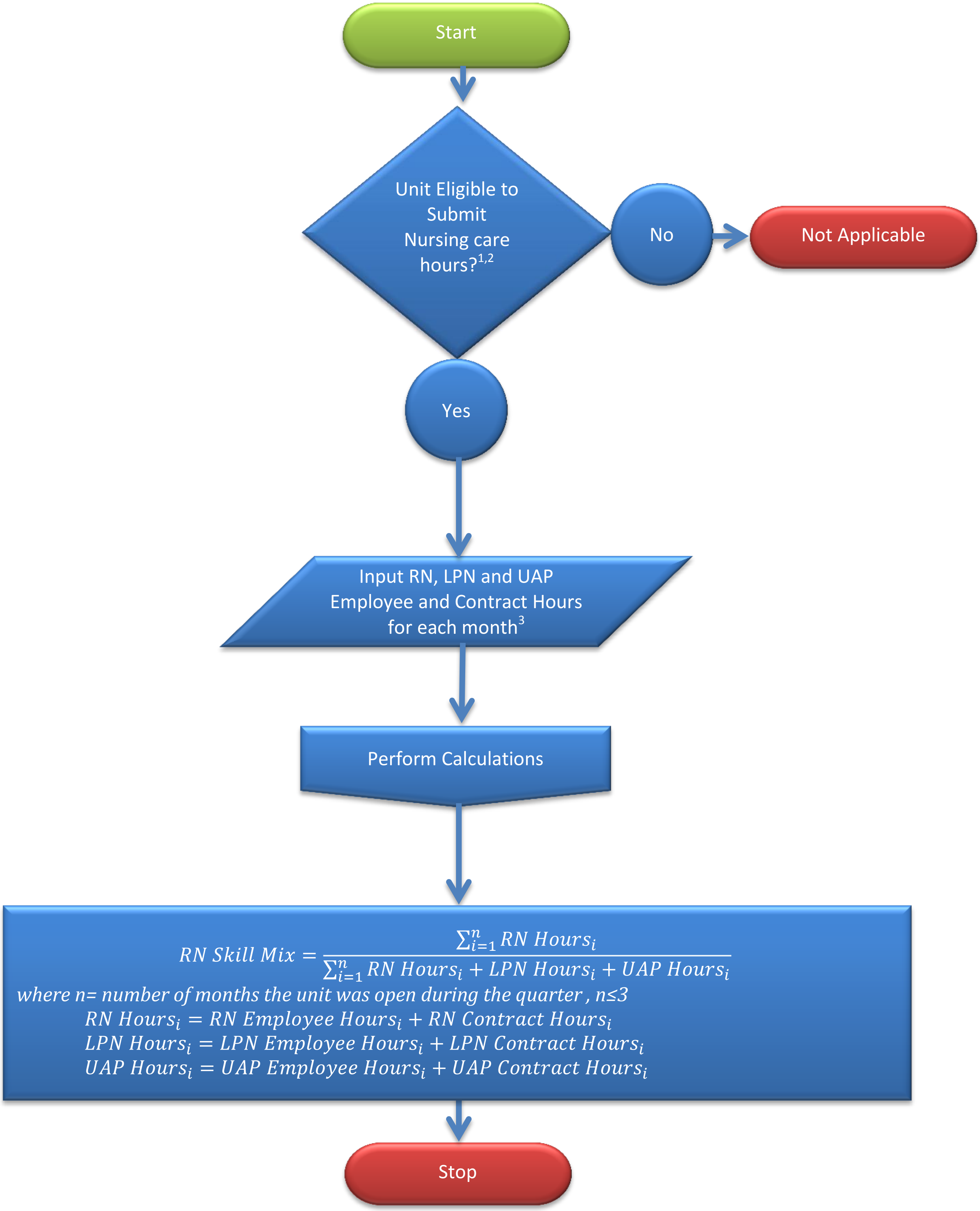
RN Skill Mix Algorithm /Measure Logic Flowchart



[1]Unit eligibility depends on NDNQI unit type designations. Eligible unit type designations are:

- Level III/IV Critical Care- Neonatal
- Level II Intermediate Care- Neonatal
- Level I Continuing Care- Neonatal
- Well Baby Nursery- Neonatal
- Critical Care- Pediatric
- Step Down- Pediatric
- Medical- Pediatric
- Surgical- Pediatric
- Med-Surg- Combined -pediatric
- Critical Care- Adult
- Step Down- Adult
- Medical- Adult
- Surgical- Adult
- Med-Surg Combined- Adult
- Critical Access - Adult
- Adult - Psychiatric
- Child/Adolescent- Psychiatric
- Geripsych-Psychiatric
- Other (Behavioral Health/Specialty/Multiple Unit Types) – Psychiatric
- Adult- Rehabilitation

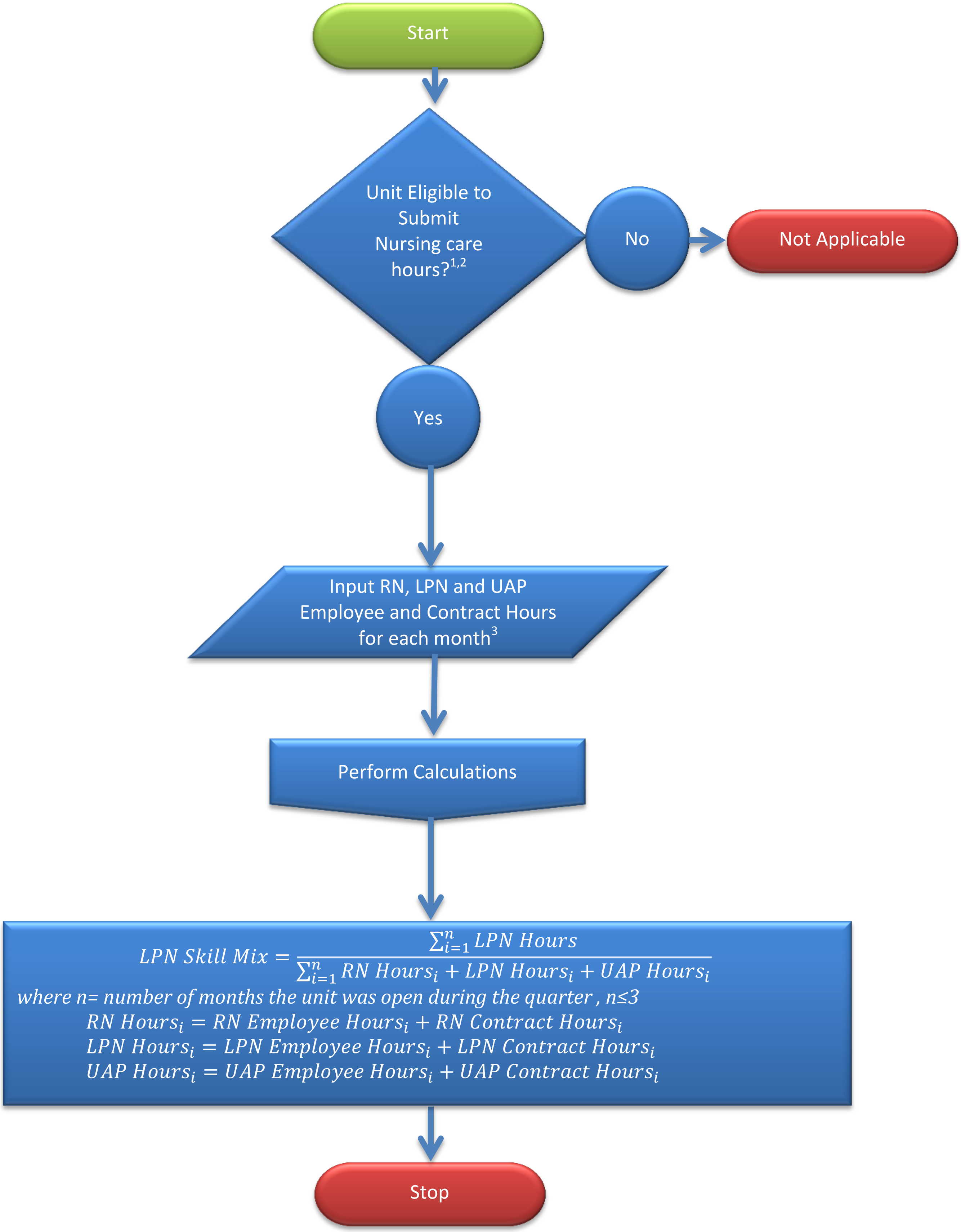
[2]Unit must have been open (patients present) at least 1 month during reporting period.

[3]The number of ***productive*** hours worked by nursing staff assigned to the unit who have ***direct patient care responsibilities*** for greater than 50% of their shift.

Include:
• Staff who are counted in the unit's staffing matrix, ***and***
• Who are replaced if they call in sick, ***and***
• Work hours are charged to the unit's cost center

Exclude:
• Persons whose primary responsibility is administrative in nature
• Specialty teams, patient educators, or case managers who are not assigned to a specific unit

LPN Skill Mix Algorithm /Measure Logic Flowchart



[1]Unit eligibility depends on NDNQI unit type designations. Eligible unit type designations are:

- Level III/IV Critical Care- Neonatal
- Level II Intermediate Care- Neonatal
- Level I Continuing Care- Neonatal
- Well Baby Nursery- Neonatal
- Critical Care- Pediatric
- Step Down- Pediatric
- Medical- Pediatric
- Surgical- Pediatric
- Med-Surg- Combined -pediatric
- Critical Care- Adult
- Step Down- Adult
- Medical- Adult
- Surgical- Adult
- Med-Surg Combined- Adult
- Critical Access - Adult
- Adult - Psychiatric
- Child/Adolescent- Psychiatric
- Geripsych-Psychiatric
- Other (Behavioral Health/Specialty/Multiple Unit Types) – Psychiatric
- Adult- Rehabilitation

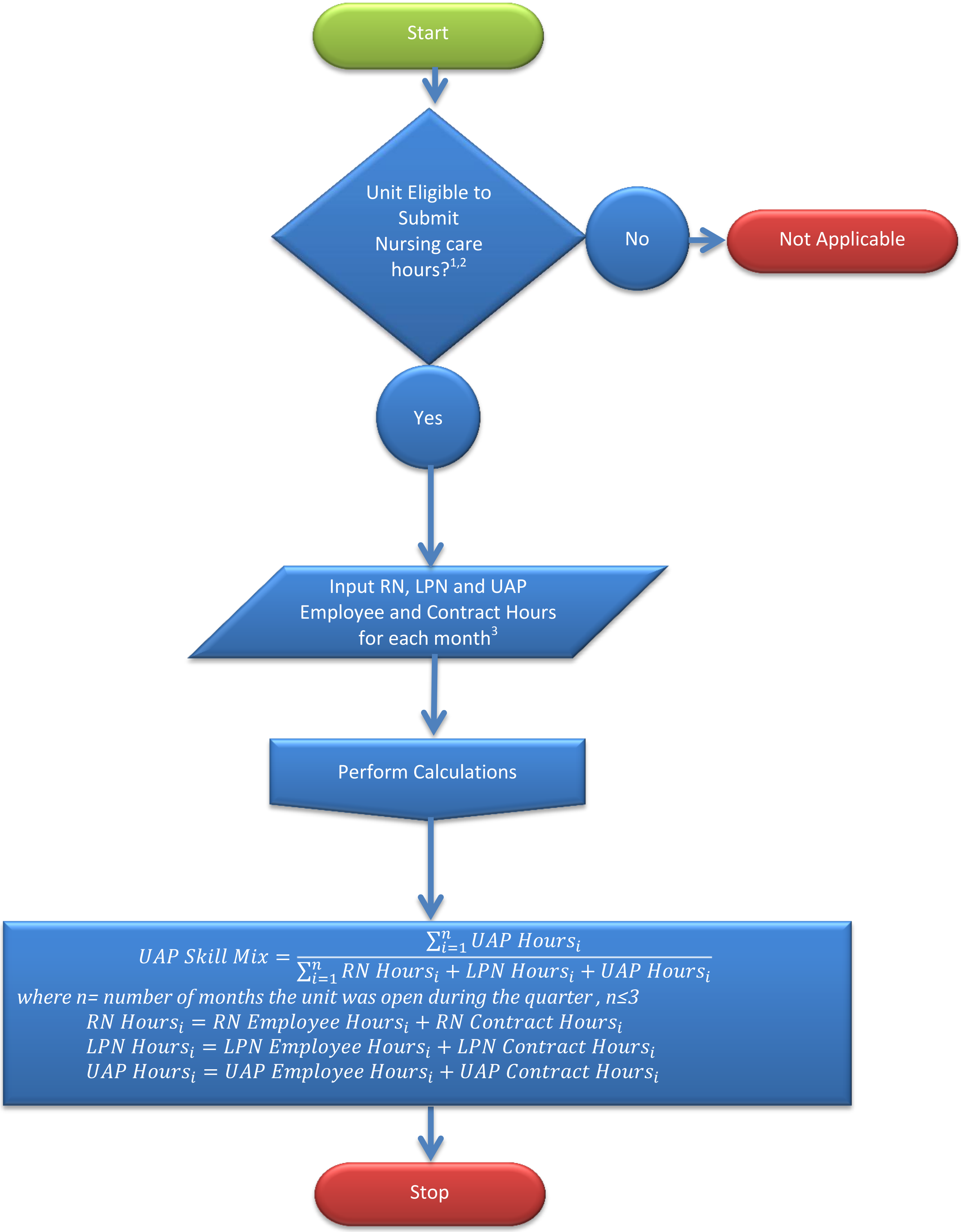
[2]Unit must have been open (patients present) at least 1 month during reporting period.

[3]The number of ***productive*** hours worked by nursing staff assigned to the unit who have ***direct patient care responsibilities*** for greater than 50% of their shift.

