

The Effect of Recent Workload Exposure on Nurses’ Perceived Workload

Carri W. Chan

Decision, Risk, and Operations, Columbia Business School, cwchan@gsb.columbia.edu

Yi Chen

Department of Industrial Engineering and Decision Analytics, Hong Kong University of Science and Technology, yichen@ust.hk

Jing Dong

Decision, Risk, and Operations, Columbia Business School, jing.dong@gsb.columbia.edu

Sarah C. Rossetti

Department of Biomedical Informatics, Columbia University sac2125@cumc.columbia.edu

Problem definition: Recent and ongoing nursing shortages have highlighted the importance of managing nursing workload in hospital inpatient care. Hospitals often rely on order-based workload measures to support staffing and assignment decisions, but these measures may not fully capture the workload experienced by frontline nurses. We study workload misalignment, defined as the difference between a nurse’s professional-judgment workload assessment and an automated order-based workload benchmark. **Methodology/results:** We leverage patient–nurse–shift level data from a medical ICU, including detailed patient-to-nurse assignments, an automated order-based workload measure, and a nurse-recorded professional-judgment workload measure for each patient during each shift. We estimate how a nurse’s prior-shift workload exposure affects current workload misalignment while addressing challenges including assignment endogeneity, missing values, and measurement error. Our results show that heavier prior-shift workload systematically shifts nurses’ professional-judgment assessments upward relative to the order-based benchmark. **Managerial implications:** Understanding workload misalignment can help managers improve assignment allocation where standardized order-based workload measures may understate frontline burden. We develop a perception-aware patient-to-nurse assignment policy that uses recent workload exposure to reduce high predicted perceived workload while monitoring objective-workload fairness. The framework complements charge-nurse judgment and supports more responsive workload balancing.

Key words: healthcare operations, intensive care unit, nursing workload, workload balancing

History:

1. Introduction

Managing nursing workload is a central operational challenge in inpatient care. Recent and ongoing nursing shortages have highlighted the valuable and skilled work that nurses provide around the

clock ([Berlin et al. 2022](#), [Stiegler et al. 2021](#)). At the same time, intense and sustained nursing workload has been associated with nurse burnout, job dissatisfaction, turnover intentions, and patient safety concerns ([Aiken et al. 2002](#), [Vahey et al. 2004](#), [Halbesleben et al. 2008](#)). These challenges are particularly salient in intensive care units (ICUs), where patients often require frequent monitoring, coordination, and rapid clinical response.

In practice, hospitals often rely on automated workload indicators, which typically capture orders-based measures, to support staffing and patient-to-nurse assignment decisions. These indicators are useful because they are standardized, scalable, and available at the time assignments are made. However, nursing workload is not determined only by physician orders or measurable acuity. It also includes non-order-driven care demands, such as communication needs, behavioral support, family interactions, coordination burden, and other contextual demands that may not be fully captured in structured data. Nurses' professional judgment therefore provides a complementary assessment of patient workload. Because nurses observe patients directly at the bedside and integrate clinical and contextual information, their professional judgment can be interpreted as a measure of clinically perceived workload: the workload burden they assess based on their expertise and experience ([Levy et al. 2024](#), [van der Mark et al. 2021](#)).

Automated workload indicators and professional-judgment assessments may therefore diverge because they capture different dimensions of care burden. We refer to this divergence as workload misalignment, and focus especially on scenarios where a nurse's professional-judgment assessment exceeds the order-based benchmark (positive workload misalignment). Such misalignment can serve as an operational risk signal, identifying assignments for which standardized measures may understate the burden experienced by frontline nurses. This is important because high perceived workload has been associated with nurse burnout, job dissatisfaction, and patient-safety concerns ([Holland et al. 2019](#), [Jourdain and Chênevert 2010](#), [MacPhee et al. 2017](#), [Phillips 2020](#)). Understanding when and why positive workload misalignment occurs can help managers identify potentially burdensome assignments and design more responsive patient-to-nurse assignment policies.

In this work, we study this issue through the lens of workload carryover. Prior nursing and occupational-health studies have documented that fatigue, stress, and performance impairment can persist across consecutive or compressed shifts ([Thompson et al. 2017](#), [Thompson 2019](#), [Haluza et al. 2019](#), [Cockerham et al. 2023](#)). Building on these evidences, we ask whether the recent workload exposure also carries over into nurses' professional-judgment workload assessments and contributes to positive workload misalignment. The direction of this effect, if any, is not obvious a priori. On one hand, a demanding prior shift may deplete physical and cognitive resources, increase fatigue or stress, and make a current assignment feel more burdensome than suggested by an automated order-based benchmark ([Greenglass and Burke 2001](#), [Stimpfel et al. 2012](#), [Khamisa et al. 2015](#)).

This mechanism predicts a positive relationship between recent workload exposure and workload misalignment. On the other hand, perceived workload may also be shaped by anchoring: a nurse who has recently experienced an especially demanding shift may perceive a subsequent assignment as less burdensome by comparison (Sherif et al. 1958, Furnham and Boo 2011). This mechanism predicts the opposite relationship. We therefore empirically test the hypothesis that *greater recent workload exposure increases the gap between a nurse's clinically perceived workload and the order-based workload*, and quantify the magnitude of this effect.

Our empirical setting is a medical intensive care unit (MICU) at a large academic medical center. We leverage a novel dataset from a commercial workload-management system used by the study hospital (OptiLink 2012). The data contain detailed patient-to-nurse assignment information and two workload measures for each patient-shift. DDCIntensity is an automated order-based workload indicator generated from physician orders, whereas, PJIntensity is entered by the staff nurse caring for the patient and reflects the nurse's professional judgment of workload. Both are measured on four levels: low (1), average (2), high (3), and extremely high (4). We interpret DDCIntensity as an automated order-based workload benchmark and PJIntensity as a nurse-reported measure of clinically perceived workload. Our outcome is their difference, PJIntensity minus DDCIntensity, and we measure recent workload exposure using the nurse's order-based workload in the most recent prior working shift within a seven-day lookback window. Our empirical results provide robust evidence that heavier prior-shift workload increases the discrepancy between nurses' perceived workload and the order-based benchmark.

This finding has direct implications for patient-to-nurse assignment. If nurses with heavier recent workload are more likely to experience a given assignment as more burdensome than indicated by its order-based workload, then balancing only contemporaneous workload may be insufficient. Incorporating recent workload history into assignment decisions may help smooth workload over time and reduce repeated exposure to highly burdensome assignments. Such policies may create a safer and more supportive working environment for nurses, ultimately improving both nurse satisfaction and patient care (Sir et al. 2015).

Motivated by this insight, we develop a perception-aware patient-to-nurse assignment framework. Using our empirical model, we predict the perceived workload associated with each feasible patient-nurse assignment and formulate an integer program to find the patient-nurse matching that minimizes total predicted perceived workload. Because the set of patients within a shift is fixed, the total order-based workload is fixed as well, so reducing predicted perceived workload is closely related to reducing positive workload misalignment. Relative to the hospital's observed assignments, our proposed policy reduces the incidence of extremely high predicted perceived workload by 11% while maintaining a more balanced allocation of order-based workload across nurses.

In summary, our work makes two main contributions. First, we provide causal evidence on workload carryover effect in nursing: greater recent workload exposure increases the extent to which nurses' professional-judgment workload assessments exceed an automated order-based benchmark. Identification is complicated by nonrandom missingness in PJIntensity, endogenous nurse assignments, and measurement error in prior workload. We address sample selection using a Heckman correction, endogeneity using instrumental variables, and measurement error through sensitivity analyses based on alternative imputation schemes. Across these analyses, the estimated effect remains positive and statistically significant, maintaining a similar magnitude.

Second, we translate this empirical finding into a patient-to-nurse assignment policy that explicitly accounts for predicted perceived workload. Through trace-based simulation, we show that the proposed policy reduces exposure to extremely high predicted perceived workload while maintaining comparable or improved order-based workload balance across nurses. The resulting framework provides a practical way to incorporate recent workload history into patient-assignment decisions and complements existing order-based workload measures..

1.1. Literature Review

Our work contributes to several streams of literature on nursing and healthcare operations.

Medical Literature on Nursing Workload Measurement. A large medical literature studies how nursing workload should be measured, particularly in intensive care settings. Earlier work emphasizes order-, activity-, and acuity-based measures ([Alghamdi 2016](#), [Hart and Staveland 1988](#), [Miranda et al. 2003](#)), while subsequent studies highlight care demands that may not be fully captured by such measures, including coordination, documentation, communication, and other indirect care activities([Myny et al. 2011](#)). More recent work therefore considers nurses' perceived workload and professional judgment as complementary dimensions of workload ([Hoogendoorn et al. 2021](#), [Kapu et al. 2016](#), [Levy et al. 2024](#), [van der Mark et al. 2021](#)). For example, [Dhaini et al. \(2022\)](#), [Chen et al. \(2024\)](#) examine how shift and patient characteristics are associated with perceived workload. High workload and perceived workload have also been linked to care left undone, patient-safety concerns, burnout, and other adverse outcomes ([Banda et al. 2022](#), [Liu et al. 2018](#), [MacPhee et al. 2017](#), [Phillips 2020](#), [Aiken et al. 2002](#), [Laschinger and Leiter 2006](#), [Shah et al. 2021](#)).

A related nursing and occupational-health literature documents that the effects of demanding work can persist across shifts. Studies of nurses working consecutive or compressed 12-hour shifts find deterioration in physical and cognitive performance, increased fatigue and sleepiness, and incomplete recovery between shifts ([Thompson et al. 2017](#), [Thompson 2019](#), [Geiger-Brown et al. 2012](#), [Haluza et al. 2019](#), [Blasche et al. 2017](#), [Cockerham et al. 2023](#)). Extended work hours have also been associated with reduced vigilance, patient-safety concerns, burnout, job dissatisfaction,

and intention to leave (Scott et al. 2006, Dall’Ora et al. 2015). Building on these literatures, we study whether recent objective workload exposure changes nurses’ subsequent professional-judgment workload assessments relative to an automated order-based benchmark. Our contribution is to identify this carryover relationship empirically at the patient–nurse–shift level and quantify its implications for assignment decisions.

Impact of Workload in Healthcare Operations Management. Workload has also been widely studied in healthcare operations management. Much of this literature examines how unit-level aggregated workload affects patient outcomes, service times, discharge decisions, or clinician behavior (see, e.g., Kc and Terwiesch (2009), Berry Jaeker and Tucker (2017), Luo et al. (2022), Kc and Terwiesch (2012), Long and Mathews (2018), Soltani et al. (2022)). Related work shows that clinicians adapt their behavior under high workload: they may ration discretionary services or increase referrals (Freeman et al. 2021), initiate tasks earlier (Batt and Terwiesch 2017), change patient prioritization (Li et al. 2021), or alter task selection and sequencing (Ibanez et al. 2018, Kc et al. 2020). In nursing settings, workload has likewise been associated with care processes and staff well-being (Ball et al. 2014, Liu et al. 2018, Carayon and Gurses 2008). Recent work further shows that formal assignments may understate the work nurses actually perform because nurses may help patients beyond those formally assigned to them (Jiang et al. 2025). Our study complements this literature by focusing on a different mechanism: how objective workload from a prior shift carries over into a nurse’s assessment of current patient workload.

Nursing Workload Management It has long been recognized that it is important and highly nontrivial to properly manage the nursing workload. A large operations literature studies nurse staffing and scheduling, typically balancing staffing cost against service quality or staffing uncertainty (see, e.g., Green et al. (2013), Wang and Gupta (2014), Kim and Mehrotra (2015)). A more closely related stream studies patient-to-nurse assignment and workload balancing. Mullinax and Lawley (2002) formulate an acuity-based assignment model for neonatal intensive care, and Schaus et al. (2009) develop scalable workload-balancing methods for nurse-to-patient assignment. Most closely related, Sir et al. (2015) incorporate both patient acuity and nurses’ perceived workload into nurse-patient assignment models. Our approach differs in two respects. First, our perceived-workload measure is observed during actual care delivery at the patient–nurse–shift level rather than constructed from survey-based ratings of acuity indicators. Second, we explicitly model temporal carryover: prior objective workload affects current perceived workload and therefore enters the assignment decision through nurses’ workload histories.

Related work studies other dimensions of matching and scheduling. Niewoehner III et al. (2022) examine how familiarity affects physician patient selection and multitasking, while Shi (2022) studies matching in broader two-sided marketplaces. Staffing and scheduling decisions can also affect

worker behavior and well-being: schedule volatility has been linked to nurse turnover (Bergman et al. 2022), and scheduling practices affect productivity and employee well-being in service and retail settings (Kamalahmadi et al. 2021, Kesavan et al. 2022). Our work contributes to this stream of literature by showing how recent workload exposure can be incorporated into assignment decisions to reduce predicted perceived workload while maintaining objective-workload fairness.

2. Setting and Data

Our research setting is the Medical Intensive Care Unit (MICU) at New York Presbyterian/Columbia University Irving Medical Center, a 738-bed academic medical center. The MICU contains 24 beds and operates under two nursing shifts per day. One shift is from 7:00 am to 7:00 pm. The other is from 7:00 pm to 7:00 am.

We combine data from multiple sources to construct a comprehensive list of variables that may affect nursing workload. First, we obtain hospitalization data, which includes patient-visit level information such as patient demographics and comorbidities. Second, we obtain patient-flow data, which includes detailed time stamps on patient admission and discharge activities to different units within the hospital. Third, we collect hourly Laboratory-based Acute Physiology Score (LAPS) (Escobar et al. 2008) for each patient, which integrates information from 14 laboratory tests and serves as a measure of patient acuity. Fourth, we obtain data from a commercial workload management system – Kronos OptiLink, which further includes two sets of patient-shift level data: assignment data which contains detailed patient-to-nurse assignment information for each nursing shift, and workload intensity data which contains two workload intensity measures associated with each patient in each nursing shift. Lastly, we obtain data from the human resources system that records the nurses' working schedules and actual punch-in and punch-out times.

There are several advantages of utilizing data from multiple sources. In addition to being able to extract a rich collection of covariates, we utilize the overlapping parts of the data sets to ensure better data accuracy. For example, the patient-flow data and workload management data both contain information on patient admission and discharge times. The human resources data and workload management software data both contain information on the nurses' shift start and end times. Inconsistency in these time stamps would flag a potential error in the data (see Appendix B for more details). Lastly, the detailed patient-to-nurse assignment data allow us to measure the workload at the individual nurse level, rather than the aggregated unit level.

We collect data from the MICU in 2018. Our analysis is at the nurse-patient-shift level, (n, p, t) , i.e., nurse n taking care of patient p in shift t . After data processing, including data imputation and removal (see Appendix B for details), our final cohort contains 11,736 nurse-patient-shift observations, associated with 939 unique patients and 72 unique nurses.

2.1. Main Outcome Measure

Our workload measures are extracted from Kronos OptiLink, a commercial healthcare workload-management and patient-assignment system deployed by the study hospital, which is designed to support acuity-based staffing and workload management in healthcare settings (OptiLink 2012). This system matches patient care demand with clinical labor supply by measuring each patient's level-of-care intensity, compiling unit-level workload, validating staffing levels, and generating balanced patient-to-staff assignments for review and approval. The workload intensity data are recorded at the patient-shift level and contain two measures central to our analysis. The first is DDCIntensity, where DDC stands for "data-driven classification". DDCIntensity is an automated, order-based workload intensity measure derived from the weighted sum of physician orders placed for a patient. This measure reflects the system's data-driven assessment of the patient's workload intensity based on anticipated care demands implied by clinical orders, and does not necessarily capture a complete measure of true nursing workload. Although the proprietary order weights and classification thresholds are not observed, the approach to generate DDCIntensity is consistent throughout our study and serves as a standardized order-based benchmark. The second is PJIntensity, where PJ stands for "professional judgment." PJIntensity is entered by the staff nurse who cares for the patient and reflects the clinical or psychosocial situation(s) that, in the nurse's professional judgment, best describe the patient's workload. Accordingly, in our setting, DDCIntensity serves as a system-generated benchmark for order-based workload, whereas PJIntensity captures the nurse's subjective assessment of the same patient-shift workload.

