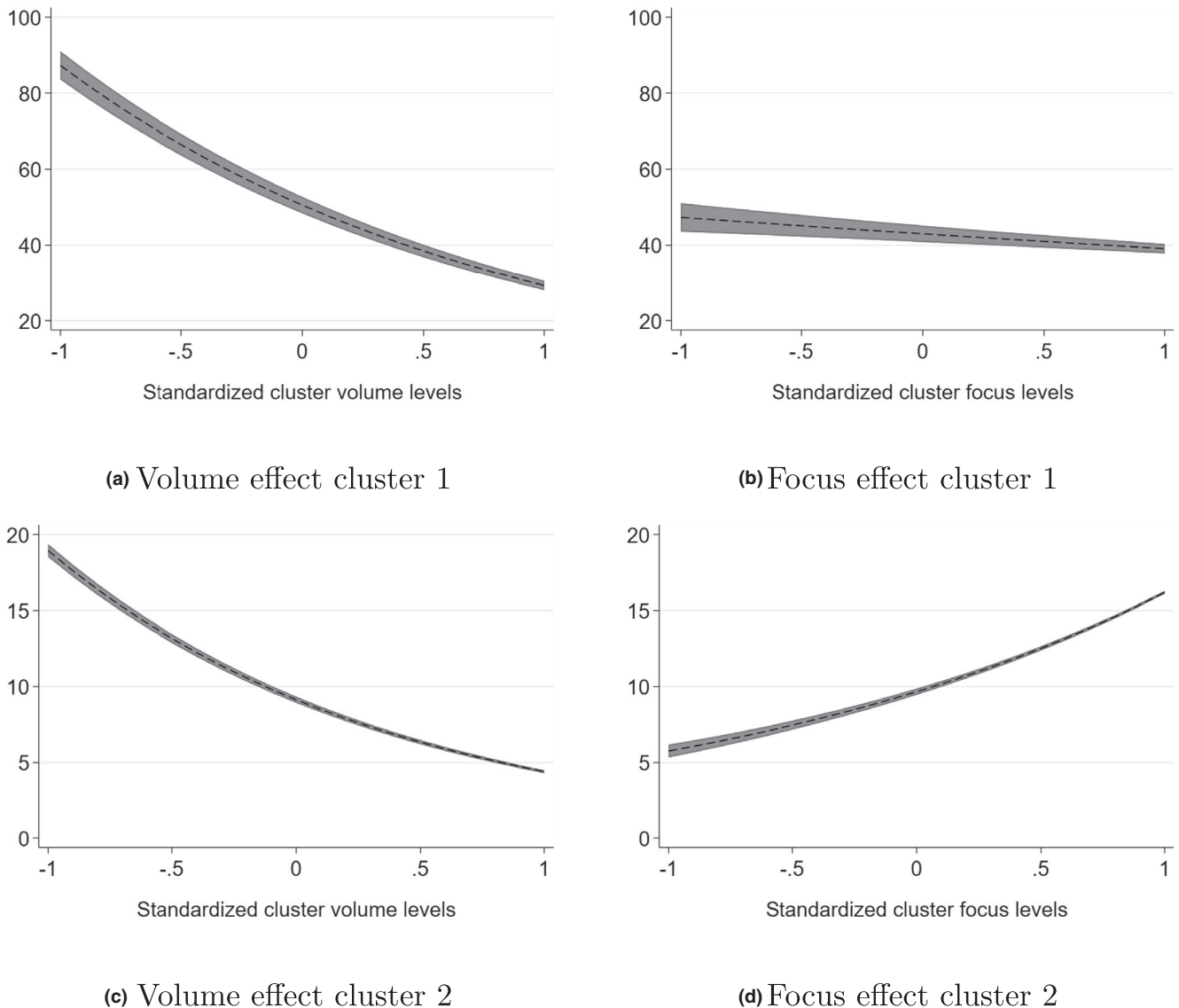
5. Robustness and Limitations

Several tests were conducted to check the robustness of our results. We provide the details in Appendix S1 and present the high-level results here. Firstly, we used different model specifications and clustering levels (chapter 3, Appendix S1), and the significant results of these different model specifications are in line with our main results reported herein. In addition, we conducted sub-sample analyses to test the nonlinear effects of cluster volume and cluster focus (chapter 4, Appendix S1). We opted for sub-sample analyses because nonlinearity patterns might differ between the two clusters we considered. While there is no evidence of nonlinearity for the sub-sample of VLBW infants (cluster 1), there is some indication of nonlinearity for focus in the sub-sample of non-VLBW infants (cluster 2); however, we shall apply prudence here since the group level variation in this sub-sample analysis is only based upon $N = 18$ groups.

Secondly, the patient's underlying health status is most likely not observed in its completeness. As long as the health status is not related to cluster volume and cluster focus, this will not affect our results. However, if the choice of the NICU and, subsequently, the levels of volume and focus the patient is exposed to are affected by the patient's underlying health status, two potential situations might occur. In situation one, high-volume (high-focus) NICUs are more likely to be admitting sicker newborns. Sicker newborns, however,

require a longer length of stay, and if we ignore this potential selection effect, volume and length of stay are spuriously related, with a higher volume correlating with a higher length of stay. Our findings are, however, the opposite, meaning that if NICUs with a higher volume do indeed attract sicker infants, the effect that we find is underestimated. In situation two, high-volume (high-focus) NICUs attract healthier infants. Healthier newborns are more likely to require a shorter length of stay, and if we ignore this selection effect, volume and length of stay are spuriously related, with a higher volume correlating with a decreasing length of stay, which would be in line with our findings. A similar concern arises if people who are at risk for more complicated or severe VLBW infants (lower socio-economic status and incomplete prenatal care) opt for NICUs with lower volumes of VLBW infants. While our data does not allow us to control for differences in socio-economic status, we seek to tackle this