Include:
• Staff who are counted in the unit's staffing matrix, ***and***
• Who are replaced if they call in sick, ***and***
• Work hours are charged to the unit's cost center

Exclude:
• Persons whose primary responsibility is administrative in nature
• Specialty teams, patient educators, or case managers who are not assigned to a specific unit

UAP Skill Mix Algorithm /Measure Logic Flowchart



[1]Unit eligibility depends on NDNQI unit type designations. Eligible unit type designations are:

- Level III/IV Critical Care- Neonatal
- Level II Intermediate Care- Neonatal
- Level I Continuing Care- Neonatal
- Well Baby Nursery- Neonatal
- Critical Care- Pediatric
- Step Down- Pediatric
- Medical- Pediatric
- Surgical- Pediatric
- Med-Surg- Combined -pediatric
- Critical Care- Adult
- Step Down- Adult
- Medical- Adult
- Surgical- Adult
- Med-Surg Combined- Adult
- Critical Access - Adult
- Adult - Psychiatric
- Child/Adolescent- Psychiatric
- Geripsych-Psychiatric
- Other (Behavioral Health/Specialty/Multiple Unit Types) – Psychiatric
- Adult- Rehabilitation

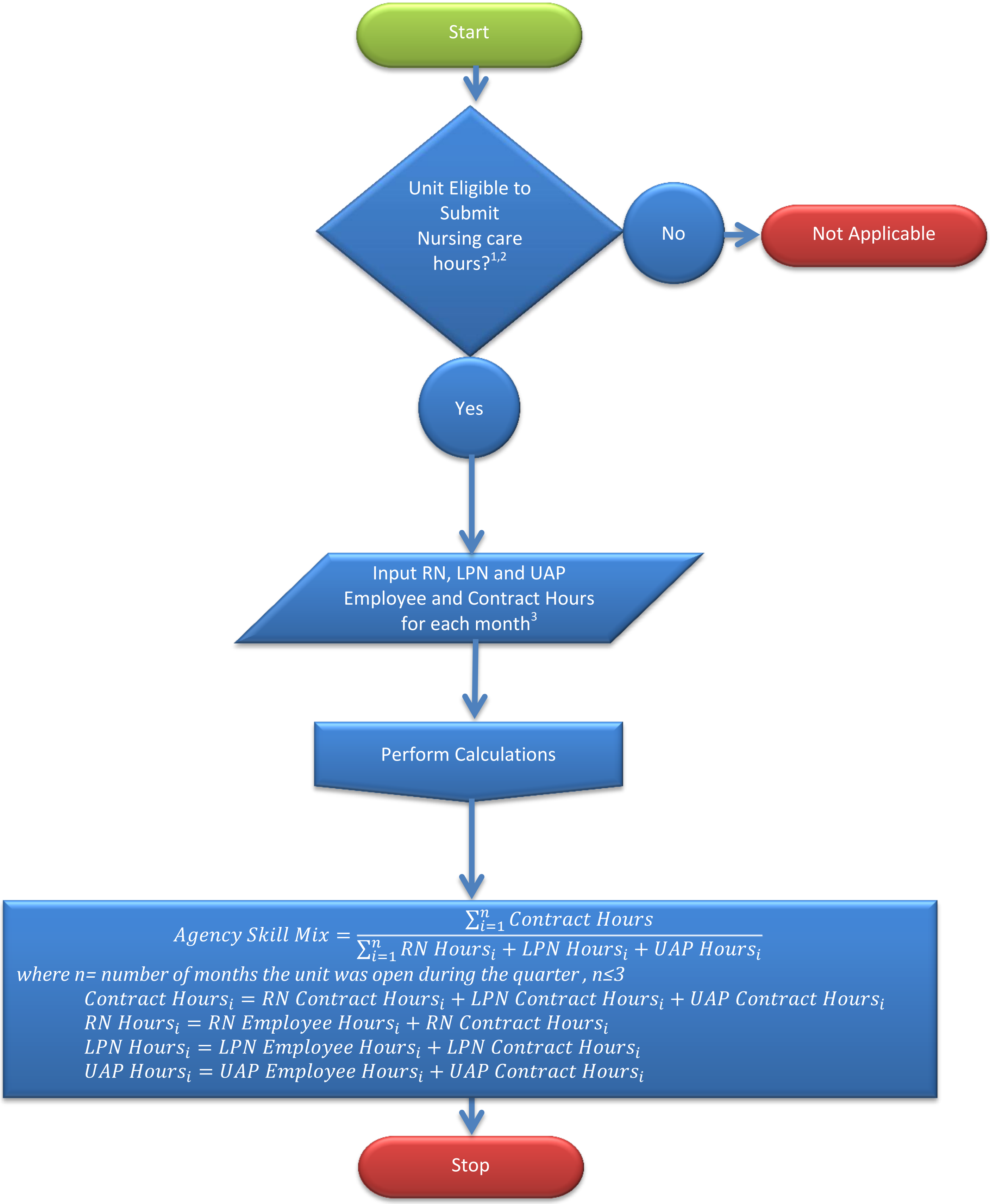
[2]Unit must have been open (patients present) at least 1 month during reporting period.

[3]The number of ***productive*** hours worked by nursing staff assigned to the unit who have ***direct patient care responsibilities*** for greater than 50% of their shift.

Include:
- Staff who are counted in the unit's staffing matrix, ***and***
- Who are replaced if they call in sick, ***and***
- Work hours are charged to the unit's cost center

Exclude:
- Persons whose primary responsibility is administrative in nature
- Specialty teams, patient educators, or case managers who are not assigned to a specific unit

Agency Skill Mix Algorithm /Measure Logic Flowchart



[1]Unit eligibility depends on NDNQI unit type designations. Eligible unit type designations are:

- Level III/IV Critical Care- Neonatal
- Level II Intermediate Care- Neonatal
- Level I Continuing Care- Neonatal
- Well Baby Nursery- Neonatal
- Critical Care- Pediatric
- Step Down- Pediatric
- Medical- Pediatric
- Surgical- Pediatric
- Med-Surg- Combined -pediatric
- Critical Care- Adult
- Step Down- Adult
- Medical- Adult
- Surgical- Adult
- Med-Surg Combined- Adult
- Critical Access - Adult
- Adult - Psychiatric
- Child/Adolescent- Psychiatric
- Geripsych-Psychiatric
- Other (Behavioral Health/Specialty/Multiple Unit Types) – Psychiatric
- Adult- Rehabilitation

[2]Unit must have been open (patients present) at least 1 month during reporting period.

[3]The number of ***productive*** hours worked by nursing staff assigned to the unit who have ***direct patient care responsibilities*** for greater than 50% of their shift.

Include:
• Staff who are counted in the unit's staffing matrix, ***and***
• Who are replaced if they call in sick, ***and***
• Work hours are charged to the unit's cost center

Exclude:
• Persons whose primary responsibility is administrative in nature
• Specialty teams, patient educators, or case managers who are not assigned to a specific unit

Measure #0204: Nursing Skill Mix

Scientific Supplement

4/10/2015

Table of Contents

Nursing Care Hours Reliability Study

Prepared for Press Ganey Associates, Inc.

By

Chenjuan Ma, PhD

Emily Cramer, PhD

Danielle Olds, PhD, RN

Karen Wambach, PhD, RN, IBCLC

Introduction

The National Database of Nursing Quality Indicators® (NDNQI®) is a proprietary database of Press Ganey Associates, Inc. NDNQI® was established in 1998 to support the American Nurses Association's Safety and Quality Initiative (Montalvo, 2007). The goals of NDNQI® are to promote and facilitate the standardization of information submitted by hospitals across the United States on nursing quality and patient outcomes. The NDNQI® database is one of the nation's largest repositories of data on nurse-sensitive quality and safety measures. One of the defining features of the NDNQI® database is that unit-level data are collected and analyzed to provide hospitals with unit-level benchmarks. Because of the important role that the NDNQI® database plays in improving unit-level quality and safety, it is important that the data are reliable and valid. NDNQI® evaluates the reliability of its measures on a regular basis to ensure continued confidence in the effectiveness of the measures. These measures include nursing care hours, patient days, and patient falls, which are submitted to NDNQI® quarterly by member hospitals. The reliability of the Nursing Care Hours (NCH) measure was evaluated in 2007, 2011, and most recently in 2014. The specific aims of the 2014 reliability evaluation were:

Aim 1: To establish the reliability (consistency) of the NDNQI® NCH measure by estimating reliability coefficients between unit types and within unit types over a 2-year period when adjusting for patient days

Aim 2: To determine the convergent validity of the NDNQI® NCH measure by correlating the NCH measure to nurse staffing information reported by registered nurses (RN) on the NDNQI® RN Survey.

To accomplish each of these aims, two separate analyses were performed and will be presented in this report.

Background

Nurse staffing measures have taken on greater importance as more research has demonstrated relationships between staffing and clinical outcomes. Nurse staffing has been associated with patient mortality (Aiken et al., 2011, 2014; Aiken, Clarke, Sloane, Sochalski, & Silber, 2002; Cho et al., 2014; Kane, Shamliyan, Mueller, Duval, & Wilt, 2007; Needleman et al., 2011; Rafferty et al., 2007; Twigg, Duffield, Bremner, Rapley, & Finn, 2011; Wiltse Nicely, Sloane, & Aiken, 2013), readmission (Ma, McHugh, & Aiken, 2015; Tubbs-Cooley, Cimiotti, Silber, Sloane, & Aiken, 2013), length of stay (Needleman, Buerhaus, Mattke, Stewart, & Zelevinsky, 2002; Tschannen & Kalisch, 2009; Twigg et al., 2011), hospital-acquired infections (Cimiotti, Haas, Saiman, & Larson, 2006; Hugonnet, Chevrolet, & Pittet, 2007; Hugonnet, Uçkay, & Pittet, 2007; Needleman et al., 2002; Rogowski et al., 2013), and falls (Lake, Shang, Klaus, & Dunton, 2010).

A variety of staffing measures are used in health services research including nurse to patient ratios (NPR), nursing hours per patient day (NHPPD), and skill mix (Butler, 2008; Butler et al., 2011; Choi & Staggs, 2014; Lake & Cheung, 2006; Unruh, Russo, Jiang, & Stocks, 2009). NHPPD and skill mix are included as NDNQI® measures and were developed as structural measures within Donabedian's structure-process-outcome framework (Donabedian, 1988). The National Quality Forum also endorses these two measures and both incorporate nursing care hours. The numerator for NHPPD is the total number of productive hours worked by nursing staff with direct patient care responsibilities for each hospital in-patient unit during the calendar month. The denominator is the total number of patient days for each in-patient unit during the calendar month (“National Quality Forum,” 2014a). For RN skill mix, the numerator is the productive nursing care hours worked by RNs with direct patient care responsibilities for each hospital in-patient unit during the calendar month. The denominator is the total number of productive hours worked by employee or contract nursing staff with direct patient care responsibilities (RNs, LPNs/LVNs, and UAPs) for each hospital in-patient unit during the calendar month (“National Quality Forum,” 2014b). Because of the key role that nursing care hours play in these measures, it is important to establish reliability for research and quality benchmarking.

Reliability Studies of Nursing Care Hours

Much like nurse staffing measures, there is a wide variety of ways nursing care hours are defined in the literature. These variations include the criteria of nursing staff type or role, included hours, and hospital or unit settings. In the NDNQI® measure, nursing care hours are defined as, “the number of productive hours worked by nursing staff assigned to the unit who have direct patient care responsibilities for greater than 50% of their shift.” The measure includes “staff who are counted in the unit’s staffing matrix and who are replaced if they call in sick and whose work hours are charged to the unit’s cost center.” The measure excludes APRNs in the role of an institutionally credentialed provider, persons whose primary role is administrative in nature, and specialty teams, patient educators, or case managers who are not assigned to a particular unit.” Productive hours are defined as the, “actual direct patient care hours worked including overtime, not budgeted or scheduled hours.” Nursing care hours are used in the following NDNQI® staffing measures: skill mix, total nursing hours per patient day (inpatient units), nursing hours per patient visit (ED and pre/postoperative units), RN hours per patient day (inpatient units), and percent of total nursing hours supplied by contract or agency staff (Press Ganey Associates, 2014).

The reliability of the NDNQI® nursing care hours measure has been evaluated in two previous studies, one in 2007 and one in 2011 (Choi, Boyle, & Dunton, 2014; Klaus, Dunton, Gajewski, & Potter, 2013). The 2007 study was conducted in two phases (Klaus et al., 2013). In the first phase, the researchers estimated reliability of the indicator by evaluating the compliance of member hospitals with NDNQI® data collection guidelines. The researchers found that 70% of participating hospitals correctly identified the NDNQI® definition of nursing care hours, including the inclusion and exclusion criteria (Klaus, Dunton, Forbis, & Gajewski, 2008). However, it was also determined that site coordinators should review the guidelines periodically and clarification should be provided about the inclusion of unit secretaries, monitor technicians, sitters, and nurses in orientation. In Phase 2, researchers estimated the reliability of the NDNQI® nursing care hours indicator by comparing it to nursing care hours obtained in electronic data sources from participating hospitals in this reliability study (Klaus, Dunton, Forbis, Gajewski, et al., 2008). The intraclass correlations coefficients (ICCs) for each hospital's nursing care hours ranged from .84 to .99. The high ICCs confirm the reliability of the NDNQI® nursing care hours measures (Klaus et al., 2013).

A reevaluation of the reliability of the nursing care hours indicator was conducted by NDNQI® researchers in 2011 using two studies (Choi, Boyle, & Dunton, 2012). The aim of the first study was to compare nursing care hours across two datasets, the NDNQI® and the California Office for Statewide Health Planning and Development (OSHPD). The aim of the second study was to identify issues and compliance with the nursing care hours data collection process. For the first study, the ICCs between the NDNQI®-reported nursing care hours and the OSHPD-reported nursing care hours were 0.96 for RN nursing hours and 0.72 for non-RN nursing hour. All of these are above the 0.60 threshold for considering a measure to be reliable (Choi et al., 2014). The results of the second study indicate that there is high compliance with the NDNQI data collection procedures (Choi et al., 2012).