Table 1 summarizes the distribution of the two intensity scores where NA denotes missing values. Recall that both DDCIntensity and PJIntensity take on four different levels: Low (1), Medium/Average (2), High (3), and Extreme (4). Note that 19.8% of the PJIntensity scores are missing (None of the DDCIntensity scores are missing). Excluding those missing data, among the remaining observations, 52.9% of the patient-shifts have a PJIntensity that is higher than their corresponding DDCIntensity ($PJ > DDC$), while only 5.2% of the patient-shifts have a PJIntensity that is lower than their DDCIntensity. The remaining shifts have equal PJ and DDC intensities. For nurse n taking care of patient p in shift t , our main outcome variable is the difference between the PJ and DDC intensities:

$$Intensity_Gap_{n,p,t} = PJIntensity_{n,p,t} - DDCIntensity_{n,p,t}.$$

This quantity captures how the focal nurse's perceived workload deviates from the corresponding order-based workload benchmark for the same patient and shift. We focus on this gap, rather than PJIntensity alone, because our research question is whether recent workload exposure – as

measured by the workload in the most recent shift – affects nurses' perceived burden beyond what is explained by the current patient's order-based workload. A high PJIntensity may simply reflect that the patient requires intensive care, which is common in the ICU. In contrast, a higher PJIntensity relative to DDCIntensity suggests that the assignment is perceived as more burdensome than what is captured by the system-generated benchmark. This distinction is important for patient-to-nurse assignment decisions. While some nurses must care for high-intensity patients, managers may still want to understand when the same order-based workload is experienced as more burdensome, as such discrepancies may reflect fatigue and mental stress of the nurses. Thus, the gap provides an operationally meaningful measure. As a robustness check, Appendix E.1.2 re-estimates our main specification using PJIntensity directly as the outcome variable.

Table 1 Distribution of PJIntensity and DDCIntensity Pairs

PJIntensity \ DDCIntensity	1 (Low)	2 (Average)	3 (High)	4 (Extreme)	Row sum
1 (Low)	40 (0.34%)	107 (0.91%)	2 (0.02%)	0 (0.00%)	149 (1.27%)
2 (Average)	621 (5.29%)	2295 (19.56%)	278 (2.37%)	9 (0.08%)	3203 (27.29%)
3 (High)	291 (2.48%)	3366 (28.68%)	1525 (12.99%)	94 (0.80%)	5276 (44.96%)
4 (Extreme)	5 (0.04%)	283 (2.41%)	416 (3.54%)	86 (0.73%)	790 (6.73%)
NA (Missing)	125 (1.07%)	1585 (13.51%)	555 (4.73%)	53 (0.45%)	2318 (19.75%)
Column sum	1082 (9.22%)	7636 (65.06%)	2776 (23.65%)	242 (2.06%)	11736 (100%)

2.2. Treatment Variable

We are interested in how nurses' recent workload exposure, measured by prior-shift workload, affects the discrepancy between their perceived workload and the order-based workload. For nurse n caring for patient p in shift t , our main treatment variable $RecWork_{n,p,t}$, is defined as the total DDCIntensity of patients assigned to nurse n in the nurse's most recent prior working shift within the previous seven days. Specifically, let $P_t(n)$ denote the set of the patients assigned to nurse n in shift t , and let k_n denote the number of shifts between nurse n 's most recent prior working shift and the focal shift. We set $k_n = \infty$ if nurse n has no prior working shift. Recent workload exposure is then defined as

$$RecWork_{n,p,t} = \sum_{q \in P_{t-k_n}(n)} DDCIntensity_{q,t-k_n} \cdot 1(k_n \leq 14), \quad (1)$$

where $DDCIntensity_{q,t-k_n}$ is the DDCIntensity of patient q in shift $t - k_n$. If the focal nurse had no assignment over the past seven days, $Rec_Work_{n,p,t} = 0$. In our data, 18% of nurse-shifts involve one assigned patient and 82% involve two patients, so the measure typically corresponds to either one DDCIntensity score or the sum of two scores. Figure 1 plots a histogram of the recent workload exposure defined in (1). We observe that Rec_Work mostly takes value 4 (33.7%) or 3 (22.3%). The mean and standard deviation are 3.47 and 1.32 respectively.

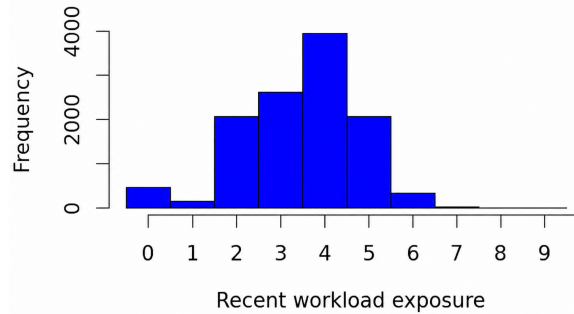


Figure 1 Distribution of treatment Rec_Work

We make several remarks. First, in our main specification, we measure recent workload exposure using DDCIntensity because it is automatically generated from physician orders and provides a relatively objective, system-based measure of patient care demand. This choice also avoids mechanically incorporating nurses' subjective workload assessments into the treatment variable. Since DDCIntensity may not capture all dimensions of nursing workload, Appendix E.1.1 considers alternative measures based on LAPS and PJIntensity. Second, we define recent workload exposure using the nurse's most recent prior working shift within a one-week lookback window. This truncation reflects the idea that workload carryover is likely to be strongest shortly after a demanding shift and may attenuate after sufficient recovery time. The one-week window therefore focuses on recent workload exposure that is more plausibly relevant to the focal shift. In Appendix E.1.1, we test alternative truncation windows. Relatedly, Appendix E.3.3 examines heterogeneity by rest time and shows that the estimated effect of recent workload is larger when rest time is shorter, suggesting recovery mitigates workload carryover. Third, we focus on the most recent prior working shift because it provides a parsimonious and operationally interpretable measure of recent workload exposure. Aggregating over multiple prior shifts would require additional assumptions about how workload accumulates and decays, and averaging may obscure variation when high- and low-workload shifts offset one another. As a robustness check, Appendix E.1.1 considers multi-shift measures, including indicators of sustained high workload over the previous two or three shifts, and obtains results consistent with our main findings.

2.3. Control Variables

We include a number of other variables in our empirical analysis to control for their potential effects on *Intensity_Gap*. Recall that our analysis is at the nurse-patient-shift level. We group the control variables into four categories: shift factors, patient factors at the patient record level, patient factors at the patient-shift level, and nursing factors.

Shift factors. For shift t , we control for weekday versus weekend, day versus night, busyness of the shift at the unit level, which includes the total number of admissions in the shift, the total number of discharges in the shift, hourly averaged census, average DDCIntensity scores over all the patients in the unit. We also control for the unit-level patient-to-nurse ratio, i.e., the total number of patients divided by the total number of nurses.

Patient factors at the patient-visit level. For patient p , we control for age, gender, insurance (commercial, Medicaid, Medicare, and self-pay), comorbidity burden as measured by the Elixhauser score, and whether the patient is admitted to the MICU directly from the emergency department.

Patient factors at the patient-shift level. For each patient p in shift t , we control for the severity and workload intensity of the patient, including the DDCIntensity, whether the shift is within 24 hours of the patient's ICU admission, length of stay (LOS) in hours in the ICU prior to shift t , whether the patient is discharged during the shift, and average hourly LAPS during the shift. We also control for the relative severity of patient p , including the relative LAPS, DDCIntensity, and Elixhauser scores, which are defined as the ratios between the patient's average hourly LAPS, DDCIntensity, Elixhauser, and the average over all the patients in the unit during shift t , respectively.

Nursing factors. For nurse n taking care of patient p in shift t , we control for the number of patients assigned to nurse n in shift t , whether nurse n has taken care of patient p before shift t , i.e., a continuity of care flag, as well as the number of shifts between nurse n 's previous working shift and shift t , which we refer to as the *rest time*. We also include the nurse fixed effects. These fixed effects account for the intrinsic likelihood of a nurse adjusting the PJIntensity relative to the DDCIntensity for the assigned patients.

3. Econometric Model and Estimation Approach

Our goal is to estimate the causal effect of nurses' recent workload exposure, as measured by the most recent shift workload, on the discrepancy between the nurses' perceived workload and order-based workload, $Intensity_Gap_{n,p,t}$. The hypothesis is that heavier recent workload exposure causes a larger $Intensity_Gap_{n,p,t}$, possibly due to stress or fatigue.

The response variable $Intensity_Gap_{n,p,t}$ takes discrete ordered values. We use a linear response model as our first main specification because it yields a directly interpretable summary of the

average change in the reported workload-gap score associated with a one-unit increase in recent workload (Angrist and Imbens 1995). The linear specification also provides a tractable estimation framework for various sensitivity analyses (see, e.g., Section 5.1). Specifically, we consider

$$Intensity_Gap_{n,p,t} = \theta \cdot Rec_Work_{n,p,t} + X_{n,p,t}^\top \beta + u_{n,p,t}, \quad (2)$$

where $X_{n,p,t}$ denotes control variables specified in Section 2.3 (including the nurse fixed effects) and $u_{n,p,t}$ is the unobserved determinant of the workload intensity gap. Treating ordinal outcomes as continuous is common in empirical research, particularly when the outcome is derived from subjective rating scales, such as cognitive burden or perceived workload (Di Tella et al. 2001, Card et al. 2012).

We also consider another main specification that explicitly models the ordered nature of the outcome using an ordered probit model. Specifically, we define a latent continuous variable, $Intensity_Gap_{n,p,t}^*$, formulated as

$$Intensity_Gap_{n,p,t}^* = \theta \cdot Rec_Work_{n,p,t} + X_{n,p,t}^\top \beta + u_{n,p,t}.$$

The observed discrete response, $Intensity_Gap_{n,p,t}$, is then determined by the latent variable and a set of strictly increasing unknown thresholds $\kappa_1 < \kappa_2 < \dots < \kappa_{J-1}$ as

$$Intensity_Gap_{n,p,t} = j \quad \text{if} \quad \kappa_{j-1} < Intensity_Gap_{n,p,t}^* \leq \kappa_j \quad (3)$$

where $j \in \{1, 2, \dots, J\}$ represents the ordered discrete values the intensity gap can take, $\kappa_0 = -\infty$, $\kappa_J = +\infty$, and the error term $u_{n,p,t}$ is assumed to follow a standard normal distribution.

3.1. Estimation Challenges

We are interested in estimating θ , which measures the impact of nurses' recent workload exposure on the gap between the perceived workload and the order-based workload. However, a naive estimation of (2) or (3) via ordinary least squares (OLS) or maximum likelihood estimation (MLE) suffers from severe biases due to several empirical challenges that arise in our data. In particular, there are four main identification challenges: 1) missing outcome measures due to missing PJIntensity, which may not be missing at random; 2) endogeneity due to omitted variables that are correlated with both $Intensity_Gap_{n,p,t}$ and $Rec_Work_{n,p,t}$; 3) measurement error of $Rec_Work_{n,p,t}$ due to missing admission shift assignment; and 4) censored responses since the workload intensity scores only have four levels. We next describe these estimation challenges in more detail and the econometric approaches we take to address them. To ease exposition and preserve analytical tractability, all subsequent derivations are based on the linear response model (2). The ordered probit specification in (3) is handled analogously.

3.1.1. Sample Selection In our data, 19.8% of the PJIntensity scores are missing, and thus we are unable to observe the outcome variable $Intensity_Gap_{n,p,t}$ for these observations. If the PJIntensity scores are not missing at random, it may lead to estimation bias. For example, since the PJIntensity scores are entered manually by the nurses, a nurse who has a very high perceived workload may be too busy to report the PJIntensity, resulting in an underestimate of the true treatment effect; alternatively, a nurse who has a low perceived workload may not bother to report the PJIntensity, resulting in an overestimate of the true treatment effect.

We address the potential sample selection bias using the Heckman selection model (Wooldridge 2010). Let $Report_PJ_{n,p,t}$ denote the binary variable indicating whether the PJIntensity is reported by nurse n for patient p in shift t . Note that the response $Intensity_Gap_{n,p,t}$ is observable only if $Report_PJ_{n,p,t} = 1$. We assume a probit model for sample selection:

$$Report_PJ_{n,p,t} = 1(\tilde{X}_{n,p,t}^\top \alpha + v_{n,p,t} > 0), \quad (4)$$

where $\tilde{X}_{n,p,t}$ denotes the control variables which is a superset of $X_{n,p,t}$, i.e., $\tilde{X}_{n,p,t}$ may include more variables than $X_{n,p,t}$; $v_{n,p,t}$ is assumed to follow a standard normal distribution that is independent of $\tilde{X}_{n,p,t}$. We estimate (2) and (4) jointly, assuming $u_{n,p,t}$ in (2) is also independent of $\tilde{X}_{n,p,t}$, and $u_{n,p,t}, v_{n,p,t}$ satisfy the condition $\mathbb{E}[u_{n,p,t}|v_{n,p,t}] = \gamma v_{n,p,t}$. The Heckman selection model introduces a correction term to account for sample selection. To see this, we first rewrite (2) as

$$\begin{aligned} Intensity_Gap_{n,p,t} &= \theta \cdot Rec_Work_{n,p,t} + X_{n,p,t}^\top \beta + \mathbb{E}[u_{n,p,t}|\tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1] \\ &\quad - \mathbb{E}[u_{n,p,t}|\tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1] + u_{n,p,t} \\ &= \theta \cdot Rec_Work_{n,p,t} + X_{n,p,t}^\top \beta + \mathbb{E}[u_{n,p,t}|\tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1] + e_{n,p,t}. \end{aligned}$$

It is easy to see that $\mathbb{E}[e_{n,p,t}|\tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1] = 0$. We also have

$$\begin{aligned} &\mathbb{E}[u_{n,p,t}|\tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1] \\ &= \mathbb{E}[\mathbb{E}[u_{n,p,t}|\tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1, v_{n,p,t}]|\tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1] \\ &= \mathbb{E}[\mathbb{E}[u_{n,p,t}|v_{n,p,t}]|\tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1] = \gamma \mathbb{E}[v_{n,p,t}|\tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1] \\ &= \gamma \mathbb{E}[v_{n,p,t}|v_{n,p,t} > -\tilde{X}_{n,p,t}^\top \alpha] = \gamma \lambda(\tilde{X}_{n,p,t}^\top \alpha). \end{aligned}$$

where the last step follows from the independence between $v_{n,p,t}$ and $\tilde{X}_{n,p,t}$ and $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio; ϕ and Φ are the probability density function and cumulative distribution function of the standard normal distribution, respectively. As a result, we obtain the following reduced-form model of $Intensity_Gap_{n,p,t}$ when $Report_PJ_{n,p,t} = 1$:

$$Intensity_Gap_{n,p,t} = \theta \cdot Rec_Work_{n,p,t} + X_{n,p,t}^\top \beta + \gamma \cdot \lambda(\tilde{X}_{n,p,t}^\top \alpha) + e_{n,p,t}. \quad (5)$$

3.1.2. Endogeneity Like many observational studies, our estimation can suffer from omitted variable bias. In particular, there can be unobserved variables that are correlated with both nurses' recent workload and their perceived workload. Even though we control for nursing fixed effects, time-varying personal factors may still influence both a nurse's recent workload and their perceived workload. For example, personal challenges, such as managing cancer treatment for oneself or a family member, may lead nursing managers to assign a lower workload over multiple shifts, while the nurse may still perceive a higher workload burden due to the stress of these external factors which may last for a couple of weeks/shifts. In this case, the unobservable (nurse's personal challenges), captured in $u_{n,p,t}$, is negatively correlated with $Rec_Work_{n,p,t}$ but positively correlated with $Intensity_Gap_{n,p,t}$, resulting in an underestimate for the true treatment effect. Alternatively, if there are temporal patterns that correlate with busy MICU shifts, and nursing managers may proactively reduce workload in earlier shifts for nurses scheduled during these times, the true effect of recent workload on perceived workload may be underestimated. More broadly, if some managers informally attempt to balance workload across shifts, despite this not being an official (or standard) practice in our study hospital, our estimates may be biased downward. In particular, while we control for order-based workload in our analysis, such load-balancing efforts may reduce nurses' actual workload in ways not fully captured by observed measures and instead reflected in unobservable components.

To address the endogeneity issue, we apply the instrumental variable (IV) strategy (Wooldridge 2010). A valid IV in our case must be correlated with $Rec_Work_{n,p,t}$, i.e., the relevance condition; but has no direct effect on $Intensity_Gap_{n,p,t}$ other than through its effect on $Rec_Work_{n,p,t}$, i.e., the exclusion restriction. The IVs we propose are the average workload, measured by DDCIntensity and LAPS respectively, at the unit level, excluding the focal patient, in one shift prior to the focal nurse's last assignment/working shift. For example, consider the triple (n, p, t) . Suppose the last assignment shift of nurse n prior to shift t is shift $t - k_n$. Let P_{t-k_n-1} denote the set of patients who were in the ICU in shift $t - k_n - 1$. Then, the two IVs can be expressed as:

$$Avg_Last_DDC_{n,p,t} = \frac{1}{|P_{t-k_n-1} \setminus \{p\}|} \sum_{i \in P_{t-k_n-1} \setminus \{p\}} DDCIntensity_{i,t-k_n-1}, \quad (6)$$

$$Avg_Last_LAPS_{n,p,t} = \frac{1}{|P_{t-k_n-1} \setminus \{p\}|} \sum_{i \in P_{t-k_n-1} \setminus \{p\}} LAPS_{i,t-k_n-1}. \quad (7)$$

See Figure 2 for a pictorial illustration of the treatment and IVs. There, circles represent patients and squares represent nurses.