aspect by focusing on geographic areas where the patient's choice set of alternative NICUs is more limited. Based on these analyses (chapter 5, Appendix S1), we do not find strong evidence of potential selection effects for VLBW infants in cluster 1. For the majority of preterm infants, timely access to a nearby hospital with a public mandate to treat these newborns is crucial. Selection effects that might occur in a decision process for which there is less time pressure (as, for instance, in the case of elective procedures) are therefore less likely to occur. For cluster 2 patients, we cannot rule out that the volume results are affected by selection effects. One explanation could be a difference in selection and transfer procedures for these patients, yet we lack the information to test this reasoning. Our data do not provide information on whether infants were born at the NICU directly or transferred to the NICU from another health care provider; the latter implies the possibility of a more informed decision process.

Thirdly, we might be concerned with the fact that high-volume (high-focus) NICUs transfer patients more quickly to downstream units and that the shortened episode of the NICU length of stay is offset by a longer length of stay in other units. Focusing on the processes within the NICU and, subsequently, on the process outcomes in that organizational unit, we are less concerned with process outcomes of downstream units, provided that the NICU, as the leading operating unit, does not transfer patients prematurely. Premature discharges or transfers are frequently the result of capacity shortages and the need to free up beds for incoming patients of higher severity (KC and Terwiesch 2012). However, this important operational factor of premature discharge due to high occupancy has been incorporated into our econometric model.

Figure 1 Expected Length of Stay in Days Incl. 95% CI for Different Levels of Cluster Volume and Cluster Focus

Finally, the cross-sectional nature of our data requires prudence in arguing causally, as our empirical findings rather reflect associations. We test the effects of volume and focus on the NICU setting, which might restrict generalizability beyond the NICU setting because NICUs are more likely subject to stronger regulations than other clinical divisions. The restricted generalizability is, however, partially offset by the advantages of having a clearly defined patient group with a limited chance of being routed to other units and clearly defined medical clusters within this patient group.