Although the 2007 and 2011 studies indicate that the NDNQI® nursing care hours measure is reliable, we sought to re-assess the reliability (consistency) and convergent validity using two approaches: 1) using longitudinal data to compare the within and between unit nursing care hours variances using intraclass correlation coefficients (ICCs) and 2) correlating nursing care hours data with data from NDNQI® RN survey, in which nurses reported the number of patients cared for on the previous shift. While the NDNQI® nursing care hours measure has been reliable in previous studies, there remains an opportunity to corroborate these findings with different methodologies.

Methods

Data Source and Sample

This is a secondary analysis of longitudinal data to determine the reliability of the NDNQI® Nursing Care Hour Indicator. Two NDNQI® data sets were used, including the quarterly reported NDNQI® Nursing Care Hour data and the annual Registered Nurse (RN) Survey.

The nursing care hours data were collected from NDNQI® member hospitals at the unit level. These data were reported quarterly but collected on a monthly-basis. NDNQI® defines nursing care hours as the number of productive hours worked by nursing staff assigned to the unit who have direct patient care responsibilities for greater than 50% of their shift. Accordingly, nurse care hours from persons whose primary responsibility is administrative or who are not assigned to a specific unit (e.g., specialty teams, patient educators, and case managers) were not counted. The NDNQI® nursing care hours data are collected separately for each nursing licensure category (RN, LPN/LVN, and UAP) as well as employment status (hospital employees vs. agency/contract staff). **Table 1** summarizes the definition and/or calculation of each variable related to nursing care hours. All measures were calculated on a monthly-basis for each unit.

One major concern in the present study design is the fluctuation of monthly nursing care hours resulting from the changes in patient days in each month. Patient days reflect the care demand for nurse staffing. This would bias our reliability estimation of the nursing care hours because this fluctuation was not a result of inaccuracy data collection. To eliminate this limit, we adjusted for patient days for each staffing measure (e.g., total nursing care hours and RN care hours). The patient days data also were collected from units in member hospitals on a monthly-basis and reported to NDNQI® quarterly.

In the present study, we included units from NDNQI® member hospitals that submitted nursing care hours data on each month from January 2012 to December 2013. We limited units to the following types: critical care, step-down, medical, surgical, medical-surgical combined, and rehabilitation. These types of units are common in most general hospitals. The application of these inclusion/exclusion criteria resulted in 7,961 units from 1,186 hospitals.

The NDNQI® RN survey was used in approaching research aim 2, comparing the RN patient care hours to patient day in the NDNQI® nursing care hours data with the staffing levels reported by RNs in each unit from the RN survey. In the RN survey, nurses were asked to report the number of patients assigned to them on their last shift as well the maximum number of patient assigned to them. Individual RN responses in a unit were averaged to reflect the unit RN staffing levels. This measurement has been used in many studies and is considered a reliable measure in predicting patient outcomes (Ma, McHugh, & Aiken, 2015; McHugh & Ma, 2013).

Data Analysis

Characteristics of the participating hospitals and units were assessed using descriptive statistics. Trends in nursing care hours and skill mix (with and without adjusting for patient days) by nursing licensure categories were examined over the study period using descriptive statistics.

The reliability of the nursing care hours and skill mix measures (adjusted for patient days) was evaluated by assessing the consistency of reporting over time, using monthly data from 2012-2013. Critical care, step-down, medical, surgical, medical-surgical combined and rehab units were included in the analysis. Only units reporting in all 24 months were included in the analysis.

Random intercept linear mixed models, with repeated measures nested within units were estimated for each staffing measure by unit type. The intraclass correlation coefficient (ICC) was calculated as:

$$\sigma_{ij}^2 / \sigma_{ij}^2 + \sigma_i^2,$$

where, σ_(ij)^2 is the between level variance, and σ_(i)^2, is the within-level variance. The ICC is the ratio of between level variance to the total model variance. A high ICC indicates that the amount of variance between units is greater than the variance within a given unit over time, and provides support for the reliability of the nursing care hours measure. Generally, an ICC above 0.8 indicates high reliability, and ICC above 0.6 indicates an acceptable level of reliability.

Convergent validity of the nursing care hours measure was examined through correlations between the quarterly reported NDNQI® nursing care hours measure and RN reports of staffing levels from the annual NDNQI® RN Survey. The quarterly reported nursing care hours measures were annualized by averaging the monthly nursing care hours, and matched to the annual survey data for each unit. Approximately half of all NDNQI® hospitals participate in the RN survey, and the total number of units that submitted both staffing and RN survey data was 6,663 for the six unit types included in this study. Two items from the NDNQI® RN Survey ask nurses to provide information about staffing levels on their unit. Specifically these items ask how many total patients were assigned to a nurse on his or her last shift, and the maximum number of patients assigned to a nurse at any one time on his or her last shift. In previous studies, these measures have been shown to have very high correlations with NDNQI®'s quarterly reported NCH measures (Choi & Staggs, 2014).

Table 1 Measures of nursing care hours in the NDNQI® NCH data

Variable	Definition/calculation
Total nursing care hours	The total number of productive nursing care hours provided by eligible nursing staffs (i.e., RNs, LPNs/LVNs, and UAPs)
RN nursing care hours	The total number of productive nursing care hours provided by RNs
LPN/LVN nursing care hours	The total number of productive nursing care hours provided by LPNs/LVNs
UAP nursing care hours	The total number of productive nursing care hours provided by UAPs
Total nursing care hours per patient day	Total nursing care hours divided by total patient days
RN nursing care hours per patient day	RN nursing care hours provided divided by total patient days
LPN/LVN nursing care hours per patient day	LPN/LVN nursing care hours provided divided by total patient days
UAP nursing care hours per patient	UAP nursing care hours provided divided by total patient days

day	
RN skill mix	The proportion of total nursing care hours supplied by RNs
LPN/LVN skill mix	The proportion of total nursing care hours supplied by LPNs/LVNs
UAP skill mix	The proportion of total nursing care hours supplied by UAPs

*All the measures were calculated on a monthly-basis for each unit.

RN, registered nurse; LPN, licensed practical nurse; UAP, unlicensed assistive personnel

Results

Sample Description

Table 2 presents the characteristics of study hospitals. The majority of hospitals were not for profit (85%), located in a metropolitan area (population of 50,000 or more people), and Magnet designated (73%). Approximately one in five (21%) were small hospitals with less than 100 beds, and 28% were large hospitals with at least 300 beds. Roughly half (48%) were teaching hospitals.

Table 2. Characteristics of hospitals submitting administrative NCH data

	N	%
Ownership		
Not for Profit	1,005	84.74
For Profit	90	7.59
Government	91	7.68
Bed size		
< 100	252	21.25
100-199	345	29.09
200-299	249	20.99
300-399	146	12.31
400-499	88	7.42
>=500	106	8.94
Teaching status		
Academic Medical Center	130	10.96

Teaching	437	36.85
Non-teaching	619	52.19
Location		
Neither Metro nor Micro area	21	1.77
Metropolitan area	1,064	89.71
Micropolitan area	101	8.52
Magnet		
Non Magnet	857	72.62
Magnet	329	27.74

Table 3 summarizes units by unit types for both the reliability (consistency) analyses and convergent validity analysis. Of the 7,961 units with administrative nursing care hour data for reliability (consistency) analyses, critical care units and medical-surgical combined units made up 23% of the sample units; approximately one in five (19%) were medical units; one in seven were step-down (15%) or surgical (14%) units; and the rest (5%) were rehabilitation units. Of the 6,663 units with nurse staffing data from both administrative and RN survey data in the convergent validity analyses, they had similar distribution in term of unit types.

Table 3. Unit types

	Reliability Study Administrative Data Only (n=7,961)		Validity Study Administrative and Survey Data (n=6,663)	
	N	%	N	%

ICU	1,841	23.1	1,484	22.2
Step-down	1,194	15.0	1,005	15.1
Medical	1,544	19.4	1,306	19.6
Surgical	1,122	14.1	1,018	15.3
Med-surg	1,837	23.1	1,510	22.7
Rehab	423	5.3	340	5.1

Table 4 presents the trends in nursing care hours and skill mix by nursing licensure categories over the study period. The monthly total nursing care hours ranged from 5,672 to 6,343 hours. The monthly nursing care hours provided by RNs ranged from 4,200 to 4,674 and this proportion is about 73% to 74% of the total nursing care hours. The monthly nursing care hours provided by LPNs/LVPs ranged from 105 to 153 hours and this proportion is about 2% to 3% of the total nursing care hours. The monthly nursing care hours provided by UAPs ranged from 1,333 to 1,533 hours and this proportion is about 23% to 24% of the total nursing care hours.

Table 5 shows the trends in patient-day adjusted nursing care hours (or nursing care hours per patient day) over the study period. The total nursing care hours per patient day (THNPPD) ranged from 10.77 to 11.37 hours; the RN nursing care hours per patient day (RNHPPD) ranged from 8.24 to 8.74 hours; the LPN nursing care hours per patient day ranged from 0.17 to 0.23 hours; and the UAP nursing care hours per patient day ranged from 2.29 to 2.51 hours.

Table 5. Trends in nursing care hours per patient day (all units)

	THPPD		RNHPPD		LPNHPPD		UAPHPPD	
Time point	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Jan 2012	10.95	4.13	8.40	4.38	0.23	0.65	2.32	1.20
Feb 2012	10.97	4.13	8.45	4.39	0.23	0.64	2.29	1.19
Mar 2012	10.99	4.10	8.44	4.36	0.23	0.65	2.32	1.19
Apr 2012	11.13	4.16	8.55	4.44	0.22	0.61	2.36	1.20
May 2012	11.27	4.25	8.62	4.50	0.22	0.62	2.43	1.24
Jun 2012	11.29	4.32	8.60	4.54	0.22	0.64	2.46	1.28
Jul 2012	11.28	4.26	8.59	4.50	0.22	0.63	2.48	1.27
Aug 2012	11.24	4.24	8.57	4.47	0.22	0.65	2.45	1.26
Sep 2012	11.21	4.18	8.60	4.44	0.21	0.62	2.40	1.24
Oct 2012	11.19	4.26	8.59	4.49	0.21	0.65	2.40	1.22
Nov 2012	11.25	4.12	8.62	4.38	0.21	0.66	2.42	1.23
Dec 2012	11.08	4.06	8.47	4.30	0.21	0.71	2.40	1.22
Jan 2013	10.77	3.94	8.24	4.18	0.20	0.64	2.33	1.17
Feb 2013	11.04	4.10	8.49	4.33	0.20	0.63	2.35	1.20
Mar 2013	11.08	4.04	8.50	4.29	0.20	0.62	2.38	1.20
Apr 2013	11.15	4.10	8.54	4.35	0.19	0.61	2.42	1.20
May 2013	11.31	4.21	8.64	4.44	0.19	0.62	2.48	1.24
Jun 2013	11.28	4.22	8.60	4.44	0.19	0.61	2.49	1.24
Jul 2013	11.33	4.19	8.64	4.44	0.19	0.61	2.51	1.24
Aug 2013	11.34	4.21	8.66	4.45	0.19	0.62	2.49	1.24
Sep 2013	11.27	4.16	8.67	4.42	0.18	0.61	2.42	1.21
Oct 2013	11.24	4.14	8.64	4.39	0.18	0.60	2.43	1.23
Nov 2013	11.37	4.13	8.74	4.39	0.17	0.60	2.46	1.24
Dec 2013	11.12	4.06	8.53	4.29	0.17	0.64	2.42	1.22

Reliability Findings

The results of the mixed model analyses are presented in **Table 6**. The ICCs for patient-day-adjusted nursing hours measures ranged from 0.73 to 0.94 across unit types. The lowest ICCs were RN hours in critical care (ICC = 0.73) and rehabilitation units (ICC = 0.74) and the highest ICCs were for LPN/LVN hours. The ICCs for total nursing hours reached 0.7 or higher in all unit types.

Table 6. Intraclass Correlations (ICC) of Patient-Day-Adjusted Nursing Care Hours by Unit Types

	ICC					
	Critical Care	Step-down	Medical	Surgical	Med-surg	Rehab
RN hours	0.73	0.79	0.81	0.78	0.77	0.74
LPN/LVN hours	0.94	0.90	0.89	0.89	0.94	0.94
UAP hours	0.80	0.77	0.79	0.80	0.79	0.79
Total Hours	0.72	0.73	0.73	0.73	0.69	0.68
Skill Mix (% RN hours)	0.83	0.83	0.85	0.86	0.87	0.82

Note: all the nursing care hours measure were adjusted by patient days, except RN skill mix

Validity Findings

Tables 7-13 presents the results of the convergent validity study. Correlations between RNHPPD and RN reported maximum number of patients ranged from -0.46 to -0.70 across unit types, and was -0.86 for all unit types combined. Correlations between RN nursing care hours per patient day and RN reported total number of patients ranged from 0.40 to -0.74 across unit types and was -0.85 for all unit types combined.