When the unit is busy with high workload intensity patients in shift $t - k_n - 1$, nurse n is more likely to experience a higher workload in shift $t - k_n$, i.e., the relevance condition is likely to hold.

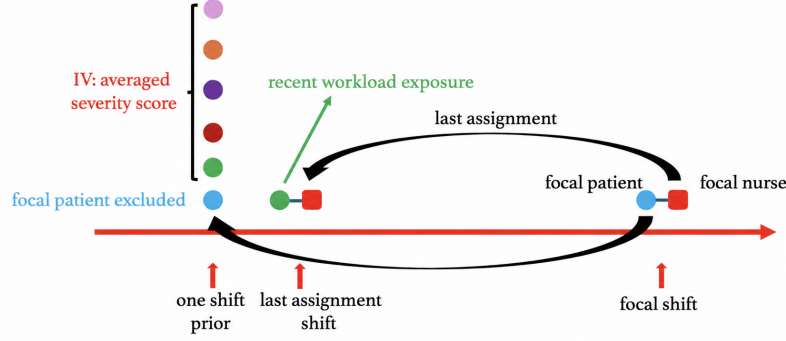


Figure 2 Calculation of IVs and the treatment variable

Conversely, since we exclude the focal patient p when calculating the IVs and nurse n did not work in shift $t - k_n - 1$ (no nurse works in two consecutive shifts in our data), the IVs are unlikely to affect the workload intensity for patient p in shift t , other than through its effect on $Rec_Work_{n,p,t}$, i.e., the exclusion restriction is also likely to be satisfied. Thus, our proposed IVs are valid.

Combining the IV mechanism with the Heckman selection, our final model becomes

$$\begin{aligned} Intensity_Gap_{n,p,t} &= \theta \cdot Rec_Work_{n,p,t} + X_{n,p,t}^\top \beta + \gamma \cdot \lambda(\tilde{X}_{n,p,t}^\top \alpha) + e_{n,p,t}, \\ Rec_Work_{n,p,t} &= \tilde{X}_{n,p,t}^\top \delta + \varepsilon_{n,p,t}, \\ Report_PJ_{n,p,t} &= 1(\tilde{X}_{n,p,t}^\top \alpha + v_{n,p,t} > 0). \end{aligned} \quad (8)$$

In this equation system, covariate $X_{n,p,t}$ contains all variables in Section 2.3. For the covariate $\tilde{X}_{n,p,t}$, it contains $X_{n,p,t}$, IVs: $Avg_Last_DDC_{n,p,t}$, $Avg_Last_LAPS_{n,p,t}$, and $Careful_Nurse_t$, which is defined as the number of “careful” nurses working in shift t . A nurse is referred to as being careful if less than 20% of their assignments’ PJIntensity is missing over the study period. The necessity of introducing this variable is explained later. In the treatment equation, the covariate $\tilde{X}_{n,p,t}$ includes $X_{n,p,t}$, the two IVs, and the inverse Mills ratio $\lambda(\tilde{X}_{n,p,t}^\top \alpha)$. A two-stage least-square estimation with robust standard error clustered by nurses will be used to estimate θ .

We make a few remarks on the specification of (8). First, all exogenous variables in $X_{n,p,t}$ appear in the selection equation $\tilde{X}_{n,p,t}$. We believe that these nurse, patient, and shift level factors are likely to impact both the selection and the outcome variable, $Intensity_Gap_{n,p,t}$. An added benefit is that by making the covariates in $\tilde{X}_{n,p,t}$ a superset of the covariates in $X_{n,p,t}$, we do not need to make additional assumptions on the excluded variables to ensure the exclusion restriction on the reduced form equations is satisfied. Second, in theory, $\tilde{X}_{n,p,t}$ can be the same as $(X_{n,p,t}, Avg_Last_DDC_{n,p,t}, Avg_Last_LAPS_{n,p,t})$. The identifiability holds due to the nonlinearity of the inverse Mills ratio. However, $\lambda(\tilde{X}_{n,p,t}^\top \hat{\alpha})$ can sometimes be approximated well by a linear combination of $\tilde{X}_{n,p,t}$. In this case, setting $\tilde{X}_{n,p,t} = (X_{n,p,t}, Avg_Last_DDC_{n,p,t}, Avg_Last_LAPS_{n,p,t})$

can lead to a multi-collinearity issue, which is known as “variance inflation” (Chapter 19.6 of Wooldridge (2010)). Thus, we introduce the variable $Careful_Nurse_t$ to ensure numerical stability. In particular, we hypothesize that the presence of more meticulous nurses in a shift may positively influence their colleagues, reducing the likelihood of others forgetting to report their PJIntensity scores. As a result, the overall missing rate for PJIntensity in that shift may be lower. On the other hand, the number of careful nurses is unlikely to be correlated with $PJIntensity_{n,p,t}$.

While we can verify the relevance condition of our IVs with data, directly testing the exclusion restriction is challenging. In Section 5.1, we conduct sensitivity analyses by introducing a hypothetical unobserved confounder correlated with both the IVs and the outcome, stress-testing its potential to invalidate our conclusions. Our analysis suggests that a moderate violation of the exclusion restriction is unlikely to alter our results. In Appendix C, we provide more data evidence from the perspective of nurse working patterns to support the validity of our IVs.

3.1.3. Measurement Error The treatment variable $Rec_Work_{n,p,t}$, which is defined as the sum of DDCIntensity scores of patients assigned to the focal nurse in their last working shift, may incur a measurement error due to the high rate of missing assignment information for patients newly admitted during a shift. In particular, consider the triple (p, n, t) , when calculating $Rec_Work_{n,p,t}$, which corresponds to nurse n 's workload in shift $t - k_n$, we exclude patients who are admitted in shift $t - k_n$ due to a large amount of missing assignment information for these patients. As such, the observed recent workload may be an underestimate of the nurse's true recent workload if nurse n is assigned one of the newly admitted patients in shift $t - k_n$.

Let $Rec_Work_{n,p,t}^*$ denote the true recent workload, including both non-admission and admission patients assigned to nurse n in shift $t - k_n$. The measurement error is therefore $\Delta Rec_Work_{n,p,t} = Rec_Work_{n,p,t}^* - Rec_Work_{n,p,t}$. As formally shown in Appendix D, this unobserved component can bias the estimated treatment effect. The direction of the bias depends on how the missing workload is correlated with our IVs. If the IVs are negatively correlated with the missing workload, the estimator tends to understate the true treatment effect; if the correlation is positive, the estimator may overstate it.

To gauge the impact of this measurement error and ensure it does not drive our results, we conduct sensitivity analyses using alternative imputation schemes for missing admission-shift assignments. The idea is to create different levels of correlation between $IV_{n,p,t}$ and $\Delta Rec_Work_{n,p,t}$. Specifically, we design three imputation mechanisms to impute the missing admission-shift assignments such that $IV_{n,p,t}$ and $\Delta Rec_Work_{n,p,t}$ are positively, negatively, and not correlated, and then re-estimated the model with the adjusted treatment variable. The estimation results for this sensitivity analysis are reported in Section 5.2. The details of data imputation mechanisms are provided in Appendix D.

3.1.4. Censored Response For the linear response model, an additional empirical challenge is that the outcome variable *Intensity_Gap* may be censored since the workload intensity scores can only take four different levels. For example, when a patient's DDCIntensity is 4, the nurse cannot report a PJIntensity above 4. As a result, the observed workload gap is mechanically capped at zero, even if the nurse experiences a level of burden that exceeds what is reflected by the maximum reported PJIntensity category.

This censoring can introduce bias in the estimated treatment effect. However, in our setting, the direction of the bias works against our main hypothesis. As shown formally in Appendix D, the censoring error is negatively correlated with the IV-induced variation in recent workload. Consequently, censoring tends to attenuate the estimated effect of recent workload on the workload gap. We therefore do not explicitly correct for this censoring in the main specification. Instead, the main estimate should be interpreted as conservative: if we find a positive effect of recent workload on the observed, censored workload gap, the corresponding effect on the latent uncensored workload gap is likely to be at least as large.

3.2. Estimation Approach

For the linear response model (2), we address endogeneity and sample selection by combining instrumental variables with a Heckman selection correction and estimating θ using the two-step procedure in Wooldridge (2010). Measurement error and outcome censoring are more difficult to incorporate directly without imposing additional strong assumptions. We therefore assess their potential impact through sensitivity analyses rather than model them explicitly in the main specification. These analyses show that while these issues are valid limitations, they do not appear to drive our results.

For the ordered probit model (3), we use a parallel strategy that accounts for both sample selection and endogeneity through a control-function approach. We first estimate the selection equation (4) using a probit model to obtain the inverse Mills ratio $\hat{\lambda}(\tilde{X}_{n,p,t}\alpha)$. We then regress $Rec_Work_{n,p,t}$ on the exogenous controls $\tilde{X}_{n,p,t}$, including the IVs, and obtain the first-stage residual $\hat{\varepsilon}_{n,p,t} = Rec_Work_{n,p,t} - \widehat{Rec_Work}_{n,p,t}$. Both the inverse Mills ratio and the first-stage residual are then included as additional controls in the ordered probit model, which we estimate by MLE.

In our primary specifications, standard errors are clustered at the nurse level to allow for within-nurse dependence arising from unobserved characteristics such as stress tolerance or reporting tendencies. Because our models already control for a rich set of patient- and shift-level characteristics, nurse-level clustering is a natural baseline choice. Appendix E reports robustness checks using more conservative multi-way clustering specifications.

4. Main Estimation Results

We present the estimation results of the linear response and ordered probit models here. In both cases, the treatment effects are positive and statistically significant with similar magnitude, demonstrating the recent workload indeed has a positive causal impact on nurses' perceived workload.

4.1. Estimation Results of Linear Response Model

Table 2 summarizes the estimation results of the treatment effect based on our main model (8). The full table of regression results are reported in Appendix A. We also provide the estimation results without the Heckman selection part, i.e., using IVs only (**IVs-only**) on non-missing data, and direct estimation of (2) using OLS (**OLS**) without considering endogeneity and sample selection.

We observe from the main model estimation that a higher recent workload exposure is associated with a higher deviation of the nurse's perceived workload from the order-based workload. In particular, one level of increase in recent workload is associated with a 0.273 level increase in *Intensity_Gap*. Given the ordered nature of *Intensity_Gap*, we interpret this coefficient primarily as a linear average treatment effect: a one-unit increase in recent workload exposure shifts the nurse's professional-judgment assessment upward relative to the order-based benchmark by 0.273 categories on average. It provides a summary of the direction and average magnitude of the effect rather than a precise cardinal measure of workload burden. Across different model specifications, we note that the estimated treatment effects are quite similar in magnitude for the main and IVs-only models. When comparing the main and IVs-only models to OLS, we note that direct estimation of (2) leads to an underestimation of the treatment effect due to the endogeneity of *Rec_Work*. This supports one of the potential endogeneity mechanisms discussed in Section 3.1.2.

As shown in Table 2, the first-stage results further support the relevance of our instrumental variables. In the treatment equation, the coefficients on the IVs are positive and statistically significant in both the main and IVs-only specifications. We also note that in the main model, the coefficient for the inverse Mills ratio is not significantly different from zero at the 0.1 level. This suggests that, after accounting for observed covariates, sample selection does not appear to be a major source of bias in this setting.

4.2. Estimation Results of Ordered Probit Model

Table 2 presents the estimation results for the ordered probit specification. Mirroring the structure of our previous analysis, we compare the full ordered probit model (main model), which accounts for both endogeneity and sample selection, with an IVs-only specification (ignoring sample selection) and a naive ordered probit model.

Consistent with the findings from the linear response model, the coefficient for the most recent shift workload is positive and statistically significant. In the ordered probit model, this positive

Table 2 Treatment Effect Estimation: Comparison across Models and Specifications

	Linear Response			Ordered Probit		
	Main	IVs-Only	OLS	Main	IVs-Only	Naive MLE
<i>Rec_Work</i>	0.273*** (0.060)	0.294*** (0.064)	-0.010* (0.005)	0.595*** (0.167)	0.643*** (0.156)	-0.024 (0.102)
<i>Last_Avg_DDC</i>	0.563*** (0.075)	0.587*** (0.085)	—	0.563*** (0.075)	0.587*** (0.085)	—
<i>Last_Avg_LAPS</i>	0.692*** (0.015)	0.651*** (0.169)	—	0.692*** (0.015)	0.651*** (0.169)	—
Inverse Mills Ratio	0.238 (0.172)	—	—	0.573 (0.555)	—	—
Observations	11,736	9,427	9,427	11,736	9,427	9,427

Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

coefficient indicates that a heavier recent workload systematically increases the latent workload perception, which then increases the probability of it falling into a higher category of *Intensity_Gap*. The estimated effect remains positive and significant in the IVs-only specification. By contrast, the naive ordered probit estimate is close to zero and statistically insignificant, suggesting that failing to instrument for recent workload substantially attenuates the estimated effect. The similarity between the main and IVs-only estimates, together with the lack of statistical significance of the selection correlation parameter in the main model, suggests that the estimated effect is not materially affected by sample selection.

Beyond the point estimates, the ordered probit model provides a granular decomposition of how the recent workload shifts the entire distribution of perceived deviations. By calculating the marginal effect of the recent shift's workload, we find that a one-unit increase of *Rec_Work* reduces the probability of reporting a zero *Intensity_Gap* by 11.97 percentage points and the probability of a negative *Intensity_Gap* by 4.05 percentage points (combining levels -2 and -1). Meanwhile, the probability of perceiving a positive gap increases substantially, with the likelihood of a Level-1 gap rising by 11.43 percentage points and a Level-2 gap by 4.51 percentage points. Aggregating these category-specific marginal effects, the implied change in the expected perceived workload deviation is approximately 0.249. This magnitude is close to the estimate from the linear response model, 0.273. This alignment validates our core hypothesis that a heavier recent workload causally inflates a nurse's current perceived workload deviation.

5. Additional Estimation Analysis

In this section, we conduct several sensitivity analyses to assess the consistency and reliability of our estimation results, with particular attention to endogeneity and measurement error. We

focus on these two estimation challenges because 1) we do not find evidence that selection bias significantly impacts our results and 2) our discussion in Section 3.1.4 suggests that the estimates in Table 2 are conservative when the outcome variable is censored.

We also conduct additional robustness checks along several additional dimensions, as reported in Appendix E. There, we first consider alternative model specifications, including different constructions of the treatment variable and outcome variable, alternative sets of control variables, and different approaches to calculating clustered standard errors. Second, we examine the robustness of our findings to alternative data selection procedures, i.e., study populations. Third, we investigate potential heterogeneous treatment effects along several clinically relevant dimensions, including patient workload intensity, nurse experience, and nurses' resting time. Overall, these analyses show that our main findings remain stable.

5.1. Sensitivity Analysis on the Validity of the IVs

As discussed in Section 3.1.2, it is in general hard to verify the exclusion restriction for the IVs. The main concern is that there can be unobserved patient or nurse characteristics that are correlated with both the instruments and the outcome. For example, when the MICU is busy with a lot of high-acuity patients, the nursing manager may tend to schedule more experienced nurses; these nurses may also be less likely to report a higher clinically perceived workload than the order-based workload because their experience has provided them the skills to navigate these high-stress shifts. Alternatively, the more experienced nurses may be more likely to report a higher clinically perceived workload as their experience provide them with a better assessment of the workload. In addition, when the MICU is busy, the hospital may be more selective in terms of who to admit to the ICU, i.e., they may tend to reserve the ICU beds for the more severe patients. Severe patients are also more likely to be associated with a higher perceived workload. In this section, we conduct a sensitivity analysis of the IVs regarding the exclusion restriction.