6. Discussion and Conclusion

This paper is concerned with analyzing the impacts of volume and focus for complex patients in different

medical clusters. In the context of NICUs, we distinguish between two clusters that differ based on their inherent risk of medical complications and heterogeneity in disease pattern. Our first finding suggests that an increase in cluster volume is associated with better process outcomes, and this finding is in line with the body of literature supporting positive volume outcome relationships (e.g., Birkmeyer et al. 2002, Gaynor et al. 2005, KC and Staats 2012, Profit et al. 2013, 2016, Reagans et al. 2005, Theokary and Ren 2011). Our results also indicate that this positive relationship holds for both cluster. Integrating arguments of learning from related and unrelated variety (Boh et al. 2007, KC and Staats 2012, Schilling et al. 2003, Staats and Gino 2012), forgetting (Ramdas et al. 2018), team familiarity (Huckman and Staats 2011, Huckman et al. 2009, Reagans et al. 2005, Staats 2012)

and process uncertainty (Kuntz et al. 2019), we expected the volume outcome effect to be weaker for the cluster with lower medical severity and higher operational heterogeneity. Despite the fact that the theoretical arguments indeed stipulate weaker volume outcome effects for the heterogeneous cluster, empirically we do not find any evidence of a volume effect difference, that is, the volume outcome effect was equally strong for both clusters. We acknowledge that our empirical analysis focuses on the aggregated effect modification and cannot distinguish between the different theoretical mechanisms. Consequently, we cannot empirically assess potential inter-dependencies between these mechanisms.

Our second result shows that an increase in the cluster focus does not seem to affect the process outcomes for complex patients with high medical severity and low operational heterogeneity. This implies that, for this cluster, process outcomes are driven by volume and not by focus. For patients with lower medical risk but higher operational heterogeneity, we find that an increase in cluster focus is associated with worse process outcomes. In line with arguments concerning reduction of work routines (Huckman and Zinner 2008) and availability of complementary services outside the focal activity (Clark and Huckman 2012), we were indeed expecting a weaker focus outcome effect for the heterogeneous lower-risk cluster. We do, however, not only find a weaker focus outcome effect, but we also find detrimental focus effects for this patient group. Our results suggest that as long as the unit has moderate levels of cluster focus for the heterogeneous lower-risk patients, there are still sufficient complementary services outside the cluster available. Complementary services outside the focal activity can generate spillovers, which seem to benefit the heterogeneous lower-risk patients. Akin to Clark and Huckman (2012), we cannot identify whether the spillover effect is driven by knowledge transfer, information exchange or physical proximity. Expanding this knowledge base is thus not only of interest to scholars but also for practitioners who seek to ensure or stimulate relevant spillovers between patient clusters within their organizational context.

Hospitals provide a variety of services for various patient groups and not all patient groups benefit equally from operational factors such as volume and focus. The work by Kuntz et al. (2019) shows that it is beneficial for complex patients if they are routed to the same department instead of experiencing fragmented service provision. Our analysis of complex neonatal patients indicate equivalent implications; hospitals should also avoid separation of complex patients within clinical departments. The implications are also relevant for hospital networks and favor pooling complex patients and thereby increasing the

volume rather than providing services at multiple locations. Obviously, the effectiveness of such an organizational design depends on effective co-operation between professional groups and participating hospitals. A more substantial and determining factor, however, is the location of a hospital. Clearly, such agreements are more feasible in areas of high population density where multiple hospitals exist in close proximity. If the distance between the collaborating hospitals is too large, this will impede timely access to care provision, which is particularly relevant for neonatal intensive care. Another implementation challenge is the question of how profit and risk-sharing could be arranged between the collaborating entities and how the streaming of patients should occur to minimize inter-organizational transfers. How to design such collaborations and what distance can be deemed acceptable are important research questions in themselves and provide a fruitful avenue for future research to expand upon our study.

Acknowledgments

We are indebted to the participating infants, their parents, and to the medical and nursing teams in participating NICUs, as well as all colleagues of the HSR-NICU research project. For their excellent cooperation, we thank Professor Ludwig Kuntz as head of the HSR-NICU research project, Professor Christiane Woopen as head of the sub-project “Ethical Aspects,” Professor Bernhard Roth as head of the sub-project “Medical Aspects,” Professor Rainer Riedel as head of the sub-project “Economic Outcomes,” and Professor Holger Pfaff as head of the collaborating Institute of Medical Sociology, Health Services Research and Rehabilitation Science of the University of Cologne. We thank Michael Becker-Peth and our colleagues from the HSMO Science Club for their helpful comments on earlier versions of the manuscript. Finally, we thank two anonymous reviewers, the senior editor, and the department editor for their constructive guidance during the review process. This study was funded by the Federal Ministry of Education and Research (project grant 01GY1152) and we declare no conflict of interest.