Table 7. Correlations among Clinical Staffing (Administrative) Measure and RN Report of Staffing

	1	2	3	4	5	6	7
1. RN Hours	1.00						
2. LPN/LVN Hours	-0.17	1.00					
3. UAP Hours	-0.19	-0.13	1.00				
4. Total Hours	0.96	-0.07	-0.06	1.00			
5. Skill Mix	0.74	-0.33	-0.69	0.55	1.00		
6. RN Report Max Patients	-0.86	0.25	0.30	-0.81	-0.77	1.00	
7. RN Report Total Patients	-0.85	0.24	0.29	-0.80	-0.75	0.98	1.00

Table 8. Correlations among Clinical Staffing (Administrative) Measure and RN Report of Staffing (Survey) in Critical Care Units

Variables	1	2	3	4	5	6	7
1. RN Hours	1.00						
2. LPN/LVN Hours	-0.04	1.00					
3. UAP Hours	-0.01	-0.07	1.00				
4. Total Hours	0.91	0.09	0.38	1.00			
5. Skill Mix	0.21	-0.22	-0.92	-0.20	1.00		
6. RN Report Max Patients	-0.46	0.03	0.01	-0.43	-0.12	1.00	
7. RN Report Total Patients	-0.40	0.02	0.00	-0.37	-0.10	0.92	1.00

Table 9. Correlations among Clinical Staffing (Administrative) Measure and RN Report of Staffing (Survey) in Step Down Units

Variables	1	2	3	4	5	6	7
1. RN Hours	1.00						
2. LPN/LVN Hours	-0.08	1.00					
3. UAP Hours	-0.04	-0.17	1.00				
4. Total Hours	0.96	0.02	0.20	1.00			
5. Skill Mix	0.26	-0.31	-0.74	0.01	1.00		
6. RN Report Max Patients	-0.74	0.24	-0.05	-0.61	-0.39	1.00	

7. RN Report Total Patients	-0.69	0.22	-0.04	-0.56	-0.10	.95	0 1.00

Table 10. Correlations among Clinical Staffing (Administrative) Measure and RN Report of Staffing (Survey) in Medical Units

Variables	1	2	3	4	5	6	7
1. RN Hours	1.00						
2. LPN/LVN Hours	-0.15	1.00					
3. UAP Hours	-0.05	-0.20	1.00				
4. Total Hours	0.79	0.06	0.50	1.00			
5. Skill Mix	0.62	-0.29	-0.70	0.03	1.00		
6. RN Report Max Patients	-0.70	0.23	0.06	-0.52	-0.52	1.00	
7. RN Report Total Patients	-0.66	0.22	0.05	-0.49	-0.49	0.96	1.00

Table 11. Correlations among Clinical Staffing (Administrative) Measure and RN Report of Staffing (Survey) in Surgical Units

Variables	1	2	3	4	5	6	7
1. RN Hours	1.00						
2. LPN/LVN Hours	-0.11	1.00					
3. UAP Hours	0.02	-0.19	1.00				
4. Total Hours	0.81	0.08	0.53	1.00			
5. Skill Mix	0.54	-0.28	-0.71	-0.03	1.00		
6. RN Report Max Patients	-0.62	0.26	0.08	-0.43	-0.49	1.00	
7. RN Report Total Patients	-0.59	0.22	0.10	-0.41	-0.47	0.95	1.00

Table 12. Correlations among Clinical Staffing (Administrative) Measure and RN Report of Staffing (Survey) in Medical-Surgical Combined Units

Variables	1	2	3	4	5	6	7
1. RN Hours	1.00						

2. LPN/LVN Hours	-0.21	1.00					
3. UAP Hours	-0.01	-0.30	1.00				
4. Total Hours	0.77	0.14	0.44	1.00			
5. Skill Mix	0.63	-0.44	-0.56	0.01	1.00		
6. RN Report Max Patients	-0.68	0.19	0.08	-0.49	-0.51	1.00	
7. RN Report Total Patients	-0.64	0.17	0.08	-0.46	-0.47	0.95	1.00

Table 13. Correlations among Clinical Staffing (Administrative) Measure and RN Report of Staffing (Survey) in Rehabilitation Units

Variables	1	2	3	4	5	6	7
1. RN Hours	1.00						
2. LPN/LVN Hours	-0.04	1.00					
3. UAP Hours	-0.09	-0.32	1.00				
4. Total Hours	0.71	0.13	0.53	1.00			
5. Skill Mix	0.64	-0.22	-0.65	-0.05	1.00		
6. RN Report Max Patients	-0.64	-0.06	0.04	-0.48	-0.38	1.00	
7. RN Report Total Patients	-0.62	-0.06	0.03	-0.48	-0.36	0.98	1.00

Conclusions

The NDNQI® nursing care hours indicator was reevaluated for its reliability in two approaches, aim 1) to assess the reliability (consistency) of the nursing care hours measure by ICCs when adjusting for patient days and aim 2) to compare the nursing care hours measure to nurse-reported staffing measures from the NDNQI® RN Survey.

In aim 1, our estimates from the multilevel mixed model yielded ICCs for patient-day-adjusted nursing care hours measures ranged from 0.73 to 0.94 across unit types. Furthermore, we also identified the ICCs for each nursing care hours measure by unit types and found that all the ICCs were at least 0.69 or higher. Recommendations for acceptable levels of reliability differ across sources. Center for Medicare and Medicaid Services (CMS) has suggested value above 0.4 indicate an acceptable level of reliability (CMS, 2012), while others suggest thresholds from 0.7 to 0.9 in more conservative estimates (Adams et al, 2010). There are some slight variations in the reliability when nursing care hours measures were stratified by nurse licensure status (RNs, LPNs/LVNs, and UAPs) or grouped by unit types (e.g., critical care units and medical units). The ICC ranged from 0.97 for LPNs/LVNs on critical care, medical-surgical combined, and rehabilitation units to 0.73 for RN hours on critical care units. Therefore our findings indicate that the NDNQI nursing care hours indicator has acceptable to high reliability, and is similar to findings in the previous studies. (Choi, Boyle, & Dunton, 2014; Klaus, Dunton, Gajewski, & Potter, 2013). Unlike previous reliability studies, however, this study assessed longitudinal consistency of the nursing care hours measures across time.

In aim 2, the correlation coefficients between the RN care hours measure (adjusted for patient days) and RN reported nurse staffing measures were -0.86 for RN reported maximum number of patients on last shift, and -0.85 for RN reported total number of patients on last shift, indicating strong convergent validity. These findings were very similar to Choi and Staggs' study (r= -0.86 for total number of patients on last shift) (Choi & Staggs, 2014). There were some variations by unit types. We found that when stratified by unit types, the correlation coefficients between RN care hours measure and RN reported maximum number of patients on last shift ranged from -0.46 (critical care units) to -0.74 (step-down units); and the correlation coefficients between RN care hours measure and RN reported total number of patients on last shift ranged from -0.40 (critical care units) to -0.69 (step-down units). These findings indicate moderate to strong correlations between the RN care hours measure and RN-reported nurse staffing measures.

There are limitations in this study that should be noted. First, we only included five types of units in this study. Therefore we cannot generalize to other unit types that were not assessed. Second, the nurse staffing measures from the RN survey were reported only by RNs; therefore we were only able to compare the RN care hours to RN-reported nurse staffing measures from the RN survey. Despite these limitations, the design of this reliability study offers many advantages over previous study designs. The benefit of examining the intraclass correlations of existing longitudinal data is that the reliability of the nursing care hours measure can be evaluated across a wider variety of unit types beyond the critical care units used in previous work (Choi et al., 2014). In the reliability study by Choi and colleagues (2014), unit types in the NDNQI® database had to be matched to those in the OSHPD database. Due to definitional constraints, only data on critical care units were used. The design of this reliability study addressed that limitation.

The other advantage of the design of the current study is that it overcomes potential response bias that was a limitation of the 2007 reliability study (Klaus et al., 2013). In that reliability study, 1,069 hospitals were invited to participate in the quasi-interrater reliability portion of the study, however only 11 hospitals submitted data that could be useful for the analyses. The authors point out that those participating could be "high-achievers" and fundamentally different than those who did not participate (Klaus et al., 2013). The current reliability study has less self-selection bias because all of the hospitals in the database meeting inclusion criteria were included. Another potential limitation of the 2007 reliability study relates to the survey portion of the study in which site coordinators from 714 hospitals completed a multiple-choice survey about the processes used to collect nursing care hours data and compliance with standardized data collection protocols. Because the survey was self-report, there is the potential for social desirability bias such that some participants could have exaggerated or misrepresented their compliance with data collection protocols. The current reliability study includes analyses comparing RN reported staffing with site coordinator reported staffing. Because the two data sources are separate, there is a lower likelihood of bias.

The reliability of the NDNQI® nursing care hours measure is important for clinical users, researchers, and policy makers. Clinical users, including administrators and unit managers in NDNQI® member hospitals, use the staffing measures based on nursing care hours (skill mix, total NHPPD, RN hours per patient day, and the percent of total nursing hours supplied by contract/agency staff) as benchmarks. The unit-level staffing data that hospitals receive in NDNQI® reports are often used to make operational decisions. Establishing the reliability of the nursing care hours measure is also critical for research. In 2005 and 2006 a Delphi study was conducted to assess variables used in the research linking nurse staffing and patient outcomes. Of three variables that measure the number of nurse staff in relation to patient volume (NHPPD, nurse-to-patient ratio, and full-time equivalents employed by an organization or unit), NHPPD received the highest consensus for use in health services research (Van den Heede, Clarke, Sermeus, Vleugels, & Aiken, 2007). Because nursing care hours serve as the basis for NHPPD, it is critical that the measure be reliable to ensure research quality.

Finally, in an era of greater accountability and transparency in healthcare, NHPPD and skill mix, which incorporate nursing care hours, are under consideration for inclusion in the Centers for Medicare and Medicaid Services' Hospital Inpatient Quality Reporting Program ("National Quality Forum: Measure Applications Partnership," 2014). The program provides a financial incentive to hospitals for reporting quality measures and provides consumers with a website, Hospital Compare, with the data so that they can make decisions about their healthcare. Because of this, the reliability of nursing care hours is of the utmost importance to hospitals, the public, and policy makers.

In summary, this is the third study to find the NDNQI® nursing care hours measure to be sufficiently reliable. Because three studies using different methodologies have all found the measure to be reliable, we believe that it can serve as the basis for NHPPD and skill mix indicators in clinical operations, research, and policy.

References

Adams, J. L., Mehrotra, A., Thomas, J. W., & McGlynn, E. A. (2010). Physician cost profiling – reliability and risk of misclassification. *New England Journal of Medicine*, *362*(11), 1014-1021.

Aiken, L. H., Cimiotti, J. P., Sloane, D. M., Smith, H. L., Flynn, L., & Neff, D. F. (2011). Effects of nurse staffing and nurse education on patient deaths in hospitals with different nurse work environments. *Medical Care*, *49*(12), 1047–1053. doi:10.1097/MLR.0b013e3182330b6e

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., & Silber, J. H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA: The Journal of the American Medical Association*, *288*(16), 1987–1993.

Aiken, L. H., Sloane, D. M., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., … RN4CAST consortium. (2014). Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. *Lancet*, *383*(9931), 1824–1830. doi:10.1016/S0140-6736(13)62631-8

Butler, M. (2008). Hospital nurse staffing models and patient and staff-related outcomes. *Cochrane Database of Systematic Reviews*, (1).

Butler, M., Collins, R., Drennan, J., Halligan, P., O'Mathúna, D. P., Schultz, T. J., … Vilis, E. (2011). Hospital nurse staffing models and patient and staff-related outcomes. *The Cochrane Database of Systematic Reviews*, (7), CD007019. doi:10.1002/14651858.CD007019.pub2

Cho, E., Sloane, D. M., Kim, E.-Y., Kim, S., Choi, M., Yoo, I. Y., … Aiken, L. H. (2014). Effects of nurse staffing, work environments, and education on patient mortality: An observational study. *International Journal of Nursing Studies*. doi:10.1016/j.ijnurstu.2014.08.006

Choi, J., Boyle, D., & Dunton, N. (2012, March). Nursing Care Hours: Reliability Study. National Database of Nursing Quality Indicators.

Choi, J., Boyle, D. K., & Dunton, N. (2014). A Standardized measure: NDNQI nursing care hours indicator. *Western Journal of Nursing Research*, *36*(1), 105–116. doi:10.1177/0193945913501723

Choi, J., & Staggs, V. S. (2014). Comparability of nurse staffing measures in examining the relationship between RN staffing and unit-acquired pressure ulcers: a unit-level descriptive, correlational study. *International Journal of Nursing Studies*, *51*(10), 1344–1352. doi:10.1016/j.ijnurstu.2014.02.011

Cimiotti, J. P., Haas, J., Saiman, L., & Larson, E. L. (2006). Impact of staffing on bloodstream infections in the neonatal intensive care unit. *Archives of Pediatrics & Adolescent Medicine*, *160*(8), 832–836. doi:10.1001/archpedi.160.8.832

Centers for Medicare & Medicaid Services. (2012). Memorandum: Results of reliability analysis from Mathematica Policy Research. Baltimore, MD: U.S. Department of Health & Human Services. Available at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf

Donabedian A. (1988). The quality of care: How can it be assessed? *JAMA*, *260*(12), 1743–1748. doi:10.1001/jama.1988.03410120089033

Hugonnet, S., Chevrolet, J.-C., & Pittet, D. (2007). The effect of workload on infection risk in critically ill patients. *Critical Care Medicine*, *35*(1), 76–81. doi:10.1097/01.CCM.0000251125.08629.3F

Hugonnet, S., Uçkay, I., & Pittet, D. (2007). Staffing level: a determinant of late-onset ventilator-associated pneumonia. *Critical Care (London, England)*, *11*(4), R80. doi:10.1186/cc5974

Kane, R. L., Shamliyan, T. A., Mueller, C., Duval, S., & Wilt, T. J. (2007). The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. *Medical Care*, *45*(12), 1195–1204. doi:10.1097/MLR.0b013e3181468ca3

Klaus, S., Dunton, N., Forbis, D., & Gajewski, B. (2008, January 23). *Nursing Care Hours Reliability Study Phase I Report.* National Database of Nursing Quality Indicators.

Klaus, S., Dunton, N., Forbis, D., Gajewski, B., Potter, C., & Leiker, K. (2008, August 13). *Nursing Care Hours Reliability Study Phase II Report*. National Database of Nursing Quality Indicators.