For a linear response model, IV sensitivity analysis are well-developed and extensively applied in empirical research. Consequently, we focus the IV sensitivity analysis exclusively on our linear specification. Following [Baiocchi et al. \(2014\)](#), suppose there exists an unobserved confounder $\zeta_{n,p,t}$ with mean zero and variance one that is correlated with the outcome $Intensity_Gap_{n,p,t}$ and the IVs, i.e., $Avg_Last_DDC_{n,p,t}$ and $Avg_Last_LAPS_{n,p,t}$, but is uncorrelated with the other measurable covariates $X_{n,p,t}$. We assume a linear model for this relation:

$$\begin{aligned} Intensity_Gap_{n,p,t} &= \theta \cdot Rec_Work_{n,p,t} + X_{n,p,t}^\top \beta + \gamma \cdot \lambda(\tilde{X}_{n,p,t}^\top \alpha) + \psi \zeta_{n,p,t} + e_{n,p,t}, \\ \zeta_{n,p,t} &= \xi_1 \cdot Avg_Last_DDC_{n,p,t} + \xi_2 \cdot Avg_Last_LAPS_{n,p,t} + \omega_{n,p,t}, \end{aligned} \quad (9)$$

where the noise terms $e_{n,p,t}$ and $\omega_{n,p,t}$ satisfy

$$\begin{aligned} \mathbb{E}[e_{n,p,t} | X_{n,p,t}, Report_PJ_{n,p,t} = 1, Avg_Last_DDC_{n,p,t}, Avg_Last_LAPS_{n,p,t}] &= 0, \\ \mathbb{E}[\omega_{n,p,t} | X_{n,p,t}, Report_PJ_{n,p,t} = 1, Avg_Last_DDC_{n,p,t}, Avg_Last_LAPS_{n,p,t}] &= 0. \end{aligned}$$

Note that ψ, ξ_1 , and ξ_2 are sensitivity parameters, where ψ measures the effect of a one standard deviation increase in the unobserved confounder on $Intensity_Gap_{n,p,t}$, and ξ_1, ξ_2 measure the effect of one unit of increase in the IVs on the unobserved confounder respectively. Under this model, if we can control for ζ , Avg_Last_DDC and Avg_Last_LAPS would be valid IVs. Thus, in the sensitivity analysis, we treat $Intensity_Gap_{n,p,t} - (\psi\xi_1 \cdot Avg_Last_DDC_{n,p,t} + \psi\xi_2 \cdot Avg_Last_LAPS_{n,p,t})$ as the modified outcome variable and apply 2SLS to estimate θ . In particular, the outcome equation takes the form

$$\begin{aligned} & Intensity_Gap_{n,p,t} - (\psi\xi_1 \cdot Avg_Last_DDC_{n,p,t} + \psi\xi_2 \cdot Avg_Last_LAPS_{n,p,t}) \\ &= \theta \cdot Rec_Work_{n,p,t} + X_{n,p,t}^\top \beta + \gamma \cdot \lambda(\tilde{X}_{n,p,t}^\top \alpha) + (e_{n,p,t} + \psi\omega_{n,p,t}), \end{aligned}$$

where $(e_{n,p,t} + \psi\omega_{n,p,t})$ has mean zero conditional on controls $X_{n,p,t}$, $\lambda(\tilde{X}_{n,p,t}^\top \alpha)$, $Avg_Last_DDC_{n,p,t}$, $Avg_Last_LAPS_{n,p,t}$, and $Report_PJ_{n,p,t} = 1$. We estimate the model with different values of ψ , ξ_1 , and ξ_2 to determine the parameter regimes where the estimate of θ is a) positive and statistically significant at the 5% level, b) negative and statistically significant at the 5% level, or c) statistically not significantly different than 0 at the 5% level. If the regime where a) holds is large, we are more confident that our estimate for θ is reasonably robust against the potential violation of exclusion restriction for the IVs.

Figure 3 summarizes the results of our sensitivity analysis. For this analysis, we standardize all the control variables (i.e., subtract the mean and normalized by the standard deviation accordingly) to ensure that the coefficients are comparable. In the first row of Figure 3, we set $\xi_1 = -0.08, 0, 0.1$, where -0.08 and 0.1 correspond to the minimal and maximal coefficients respectively when regressing the control variables X against Avg_Last_DDC , and vary the values of ξ_2 and ψ . In the second row of Figure 3, we set $\xi_2 = -0.12, 0, 0.08$, where -0.12 and 0.08 correspond to the minimal and maximal coefficients respectively when regressing the control variables X against Avg_Last_LAPS , and vary the values of ξ_1 and ψ . The white region depicts the area where the estimate for θ is positive and statistically significant at the 5% level. The grey region is where the estimate for θ is statistically not different than 0, and the black region is where the estimate for θ is negative and statistically significant. To construct Figure 3, we need to run 6×60^2 different regressions (with different values of ξ_1 , ξ_2 , and ψ). In this case, using the bootstrap method to calculate the robust standard errors clustered by nurses is computationally prohibitive given our computational resources. To address this, we use a conservative approximation of the robust standard errors instead. In particular, we first calculate the standard non-clustered standard errors, and then multiply these standard errors by a factor of 20. The value 20 is the largest ratio between the robust standard error and the standard non-clustered standard error we see in all the other estimations we run in Section 4 and Sections 5.2 – E.1; the average was 8.2. Note that such a

conservative approximation of the robust standard errors could make the gray region a bit larger than its actual size, and it could also make the white region smaller than its actual size.

We observe in Figure 3 that the white region is quite large. The dots in the plot correspond to the (ξ_1, ψ) -values (in the first row) or (ξ_2, ψ) -values (in the second row) for all observed covariates. Note that most of the dots are in the white region, suggesting that in order for there to be an unobserved confounder that would explain away our result, the effect size of this unobserved confounder would have to be much larger than those of most of the observed covariates. The only dot that falls into the grey region is the dot corresponding to DDCIntensity in Figures 3 (a) and (d). DDCIntensity is known to be highly correlated with the outcome, and it is quite unlikely that the unobserved confounder has as large an effect on the outcome as DDCIntensity.

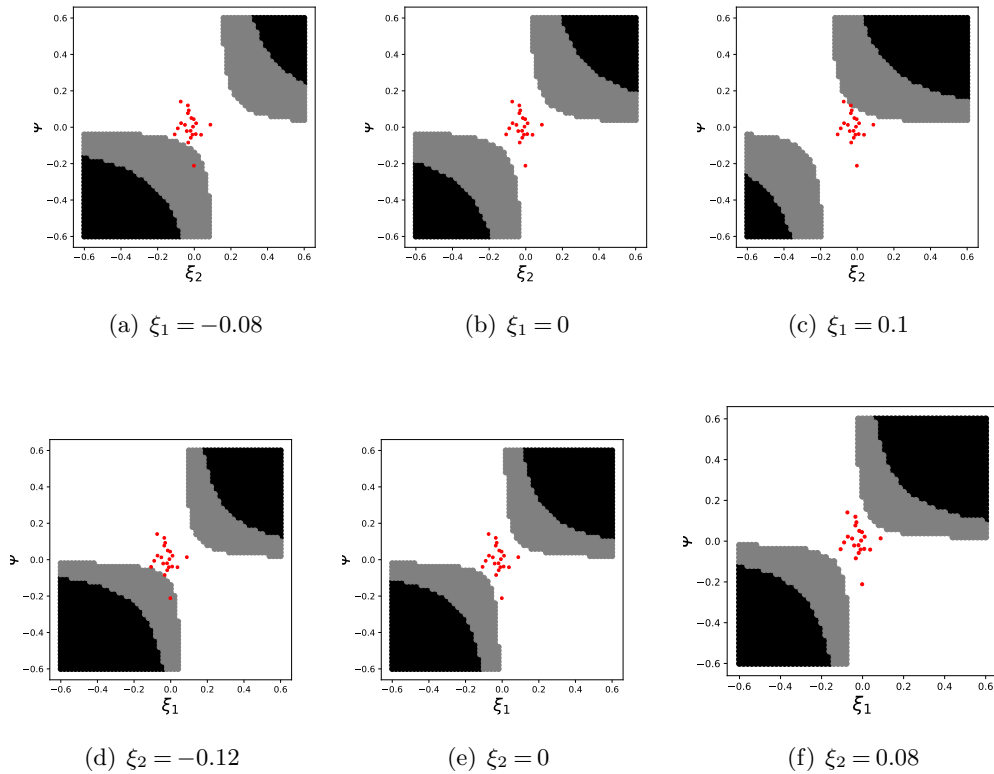


Figure 3 Sensitivity analysis for the validity of IVs: ψ is the impact of unobserved confounder ζ on the outcome; ξ_1 and ξ_2 are the sensitivity parameters of IVs *Avg_Last_DDC* and *Avg_Last_LAPS* on ζ ; and the dots denote corresponding coefficients of the observed covariates.

5.2. Sensitivity Analysis for Measurement Errors

In this section, we conduct a sensitivity analysis with respect to the measurement error induced by missing admission-patient assignments. As discussed in Section 3.1.3, in the linear response model, the direction of the estimation bias depends on the correlation between the measurement

error and the IVs. To examine this issue, we consider the three imputation schemes introduced in Section 3.1.3. The first scheme imputes assignments to induce a positive correlation between the IVs and $\Delta Rec_Work_{n,p,t}$ (Positive Cor); the second induces a negative correlation between them (Negative Cor); and the third uses random assignment (No Cor). We then re-estimate the model using the imputed true recent workload $Rec_Work_{n,p,t}^*$ defined in (13).

Under the Positive Cor scheme, the estimated coefficient of $Rec_Work_{n,p,t}^*$ is 0.218 with standard error 0.057; under the No Cor scheme, the coefficient is 0.259 with standard error 0.069; and under the Negative Cor scheme, the coefficient is 0.315 with standard error 0.088. All three estimates are statistically significant at the 0.001 level, based on 11,736 observations. The pattern is consistent with our analytical discussion. When the IVs and $\Delta Rec_Work_{n,p,t}$ are positively correlated, the estimated treatment effect becomes smaller, suggesting that ignoring this type of measurement error would overestimate the treatment effect. Conversely, when the IVs and $\Delta Rec_Work_{n,p,t}$ are negatively correlated, the estimated treatment effect becomes larger, suggesting that ignoring this type of measurement error would underestimate the treatment effect. Importantly, across all three imputation schemes, the estimated effects remain close to the main estimate of 0.273 and preserve the same positive and statistically significant direction. These results suggest that our main conclusion is not driven by measurement error in the recent workload variable.

Additionally, while deriving the directional analysis for estimation bias is analytically intractable under the non-linear ordered probit specification, we leverage the theoretical insights established in the linear framework to design some stress tests. Specifically, we apply the same three missing-data imputation schemes to re-estimate the ordered probit model. By mechanically inducing these extreme correlation structures between the IVs and the measurement error, we can empirically evaluate the stability of the treatment effect. Consistent with the linear model's behavior, we observe the exact same directional pattern in the ordered probit estimates. The coefficient for Rec_Work^* is 0.454 (standard error is 0.119, average treatment effect is 0.226) under the Positive Cor scheme, increases to 0.542 (standard error is 0.150, average treatment effect is 0.245) under the No Cor scheme, and reaches its largest magnitude at 0.691 (standard error is 0.198, average treatment effect is 0.289) under the Negative Cor scheme. All estimates remain statistically significant at 0.001 level. Such observations are aligned with the linear response model case, thereby ensuring that our core empirical conclusions remain robust even when the discrete nature of the response variable is explicitly modeled.

6. Simulation Analysis: Patient-to-Nurse Assignment Policy

Our empirical analysis reveals that the recent workload exposure increases nurses' perceived workload relative to the order-based benchmark. Because psychological strain is a primary driver of

nursing burnout and compromised patient safety, it is desirable to reduce the perceived workload. Motivated by this, we explore how hospitals can proactively leverage assignment history when making assignment decisions to achieve better workload management.

6.1. Perception-Aware Assignment Optimization

We first develop a perception-aware assignment optimization that dynamically assigns patients with a goal of reducing total perceived workload. Specifically, our policy myopically solves a linear integer program for each shift t . Let N_t denote the set of nurses who work in shift t and P_t denote the set of patients in the MICU in shift t . Then, the integer program takes the form

$$\begin{aligned} \min \quad & \sum_{n \in N_t} \sum_{p \in P_t} \text{Predict_PJ}_{n,p,t} \cdot x_{n,p,t} \\ \text{s.t.} \quad & \sum_{p \in P_t} x_{n,p,t} \cdot \text{DDCIntensity}_{p,t} \leq 5, \forall n \in N_t, \quad \sum_{p \in P_t} x_{n,p,t} = \text{Patient_Num}_{n,t}, \forall n \in N_t, \\ & \sum_{n \in N_t} x_{n,p,t} = 1, \forall p \in P_t, \quad x_{n,p,t} \in \{0, 1\} \end{aligned} \quad (10)$$

where $\text{Predict_PJ}_{n,p,t}$ is the expectation of predicted PJIntensity of assigning patient p to nurse n in shift t (explained in more details below), and $x_{n,p,t}$ is the binary decision variable, with $x_{n,p,t} = 1$ indicating patient p is assigned to nurse n in shift t and 0 otherwise. The first constraint requires that the sum of DDCIntensity of patients assigned to a nurse can not exceed 5. For example, we cannot assign two patients with $\text{DDCIntensity} \geq 3$ to the same nurse in a shift. We impose this constraint to avoid assigning too heavy a workload to a nurse in a shift. The second constraint requires that we maintain the same number of patients assigned to nurse n in shift t as in the data ($\text{Patient_Num}_{n,t}$ denotes the number of patients assigned to nurse n in shift t according to raw data). We impose this constraint to make the comparison with the status quo conservative: the algorithm can only improve assignments by reallocating patients across nurses, while holding fixed each nurse's patient count as observed in the data. The third constraint requires that each patient be assigned to exactly one nurse. For patients who are assigned to more than one nurse in the data, we maintain the same nursing assignment for them and do not include them in the optimization problem, i.e., we only shuffle the patient-to-nurse assignment for patients who are originally assigned to one nurse. In addition to constraints listed in (10), more practical constraints can be easily accommodated to our framework, as we will discuss later.

Recall $\text{PJIntensity}_{n,p,t} = \text{DDCIntensity}_{p,t} + \text{Intensity_Gap}_{n,p,t}$. Since the set of patients in a shift is fixed, the total order-based workload does not depend on how patients are matched to nurses. Therefore, minimizing total predicted perceived workload is equivalent to minimizing the predicted perception gap. The framework can also accommodate alternative objectives, such as threshold or convex penalties that place greater weight on extreme perceived workload.

To implement the counterfactual assignment policies, we proceed shift by shift. The observed data policy, which corresponds to the patient-to-nurse assignments recorded in the hospital data, serves as the status quo benchmark. In each counterfactual policy, we hold fixed the set of patients, the set of working nurses, and the number of patients assigned to each nurse in every shift. The policy only changes which nurse is assigned to which patient among the feasible assignments.

A practical challenge is that the fitted empirical model cannot be used mechanically for real-time assignment. Some covariates, such as current-shift LAPS, are unavailable when assignments are made. The continuity of care indicator may be affected by the assignment itself, but its coefficient does not have a clear causal interpretation. Nurse fixed effects may also create undesirable allocation incentives. We therefore distinguish between the score used to construct assignments and the score used to evaluate counterfactual outcomes. For assignment construction, unavailable covariates are replaced with their values from the previous shift, continuity of care indicator without the clear causal interpretation is set to unit-level averages, and nurse fixed effects are replaced by a common reference value. After assignments are generated, we evaluate them using a richer information set, including realized patient acuity and nurse fixed effects. For continuity of care, we continue to use the average values. This separation ensures that the assignment rule is implementable and does not exploit noncausal or potentially unfair predictors, while the evaluation uses the best available information to approximate the realized perceived workload under each policy.

6.2. Policy Analysis

To evaluate the value of perceived-workload prediction, we compare our policy with the observed assignments and two DDCIntensity-based benchmarks. Let $W_{n,t} = \sum_{p \in P_t} DDCIntensity_{p,t} \cdot x_{n,p,t}$ denote nurse n 's current-shift order-based workload. Because nurses may be assigned different numbers of patients, both benchmarks balance workload within groups of nurses with the same patient count.

The *current-only* benchmark minimizes $\sum_k \sum_{n \in N_{t,k}} |W_{n,t} - \bar{W}_{t,k}|$, where $\bar{W}_{t,k}$ is the average current-shift workload among nurses who are assigned k patients. The *history-aware* benchmark instead balances $B_{n,t} = W_{n,t} + Rec_Work_{n,t}$, where $Rec_Work_{n,t}$ is nurse n 's total assigned DDCIntensity in the most recent prior working shift, corresponding to the main treatment variable used in our empirical analysis. In particular, this benchmark minimizes $\sum_k \sum_{n \in N_{t,k}} |B_{n,t} - \bar{B}_{t,k}|$, where $\bar{B}_{t,k}$ is the average among nurses who are assigned k patients in the current shift. Note that this policy accounts for recent objective workload history, but only through a simple additive adjustment rather than through the estimated empirical model.

Next, we consider a set of complementary metrics that capture both subjective perceived workload and objective workload allocation fairness.