Notes

¹This argument is supported by additional analyses presented in Appendix S1, chapter 2.

²To obtain predictions evaluated in units of days, we re-transform as follows: $\tilde{y} = \exp(\text{Ln}(\overline{LOS}_{icn})) \times \exp(\hat{\tau}^2/2) \times \exp(\hat{\sigma}^2/2)$.

References

- Ang, S., S. Slaughter, K. Yee Ng. 2002. Human capital and institutional determinants of information technology compensation: Modeling multilevel and cross-level interactions. *Management Sci.* 48(11): 1427–1445.
- Bartels, D. B., D. Wypij, P. Wenzlaff, O. Dammann, C. F. Poets. 2006. Hospital volume and neonatal mortality among very low birth weight infants. *Pediatrics* 117(6): 2206–2214.

- Berry Jaeker, J. A., A. L. Tucker. 2016. Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity. *Management Sci.* **63**(4): 1042–1062.
- Best, T. J., B. Sandikci, D. D. Eisenstein, D. O. Meltzer. 2015. Managing hospital inpatient bed capacity through partitioning care into focused wings. *Manuf. Serv. Oper. Manag.* **17**(2): 157–176.
- Birkmeyer, J., A. Siewers, E. Finlayson, T. Stukel, F. Lucas, I. Bastista, H. G. Welch, D. Wennberg. 2002. Hospital volume and surgical mortality in the United States. *N. Engl. J. Med.* **246**(15): 1128–1137.
- Boh, W., S. Slaughter, J. Espinosa. 2007. Learning from experience in software development: A multilevel analysis. *Management Sci.* **53**(8): 1315–1331.
- Chung, J. H., C. S. Pibbs, W. J. Boscardin, G. F. Kominski, A. N. Ortega, J. Needleman. 2010. The effect of neonatal intensive care level and hospital volume on mortality of very low birth weight infants. *Med. Care* **48**(7): 635–644.
- Clark, J., R. Huckman. 2012. Broadening focus: Spillovers, complementarities and specialization in the hospital industry. *Management Sci.* **58**(4): 708–722.
- DeHoratius, N., A. Raman. 2008. Inventory record inaccuracy: An empirical analysis. *Management Sci.* **54**(4): 627–641.
- Edmondson, A. 1999. Psychological safety and learning behavior in work teams. *Adm. Sci. Q.* **44**(2): 350–383.
- Freeman, M., N. Savva, S. Scholtes. 2019. Economies of scale and scope in hospitals. Working paper, INSEAD.
- Gaynor, M., H. Seider, W. Vogt. 2005. The volume-outcome effect, scale economies, and learning-by-doing. *Am. Econ. Rev.* **95**(2): 243–247.
- Gelman, A., J. Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.
- Green, L. V. 2012. Om forum: The vital role of operations analysis in improving healthcare delivery. *Manuf. Serv. Oper. Manag.* **14**(4): 488–494.
- Huckman, R., B. Staats. 2011. Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manuf. Serv. Oper. Manag.* **13**(3): 310–328.
- Huckman, R., D. Zinner. 2008. Does focus improve operational performance? Lessons from the management of clinical trials. *Strat. Manage. J.* **29**: 173–193.
- Huckman, R., B. Staats, D. Upton. 2009. Team familiarity, role experience, and performance: Evidence from Indian software services. *Management Sci.* **55**(1): 85–100.
- Institut für Qualitätssicherung und Transparenz im Gesundheitswesen. 2018. Strukturabfrage. Available at <https://perinatalzentren.org/strukturabfrage.php> (accessed date December 3, 2018).
- Institute for Applied Quality Improvement and Research in Health Care. 2013. Quality report 2013. Available at <http://sqg.de/sqg/upload/CONTENT/Qualitaetsberichte/2013/AQUA-Qualitaetsreport-2013.pdf> (accessed date September 15, 2017).