Klaus, S., Dunton, N., Gajewski, B., & Potter, C. (2013). Reliability of the nursing care hour measure: a descriptive study. *International Journal of Nursing Studies, 50*(7), 924–932. doi:10.1016/j.ijnurstu.2012.07.012

Lake, E. T., & Cheung, R. B. (2006). Are patient falls and pressure ulcers sensitive to nurse staffing? *Western Journal of Nursing Research*, *28*(6), 654–677. doi:10.1177/0193945906290323

Lake, E. T., Shang, J., Klaus, S., & Dunton, N. E. (2010). Patient falls: Association with hospital Magnet status and nursing unit staffing. *Research in Nursing & Health*, *33*(5), 413–425. doi:10.1002/nur.20399

Ma, C., McHugh, M. D., & Aiken, L. H. (2015). Organization of Hospital Nursing and 30-Day Readmissions in Medicare Patients Undergoing Surgery. *Medical Care*, *53*(1), 65–70. doi:10.1097/MLR.0000000000000258

McHugh, M. D., & Ma, C. (2013). Hospital nursing and 30-day readmissions among medicare patients with heart failure, acute myocardial infarction, and pneumonia. *Med Care*, *51*(1), 52-59. doi: 10.1097/MLR.0b013e3182763284

Montalvo, I. (2007). The National Database of Nursing Quality Indicators (NDNQI). *Online Journal of Issues in Nursing*, *12*(3), 13p.

National Quality Forum. (2014a, January 6). *Nursing Hours Per Patient Day*. Retrieved December 29, 2014, from http://www.qualityforum.org/QPS/0205.

National Quality Forum. (2014b, January 6). *Skill Mix (Registered Nurse [RN], Licensed Vocational/Practical Nurse [LVN/LPN], Unlicensed Assistive Personnel [UAP], and Contract).* Retrieved December 29, 2014, from http://www.qualityforum.org/Qps/204.

National Quality Forum: Measure Applications Partnership. (2014, January). Retrieved January 6, 2015, from http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=78497

Needleman, J., Buerhaus, P., Mattke, S., Stewart, M., & Zelevinsky, K. (2002). Nurse-staffing levels and the quality of care in hospitals. *The New England Journal of Medicine*, *346*(22), 1715–1722. doi:10.1056/NEJMsa012247

Needleman, J., Buerhaus, P., Pankratz, V. S., Leibson, C. L., Stevens, S. R., & Harris, M. (2011). Nurse staffing and inpatient hospital mortality. *The New England Journal of Medicine*, *364*(11), 1037–1045. doi:10.1056/NEJMsa1001025

Press Ganey Associates. (2014, October). Guidelines for Data Collection and Submission on Nursing Care Hours (NCH) Indicator. Press Ganey Associates.

Rafferty, A. M., Clarke, S. P., Coles, J., Ball, J., James, P., McKee, M., & Aiken, L. H. (2007). Outcomes of variation in hospital nurse staffing in English hospitals: cross-sectional analysis of survey data and discharge records. *International Journal of Nursing Studies*, *44*(2), 175–182. doi:10.1016/j.ijnurstu.2006.08.003

Rogowski, J. A., Staiger, D., Patrick, T., Horbar, J., Kenny, M., & Lake, E. T. (2013). Nurse staffing and NICU infection rates. *JAMA Pediatrics*, *167*(5), 444–450. doi:10.1001/jamapediatrics.2013.18

Tschannen, D., & Kalisch, B. J. (2009). The effect of variations in nurse staffing on patient length of stay in the acute care setting. *Western Journal of Nursing Research*, *31*(2), 153–170. doi:10.1177/0193945908321701

Tubbs-Cooley, H. L., Cimiotti, J. P., Silber, J. H., Sloane, D. M., & Aiken, L. H. (2013). An observational study of nurse staffing ratios and hospital readmission among children admitted for common conditions. *BMJ Quality & Safety*, *22*(9), 735–742. doi:10.1136/bmjqs-2012-001610

Twigg, D., Duffield, C., Bremner, A., Rapley, P., & Finn, J. (2011). The impact of the nursing hours per patient day (NHPPD) staffing method on patient outcomes: a retrospective analysis of patient and staffing data. *International Journal of Nursing Studies*, *48*(5), 540–548. doi:10.1016/j.ijnurstu.2010.07.013

Unruh, L., Russo, C. A., Jiang, H. J., & Stocks, C. (2009). Can state databases be used to develop a national, standardized hospital nurse staffing database? *Western Journal of Nursing Research*, *31*(1), 66–88. doi:10.1177/0193945908319992

Wiltse Nicely, K. L., Sloane, D. M., & Aiken, L. H. (2013). Lower mortality for abdominal aortic aneurysm repair in high-volume hospitals is contingent upon nurse staffing. *Health Services Research*, *48*(3), 972–991. doi:10.1111/1475-6773.12004

Evidence of the Association between **Nursing Staff Skill Mix (% of Hours Supplied by RNs)** and Patient Outcomes

Patient Outcome	Author (year)	Result
Falls	Tzeng et al (2012)	(-)
	He et al (2012)	(-)
	Dunton et al (2007)	(-)
	Dunton et al (2004)	(-)
	Cho et al (2003)	(NS)
	Blegen & Vaughn (1998)	(-)
Pressure Ulcers	Choi et al (2014)	(-)
	Yang et al (2012)[a]	(NS)
	Blegen et al (2011)	(NS)
	Cho et al (2003)	(NS)
	Dunton et al (2007)	(-)
Mortality	Yang et al (2012)[a]	(NS)
	Blegen et al (2011)	(NS)
	Estabrooks et al (2005)	(-)
Length of Stay	Esparza (2012)	(-)
	Yang et al (2012)[a]	(NS)
	Blegen et al (2011)	(NS)
	Needleman et al (2002)	(-) Medical patients only
Medication Administration Errors	Stratton (2008)	(NS)
Urinary Tract Infections	Esparza (2012)	(-)
	Yang et al (2012)[a]	(-)
	Cho et al (2003)	(NS)
	Needleman et al (2002)	(-)
Pneumonia	Cho et al (2003)	(NS)
	Needleman et al (2002)	(-) Medical patients only
Bloodstream Infections	Yang et al (2012)[a]	(+)
	Stratton (2008)	(-)
Ventilator Weaning	Yang et al (2012)[a]	(+)
Complaints	Stratton (2008)	(NS)

a: RN/Aide composition vs. 100% RN care

Note. (-) statistically inverse relationship between RN skill mix and patient outcomes (higher proportion of nursing hours provided by RNs is related to lower rates of the patient outcomes); (+) statistically positive relationship between RN skill mix and patient outcomes (higher proportion of nursing hours provided by RNs is related to higher rates of the patient outcomes);

NS = results were not significant; RN = registered nurses.

References

Blegen, M. A., & Vaughn, T. (1998). A multisite study of nurse staffing and patient occurrences. *Nursing Economic$*, *16*(4), 196–203.

Blegen, M., Goode, C., Spetz, J., Vaughn, T., & Park, S. (2011). Nurse staffing effects on patient outcomes: safety-net and non-safety-net hospitals. *Medical Care*, *49*(4), 406–414. http://doi.org/10.1097/MLR.0b013e318202e129

Choi, J., & Staggs, V. S. (2014). Comparability of nurse staffing measures in examining the relationship between RN staffing and unit-acquired pressure ulcers: a unit-level descriptive, correlational study. *International Journal of Nursing Studies*, *51*(10), 1344–1352. http://doi.org/10.1016/j.ijnurstu.2014.02.011

Cho, S.-H., Ketefian, S., Barkauskas, V. H., & Smith, D. G. (2003). The effects of nurse staffing on adverse events, morbidity, mortality, and medical costs. *Nursing Research*, *52*(2), 71–79.

Dunton, N., Gajewski, B., Klaus, S., & Pierson, B. (2007). The relationship of nursing workforce characteristics to patient outcomes. *OJIN: The Online Journal of Issues in Nursing*, *12*(3), Manuscript 3. http://doi.org/10.3912/OJIN.Vol12No03Man03

Dunton, N., Gajewski, B., Taunton, R. L., & Moore, J. (2004). Nurse staffing and patient falls on acute care hospital units. *Nursing Outlook*, *52*(1), 53–59. http://doi.org/10.1016/j.outlook.2003.11.006

Esparza, S. J., Zoller, J. S., White, A. W., & Highfield, M. E. F. (2012). Nurse staffing and skill mix patterns: Are there differences in outcomes? *Journal of Healthcare Risk Management: The Journal of the American Society for Healthcare Risk Management*, *31*(3), 14–23. http://doi.org/10.1002/jhrm.20092

Estabrooks, C. A., Midodzi, W. K., Cummings, G. G., Ricker, K. L., & Giovannetti, P. (2005). The impact of hospital nursing characteristics on 30-day mortality. *Nursing Research*, *54*(2), 74–84.

He, J., Dunton, N., & Staggs, V. (2012). Unit-level time trends in inpatient fall rates of US hospitals. *Medical Care*, *50*(9), 801–807. http://doi.org/10.1097/MLR.0b013e31825a8b88

Needleman, J., Buerhaus, P., Mattke, S., Stewart, M., & Zelevinsky, K. (2002). Nurse-staffing levels and the quality of care in hospitals. *The New England Journal of Medicine*, *346*(22), 1715–1722. http://doi.org/10.1056/NEJMsa012247

Stratton, K. M. (2008). Pediatric nurse staffing and quality of care in the hospital setting. *Journal of Nursing Care Quality*, *23*(2), 105–114. http://doi.org/10.1097/01.NCQ.0000313758.33654.49

Tzeng, H.-M., Titler, M. G., Ronis, D. L., & Yin, C.-Y. (2012). The contribution of staff call light response time to fall and injurious fall rates: an exploratory study in four US hospitals using archived hospital data. *BMC Health Services Research*, *12*, 84. http://doi.org/10.1186/1472-6963-12-84

Yang, P.-H., Hung, C.-H., Chen, Y.-M., Hu, C.-Y., & Shieh, S.-L. (2012). The impact of different nursing skill mix models on patient outcomes in a respiratory care center. *Worldviews* *on Evidence-Based Nursing / Sigma Theta Tau International, Honor Society of Nursing*, *9*(4), 227–233. http://doi.org/10.1111/j.1741-6787.2012.00246.x

NDNQI®
NATIONAL DATABASE OF NURSING QUALITY INDICATORS

National Database of Nursing Quality Indicators®

Methods Development Project

Final Report to

The American Nurses Association

August 3, 2011

Prepared by the National Database of Nursing Quality Indicators

School of Nursing

University of Kansas Medical Center

Executive Summary

NDNQI staff have developed methodologies for bringing mixed acuity units into the database for reporting on clinical and staffing indicators and for creating hospital-level indicators. The process used in the development of these methodologies involved literature review, advice from methodological experts, new data collection, and statistical simulation. Key findings included:

- There were many barriers to creating risk or acuity-adjusted unit-level measures.
    - Acuity data are not available for units.
    - Risk data for specific outcomes would require a large, and perhaps burdensome, increase in data collection and reporting for participating hospitals.
- A method for creating categories of mixed acuity units was developed based on Medicare billing days. The method has face and criterion validity.
- Six types of mixed acuity units were identified for adult or pediatric populations:
    - Mixed Acuity III: Units with at least 50% critical care patient days each month.
    - Mixed Acuity II: Units with at least 25% critical care days each month or at least 50% step-down days each month; includes only units not meeting the criterion for Mixed Acuity III.
    - Mixed Acuity I: Units not meeting the criteria for Mixed Acuity III or II.
    - Burn units
    - Bone Marrow Transplant units
    - Critical Access units
- Rolling up unit-level indicators should take into account the unit composition of a hospital, as well as the size of each unit. Results should be meaningful to users.
- Hospital level indicators can be developed from a weighted average or unit-level z-score. The weights are based on the staffed bed size of each unit. For hospital reporting, the z-scores are translated back into the original metric of the indicator.
- Data collection and reporting for mixed acuity units and reporting for hospital-level indicators may be implemented at the direction of the American Nurses Association.

Background and Objectives

The National Database of Nursing Quality Indicators® (NDNQI®) is a proprietary database of the American Nurses Association (ANA). NDNQI collects and evaluates nursing-sensitive data from 1,800 hospitals in the United States

Hospitals that participate in NDNQI receive quarterly unit-level comparative data reports. NDNQI comparative data are currently stratified by unit type and hospital characteristics. Thus, each unit can compare its nurse staffing and patient outcomes to national percentiles for similar units. Hospitals value the NDNQI reports and want to have all of their units included. Mixed acuity units are currently excluded from NDNQI as they are too diverse to be included in comparative data. NDNQI has identified a methodologically and conceptually sound method of acuity adjustment or acuity stratification to allow inclusion of mixed acuity units in NDNQI reports. To date, most risk or acuity adjustment work has been done at the hospital level, using patient characteristics, diagnostic related groups, or the hospital case mix index.

NDNQI also was asked to create a methodology to produce hospital-level measures. Hospital executives want summary measures of performance for their entire facility. Consumers and oversight organizations want information to make conclusions about a facility's nursing quality. Typically, hospital-level reporting is either based on patient-level data or weighted averages of unit-level data. The patient-level approach is divergent from NDNQI's unit-based data collection. Incorporating the unit orientation into the measurement model is consistent with NDNQI's focus on nursing unit performance. A methodology was needed that was based on unit level data and takes into account the fact that hospitals may not submit data on all eligible units and hospitals vary in unit composition.

The methods development project had two main goals:

1. Develop a method for unit-level acuity adjustment of nursing indicators
2. Develop a method to calculate hospital-level indicators

Development Process

The project originally had two goals:

1. Develop adjustment methods for unit data on nurse staffing and patient outcomes.
2. Develop a method for rolling-up unit-level data to hospital totals.

The requirements for adjustments and roll-up methodologies that are consistent with the NDNQI paradigm include:

1. The nursing care unit should continue to be the unit of observation.
2. Post-adjustment results should be interpretable by clinicians
3. Data used in adjustment must be comparable across sites.
4. If adjustment is to occur with every reporting period, data must be available for update on a quarterly basis and within 45 days of the end of a calendar quarter.
5. Within a group of effective methods, the method requiring the least additional respondent burden should be selected.
6. Adjustment and summary methods should have attributional and face validity.