Table 3 Comparison of Assignment Policies on Perceived Workload Metrics

Metric	data policy	current-only	history-aware	perception-aware
Level-3 predicted PJ	67.2%	66.9%	66.6%	65.8%
Level-4 predicted PJ	7.3%	6.9%	7.4%	6.5%
Tail excess above Level 3	0.071	0.070	0.074	0.066
Nurse-shift Level-4 exposure	10.8%	11.1%	11.5%	10.4%

6.2.1. Effectiveness in Reducing Total Perceived Workload We evaluate the impact of the proposed assignment policy on the upper tail of predicted perceived workload. First, we report the number of nurse-patient-shift observations with predicted perceived workload at Level 3 or Level 4, which captures the frequency of high and extremely high predicted perceived workload. Second, we compute the average tail-excess perceived workload above Level 3, defined as the mean of $(Predict_PJ_{n,p,t} - 3)_+$ across all nurse-patient-shift observations. Third, we evaluate extreme exposure at the nurse-shift level, defined as the fraction of nurse-shifts in which a nurse is assigned at least one patient with predicted perceived workload at Level 4 or above. The comparison of these metrics under our perception-aware assignment policy versus the three benchmark policies is reported in Table 3.

We observe from Table 3 that the perception-aware policy performs best across all four perceived-workload metrics. Relative to the observed assignments, it reduces the share of Level-4 predicted PJ from 7.3% to 6.5%, an 11% relative reduction, and also improves the Level-3 rate, tail excess above Level 3, and nurse-shift Level-4 exposure. The two DDC-based benchmarks produce mixed results: they improve some perceived-workload measures but worsen others, particularly in the extreme upper tail. This comparison highlights the incremental value of incorporating the perceived-workload prediction model into the assignment optimization. While the history-aware benchmark also accounts for prior assignments, it does so through a simple additive measure of current and prior workload. In contrast, our proposed policy combines these and other assignment characteristics according to estimated weights, providing a more refined measure of assignment burden.

6.2.2. Fairness in Assignment Allocation A natural concern is that reducing predicted perceived workload could come at the expense of objective-workload fairness. Because our policy incorporates recent workload history, it may promote temporal balancing, but this is not guaranteed. We therefore audit objective workload fairness at both the nurse-shift and the nurse levels.

Nurse-Shift Level Analysis: Let $W_{n,t}^\pi = \sum_{p \in P_{n,t}^\pi} DDCIntensity_{p,t}$ denote the total objective workload assigned to nurse n in shift t under assignment policy π . Table 4 shows that the perception-aware policy improves nurse-shift level fairness metrics relative to the observed assignments. The

Table 4 Comparison of Assignment Policies on Nurse-Shift-Level Fairness Metrics

Metric	data policy	current-only	history-aware	perception-aware
Avg. spread of $W_{n,t}$	3.362	2.813	2.803	3.088
High workload outlier rate	2.93%	0.42%	0.17%	0.12%
Top 10% workload share	15.1%	14.4%	14.8%	14.2%
Top 20% workload share	29.3%	27.2%	28.5%	28.5%

Table 5 Comparison of Assignment Policies on Nurse-Level Fairness Metrics

Metric	data policy	current-only	history-aware	perception-aware
Mean $ A_n $	0.076	0.050	0.046	0.058
SD of A_n	0.092	0.068	0.079	0.072
90th pct. $ A_n $	0.145	0.112	0.086	0.107
Max-min spread of A_n	0.416	0.382	0.500	0.372

average within-shift workload spread, $\max_{n \in N_t} W_{n,t}^\pi - \min_{n \in N_t} W_{n,t}^\pi$, falls from 3.362 to 3.088. The fraction of nurse-shift observations whose objective workload exceeds the mean by more than two standard deviations (high workload outlier rate) decreases from 2.93% to 0.12%. When nurse-shift observations are ranked by total objective workload $W_{n,t}^\pi$, the share of total workload carried by the top 10% decreases from 15.1% to 14.3%. Similarly, the top 20% share decreases from 29.3% to 28.5%. Although the DDC-based benchmarks directly optimize objective-workload balance, the perception-aware policy achieves broadly comparable fairness improvements.

Nurse Level Analysis: At the nurse level, raw average workload may confound assignment imbalance with differences in patient counts. We therefore residualize each nurse's workload relative to nurses in the same shift with the same patient count. Specifically, define $A_n^\pi = 1/|S_n| \cdot \sum_{t \in S_n} (W_{n,t}^\pi - \bar{W}_{t,|P_{n,t}|}^\pi)$, where S_n is the collection of shifts where nurse n works, $\bar{W}_{t,|P_{n,t}|}^\pi$ is the average workload among comparable nurses (nurses who are assigned $|P_{n,t}|$ patients) in shift t . We summarize the distribution of A_n^π using its mean absolute value, standard deviation, 90th percentile of $|A_n^\pi|$, and max-min spread. As shown in Table 5, the perception-aware policy achieves broadly comparable nurse-level fairness to the two DDC-balancing benchmarks and improves all four metrics relative to the observed assignments in data.

Above all, the fairness audit shows that the proposed perception-aware policy reduces exposure to high predicted perceived workload without worsening the fairness of objective workload allocation. In fact, this policy improves several objective workload balance metrics at both the nurse-shift and nurse levels.

6.3. Discussion: Implementation Considerations

We interpret above counterfactual analysis as a decision-support exercise rather than a replacement for charge-nurse judgment. Actual assignments may depend on various clinical and operational

Table 6 Comparison of Assignment Policies

Policy	Perceived workload	Nurse-shift fairness	Nurse-level fairness
	Proportion of Level-4 PJ	Avg. spread of $W_{n,t}$	Mean $ A_n $
data policy	7.3%	3.362	0.0752
unconstrained	6.5%	3.088	0.0587
random 30% nurses fixed	6.8%	3.154	0.0634
top 30% severe patients fixed	6.9%	3.230	0.0672

considerations not fully observed in our data, including nurse competencies, bed proximity, isolation requirements, and continuity of care. However, these factors can be incorporated as additional hard or soft constraints in the optimization easily. To assess robustness to such restrictions, we consider two constrained variants: one fixes the observed assignments of a random 30% of nurses, and the other fixes the assignments of the 30% highest-workload patients. As expected, these constraints reduce the achievable improvement, but the results remain directionally consistent. The comparison of different policies is reported in Table 6.

Another concern is strategic reporting in order to avoid burdensome assignment. Our design mitigates this risk because the assignment score excludes nurse fixed effects and uses objective order-based recent workload rather than nurses' past subjective reports. Thus, self-reported workload does not directly feed back into future assignments. Nevertheless, prospective implementation should monitor reporting behavior over time and periodically recalibrate the prediction model to guard against unintended behavioral responses.

Finally, our simulation modifies patient-to-nurse assignments within each shift while holding the set of working nurses fixed. It therefore does not alter nurse schedules, consecutive-shift patterns, night-shift sequences, or rest times. This design isolates the effect of patient assignment decisions and ensures that the counterfactual policy does not introduce additional scheduling-related risks. At the same time, it also limits the scope of the analysis. Integrating patient assignment with staffing and scheduling decisions is an important direction for future research.

7. Concluding Remarks

In this work, we study workload carryover in nursing assignments. Using patient–nurse–shift level data from a medical ICU, we quantify how a nurse's recent order-based workload affects the nurse's subsequent professional-judgment assessment of patient workload relative to an automated order-based benchmark. We find that heavier recent workload systematically increases perceived workload relative to the benchmark, highlighting an important temporal dimension of nursing workload that contemporaneous order-based measures may miss. We further show how this insight can inform patient-to-nurse assignment decisions. Our perception-aware assignment policy uses the

estimated workload-carryover mechanism to reduce exposure to high predicted perceived workload. In retrospective simulations, the policy shifts assignments away from the upper tail of predicted perceived workload while maintaining safeguards on objective workload allocation. The framework is intended as a decision-support tool that helps charge nurses incorporate recent workload exposure into assignment decisions, rather than as a replacement for clinical judgment.

Our study also has several limitations. First, the empirical setting is a single medical ICU, and the magnitude of workload carryover may differ across hospitals, units, and nursing specialties. Second, DDCIntensity and PJIntensity are operational measures generated within a commercial workload-management system; because the underlying DDC weighting rules are proprietary and implementation-specific, we interpret DDCIntensity as an automated order-based benchmark rather than as a complete measure of true nursing workload. Third, although we control for a rich set of patient, shift, and nurse characteristics, the gap between PJIntensity and DDCIntensity may also reflect unobserved clinical complexity or non-order-driven care demands. Relatedly, our analysis does not directly quantify how workload misalignment affects downstream outcomes such as nurse burnout, job dissatisfaction, turnover, or patient safety. These outcomes are important, but they are realized over longer horizons and cannot be reliably linked to specific patient–nurse–shift assignments in our data. Our findings should therefore be interpreted as evidence on a proximal assignment-level workload outcome. Future research could combine workload-misalignment measures with longitudinal nurse and patient outcomes to study these downstream consequences.

Finally, our simulation holds staffing levels and nurse schedules fixed. This means that our counterfactual assignment policy changes patient-to-nurse matching within a shift, but does not change consecutive workdays, night-shift sequences, or rest times. Identifying the causal effects of such scheduling patterns would require a separate empirical strategy because schedules are themselves endogenous and may depend on nurse characteristics, staffing needs, and unit-level demand. Future research could integrate workload carryover into joint staffing, scheduling, and assignment models that explicitly account for rest time, consecutive workdays, cumulative fatigue, and nurse-level workload fairness.

References

- Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH (2002) Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *Journal of the American Medical Association* 288(16):1987–1993.
- Alghamdi MG (2016) Nursing workload: a concept analysis. *Journal of nursing management* 24(4):449–457.
- Angrist JD, Imbens GW (1995) Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association* 90(430):431–442.

- Baiocchi M, Cheng J, Small DS (2014) Instrumental variable methods for causal inference. *Statistics in Medicine* 33(13):2297–2340.
- Ball JE, Murrells T, Rafferty AM, Morrow E, Griffiths P (2014) ‘care left undone’ during nursing shifts: associations with workload and perceived quality of care. *BMJ quality & safety* 23(2):116–125.
- Banda Z, Simbota M, Mula C (2022) Nurses’ perceptions on the effects of high nursing workload on patient care in an intensive care unit of a referral hospital in malawi: a qualitative study. *BMC nursing* 21(1):136.
- Batt RJ, Terwiesch C (2017) Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Science* 63(11):3531–3551.
- Bergman A, David G, Song H (2022) “I quit”: Schedule volatility as a driver of voluntary employee turnover. *Available at SSRN 3910077* .
- Berlin G, Lapointe M, Murphy M, Wexler J (2022) Assessing the lingering impact of COVID-19 on the nursing workforce. McKinsey & Company. May 11, 2022.
- Berry Jaeker JA, Tucker AL (2017) Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity. *Management Science* 63(4):1042–1062.
- Blasche G, Bauböck VM, Haluza D (2017) Work-related self-assessed fatigue and recovery among nurses. *International Archives of Occupational and Environmental Health* 90(2):197–205.
- Carayon P, Gurses AP (2008) Nursing workload and patient safety—a human factors engineering perspective. *Patient Safety and Quality: An Evidence-Based Handbook for Nurses* .
- Card D, Mas A, Moretti E, Saez E (2012) Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review* 102(6):2981–3003.
- Chen Y, Chan CW, Dong J, Jackson EM, Yip NH, Rossetti SC (2024) Beyond order-based nursing workload: A retrospective cohort study in intensive care units. *Journal of Nursing Scholarship* .
- Cockerham M, Kang DH, Beier ME (2023) Consecutive shifts: A repeated measure study to evaluate stress, biomarkers, social support, and fatigue in medical/surgical nurses. *Behavioral Sciences* 13(7):571.
- Dall’Ora C, Griffiths P, Ball J, Simon M, Aiken LH (2015) Association of 12 h shifts and nurses’ job satisfaction, burnout and intention to leave: findings from a cross-sectional study of 12 european countries. *BMJ open* 5(9):e008331.
- Dhaini SR, Abed Al Ahad M, Simon M, Ausserhofer D, Abu-Saad Huijjer H, Elbejjani M (2022) Repeated assessments and predictors of nurses’ shift-specific perceived workload. *Nursing Forum*, volume 57, 1026–1033 (Wiley Online Library).
- Di Tella R, MacCulloch RJ, Oswald AJ (2001) Preferences over inflation and unemployment: Evidence from surveys of happiness. *American economic review* 91(1):335–341.
- Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P (2008) Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. *Medical Care* 232–239.

- Freeman M, Robinson S, Scholtes S (2021) Gatekeeping, fast and slow: An empirical study of referral errors in the emergency department. *Management Science* 67(7):4209–4232.
- Furnham A, Boo HC (2011) A literature review of the anchoring effect. *The journal of socio-economics* 40(1):35–42.
- Geiger-Brown J, Rogers VE, Trinkoff AM, Kane RL, Bausell RB, Scharf SM (2012) Sleep, sleepiness, fatigue, and performance of 12-hour-shift nurses. *Chronobiology international* 29(2):211–219.
- Green LV, Savin S, Savva N (2013) “Nursevendor problem”: Personnel staffing in the presence of endogenous absenteeism. *Management Science* 59(10):2237–2256.
- Greenglass ER, Burke RJ (2001) Stress and the effects of hospital restructuring in nurses. *Canadian Journal of Nursing Research Archive* .
- Halbesleben JR, Wakefield BJ, Wakefield DS, Cooper LB (2008) Nurse burnout and patient safety outcomes: Nurse safety perception versus reporting behavior. *Western Journal of Nursing Research* 30(5):560–577.
- Haluza D, Schmidt VM, Blasche G (2019) Time course of recovery after two successive night shifts: A diary study among austrian nurses. *Journal of nursing management* 27(1):190–196.
- Hart SG, Staveland LE (1988) Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, volume 52, 139–183 (Elsevier).
- Holland P, Tham TL, Sheehan C, Cooper B (2019) The impact of perceived workload on nurse satisfaction with work-life balance and intention to leave the occupation. *Applied Nursing Research* 49:70–76.
- Hoogendoorn M, Brinkman S, Spijkstra J, Bosman R, Margadant C, Haringman J, de Keizer N (2021) The objective nursing workload and perceived nursing workload in intensive care units: Analysis of association. *International journal of nursing studies* 114:103852.
- Ibanez MR, Clark JR, Huckman RS, Staats BR (2018) Discretionary task ordering: Queue management in radiological services. *Management Science* 64(9):4389–4407.
- Jiang Z, Silberholz J, Wang Y, Costa DK, Sjoding M (2025) Can employees’ past helping behavior be used to improve shift scheduling? evidence from icu nurses. *Management Science* .
- Jourdain G, Chênevert D (2010) Job demands–resources, burnout and intention to leave the nursing profession: A questionnaire survey. *International Journal of Nursing Studies* 47(6):709–722.
- Kamalahmadi M, Yu Q, Zhou YP (2021) Call to duty: Just-in-time scheduling in a restaurant chain. *Management Science* 67(11):6751–6781.
- Kapu AN, McComiskey CA, Buckler L, Derkazarian J, Goda T, Lofgren MA, McIlvennan CK, Raaum J, Selig PM, Sicoutris C, et al. (2016) Advanced practice providers’ perceptions of patient workload: Results of a multi-institutional survey. *JONA: The Journal of Nursing Administration* 46(10):521–529.
- Kc DS, Staats BR, Kouchaki M, Gino F (2020) Task selection and workload: A focus on completing easy tasks hurts performance. *Management Science* 66(10):4397–4416.