- KC, D., B. Staats. 2012. Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manuf. Serv. Oper. Manag.* **14**(4): 618–633.
- KC, D., C. Terwiesch. 2011. The effects of focus on performance: Evidence from California hospitals. *Management Sci.* **57**(11): 1897–1912.
- KC, D., C. Terwiesch. 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *Manuf. Serv. Oper. Manag.* **14**(1): 50–65.
- Kuntz, L., R. Mennicken, S. Scholtes. 2015. Stress on the ward: Evidence of safety tipping points in hospitals. *Management Sci.* **61**(4): 754–771.
- Kuntz, L., S. Scholtes, S. Sülz. 2019. Separate & concentrate: Accounting for patient complexity in general hospitals. *Management Sci.* **65**(6): 2482–2501.
- Lee, K. S., N. Paneth, L. M. Gartner, M. Pearlman. 1980. The very low-birth-weight rate: Principal predictor of neonatal mortality in industrialized populations. *J. Pediatr.* **97**(5): 759–764.
- McDermott, C., G. Stock. 2011. Focus as emphasis: Conceptual and performance implications for hospitals. *J. Oper. Manag.* **29**(6): 616–626.
- Narayanan, S., S. Balasubramanian, J. Swaminathan. 2009. A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Sci.* **55**(11): 1861–1876.
- Payne, N. R., J. H. Carpenter, G. J. Badger, J. D. Horbar, J. Rogowski. 2004. Marginal increase in cost and excess length of stay associated with nosocomial bloodstream infections in surviving very low birth weight infants. *Pediatrics* **114**(2): 348–355.
- Phibbs, C. S., L. C. Baker, A. B. Caughey, B. Danielsen, S. K. Schmitt, R. H. Phibbs. 2007. Level and volume of neonatal intensive care and mortality in very-low-birth-weight infants. *N. Engl. J. Med.* **356**(21): 2165–2175.
- Profit, J., J. A. Zupancic, J. B. Gould, K. Pietz, M. A. Kowalkowski, D. Draper, S. J. Hysong, L. A. Petersen. 2013. Correlation of neonatal intensive care unit performance across multiple measures of quality of care. *JAMA Pediatr.* **167**(1): 47–54.
- Profit, J., J. B. Gould, M. Bennett, B. A. Goldstein, D. Draper, C. S. Phibbs, H. C. Lee. 2016. The association of level of care with NICU quality. *Pediatrics* **137**(3): e20144210.
- Ramdas, K., K. Saleh, S. Stern, H. Liu. 2018. Variety and experience: Learning and forgetting in the use of surgical devices. *Management Sci.* **64**(6): 2590–2608.
- Reagans, R., L. Argote, D. Brooks. 2005. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Sci.* **51**(6): 869–881.
- Schilling, M., P. Vidal, R. Ployhart, A. Marangoni. 2003. Learning by doing something else: Variation, relatedness, and the learning curve. *Management Sci.* **49**(1): 39–56.
- Skinner, W. 1974. The focused factory. *Harv. Bus. Rev.* **52**(3): 113–121.
- Staats, B. 2012. Unpacking team familiarity: The effects of geographic location and hierarchical role. *Prod. Oper. Manag.* **21**(3): 619–635.
- Staats, B., F. Gino. 2012. Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Sci.* **58**(6): 1141–1159.
- Theokary, C., Z. Ren. 2011. An empirical study of the relations between hospital volume, teaching status, and service quality. *Prod. Oper. Manag.* **20**(3): 303–318.
- UK Neonatal Staffing Study Group. 2002. Patient volume, staffing, and workload in relation to risk-adjusted outcomes in a random stratified sample of UK neonatal intensive care units: A prospective evaluation. *Lancet* **359**(9301): 99–107.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Boundaries of focus and volume-background information and additional analyses.