The development process consisted of four phases:

1. Literature review
2. Advisory panel input
3. Data collection
4. Statistical methods development

The report concludes with a description of a work order for implementation of the adjustment and roll-up methodologies.

Literature Review

The project began with a literature review of risk and acuity adjustment as they might apply to the inclusion of mixed acuity units into NDNQI reporting. It also covered the development of hospital-level measures that maintain NDNQI's unit-level focus. The review identified the following information:

Adjusting Staffing Measures

- Risk adjustment is a statistical method that accounts for patient characteristics known to the correlated with a particular outcome measure. In general, outcome measures are adjusted for risk.
- Acuity adjustment refers to the use of a composite variable reflecting the level of patients' need for care. In general, staffing measures are adjusted for acuity.
- Adjustment variables that are themselves measures of quality should be excluded in order to avoid over-adjustment.
- There was disagreement in the literature on which specific variables to use for adjustment.
    - A hospital's case mix index is the average of its patients' diagnosis-related group (DRG) weights. DRG weights reflect the relative resource consumption. Each patient is assigned one DRG for their entire inpatient stay. The level of measurement for DRGs is hospital, not nursing care unit. Further, the extent to which DRG weights capture *nursing* resource consumption is questionable. DRG weights do not explicitly account for variation in need for daily nursing.
    - Nursing intensity weights (NIW) were developed as a nursing-specific refinement of the DRG payment system. Nursing intensity weights reflect the relative level of nursing care needed by a typical patient in each diagnosis-related group. NIW, being based on DRGs, are at the hospital-level, rather than the nursing care unit-level.
    - Numerous proprietary software programs classify patients' severity of illness or project a unit's nursing workload. There is no predominant patient classification system in use across the U.S., so comparable data are not available.

Adjusting Patient Outcomes

- Outcome-specific adjustments are needed for pressure ulcers, falls, and other patient outcomes. Published examples of adjusted outcomes typically rely on patient-level measures of risk specific to each outcome, as opposed to more general illness severity or acuity measures.
- Adjusting outcomes based on patient characteristics would place a large data collection burden on hospitals that do not yet have electronic health records. Further, the data are protected health information which hospitals may be reluctant to release. Finally, the evidence for the use of PHI data producing effective adjustments was limited.

Hospital-Level Reporting

- Hospital level reporting typically is done by counting all of the various conditions or events that occur in the hospital and then dividing by the number of patients, days, device days, etc. across the hospital.
- This approach is patient-focused and abandons the unit-performance perspective that is the primary unit of analysis for NDNQI.
- Minnick (2000) noted that nursing-sensitive indicators typically vary more across units in the same hospital than across hospitals. Hospital summary measures of pressure ulcers, patient falls, and other indicators that mask variation between units within the organization are not as meaningful as unit-type comparisons across hospitals.
- In 2009, the National Quality Forum's report on Voluntary Consensus Standards for Nursing-Sensitive Care: An Initial Performance Measure Set

Expert Panel

ANA established an expert advisory panel to provide guidance on the data and methodological issues of adjustment and roll-up. Panelists brought forth a variety of issues, but no consensus was established on methodological approach. The following issues were identified.

- NDNQI's unit-based approach may not provide stable measures, much less enough data for unit-level risk adjustment.
- Risk adjusters must have high levels of sensitivity and specificity. More research would need to be done to identify which, if any, adjusters have these measurement properties.

- Nursing intensity weights are based on AP-DRGs, which reflect the care provided for an entire hospital stay, not the care provided on a specific unit. Current NIW have 2-3 levels within each DRG to distinguish critical vs. acute care. DRGs reflect patients at their sickest during the hospital stay.
- Use of patient classification systems may eventually be mandated nationally, but they vary across sites and are generally not well validated.
- Admissions-Transfer and Discharge data is another important predictor of workload that varies widely across units.
- Data for risk adjustment should be based on admission risk assessments. DRGs are not admission-based.
- Mixed acuity units use billing levels that correspond to existing NDNQI unit type strata. Validity of unit type as a proxy for acuity is uncertain.
- Rolling up different sized units in a hospital creates a weighted average, which complicates statistical issues. Some units will have a higher put-through than others so their average occupancy rate will be lower but the total number of people seen might be high.
- The statistical issues for dealing with weighted averages are pretty straight forward but the average needs to be interpretable.
- Any adjustment method will make mistakes. How high you set the standard for accuracy is subjective. Public reporting requires a higher standard than internal use for quality improvement.

Revised Study Purpose

After the completion of the discovery portion of the project, NDNQI researchers concluded that the original scope of work should be modified to reflect the information from the literature review and expert panel.

Due to questions about feasibility and advisability of various approaches to adjustment, we limited our objective to the primary goal of our adjustment work, namely to have a way to make meaningful comparisons of mixed acuity units so that these units can begin submitting data on NDNDQI clinical and staffing indicators.

The method we propose for hospital roll-up is a straightforward weighting technique. The original scope of work stated that we would collect data from a variety of hospitals to test alternative approaches. Because there are a wide variety of unit profiles within hospitals and unpredictable patterns of non-reporting, the sample size required to test all combinations of circumstances would be very large. Yet, even a large sample might not encounter all possible situations. A more satisfactory, and efficient, approach was to conduct the development tests with simulated data. It is quite feasible to vary systematically the composition of unit profiles and patterns of missing data.

Results

Mixed Acuity Units

The purpose of this study was to develop acuity adjustment methods to allow meaningful comparisons of mixed acuity units' patient outcomes, and to compare the feasibility and effectiveness of these methods using data collected from a sample of NDNQI hospitals. Due to questions about the feasibility and advisability of various approaches to adjustment (discussed below), we limited our objectives to the primary goal of finding a way to make meaningful comparisons of mixed acuity units so that these units can begin submitting data on NDNDQI quarterly indicators. During the discovery activities for unit-based acuity adjustments a variety of issues were uncovered. We concluded the following:

- Adjustment of staffing measures is not warranted. Staffing measures themselves are a measure of acuity, so adjusting them for acuity is analogous to adjusting SAT scores for test-taker IQ.
- Device-related infection measures are based on at-risk patients and are thus already adjusted for acuity to a large degree.
- Gathering patient clinical admission data for use in risk or acuity adjustment would represent a sizeable data collection burden for hospitals, as well as a substantial expansion in the scope of NDNQI's data collection. Even if certain clinical variables proved useful for adjusting NDNQI outcome measures, any adjustment method based on these variables could be used only for hospitals willing and able to provide these data to NDNQI; other hospitals' mixed acuity units would continue to be excluded from participation in quarterly indicators. Moreover, a patient's risk and acuity often change over the course of the hospital stay due to surgery, healthcare-acquired infection, effectiveness of treatment, etc.
- It was argued in the first Advisory Council meeting that one-day prevalence measures, which are unstable because they are based on data for only one day out of every 90, should not undergo unit-level risk adjustment. We disagree that the instability of these measures is a reason not to adjust them for risk or acuity, but we advise against adjustment using patient risk or acuity variables for the reasons stated above.

The mixed acuity unit study is described below. We propose creating new acuity-based unit types using the proportions of patient billing days at the critical, step-down, and standard levels of care, as well as comparison groups based on unit sub-specialty for unit types for which this is practicable.

NDNQI member hospitals with at least one mixed acuity unit enrolled were invited to participate in a survey on a volunteer basis. No specific sample size was targeted, as the study does not involve statistical inference. Hospitals submitted data for March, 2011, on nursing care hours, patient days, and (in the case of adult units) patient falls. Data were also submitted for a single pressure ulcer prevalence survey conducted in 2011 (prior to the data submission deadline of May 6).

In addition, hospitals reported for each mixed acuity unit the number of patient days billed by level of care/acuity for March, 2011. NDNQI sent a customized Excel spreadsheet to each participating site coordinator for recording these data. Patient days were classified using the following billing levels:

1. Critical/intensive care (highest level of care).
2. Step-down/intermediate/progressive care. This may also be called transitional care.
3. Standard/routine care (e.g. medical or surgical care).
4. Rehabilitation care.
5. Skilled nursing/sub-acute care.
6. Inpatient hospice care. This level applies only to patients who have been discharged from acute care.
7. Short-stay observational care. This includes 23-hour observational care.

Adult Mixed Acuity Units. Patient days by billing level were reported by 52 adult mixed acuity units. Critical, step-down, and standard care days made up the bulk of the days for these units. Five units reported at least one hospice or rehabilitation day, but these days did not make up more than 1.1% of the total patient days for any unit. One unit reported skilled nursing days, which accounted for 27.8% of its total days, the remainder being standard care days. Most units billed one or more short stay days; the average percentage of short stay days was 2.3%.

All 52 units billed at least 45% of their patient days at a single level, and 48 billed over 50% of their days at one level. Three units met NDNQI's 90% criterion for an existing unit type for the month of the study. For all 52 units, one or two billing levels accounted for at least 75% of patient days. Some units combined primarily critical and step-down care days, some combined step-down and standard care days, and some combined critical and standard care days (in some cases with no step-down days; some hospitals do not have a billing level between critical and standard care).

Pediatric Mixed Acuity Units. Of the ten pediatric mixed acuity units that submitted data on patient billing days, nine billed at least 60% of their patient days at either the critical, step-down, or standard care level. No rehabilitation, skilled nursing, or hospice days were billed. Short stay days accounted for less than 8% of patient days for all units save one, for which they accounted for 36%.

Other Unit Types. Four critical access units submitted data for the study. Their proportions of days at the critical, step-down, standard, skilled nursing, and short stay levels varied widely. Data were also submitted for one bone marrow transplant unit, one neonatal mixed acuity unit, and one obstetrics unit.

Approaches to Acuity Adjustment

Adjusting Outcomes Using Patient Days by Billing Level. If NDNQI collected data on patient days at various billing levels from all participating units, these data could be used to adjust each unit's scores on the NDNQI measures using an appropriate regression model. Each measure could be regressed on the proportions of patient days billed at the critical, step-down, and standard levels to obtain a predicted value on the measure for each unit, and this could be compared to the unit's observed value on the measure to compute an acuity-adjusted score. This would require regular collection of data on billing days so that adjusted scores could be reported to hospitals each month/quarter.

There are two problems with this approach. First, not all hospitals use the same billing levels, or even the same number of billing levels, and reconciling the various sets of levels would be difficult. And second, it would be burdensome for hospitals would submit these data on a regular basis.

Using RNHPPD as a Proxy for Acuity. Another option is to use RN hours per patient day (RNHPPD)—a variable on which NQNDI already collects data—as a proxy for acuity to classify units and/or compute acuity-adjusted scores on NDNQI measures. In an internal study conducted in 2010 using four quarters of data from critical, step-down, medical, surgical, and medical/surgical units, RNHPPD was shown to contain much of the same information as unit type. As shown in Table 1, critical, step-down, and standard care units differ markedly in RNHPPD. Using a general linear model, unit type was found to account for 82% of the variation in RNHPPD. Moreover, in linear models in which unit type was included as a predictor of fall rate or rate of hospital-acquired pressure ulcers, the addition of RNHPPD as a predictor resulted in virtually no change in the proportion of variance accounted for by the models.

Table 1

RN Hours per Patient Day by Unit Type (Data from 2009-2010)

Unit Type	**Mean**	**SD**
Critical Care	15.3	*2.6*

Step Down	7.7	*1.9*
Med	5.8	*1.4*
Surg	6.0	*1.4*
Med-Surg	5.9	*1.5*

There is further evidence from the present study that RNHPPD is linked to patient acuity. As part of the study, units submitted data on three staffing variables—RNHPPD, total nursing care hours per patient day (TNHPPD), and skill mix (percentage of nursing care hours provided by RNs). The correlations between these variables and the proportions of patient days billed at the critical, step-down, standard, skilled nursing, hospice, and short stay levels are shown in Table 2. Both RNHPPD and TNHPPD were positively correlated with the proportions of days at the critical and hospice care levels and negatively correlated with the proportions at the standard, skilled nursing, and short stay levels. Correlations with skill mix were generally much weaker.

Table 2

Correlations of Staffing Variables with Proportions of Patient Days at Billing Levels (N = 59)

Billing Level	**Critical Care**	**Step-down**	**Standard**	**Skilled Nursing**	**Hospice**	**Short Stay**
Correlation with RNHPPD	0.73	0.11	–0.59	–0.42	0.38	–0.37

Correlation with TNHPPD	0.65	0.17	–0.56	–0.44	0.25	–0.40
Correlation with Skill Mix	0.44	–0.07	–0.25	–0.27	0.30	–0.09

Given the strong associations between RNHPPD and both unit type and patient acuity, RNHPPD could be used to adjust scores on NDNQI outcome measures in the same way as proportions of patient days at various billing levels (as described above). However, RNHPPD is associated not only with acuity but also with quality of care, and adjusting quality measures for RNHPPD (or other staffing variables) would involve more than adjustment for acuity. The unintended consequence of such an adjustment would be that units with higher RNHPPD would have their scores favorably adjusted regardless of the acuity of their patients or the quality of care provided, and whereas current NDNQI outcome measures reflect both quality of care and patient acuity, these adjusted scores would reflect some combination of quality of care, patient acuity, and staffing, making them difficult to interpret as indicators of quality.

Alternatively, RNHPPD could be used to define new unit types for mixed acuity units. For example, a mixed acuity unit with RNHPPD between the critical care and step-down RNHPPD means could be assigned to a critical/step-down mixed unit type. While this would be less problematic than using RNHPPD to adjust outcome measures, defining unit types based on staffing would tend to favor units with high RNHPPD relative to patient acuity by placing them in comparison groups with higher-acuity units that have RNHPPD in the same range.