- Kc DS, Terwiesch C (2009) Impact of workload on service time and patient safety: an econometric analysis of hospital operations. *Management science* 55(9):1486–1498.
- Kc DS, Terwiesch C (2012) An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management* 14(1):50–65.
- Kesavan S, Lambert SJ, Williams JC, Pendem PK (2022) Doing well by doing good: Improving retail store performance with responsible scheduling practices at the gap, inc. *Management Science* 68(11):7818–7836.
- Khamisa N, Oldenburg B, Peltzer K, Ilic D (2015) Work related stress, burnout, job satisfaction and general health of nurses. *International journal of environmental research and public health* 12(1):652–666.
- Kim K, Mehrotra S (2015) A two-stage stochastic integer programming approach to integrated staffing and scheduling with application to nurse management. *Operations Research* 63(6):1431–1451.
- Laschinger HKS, Leiter MP (2006) The impact of nursing work environments on patient safety outcomes: The mediating role of burnout engagement. *Journal of Nursing Administration* 36(5):259–267.
- Levy DR, Withall JB, Mishuris RG, Tiase V, Diamond C, Douthit B, Grabowska M, Lee RY, Moy AJ, Sengstack P, et al. (2024) Defining documentation burden (docburden) and excessive docburden for all health professionals: A scoping review. *Applied Clinical Informatics* 15(05):898–913.
- Li W, Sun Z, Hong LJ (2021) Who is next: Patient prioritization under emergency department blocking. *Operations Research* .
- Liu X, Zheng J, Liu K, Baggs JG, Liu J, Wu Y, You L (2018) Hospital nursing organizational factors, nursing care left undone, and nurse burnout as predictors of patient safety: A structural equation modeling analysis. *International Journal of Nursing Studies* 86:82–89.
- Long EF, Mathews KS (2018) The boarding patient: effects of icu and hospital occupancy surges on patient flow. *Production and Operations Management* 27(12):2122–2143.
- Luo Q, Wang R, Chen H, Xie X, Tang CS (2022) Going beyond occupancy rate: The impact of doctors' and nurses' task workload on health outcomes. *Available at SSRN* .
- MacPhee M, Dahinten VS, Havaei F (2017) The impact of heavy perceived nurse workloads on patient and nurse outcomes. *Administrative Sciences* 7(1):7.
- Miranda DR, Nap R, de Rijk A, Schaufeli W, Iapichino G, of the TISS Working Group M, et al. (2003) Nursing activities score. *Critical Care Medicine* 31(2):374–382.
- Mullinax C, Lawley M (2002) Assigning patients to nurses in neonatal intensive care. *Journal of the operational research society* 53(1):25–35.
- Myny D, Van Goubergen D, Gobert M, Vanderwee K, Van Hecke A, Defloor T (2011) Non-direct patient care factors influencing nursing workload: A review of the literature. *Journal of advanced nursing* 67(10):2109–2129.

- Niewoehner III RJ, Diwas K, Staats B (2022) Physician discretion and patient pick-up: How familiarity encourages multitasking in the emergency department. *Operations Research* .
- OptiLink K (2012) Acuity-based scheduling supports quality-conscious and cost-effective care.
- Phillips C (2020) Relationships between workload perception, burnout, and intent to leave among medical-surgical nurses. *JBHI Evidence Implementation* 18(2):265–273.
- Schaus P, Van Hentenryck P, Régim JC (2009) Scalable load balancing in nurse to patient assignment problems. *International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, 248–262 (Springer).
- Scott LD, Rogers AE, Hwang WT, Zhang Y (2006) Effects of critical care nurses' work hours on vigilance and patients' safety. *American Journal of critical care* 15(1):30–37.
- Shah MK, Gandrakota N, Cimiotti JP, Ghose N, Moore M, Ali MK (2021) Prevalence of and factors associated with nurse burnout in the US. *JAMA network open* 4(2):e2036469–e2036469.
- Sherif M, Taub D, Hovland CI (1958) Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of experimental psychology* 55(2):150.
- Shi P (2022) Optimal matchmaking strategy in two-sided marketplaces. *Management Science* .
- Sir MY, Dundar B, Steege LMB, Pasupathy KS (2015) Nurse-patient assignment models considering patient acuity metrics and nurses' perceived workload. *Journal of biomedical informatics* 55:237–248.
- Soltani M, Batt RJ, Bavafa H, Patterson BW (2022) Does what happens in the ED stay in the ED? the effects of emergency department physician workload on post-ED care use. *Manufacturing & Service Operations Management* 24(6):3079–3098.
- Stiegler K, Martiniano R, Forte G (2021) Health care employment projections, 2019–2029: An analysis of bureau of labor statistics projections by setting and by occupation. *Rensselaer, NY: Center for Health Workforce Studies, School of Public Health, SUNY Albany* .
- Stimpfel AW, Sloane DM, Aiken LH (2012) The longer the shifts for hospital nurses, the higher the levels of burnout and patient dissatisfaction. *Health affairs* 31(11):2501–2509.
- Thompson BJ (2019) Does work-induced fatigue accumulate across three compressed 12 hour shifts in hospital nurses and aides? *PloS one* 14(2):e0211715.
- Thompson BJ, Stock MS, Banuelas VK (2017) Effects of accumulating work shifts on performance-based fatigue using multiple strength measurements in day and night shift nurses and aides. *Human factors* 59(3):346–356.
- Vahey DC, Aiken LH, Sloane DM, Clarke SP, Vargas D (2004) Nurse burnout and patient satisfaction. *Medical Care* 42(2 Suppl):II57.
- van der Mark CJ, Vermeulen H, Hendriks PH, van Oostveen CJ (2021) Measuring perceived adequacy of staffing to incorporate nurses' judgement into hospital capacity management: a scoping review. *BMJ open* 11(4):e045245.

Wang WY, Gupta D (2014) Nurse absenteeism and staffing strategies for hospital inpatient units. *Manufacturing & Service Operations Management* 16(3):439–454.

Wooldridge JM (2010) *Econometric analysis of cross section and panel data* (MIT press).

Appendix A: Full Estimation Results of Main Models

Table 7 Full Estimation Results: Comparison of Linear Response and Ordered Probit Models

Variable name	Linear Response Model		Ordered Probit Model	
	Coefficient	Std. Error	Coefficient	Std. Error
Recent workload exposure	0.2726***	0.060	0.5952***	0.1677
Day shift indicator	0.0297	0.0996	0.0576	0.2245
Weekday indicator	-0.0447*	0.0219	-0.1040*	0.0488
Elixhauser comorbidity score	0.0028	0.0023	0.0055	0.0051
Average hourly LAPS	0.0024	0.0013	0.0052	0.0031
Current order-based workload intensity	-0.9933***	0.0830	-2.1957***	0.1947
Patient discharged during current shift	-0.1457***	0.0344	-0.3107***	0.0756
ICU length of stay before shift, in hours	-0.000180**	0.000057	-0.000412**	0.000127
First 24 hours in ICU indicator	0.1893***	0.0451	0.4199***	0.1007
Patient age	-0.0034***	0.0008	-0.0076***	0.0017
Male patient indicator	0.0915***	0.0157	0.2032***	0.0341
Admitted from emergency department	-0.0862***	0.0195	-0.1884***	0.0429
Number of admissions during shift	0.0201*	0.0079	0.0445*	0.0174
Number of discharges during shift	-0.0183**	0.0070	-0.0394*	0.0154
Average unit census during shift	0.0120	0.0106	0.0293	0.0244
Average patient-to-nurse ratio	-0.0483	0.0627	-0.1221	0.1420
Relative DDC intensity	0.4179**	0.1423	0.8815**	0.3216
Relative Elixhauser score	-0.1575*	0.0669	-0.3278*	0.1500
Relative LAPS	-0.0973	0.1008	-0.2114	0.2282
Number of patients assigned to nurse	-0.4876***	0.0466	-1.0970***	0.1073
Continuity-of-care indicator	0.0436*	0.0205	0.0965*	0.0454
Rest time since prior working shift	0.0237***	0.0066	0.0516***	0.0148
Insurance: Medicaid	-0.0029	0.0209	-0.0109	0.0469
Insurance: Medicare	0.0033	0.0215	0.0047	0.0483
Insurance: self-pay or none	0.2036*	0.0809	0.5355**	0.1815
Inverse Mills ratio	0.2382	0.1726	0.5736	0.5549

Notes: Nurse fixed effects are included but not reported. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix B: Details of Data Cleaning Procedure

In this study, we combine data from multiple sources: patient-flow data, hospitalization data, LAPS data, OptLink assignment data, OptLink intensity data, and nurse staffing data. To process the data, we first correct erroneous records by cross-checking different data sources. When inconsistency arises, we use the majority-vote rule to correct the records. Then we impute some missing non-admission shift assignment data leveraging the nurse staffing data. For this part, there are 83 patient-shifts belonging to 62 shifts that do not have the nurse assignment information. By checking the staffing data, we find that in 33 of the 62 shifts, there is only a single patient missing the nurse assignment information and a single “idle” nurse, i.e., a nurse who is working in that shift according to the staffing data but does not have an assignment recorded in the assignment data. For these 33 shifts, we match the idle nurse and the patient who does not have an assignment. In 16 of the 62 shifts, there is a single patient missing nurse assignment information but more than one idle nurse. For these 16 shifts, we match all the idle nurses with the patient that does not have

an assigned nurse. Lastly, in 13 of the 62 shifts, there is no idle nurse. For these shifts, we randomly pick a nurse among those who have only one patient assigned and match the nurse with the patient that does not have an assigned nurse.

All shifts in January and December are excluded due to missing data: some information for patients admitted before January 2018 or discharged after December 2018 is missing. Thus, our final cohort for analysis is from Feb 1, 2018 to Nov 30, 2018. We further exclude all patient-shifts that correspond to the admission shifts of the patients (1,079 observations) due to a large amount of missing assignment information. (28.7% of the patients do not have the nurse assignment information during the shift at which they are admitted to the MICU.) However, we do include the admitted patients in the calculation of the unit census. We also control for the number of admissions in each shift in our regression analysis. Lastly, we conduct sensitivity analysis by including the imputed admission shift assignments (see Section 5.2). We exclude 94 observations associated with four shifts that have a large amount of missing assignment information. Next, we exclude 454 observations associated with 32 unique patients whose patient-level information like LAPS or Elixhauser scores are missing. These observations are included in the unit census calculation though. Lastly, we classify a nurse as “primary” versus “non-primary” based on whether the nurse worked for more than 30 shifts during the study period (30 corresponds to the lower 5% percentile). The “non-primary” nurses, which includes all the agency nurses (each of whom works for at most 13 shifts during the study period), may primarily work at other units rather than the MICU, and thus we may not be able to measure their recent workload exposure accurately. Based on this classification rule, our data contains 72 primary nurses and 77 non-primary nurses. The primary nurses worked on average 112 shifts during the study period. We exclude 700 observations that are associated with the 77 non-primary nurses. Again, these observations are included in the unit census calculation. In addition, we exclude 273 observations in the first week of the study period since the recent workload cannot be accurately evaluated for these observations. Lastly, we exclude 916 observations for which multiple nurses (mostly two nurses) are assigned to a single patient. This assignment pattern is relatively rare, only 5%. The remaining patient-to-nurse assignments are either one patient to one nurse or two patients to one nurse. This is because our analysis is at the nurse-patient-shift level, while workload intensity scores are measured at the patient-shift level. When multiple nurses are assigned to a single patient, we can not determine which nurse reported the clinically perceived workload. We also run a robustness check by including these patient-shifts and duplicating the corresponding entries so that we have the workload intensities scores measured at the nurse-patient-shift level (see Appendix E). More details of data processing and cleaning are provided in Appendix B. Our final cohort contains 11,736 nurse-patient-shift observations, associated with 939 unique patients and 72 unique nurses. See Figure 4 for a summary of the final cohort construction.

Table 8 also provides an overview and some summary statistics of the control variables used in our estimation model.

Appendix C: Data Evidences to Support Validity of IVs

In this section, we take several precautions to enhance the validity of our IVs. Since we exclude the focal patient and the focal nurse does not work in the “IV” shift, the likelihood of the IVs being correlated with

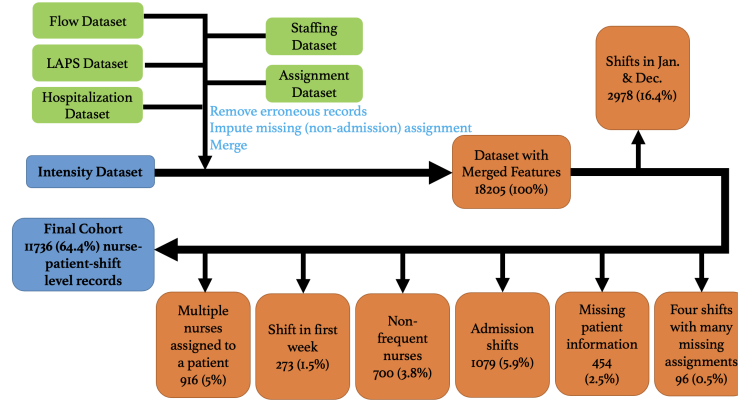


Figure 4 Flow chart of data cleaning procedure

Table 8 Summary statistics of control variables

Variable category	Variable	Mean/Proportion	Stdev
Shift factors (unit level)	day shift flag	48.8%	
	weekday flag	71.6%	
	# of admissions	1.72	1.10
	# of discharges	1.78	1.15
	census	22.36	1.14
	patient-to-nurse ratio	1.58	0.125
Patient factors (visit level)	age	57.19	17.55
	gender-male	52.4%	
	from emergency department	35.2%	
	insurance: commercial	37.3%	
	insurance: Medicaid	21.1%	
	insurance: Medicare	40.9%	
	insurance: self-pay	0.6%	
Patient factors (shift level)	Elixhauser	30.88	15.52
	average hourly LAPS	83.27	32.52
	DDCIntensity	2.18	0.61
	average unit DDCIntensity	2.32	0.56
	first 24-hour flag	7.5%	
	discharge during the shift	1.82	0.68
	LOS so far (hours)	186.48	225.3
	relative LAPS	1.006	0.407
	relative DDCIntensity	1.002	0.285
Nursing factors	relative Elixhauser	1.013	0.531
	continuity of care flag	44.8%	
	rest time	4.76	5.94
	# of patients assigned per nurse	1.84	0.43

the characteristics of the focal nurse or patient is reduced. However, unit-wide busyness could still pose a concern, as a busy unit may lead the focal nurse to perceive a higher workload, and workload levels can exhibit temporal correlation. To address this, our main model controls for unit-level busyness and the focal patient's relative severity compared to the shift average. In addition, the empirical correlation between the busyness in shift $t - k_n - 1$ and shift t is only 0.172, likely due to relatively large k_n 's. The average value of k_n

is 4.75 shifts (approximately 57 hours). Very few, i.e. less than 0.5% of the nurse-patient-shift observations have a k_n less than 2 shifts (24 hours). We also note that a nurse who previously managed a high-workload patient in shift $t - k_n$ may be more likely to be assigned to the same patient in shift t for continuity of care. Meanwhile, this familiarity may help reduce the perceived workload. To account for this, our main model includes a continuity-of-care flag.

We also explore additional potential violations of the exclusion restriction by analyzing nurses' working patterns below. We rule out scenarios such as most nurses working in the focal shift having had the same previous assignment shift, a nurse consistently working in the same type of shifts, or two nurses regularly being assigned to the same shift.

Specifically, we report some summary statistics related to nurse working patterns. For nurse-to-patient matching patterns, 2087 (17.8%) observations (nurse-patient-shifts) correspond to one-nurse-to-one-patient matching; 9320 (79.47%) observations correspond to one-nurse-to-two-patients matching; and the remaining 329 (2.8%) observations correspond to one-nurse-to-three-patients matching. The number of nurses classified into levels 1, 2, and 3 is 53 (73.6%), 5 (6.9%), and 14 (19.5%), respectively.

Figure 5 depicts the histogram of nurses' rest times (the number of shifts between the focal shift and the focal nurse's last assignment shift).

For different nurses working in the same shift, there is significant heterogeneity in the timing of their last assignment shifts. To demonstrate this, we calculate the entropy (a classic metric of variability, where a larger value indicates less concentration and higher variability) of the distribution of nurses' rest times for each shift. Figure 6 shows the distribution of these entropy values. Note that most of these entropy values are significantly greater than zero.

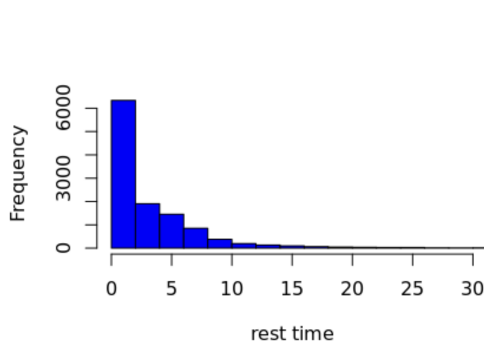


Figure 5 Histogram of nurses' rest time

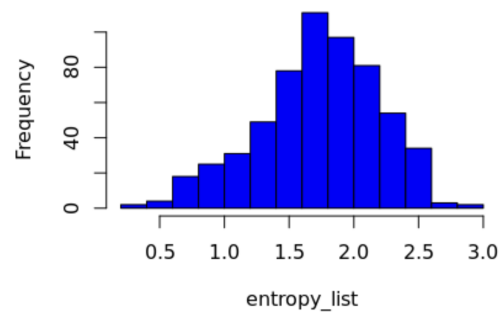


Figure 6 Histogram of the entropy of different nurses' rest times in each shift

Figure 7 shows the histogram of the total number of shifts each nurse worked during the studied period.