Proposed Method for Mixed Acuity Units. We propose using the billing-days data collected in this study to create new acuity-based unit types for mixed acuity units, and to define these unit types using simple rules similar to the 90% classification rule we currently use.

A number of factors were considered in creating these new unit types. First, it is important to have enough types to ensure within-type homogeneity, but not so many types that some have too few units to serve as a meaningful comparison group. Second, the rules for classifying units must be easy to understand and based on numbers that site coordinators can accurately estimate if they do not have exact data. Third, the new unit types should reflect the proportion of billing days at the critical, step-down, and standard care levels, as well as the RNHPPD, reported by the units in this study. And fourth, the mean RNHPPD and ulcer rates for the units assigned to the new unit types should fit reasonably well in an acuity-ordered list of unit type means, including the means for the existing critical care, step-down, and medical/surgical types.

With these considerations in mind, three new unit types were created for adult units that do not meet the 90% criterion for an existing unit type. They are defined as follows:

1. Mixed Acuity III: Units with at least 50% critical care patient days each month.
2. Mixed Acuity II: Units with at least 25% critical care days each month or at least 50% step-down days each month; includes only units not meeting the criterion for Mixed Acuity III.
3. Mixed Acuity I: Units not meeting the criteria for Mixed Acuity III or II.

We propose to use the same classification scheme to create three new unit types for pediatric mixed acuity units.

The definitions were used to assign each of the adult units in this study to a new unit type. Descriptive statistics for RNHPPD and unit-acquired pressure ulcer (UAPU) rate were calculated by unit type both for the new unit types and, using data from the first quarter of 2011, for the existing critical care, step-down, and med/surg unit types. As shown in Table 3, RNHPPD and UAPU rates are generally ordered as one would expect, with higher numbers for higher-acuity unit types.

Table 3

Descriptive Statistics for RNHPPD and UAPU by Unit Type

Unit Type		***RNHPPD***			***UAPU Rate (%)***	
	N	**Mean**	***SD***	**N**	**Mean**	***SD***
Critical Care	2267	15.1	*2.8*	2121	6.4	*9.5*
Mixed Acuity III	11	14.3	*2.7*	10	4.2	*5.7*
Mixed Acuity II	19	11.1	*2.2*	15	3.7	*8.8*
Step-down	1547	7.7	*2.0*	1417	2.6	*5.3*
Mixed Acuity I	13	7.3	*1.6*	12	0.7	*1.6*
Med/Surg	2429	5.8	*1.5*	2194	2.0	*4.0*

In addition to the six new unit types (three adult, three pediatric) defined above, we propose to introduce the following: Burn unit (adult), burn unit (pediatric), bone marrow transplant unit (adult), and bone marrow transplant (pediatric). These units, along with critical access units, should be allowed to submit data on appropriate indicators and to receive quarterly reports with the understanding that there may be a high degree of within-group heterogeneity for their unit type.

Hospital-Level Indicators

The purpose of this portion of the study was to develop and compare several methods for measuring hospital-level performance on NDNQI indicators. The primary challenge in measuring NDNQI hospital performance is that hospitals differ in the number and type of nursing units they comprise. This makes meaningful comparison of hospitals difficult, even among hospitals of the same size and teaching status. Several hospital-level measures are presented below. We propose a method in which indicator scores are adjusted for unit type and unit size before being aggregated to the hospital level.

The rate of hospital-acquired pressure ulcers (HAPUs) is used as an example in this study without loss of generality; the methods described here can be applied to any indicator computed as a rate or proportion. For method demonstration and comparison, ulcer data were simulated for three fictitious hospitals, each with 1,000 patients assessed for pressure ulcers.

The simulated data are shown in Table 4. Note that the hospitals differ in both the types and sizes of their component units. Hospital 1 has one ICU, one step-down unit, and no rehabilitation unit. Hospital 2 has more ICU and step-down patients than Hospital 1 and has rehabilitation patients. Hospital 3 has the most ICU, step-down, and rehabilitation patients of the three hospitals. In addition, there are differences among the hospitals in the number and size of their medical, surgical, and medical/surgical units.

Table 4

Hospital-Acquired Pressure Ulcers by Unit

	Hospital 1			Hospital 2			Hospital 3		
Unit	**HAPUs**	**Patients**	**Rate**	**HAPUs**	**Patients**	**Rate**	**HAPUs**	**Patients**	**Rate**
ICU 1	9	50	0.180	15	65	0.231	7	40	0.175
ICU 2	-	-	-	-	-	-	6	40	0.150
Step-down 1	12	120	0.100	7	74	0.095	8	78	0.103
Step-down 2	-	-	-	10	74	0.135	7	78	0.090
Medical 1	4	105	0.038	3	86	0.035	6	78	0.077
Medical 2	9	105	0.086	4	86	0.047	5	78	0.064
Medical 3	-	-	-	-	-	-	3	78	0.038
Surgical 1	1	95	0.011	1	84	0.012	0	69	0.000
Surgical 2	-	-	-	0	84	0.000	0	69	0.000
Med/Surg 1	7	105	0.067	5	93	0.054	5	78	0.064
Med/Surg 2	5	105	0.048	7	93	0.075	6	78	0.077
Med/Surg 3	8	105	0.076	8	95	0.084	5	78	0.064
Med/Surg 4	8	105	0.076	2	95	0.021	2	78	0.026

Med/Surg 5	7	105	0.067	-	-	-	-	-	-
Rehab	-	-	-	8	71	0.113	10	80	0.125
Totals	70	1000	-	70	1000	-	70	1000	-

Several methods for calculating a hospital-level ulcer rate measure are described below. In Table 5, which is intended to serve as a visual aid, unit-level data and the values used to compute the hospital measures for Hospital 2 are shown, along with ulcer rate means and standard deviations for the six unit types (based on third quarter 2010 NDNQI data). A comparison of the sample hospitals on the various measures is provided in Table 3.

Method 1. An overall hospital ulcer rate can be computed by dividing the total number of patients in the hospital who have a hospital-acquired pressure ulcer by the total number of patients assessed:

$$(HAPUs_1 + HAPUs_2 + . . . + HAPUs_M)/(n_1 + n_2 + . . . + n_M),$$

where the units in the hospital are numbered $j = 1, 2, . . . , M$; $HAPUs_j$ is the count of HAPUs on the *j*th unit; and n_j is the count of patients assessed on the *j*th unit.

As shown in Table 6, the overall ulcer rate for the three hospitals in the study is 7.0%. Under this method differences among hospitals in the types of units they comprise are ignored, and performance of the three hospitals in preventing pressure ulcers appears to be equal.

Method 2. A raw average of unit ulcer rates can be computed by summing the unit ulcer rates within a hospital and dividing by the number of units:

$$(HAPUs_1/n_1 + HAPUs_2/n_2 + . . . + HAPUs_M/n_M)/M.$$

As with Method 1, the three hospitals appear to be performing equally, each having an average unit ulcer rate of 7.5%. This number is higher than the overall hospital rate because the measure does not account for differences among units in the number of patients assessed, allowing small units with high rates to exert disproportionate influence on the average unit rate. For example, in Hospital 2 the ICU ulcer rate of 23.1%, which is based on 65 patients, is given the same weight as the low surgical unit rates, which are based on 84 patients each (see Table 2).

Moreover, like the overall hospital rate (Method 1), this measure ignores differences among hospitals in the types of units they comprise. For example, whereas the averages for Hospitals 2 and 3 include a rehabilitation unit rate, the average for Hospital 1 does not.

Method 3. Method 2 can be adjusted to control for differences in unit size by weighting each unit's ulcer rate by its number of patients assessed, summing these weighted rates, and dividing by the total number of patients assessed for the hospital. This is equivalent to weighting each unit ulcer rate by that unit's proportion of patients assessed and summing these weighted rates**.** This measure is identical to the overall hospital rate described under Method 1 and does not account for differences among hospitals in the types of units they comprise.

Method 4. The ulcer rate for each unit can be adjusted for unit type by subtracting the average ulcer rate for units of that type and then dividing by the standard deviation of the ulcer rates for units of that type. The resulting z-score is the difference, in standard deviations, of the unit's ulcer rate from the average ulcer rate for units of that type. For example (see Table 4), the ICU in Hospital 2 has an ulcer rate of 23.1%, which is about 1.7 standard deviations above the average ulcer rate for ICUs (6.7%).

These z-scores, which are all on the same metric, can be averaged for each hospital to yield an average unit z-score: $(z_1 + z_2 + . . . + z_M)/M$, where z_j is the z-score for the *j*th unit. As shown in Table 5, the average unit in Hospital 2 has an ulcer rate slightly over one-half a standard deviation above the mean rate for its unit type, while unit rates in Hospitals 1 and 3 average 0.69 and 0.54 standard deviations, respectively, above their unit type means (see Table 5). Like the raw average of unit ulcer rates (Method 2), the average unit z-score does not account for differences in unit size.

Method 5. A weighted average of unit z-scores can be computed by weighting each unit's z-score (defined under Method 4) by its number of patients assessed, summing these weighted scores, and dividing by the total number of patients for the hospital:

$$(z_1n_1 + z_2n_2 + . . . + z_Mn_M)/(n_1 + n_2 + . . . + n_M).$$

Under this method, which takes into account both the types and sizes of each hospital's units, Hospital 1 loses its advantage of having the fewest ICU patients and rehabilitation patients, and Hospital 3 is no longer penalized for having the greatest number of ICU and rehabilitation patients. As shown in Table 3, the score on this measure for Hospital 1 was 0.67, while the score for the other two hospitals was 0.52.

The weighted average of z-scores can be converted to the ulcer rate metric by multiplying by the ulcer rate standard deviation for units of all types (equal to 0.070 for quarter three of 2010) and adding the overall unit ulcer rate mean (0.038). Hospitals 2 and 3 have similar adjusted rates (7.4% and 7.5%, respectively), while the rate for Hospital 1 is a full percentage point higher (8.5%). These rates are higher than the overall hospital ulcer rates, reflecting adjustments for unit type and unit size.

Table 5

Examples of Unit and Hospital Measures for Hospital 2

Unit	Pts	Rate	Z-score	Weighted Z-score	Avg Rate for Unit Type	SD for Unit Type
ICU 1	65	0.231	1.67	0.109	0.067	0.098
Step-down 1	74	0.095	0.87	0.064	0.038	0.065
Step-down 2	74	0.135	1.49	0.111		
Medical 1	86	0.035	0.10	0.009	0.030	0.049

Medical 2	86	0.047	0.34	0.029		
Surgical 1	84	0.012	–0.20	–0.017	0.020	0.040
Surgical 2	84	0	–0.50	–0.042		
Med/Surg 1	93	0.054	0.49	0.045	0.029	0.051
Med/Surg 2	93	0.075	0.91	0.084		
Med/Surg 3	95	0.084	1.08	0.103		
Med/Surg 4	95	0.021	–0.16	–0.015		
Rehab	71	0.113	0.57	0.041	0.048	0.113
Average Unit Rate		0.075				
Average Unit Z-score			0.555			
Weighted Z-score Average				0.520		
Weighted Z-score Average on Ulcer Rate Metric				0.074		

Table 6

Comparison of Hospitals by Measure

	Measure	Hospital 1	Hospital 2	Hospital 3
Method 1	Hospital Rate	0.070	0.070	0.070
Method 2	Average Unit Rate	0.075	0.075	0.075
Method 4	Average Unit Z-score	0.69	0.56	0.54
Method 5	Weighted Z-score Average	0.67	0.52	0.52
	Weighted Z-score Average (Ulcer Rate Metric)	0.085	0.074	0.075

Proposed Method. We propose reporting to hospitals the weighted z-score average (Method 5). While this adjusted rate may not be transparent to all hospital users, since it will not be the same as the raw rate they have internally, it can be used to track a hospital's performance across time and, unlike the other measures considered, allows for meaningful comparison (e.g. percentile ranking) of hospitals within a given comparison group. It should be noted that this adjusted rate is a relative measure, affected not only by the performance of the other NDNQI hospitals but also by changes in the set of hospitals and units reporting data to NDNQI; however, given the number of units participating in NDNQI, we expect unit type averages and standard deviations to be quite stable across time for most indicators.

Discussion

This study resulted in a method for incorporating mixed acuity units into NDNQI data collection and reporting and producing hospital-level indicators, capabilities desired by participating hospitals. These enhancements can be implemented with minimal increase to respondent burden.

The information gathered for the methods development study underscored the difficulty of creating risk adjusted patient outcome measure for acute care units, while maintaining a low respondent burden and transparency for users. The analysis of billing days data from the special study illustrated that mixed acuity unit types can be created that are consistent with NDNQI's existing unit classification scheme. The consistency indicates criterion validity. The stratification of mixed acuity units into types I, Ii and III produced indicator results that were intermediate between existing unit types. Implementation of the mixed acuity unit types required hospitals to go through unit enrollment with the NDNQI liaisons. Hospitals should monitor changes to patient populations on mixed acuity units as signified by changes in the proportion of billing days by payment level. Updates to mixed acuity unit classifications would become part of the site coordinators' responsibility for unit classification maintenance.

The method developed for producing hospital-level indicators accounts for both the unit composition of hospitals, as well as variation in the size of the unit. Implementation of the method requires only that hospitals report on the number of staffed beds per unit. The resulting indicators are in the native metric of the original indicator and thus have a high level of transparency for report users.

Report of Patient Days Reliability Study

Prepared for the American Nurses Association

By

Chenjuan (Tina) Ma, PhD

Byron Gajewski, PhD

Emily Cramer, PhD

Nancy Dunton, PhD

January 2014

Table of Content

Abstract

Patient days measure the length of patients' exposure to the health care settings and reflect patient care workload for nurses and other healthcare professionals. A review of patient days literature highlighted the importance of collecting accurate patient days data for the sake of both clinical practice and health service research. The National Database of Nursing Quality Indicators (NDNQI) has been collecting patient days data from member hospitals since its establishment in 1998. Researchers from the NDNQI have made tremendous efforts in order to ensure high quality of its patient days data, including reliability studies.