Nurses work in a variety of shifts, with no nurse restricted to only day/night shifts or specific days of the week. There are 14 different types of shifts (e.g., Monday day shift, Monday night shift, etc). We calculate the entropy of each nurse's working shift distribution across these 14 shifts. Figure 8 plots the histogram of these entropy values. Note that in the extreme case where a nurse only ever worked in 3 types of shifts (on average, a nurse worked 3 shifts per week in our data), the corresponding entropy would be $\log(3) \approx 1.1$. We

observe that for most nurses, the entropy is much larger than 1, indicating that nurses do not always follow the same working schedule.

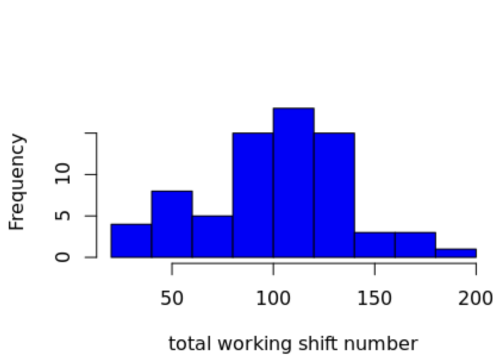


Figure 7 Histogram of nurses' total working shift numbers

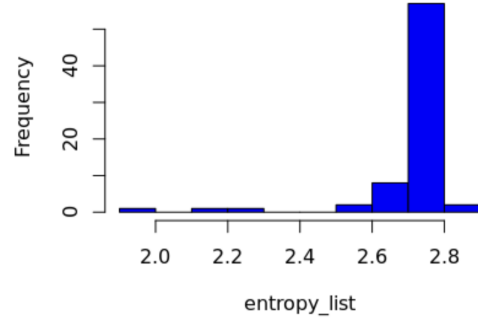


Figure 8 Histogram of the entropy of nurse working shift distribution over 14 different types of shift

Lastly, we examine the overlap between nurses' working shifts. For each pair of nurses, we calculate their overlapping rates, defined as the number of shifts they worked together divided by the total number of working shifts for each nurse. Specifically, assume the working shift sets of nurse n_1 and n_2 be T_{n_1} and T_{n_2} . The overlapping rates for nurse n_1 and n_2 are defined as $r_{n_1} = |T_{n_1} \cap T_{n_2}| / |T_{n_1}|$ and $r_{n_2} = |T_{n_1} \cap T_{n_2}| / |T_{n_2}|$. Figure 9 shows the histogram of overlapping rates of all nurse pairs, i.e., $\{r_{n_i}, r_{n_j} : i \neq j, \forall i, j \in \mathcal{N}\}$, where $\mathcal{N}$ denotes the index set of nurses. We observe that for most pairs, the overlapping rates are quite small.

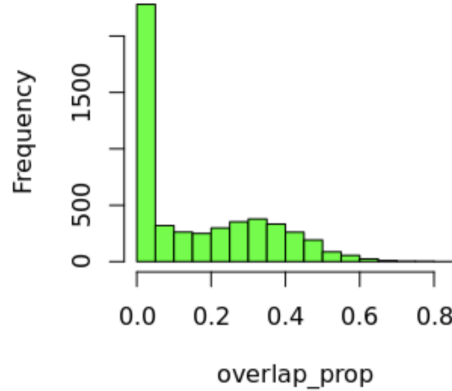


Figure 9 Histogram of overlapping rate between different nurses' working shifts

Appendix D: More Details of Sensitivity Analysis

In this section, we provide more analysis of measurement error and censored response, including detailed analytical derivations and some empirical verifications.

D.1. Measurement Error Analysis

D.1.1. Analytical Insights Let $Rec_Work_{n,p,t}^*$ denote the true recent workload of nurse n who cares for patient p in shift t , which may contain the workload associated with some assigned patient(s) who is/are admitted in the nurse's last assignment shift, i.e., shift $t - k_n$. Recall that $Rec_Work_{n,p,t}$ is the recent workload we observe, which does not contain the workload associated with the newly admitted patients. We define

$$\Delta Rec_Work_{n,p,t} = Rec_Work_{n,p,t}^* - Rec_Work_{n,p,t}.$$

Then we have

$$\begin{aligned} Rec_Work_{n,p,t}^* &= Rec_Work_{n,p,t} + \Delta Rec_Work_{n,p,t} \\ &= X_{n,p,t}^\top \delta + \eta_1 \cdot Avg_Last_DDC_{n,p,t} + \eta_2 \cdot Avg_Last_LAPS_{n,p,t} + \varepsilon_{n,p,t} + \Delta Rec_Work_{n,p,t}, \end{aligned} \quad (11)$$

and the outcome equation takes the form:

$$Intensity_Gap_{n,p,t} = \theta \cdot Rec_Work_{n,p,t}^* + X_{n,p,t}^\top \beta + \gamma \lambda (\tilde{X}_{n,p,t}^\top \alpha) + e_{n,p,t},$$

where the noise $e_{n,p,t}$ has mean zero conditional on $X_{n,p,t}$ and $Report_PJ_{n,p,t} = 1$. By plugging (11) in the outcome equation, we obtain

$$\begin{aligned} Intensity_Gap_{n,p,t} &= X_{n,p,t}^\top (\beta + \theta \delta) + \gamma \cdot \lambda (\tilde{X}_{n,p,t}^\top \alpha) + \theta \eta_1 \cdot Avg_Last_DDC_{n,p,t} \\ &\quad + \theta \eta_2 \cdot Avg_Last_LAPS_{n,p,t} + (e_{n,p,t} + \theta \varepsilon_{n,p,t}) + \theta \cdot \Delta Rec_Work_{n,p,t}. \end{aligned}$$

To simplify the discussion, we assume that there is only a single IV, which is denoted as $IV_{n,p,t}$, and the IV is positively correlated with $Rec_Work_{n,p,t}$, i.e., the IV has coefficient $\theta\eta$ with $\eta > 0$ in the outcome equation. We further assume that the effect of covariates $Z_{n,p,t} := (X_{n,p,t}, \lambda(\tilde{X}_{n,p,t}^\top \alpha))$ is negligible, i.e.,

$$Intensity_Gap_{n,p,t} = \beta_0 + \theta\eta \cdot IV_{n,p,t} + \theta \cdot \Delta Rec_Work_{n,p,t} + (e_{n,p,t} + \theta \varepsilon_{n,p,t}).$$

In this case, when using 2SLS to estimate the treatment effect, the estimator $\hat{\theta}$ converges to

$$\begin{aligned} \theta &+ \frac{Cov(IV_{n,p,t}, e_{n,p,t} + \theta \varepsilon_{n,p,t} | Report_PJ_{n,p,t} = 1)}{\eta Var(IV_{n,p,t} | Report_PJ_{n,p,t} = 1)} \\ &+ \theta \cdot \frac{Cov(IV_{n,p,t}, \Delta Rec_Work_{n,p,t} | Report_PJ_{n,p,t} = 1)}{\eta Var(IV_{n,p,t} | Report_PJ_{n,p,t} = 1)}. \end{aligned} \quad (12)$$

We next show that the second term in (12) is equal to zero. First, $Cov(IV_{n,p,t}, \varepsilon_{n,p,t} | Report_PJ_{n,p,t} = 1) = 0$ because by assumption, $IV_{n,p,t}$ is independent with $\varepsilon_{n,p,t}$. In addition,

$$\begin{aligned} &Cov(IV_{n,p,t}, e_{n,p,t} | Report_PJ_{n,p,t} = 1) \\ &= E[IV_{n,p,t} \cdot e_{n,p,t} | Report_PJ_{n,p,t} = 1] - E[IV_{n,p,t} | Report_PJ_{n,p,t} = 1] \cdot E[e_{n,p,t} | Report_PJ_{n,p,t} = 1] \\ &= E[IV_{n,p,t} \cdot (u_{n,p,t} - E[u_{n,p,t} | \tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1]) | Report_PJ_{n,p,t} = 1] \\ &= E[IV_{n,p,t} \cdot u_{n,p,t} | Report_PJ_{n,p,t} = 1] - E[E[IV_{n,p,t} \cdot u_{n,p,t} | \tilde{X}_{n,p,t}, Report_PJ_{n,p,t} = 1] | Report_PJ_{n,p,t} = 1] \\ &= 0, \text{ because } IV_{n,p,t} \text{ is contained in } \tilde{X}_{n,p,t}. \end{aligned}$$

Then, we incur an estimation bias of

$$\theta \cdot \frac{Cov(IV_{n,p,t}, \Delta Rec_Work_{n,p,t} | Report_PJ_{n,p,t} = 1)}{\eta Var(IV_{n,p,t} | Report_PJ_{n,p,t} = 1)},$$

which is the third term in (12)

It is easy to see that when $IV_{n,p,t}$ and $\Delta Rec_Work_{n,p,t}$ are negatively correlated, we underestimate the treatment effect. Otherwise, we overestimate the treatment effect. When considering the effect of other covariates $Z_{n,p,t}$, the above rationale should still hold by considering the residuals of $IV_{n,p,t}$ and $Admin_Work_{n,p,t}$ regressed on $Z_{n,p,t}$ respectively.

D.1.2. Details of Imputation Algorithm Let $t' = t - k_n$ denote nurse n 's last assignment shift prior to shift t . We first impute the missing admission assignments in shift t' and then calculate $Rec_Work_{n,p,t}^*$ based on the full assignment information for nurse n in shift t' . Recall that $P_{t'}(n)$ is the set of non-admission shift patients that nurse n is taking care of in shift t' . Define the non-admission assignment workload in shift t' as

$$Non_Admin_Work_{n,p,t'} = \sum_{p' \in P_{t'}(n)} DDCIntensity_{p',t'}.$$

Note that $Rec_Work_{n,p,t} = Non_Admin_Work_{n,p,t'}$. When imputing the missing admission assignments in shift t' , we assume a nurse can be assigned at most one newly admitted patient. In the first scheme, we try to match an admission-shift patient with a higher DDCIntensity with a nurse with a higher non-admission assignment workload. This creates a positive correlation between $IV_{n,p,t}$ and $\Delta Rec_Work_{n,p,t}$. In the second scheme, we try to match an admission-shift patient with a higher DDCIntensity with a nurse with a lower non-admission assignment workload. In the third scheme, we randomly match the admission-shift patients with nurses.

After imputing the missing admission assignments, let $AP_{t'}(n)$ denote the set of admission patients assigned to nurse n in shift t' (note that $AP_{t'}(n)$ contains at most one patient). Then, we can calculate the true recent workload exposure $Rec_Work_{n,p,t}^*$ as the sum of DDCIntensity scores of all patients assigned to nurse n in shift t' , i.e.,

$$Rec_Work_{n,p,t}^* = \sum_{q \in P_{t-k_n}(n) \cup AP_{t-k_n}(n)} DDCIntensity_{q,t-k_n} \cdot 1(k_n \leq 14). \quad (13)$$

Algorithm 1 illustrates the missing admission shift assignment imputation mechanism details.

D.2. Censored Response Analysis

D.2.1. Analytical Insights Suppose we can remove the cap at 4 and the floor at 1 and let $PJIntensity_{n,p,t}^*$ be true PJIntensity a nurse wants to report. Thus, the true workload gap can be represented as $Intensity_Gap_{n,p,t}^* = PJIntensity_{n,p,t}^* - DDCIntensity_{n,p,t}$ and the censored workload gap becomes

$$Intensity_Gap_{n,p,t}^* = \theta \cdot Rec_Work_{n,p,t} + X_{n,p,t}^\top \beta + \gamma \lambda(\tilde{X}_{n,p,t}^\top \alpha) + e_{n,p,t},$$

$$Rec_Work_{n,p,t} = X_{n,p,t}^\top \delta + \eta_1 \cdot Avg_Last_DDC_{n,p,t} + \eta_2 \cdot Avg_Last_LAPS_{n,p,t} + \varepsilon_{n,p,t},$$

$$Intensity_Gap_{n,p,t} = \mathcal{C}(PJIntensity_{n,p,t}^*) - DDCIntensity_{n,p,t}$$

where the censoring operator $\mathcal{C}$ is defined as $\mathcal{C}(x) = 1 \cdot (x < 1) + x \cdot 1(1 \leq x \leq 4) + 4 \cdot (x > 4)$. Let $\Delta Gap_{n,p,t} = Intensity_Gap_{n,p,t} - Intensity_Gap_{n,p,t}^*$, which is the gap between the censored response and true response. Then,

$$\begin{aligned} Intensity_Gap_{n,p,t} &= \theta \cdot Rec_Work_{n,p,t} + X_{n,p,t}^\top \beta + \gamma \lambda(\tilde{X}_{n,p,t}^\top \alpha) + e_{n,p,t} + \Delta Gap_{n,p,t} \\ &= X_{n,p,t}^\top (\beta + \theta \delta) + \gamma \cdot \lambda(\tilde{X}_{n,p,t}^\top \alpha) + \theta \eta_1 \cdot Avg_Last_DDC_{n,p,t} \\ &\quad + \theta \eta_2 \cdot Avg_Last_LAPS_{n,p,t} + (e_{n,p,t} + \theta \varepsilon_{n,p,t}) + \Delta Gap_{n,p,t}. \end{aligned}$$

Algorithm 1 Pseudo-code of missing admission assignment imputation strategy

for shift $t' = 1, 2, \dots, T$ **do**

Find $\mathcal{AMP}_{t'}$, the set of admission patients in shift t who do not have the nurse assignment information. Sort $\mathcal{AMP}_{t'}$ in the decreasing order of $DDCIntensity_{p,t'}$.

Let $\mathcal{NAN}_{t'}$, denote the set of nurses working in shift t' who do not have an admission assignment. Calculate their current non-admission assignment workload, i.e., $Non_Admin_Work_{n,t'}$ for all $n \in \mathcal{NAN}_{t'}$. Sort $\mathcal{NAN}_{t'}$ in the decreasing order of $Non_Admin_Work_{n,t'}$.

while $\mathcal{AMP}_{t'}$ is not empty **do**

if Want to achieve a positive correlation **then**

Match the first patient in $\mathcal{AMP}_{t'}$ with the first nurse in $\mathcal{NAN}_{t'}$.

else if Want to achieve a negative correlation **then**

Match first patient in $\mathcal{AMP}_{t'}$ with the last nurse in $\mathcal{NAN}_{t'}$.

else

Match the first patient in $\mathcal{AMP}_{t'}$ with a randomly selected nurse from $\mathcal{NAN}_{t'}$.

end if

Remove the matched patient from $\mathcal{AMP}_{t'}$ and remove the matched nurse from $\mathcal{NAN}_{t'}$.

end while

end for

where the noise terms satisfy

$$\mathbb{E}[e_{n,p,t} + \theta \varepsilon_{n,p,t} | Z_{n,p,t}, Avg_Last_DDC_{n,p,t}, Avg_Last_LAPS_{n,p,t}, Report_PJ_{n,p,t} = 1] = 0,$$

and $Z_{n,p,t} = (X_{n,p,t}, \lambda(\tilde{X}_{n,p,t}^\top \alpha))$. To simplify the discussion, we assume there is only one IV, which is denoted as $IV_{n,p,t}$ with corresponding coefficient $\theta\eta$ for $\eta > 0$ (i.e., the IV is positively correlated with $Rec_Work_{n,p,t}$), and the effect of the covariates $Z_{n,p,t}$ is negligible. Similar to the analysis in Section 3.1.3, when using 2SLS to estimate the treatment effect, the estimator $\hat{\theta}$ converges to

$$\theta + \theta \cdot \frac{Cov(IV_{n,p,t}, \Delta Gap_{n,p,t} | Report_PJ_{n,p,t} = 1)}{\eta Var(IV_{n,p,t} | Report_PJ_{n,p,t} = 1)}.$$

Note that $\mathcal{C}(x) - x$ is a monotonically decreasing function of x . Since $IV_{n,p,t}$ is positively correlated with $Rec_Work_{n,p,t}$, which is hypothesized to have a positive impact on $PJIntensity_{n,p,t}^*$, we have $Cov(IV_{n,p,t}, \Delta Gap_{n,p,t} | Report_PJ_{n,p,t} = 1) < 0$. This implies that the censored response leads to an underestimation of the treatment effect. As such, we do not explicitly correct for this potential bias, but note that it may result in *conservative* estimates of the true treatment effect. In particular, if we find a positive and significant effect of $Rec_Work_{n,p,t}$ on $Intensity_Gap_{n,p,t}$, this implies the true treatment effect is positive and significant, possibly with an even larger effect size.

Appendix E: Robustness Check and Heterogeneity Analysis

E.1. Alternative Model Specifications

In this section, we consider several alternative model specifications to further verify the robustness of our estimation results. The primary objective is to validate the qualitative consistency and directional stability of our main conclusions across various empirical settings. Since the previous analysis already demonstrates a strong qualitative and quantitative alignment between the linear response model and the ordered probit model, we conduct subsequent analysis exclusively within the linear framework. This ensures computational tractability and allows for a straightforward interpretation of coefficients across various specifications, while sufficiently capturing the underlying directional effects.

E.1.1. Alternative Recent Workload Measures In our main model, $Rec_Work_{n,p,t}$ is defined as the sum of the DDCIntensity scores among all the patients assigned to the focal nurse in the nurse's last assignment shift. In our primary analysis, we treat this as a continuous variables. However, as shown in Figure 1, our recent workload is distributed across 10 discrete levels (0–9). Ideally, one might treat each discrete level as a separate indicator variable. However, in a standard IV framework, identifying these indicators would require a correspondingly rich set of IVs. Because our identification strategy uses only two IVs, we address this concern by implementing a binary treatment model as a robustness check. Specifically, we bifurcate the recent workload into high and low categories using a median split (threshold = 4). We then re-estimate our baseline model using this binary treatment while maintaining sample selection and IV framework. The estimated treatment effect remains positive and highly significant (point estimate is 0.653, SE is 0.182, $p < 0.01$), which confirms that our findings are not an artifact of the continuous assumption of treatment. It reflects a robust workload memory effect where transitioning from a normal workload to a high-intensity workload significantly amplifies the nurse's subsequent perceived burden.

In addition to a DDCIntensity-based recent workload measure, we also consider three alternative measures of the recent workload:

1. *Adjusted Maximum DDCIntensity Score.* We calculate the maximum DDCIntensity score among all patients assigned to the nurse during their last assignment shift and add one point if the nurse was assigned two patients. For example, if nurse n was assigned two patients with DDCIntensity scores 2 and 3 in their last assignment shift, the adjusted maximum DDCIntensity score is $\max\{2, 3\} + 1 = 4$.
2. *Sum of Average LAPS Scores.* LAPS is updated hourly and is a measure of patient severity. We first calculate the LAPS at the patient-shift level by averaging the hourly LAPS. We then take the sum of average LAPS scores among all the patients assigned to the focal nurse in the nurse's last assignment shift. Note that a more severe patient is in general associated with a heavier workload.
3. *Sum of PJIntensity Scores.* We calculate the sum of PJIntensity scores among all the patients assigned to the focal nurse in the nurse's last assignment shift. Recall that around 20% of PJIntensity scores are missing in our dataset. We exclude any nurse-patient-shift whose last assignment shift involves missing PJIntensity (We still include observations where only the outcome is missing and use the Heckman selection model to estimate the treatment effect.). Note that this exclusion might introduce bias if PJIntensity is not missing at random. For example, if PJIntensity is more likely to be missing when the nurses are busy, and a

heavier recent workload is associated with a larger workload intensity gap (e.g., due to a larger “burn-out” effect), then we may underestimate the size of the treatment effect. However, since the inverse Mills ratio does not have a significant effect in our main model, we expect the sample selection bias to be small.

Table 9 summarizes the estimation results using the three alternative treatment variables. We also include the treatment effect in the main model (i.e., the “sum of DDC” column) to facilitate comparison. We observe that in all cases, the recent workload has a statistically significant positive effect on $Intensity_Gap_{n,p,t}$. The coefficient in the “sum of LAPS” column is smaller, because the LAPS scores (mean 83.3 and standard deviation 32.5) are in general much larger than the DDCIntensity scores (mean 2.2 and standard deviation 0.61). When considering the effect of a one-standard-deviation increase of the treatment variable, we observe more comparable treatment effect sizes as demonstrated in the “standardized effect” row in the table.

Table 9 Estimation with Alternative Recent Workload Exposure Measures

	Treatment Specification			
	Sum of DDC	Adjusted Max DDC	Sum of LAPS	Sum of PJ
<i>Rec_Work</i>	0.273*** (0.060)	0.321*** (0.159)	0.008*** (0.003)	0.296*** (0.080)
Standardized Effect	0.359	0.299	0.503	0.459
Observations	11736	11736	11736	8326

standard error in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standardized effect denotes the change of outcome when the treatment increases by one unit of standard deviation.

To ensure our model captures temporally relevant workload memory, we implement a look-back truncation window size when constructing *Rec_Work*. Our primary specification uses a one-week truncation window, which effectively filters out assignments separated by more than seven days of absence. To verify that our results are not sensitive to this specific choice of cutoff, we re-estimate the model using two-week and three-week truncation window size. Moreover, for the vast majority of our sample ($> 96\%$), these different windows lead to identical *Rec_Work* values because nurses typically have at least one assignment every seven days. Expanding the truncation window reduces the zero-*Rec_Work* proportion from 3.9% to 0.3%.

As shown in Table 10, the estimated treatment effects remain consistent across the three specifications (the sample sizes in the three columns are slightly different because we have to remove the first one/two/three weeks of data respectively to properly measure *Rec_Work*). A statistical test for the difference between the most extreme estimates (one-week vs. three-week) yields a p -value of approximately 0.70, indicating that the variations between these coefficients are not statistically significant. The slight, non-significant downward shift in the point estimates as the window size increases likely reflects the inclusion of additional recent workload data points. As the truncation boundary expands, the cumulative workload variable incorporates

shifts from more distant periods, which may naturally modulate the average perceived impact per unit of workload compared to the most immediate assignments. Ultimately, the high degree of stability in both the magnitude and significance of these estimates demonstrates that our primary conclusion is not sensitive to the specific choice of the truncation window size.

Table 10 Estimation with Different Truncation Window Size when Calculating *Rec.Work*

	Truncation Window Size		
	One week	Two weeks	Three weeks
<i>Rec.Work</i> *	0.273*** (0.060)	0.252*** (0.063)	0.241*** (0.059)
Observations	11736	11434	11144

standard error in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Lastly, in our main specifications, recent workload is characterized using the objective workload of a nurse's single most recent shift. While this captures immediate spillover effects, it may not fully encapsulate the dynamic nature of cumulative fatigue over time. To more comprehensively address this and capture long-term workload memory, we construct alternative recent workload measures that incorporate information from multiple recent shifts. We define a binary treatment variable to evaluate the impact of sustained high workload. Specifically, for the K -recent-shift specification ($K = 2, 3$), this binary variable equals one if the nurse's order-based workload exceeds the high-workload threshold (defined as the median of the variable, which is 4) in each of the nurse's K most recent working shifts before the focal shift, and equals zero otherwise. This variable indicates whether a nurse experienced a consistently high objective workload across their multiple consecutive assignments. We then re-estimate our primary model using this binary treatment, maintaining a similar IV strategy as before. We find that the effect of sustained recent workload on the perceived workload gap remains positive and statistically significant for both the two-recent-shift specification (point estimate is 0.745, SE is 0.234, $p < 0.01$, marginal effect is 0.256) and the three-recent-shift specification (point estimate is 0.964, SE is 0.320, $p < 0.01$, marginal effect is 0.212). These consistent results imply that the misalignment between perceived and objective workload is not merely a short-term reaction to the single past shift, but rather a genuine consequence of cumulative, sustained fatigue built up over consecutive high-stress assignments. Nevertheless, a more granular investigation into the decay dynamics of this cumulative impact—specifically how it dissipates over subsequent rest periods—is left for future research.

E.1.2. Alternative Outcome Variables Our main outcome variable is the gap between perceived workload and order-based workload. This gap is well aligned with our research question. However, one may also be interested in whether recent workload exposure affects the level of perceived workload itself. To address this question, we re-estimate our model using *PJIntensity* directly as the dependent variable. The coefficient for recent workload remains positive and statistically significant (point estimate is 0.272, SE is

0.070, $p < 0.01$). This estimate is very close to the corresponding estimate when *Intensity_Gap* is used as the dependent variable. The similarity is expected because our main specification already controls for the focal patient's current *DDCIntensity* and other patient- and shift-level acuity measures. Thus, using the gap as the outcome can be viewed as a transparent way to residualize perceived workload relative to the order-based benchmark, while using *PJIntensity* directly provides a complementary clinical interpretation.

E.1.3. Alternative Control Variables To ensure our estimates are not sensitive to omitted variables or restrictive functional form assumptions regarding our covariates, we re-estimate our model using two alternative control strategies to address the following concerns.

First, a potential concern is the treatment of the ordinal objective workload *DDCIntensity* as a continuous variable, which inherently imposes a strict linear functional form assumption. To address this, we replace the current continuous *DDCIntensity* with a set of discrete indicator variables. Under this specification, the coefficient for recent workload remains positive and statistically significant (point estimate is 0.273, SE is 0.073, $p < 0.01$), which is same as the main analysis. This alleviates the concerns regarding the functional form of control variable *DDCIntensity*, demonstrating that our baseline estimates are reliable.

Second, another concern is that the estimated effect of recent workload merely captures a nurse's time-invariant reporting habit or a persistent emotional baseline (e.g., a baseline tendency to complain or over-report stress), rather than the actual carry-over of past true workload. Although this issue can be partially controlled by including nurse fixed effect, to further isolate the impact of recent objective workload from such intrinsic reporting tendencies, we introduce the nurse's perceived workload from their most recent shift as an additional control variable. The rationale is that if a nurse possesses an intrinsic tendency to overreport fatigue, this disposition should be largely captured by their prior subjective report. By controlling for this lagged perception, we effectively absorb the time-invariant reporting bias. However, an empirical challenge in implementing this specification is the presence of missing *PJIntensity* data, meaning this covariate is not universally available. For simplicity, we estimate this specific regression restricted to the subsample of observations where this covariate is well-defined. Even with this stringent control, the effect of recent workload remains positive and statistically significant point estimate is 0.295, SE is 0.085, $p < 0.01$, reinforcing the robustness of our empirical conclusions.

E.1.4. Alternative Standard Error Calculation Lastly, we evaluate the robustness of our statistical inferences by employing alternative clustering strategies to account for potential multi-dimensional dependencies in the error structure. While our primary analysis clusters standard errors at nurse level to address serial correlation, we further implement several two-way clustering via bootstrap to calculate the standard error.

First, we consider a two-way clustering strategy at both the nurse and week-of-year levels. Given that our study only covers a one-year period, we utilize the weekly rather than the monthly dimension to satisfy the asymptotic cluster size requirement. This specification accounts for shared environmental shocks such as seasonal patient surges or unit-wide operational changes that may simultaneously affect all nurses within the same week. Comparing to the one-way clustering standard error reported in the first column of Table 2, the two-way standard error here is 0.095, maintaining a similar magnitude.

In addition, we implement a two-way clustering strategy at both the nurse and shift levels. This specification directly addresses the concern that observations may be correlated not only within nurses over time, but also within the same shift. Such within-shift dependence may arise because nurses sometimes help patients who are not formally assigned to them, and because shared unit-level conditions in a shift may simultaneously affect multiple nurse-patient observations. Therefore, clustering by both nurse and shift allows arbitrary correlation among observations from the same nurse and among observations from the same shift. In this case, the two-way clustered standard error is 0.088, again very close to that in main table.

Lastly, we implement two-way clustering at the nurse and patient level. This test guards against the possibility that unobserved patient-level idiosyncrasies (e.g., behavioral complexity or unmeasured clinical instability) might induce correlated psychological strain among different nurses who treat the same patient. We find that the standard errors in this case is 0.062, almost same as the main result. This suggests that our comprehensive set of patient-acuity controls (e.g., LAPS and Elixhauser scores) effectively captures idiosyncratic patient-level heterogeneity, leaving little residual correlation within patient clusters.

Overall, our estimated treatment effect remains positive and statistically significant under these conservative multi-way clustering specifications. Such a persistence underscores that the workload memory effect is a robust psychological phenomenon, rather than an artifact of unobserved temporal or individual-level correlations.

E.2. Alternative Data Population

In our main model estimation, we exclude 7% data where more than one nurses are assigned to a single patient during a shift. We next run a sensitivity analysis by including these patient-shifts. Since our analysis is at the nurse-patient-shift levels, we duplicate these patient-shifts so we have a copy for each nurse-patient-shift. For example, suppose nurse n_1 and n_2 are assigned to patient p in shift t , we create two observations corresponding to tuples (n_1, p, t) and (n_2, p, t) . The final cohort has 12510 observations and the estimated treatment effect is 0.285 (standard error: 0.081), which is very similar to 0.273 in our main model estimation.

We also conduct a robustness check by only including the nurse-patient-shifts with DDCIntensity equals 2 or 3, which accounts for 90% of the observations. This subpopulation is less likely to incur censored responses. In this case, the estimated treatment effect is 0.290 (standard error: 0.056), which is again very similar to 0.273 in our main model estimation.

E.3. Heterogeneous Treatment Effect

In this section, we examine potential variations in the treatment effect by focusing on three distinct aspects: patient workload intensities, nurses' experience, and rest time between shifts.

E.3.1. Heterogeneity in patient workload intensities We divide the observations into two subsets: high versus low-intensity patients. We define a focal nurse-patient-shift as high-intensity if its DDCIntensity is larger than or equal to 3, which according to the Optilink manual indicates a 'High' or 'Extreme' order-based workload; otherwise, the nurse-patient-shift is low-intensity. Then, we estimate the impact of recent workload exposure on the gap between the PJIntensity and DDCIntensity scores for the high and low-intensity groups separately.

For the low-intensity group, the estimated treatment effect is 0.278 (standard error: 0.081); for the high-intensity group, the estimated treatment effect is 0.425 (standard error: 0.072). Our analysis in Section 3.1.4 shows that censored responses can lead to an underestimation of the true treatment effect. We anticipate this censoring bias to be more pronounced for the high-intensity patient group. This is because when DDCIntensity is 4, nurses are unable to report a higher PJIntensity than the DDCIntensity. The combination of the censoring bias and the estimated treatment effects suggests that the impact of recent workload is likely to be larger for the high-intensity group. This is likely because multiple shifts of high workload would lead to more stress and fatigue, which in turn increases the intensity gap. Consequently, it becomes even more crucial to balance the nursing workload over time.

E.3.2. Heterogeneity in nurses' experience We consider two proxies for the nurses' experience: one is based on the level in the nurse's title; the other is based on the number of shifts the nurse worked during the study period.

Our data record the title of each nurse as registered nurse level 1, 2, or 3, where a higher level indicates a higher qualification and seniority. Among the 72 nurses in our data, 53, 5, and 14 nurses belong to levels 1, 2, and 3, respectively. We combine levels 2 and 3 as one group and consider two nurse groups. There are 8958 (76.5%) observations belonging to the level 1 group and 2778 (23.5%) observations belonging to the levels 2 & 3 group.

For the second grouping, we divide the nurses into two groups based on the number of shifts they worked during the study period. Nurses who have worked 120 shifts or more – equivalent to an average of three shifts per week – are considered more experienced, under the assumption that greater exposure to the MICU environment enhances familiarity and expertise. Among the 72 nurses, 43 worked less than 120 shifts, accounting for 5489 (46.7%) observations (nurse-patient-shifts). The remaining 29 nurses worked more than 120 shifts, accounting for 6247 (53.3%) observations.

We estimate the impact of recent workload on the gap between the PJIntensity and DDCIntensity for each group separately. The estimation results are summarized in Table 11. We observe that the treatment effects are positive and statistically significant at the 5% significance level for all groups except the levels 2 & 3 nurse group. For levels 2 & 3 nurse group, even though the estimated effect size, 0.228, is similar in magnitude to our main model estimation, the standard error is high, making the estimated effect statistically insignificant. This may be due to the small sample size for this group. We also observe the estimated treatment effects for the more experienced nurse group tend to be lower than those of the less experienced nurse group (although none of the differences are statistically significant). This suggests that more experience may enable nurses to better handle the workload and thus, be less impacted by high recent workload.

E.3.3. Heterogeneity in nurses' rest time To examine the heterogeneity in nurses' rest time, we divide the dataset into two subsets, depending on whether the focal nurse's rest time is larger than 6 (6 shifts, i.e., 3 days). There are 2035 (17.3%) observations for which the rest time is larger than 6. For this group, the estimated treatment effect is 0.124 (standard error: 0.061). There are 9701 (82.7%) observations for which the rest time is less than or equal to 6. For this group, the estimated treatment effect is 0.336 (standard error: 0.103). For both groups, the estimated treatment effects are positive and statistically significant at the

Table 11 Heterogeneity of Treatment Effect in Different Nurse Groups

	Nurse Group			
	Level 1 Nurse	Level 2 & 3 Nurse	Less than 120 shifts	More than 120 shifts
<i>Rec_Work</i>	0.276*** (0.089)	0.228 (0.272)	0.442* (0.206)	0.209*** (0.067)
Observations	8958	2778	5489	6247

standard error in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5% significance level. This demonstrates that the impact of recent workload on nurses' perceived workload exists in both groups. However, the effect size is larger for the group with a shorter rest time, which indicates that a longer rest time may help alleviate the impact of recent workload.