The study described in this report is the second patient days reliability study assessing the quality of the NDNQI patient days. Our patient days reliability study consist of two phases: (1) a multi-site patient census study and (2) a site coordinator survey. The multi-site patient census survey was conducted to address specific aim one, to identify the agreement between the NDNQI quarterly submitted patient days data and the patient days data from the multi-site patient census study (gold standard). The site coordinator survey was designed to investigate the current practices and other issues related to patient days data collection and data accuracy in NDNQI member hospitals. Comparisons of the two survey samples to the NDNQI population suggested both the survey samples well represented the NDNQI population.

Participating units in the multi-site patient census study were asked to count the number of patients on their units every 2 hours over a 24-hour period on each of the 7 randomly assigned data collection days in June 2013. 224 units from 58 hospitals that submitted patient days data at least for 4 days were included in our final analyses.

Among these 224 units, 41% used midnight census (Method 1) to collection patient days data for NDNQI quarterly patient days data submission, 40% used midnight censes for inpatients with actual hours for short stay patients (Method 2), 5% used actual hours for both inpatients and short stay patients (M4), and 13% used multiple census reports.

Our reliability analyses using one-way random effect models were conducted overall and by data collection methods. The results indicated a high reliability of the NDNQI patient days data. The overall intra-class correlation coefficient (ICC) between the multi-site patient census study and the matched NDNQI data was 0.95 with a 95% C.I. of 0.94-0.96. The method-specific ICCs ranged from 0.94 to 0.96, and none of their lower boundary of 95% C.I. was below 0.88. Additional analysis also suggested that different data collection method can affect the accuracy of patient days data.

The site coordinator survey found that data accuracy (42%) and availability of the data sources (37%) were the two main factors in the selection of patient days data collection in NDNQI member hospitals. Surveyed site coordinators expressed that the data collection method in use was appropriateness in both units without short stay patients (86%) and units with short stay patients (76%). They also reported that the accurate level of their patient days data for NDNQI as excellent or good for units without short stay patients (87%) and for units with short stay patients (76%).

This reliability study concludes that the current data collection methods used by the NDNQI member hospitals are appropriate. The patient days data collected using these methods is with a good reliability. Meanwhile it suggests that given the increase of observational/short stay patients and the nationwide implementation of electronic healthcare records, NDNQI should modify their patient days data collection methods timely to help member hospitals collect accurate patient days data, such as including patient counts at a different time point in addition to the midnight census.

Part I: Introduction

Patient days as an indicator in health services research and clinical practice directly measures the length of a patient’s stay in a health care setting. It also reflects the patient care workload for nurses and other healthcare professionals and is related to health care cost. Furthermore, patient day indicator is an important element in generating nursing and patient outcome indicators measuring nurse staffing levels and quality of health care. Therefore there is a necessity in collecting reliable patient days data for the sake of both clinical practice and health care research.

The National Database of Nursing Quality Indicators (NDNQI) was founded in 1998 by the American Nursing Association with the mission of aiding nurses in efforts of improving care quality and patient safety (Montalvo, 1997). The NDNQI has collected patient days data since its establishment. Patient days indicator is one of the most important indicators collected by the NDNQI. It is used to calculate the nurse staffing indicator of nursing hours per patient day as well as several nurse sensitive patient outcome indicators, including patient falls and hospital infections. Nurse staffing and patient outcome data are included in the NDNQI member hospital reports to assist hospital executives in their policy-making process, and are frequently used in health care research.

To ensure high quality of the NDNQI patient days data and the indicators generated from it, a research team from the NDNQI conducted the first study to assess the reliability of the patient days data in 2008-2009. Our study described in this report is the second patient days reliability study assessing the quality of NDNQI patient days data. The specific aims of our study are:

Aim 1: To determine the reliability of the NDNQI patient days data by identifying the Intraclass Correlation Coefficient between the NDNQI patient days data and a unique dataset collected from a specially designed multi-site patient census study.

Aim 2: To investigate site coordinators' perspectives on the NDNQI patient days data collection practices and data accuracy via a site coordinator survey.

This study has been approved by the Institutional Review Board of the University of Kansas Medical Center.

Part II Background

This session begins with a synthesis of the literature on patient days with a focus of publications since NDNQI's previous patient days reliability study in 2008-2009. Both methodological studies and investigations including patient days as a study variable were reviewed. The literature review is followed by a detailed description of the NDNQI patient days indicator.

Patient days in research literature

A most basic and important question regarding an indicator is its reliability. Only two published papers were found in PubMed and CINAHL that examined methodological issues of patient days data collection methods and data accuracy at unit-level in the past 5 years (Simon, Yankovskyy, Klaus, Gajewski, & Dunton, 2011; Simon, Yankovskyy, & Dunton, 2010). Findings from these two studies indicate that different patient days collection methods, unit types, and the proportion of short stay patients (SSPs) influence the accuracy of patient days data. Simon, et al. also suggests inaccuracy in patient days data jeopardizes the reliability of reported association between nursing and patient outcomes, because patient days measure is a block stone in generating various nursing and patient outcome indicators.

A review of literature revealed that despite limited research on the quality of patient days data, patient days have been very frequently used in health services research. Researchers have used patient days as the denominator to generate nursing care hours per patient day which reflects the nurse staffing level (Kalisch, Friese, Choi, & Rochman, 2011; Lerner, 2013; Li et al., 2011; Staggs, 2013; Twigg, Duffield, Bremner, Rapley, & Finn, 2011) and to create patient outcome measures (e.g. patient fall) (Flynn, Liang, Dickson, Xie, & Suh, 2012; He, Dunton, & Staggs, 2012; Nedved, Chaudhry, Pilipczuk, & Shah, 2012). Researchers also used patient days in studies examining the association between nursing and patient outcomes (Kalisch, Tschannen, & Lee, 2012; Lake, Shang, Klaus, & Dunton, 2010; Wilson, Bremner, Hauck, & Finn, 2011). Although not conclusive, some researchers have suggested that nursing hours per patient day at a unit-level is a better indicator of nurse staffing levels with higher predictive power in patient outcomes research, compared to alternative nurse staffing measures from administrative data (Van den Heede et al., 2009).Therefore, the quality of patient days data plays an important role in nursing and patient outcomes research. To ensure the reliability of the findings from this research, accurate patient days data are required.

In summary, the findings from our literature review regarding the scarcity of research on measuring patient days and the wide use of patient days data in clinical practice and health services research highlight the importance and necessity of collecting reliable patient days data.

NDNQI patient days indicator

Patient days indicator is one of first collected indicators by the NDNQI. Patient days have been measured at different levels, such as unit-level or hospital-level (Carayon et al., 2013; Chen, Sexton, Kaye, & Anderson, 2009; Staggs & He, 2013). The NDNQI patient days data were collected at unit-level. Registered units from NDNQI member hospitals are required to report monthly patient days data once in each quarter using one of the methods defined by the NDNQI. The NDNQI patient days data have also been used to calculate other NDNQI indicators including but not limited to the nursing care hours per patient day and patient falls.

NDNQI researchers have been making continuous efforts to improve the quality of NDNQI patient days data. One example is creating and timely updating of the NDNQI Patient Days Guidelines for Data Collection. In the Guidelines, each available method for collecting patient days data is detailed. These methods are based on either patient census counts, actual patient hours, or a combination of the two. Patient Days Guidelines instruct NDNQI member hospitals to select the method that will provide the most accurate patient days data from their units, depending on the frequency of patient turnover on the unit, the presence of short stay patients, and availability of data sources.

In addition, the NDNQI researchers conducted a reliability study in 2008-2009 to assess the accuracy of patient days data. Findings from this study indicated that the NDNQI quarterly patient days data demonstrated high reliability overall as well as method-specifically for four of the five available data collection methods then. The exception was for method 3, Midnight census and patient days from average hours for short stay patients (Simon et al., 2011). Following suggestions from this study, method 3 was removed from the Patient Days Guidelines. To date, there are four patient days data collection methods used by NDNQI member units. These methods and their definitions are described in **Table1**.

Table 1 NDNQI patient days data collection methods and definitions

Meth	Definition
M1	Midnight census
M2	Midnight census + actual hours for short stay patients
M4	Patient days from actual hours for all patients

M5	Patient days from multiple census reports

M3 was removed based on the findings from a reliability study in 2008-2009.

Meanwhile, we also realize that some potential differences may exist between the NDNQI patient days data collected using the four data collection methods and the patient days data in the real world. **Figure 1** illustrates a unit with five example patients counted in two hours census counts. **Table 2** shows the unit's patient days data derived by each collection methods. According to the definitions, Method 4 reports patient days based on the actual hours and gives the most accurate estimation of patient days. All the other methods underestimate or overestimate patient days. As shown in **Figure 1**, the estimated patient days with different collection methods are affected by the occurrence of short stay or observational patients (<23 hours on the unit) and their exact admission/discharge times. In addition, the average length of patients' stay which is related to patient turnover may also play a role in accurately measuring patient days.

Figure 1. Example patient days

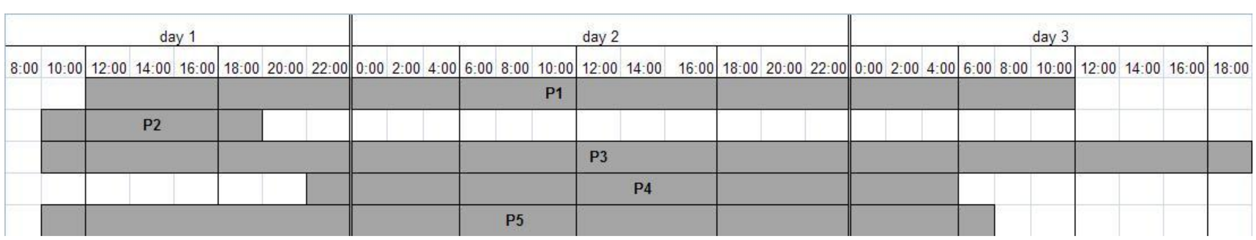


Table 2 Patient days by data collection methods of the illustrating example

Method	**Patien**	**Patie**
M1: Midnight census	192	8.0
M2: Midnight census for inpatients + actually hours for short stay patients	202	8.4

M4: Patient days from actual hours for all	194	8.1
M5: Patient days from multiple census counts	200	8.3

In addition, there have been many dynamic changes in the hospital care settings, which may have affected the ways of collecting patient days data and consequently influence the accuracy of patient days data. Among the most important changes are, for example, the increase in the number of short stay patients in some units and increasing implementation of electronic health records (EHRs). The EHRs dramatically increase the availability of more accurate patient days data within hospitals.

To summarize, both the literature review and the description of the NDNQI patient days data suggest a second reliability study assessing the quality of the NDNQI patient days data is needed.

Part III: Study Methods and Results

This reliability study was conducted in two phases, the Multi-Site Patient Census Study (Phase I) and the Site Coordinator Survey (Phase II). Phase I was designed to address the specific aim one and the phase II for specific aim two. In this session, we described the research methods and results for each phase respectively.

Phase I: Multi-Site Patient Census Study

Methods

Multi-site patient census study

To address our primary aim of assessing the reliability of the NDNQI quarterly collected patient days data, we designed a multi-site patient census survey to collect patient days data that reflects the patient days in reality. In this multi-site survey, we asked site coordinators of participating units to collect data on seven randomly assigned days during June of 2013. Each of these seven days also represented one day of a week. Therefore each hospital had a random pattern of data collection days. On data collection days, patients on each unit were counted every two hours for a period of 24 hours on each of the designated data collection day (**Appendix A**). All the collected data from each participating unit were then entered into an online database, from which researchers from the NDNQI can assess to and download the data.

Ideally the best way to assess the quality of NDNQI patient days data is to compare it to a patient days dataset based on actual check-in and check-out time of each patient cared for on a unit. However, this method will require the collection of individual patient data and substantially increase respondents' burden, which makes it not feasible. Our survey did not obtain patients' actual admission and discharge time on a unit; however, the aforementioned **Figure 1 and Table 2** display that our survey design (counting patient every two hours on seven randomly selected days) could generate very similar, if not exactly the same, data as in the real world with a huge decrease of workload to nurses. The data collected via this survey was considered as the gold standard in the reliability analyses.

Sample calculation

A sample size calculation was completed before the survey was sent out. Given the findings from last NDNQI patient days reliability study (Simon et al., 2011), an intraclass correlation coefficient (ICC) of 0.90 with an estimated error margin of 0.05 were chose to calculated the minimum desired sample size, which is 220 units (Giraudeau & Mary, 2001).

Data analysis

Patient days data from the multi-site patient census survey were calculated and aggregated for each unit to reflect the actual patient days of participating units in June of 2013. The average number of patients on each data collection day on each unit was first calculated. Then this average number was multiplied by the number of occurrences of this weekday (e.g. 5 Saturdays) in June. Finally all the products were added up to indicate the patient days of each unit for the month of June 2013. The NDNQI patient days data that are submitted quarterly was obtained from the NDNQI server for the same month of the same units and was used for comparison to the survey data in the reliability analyses.

We first described the response rates of the multi-site patient census survey. We then conducted a descriptive comparison of the survey data and the NDNQI patient days data by data collection methods, unit characteristics and hospital characteristics to identify whether the survey sample well represented the NDNQI patient days data. We identified the level of agreement presented by ICC (1,1) between multi-site patient census data and the NDNQI patient days data using one-way random effect models. Finally, we conducted an exploratory analysis to identify factors that might influence the patient days data collection and accuracy. All the analyses were completed using STATA 12.0 (StataCorp, College Station, TX).

Results

Sample

Initial call for participation in this multi-site patient census study was sent out by email to 700 randomly selected hospitals from a total of 1793 NDNQI member hospitals which are currently active and have units submitting patient days data in the past three years. Sixty six hospitals agreed to participate in this study. From these 66 hospitals, 282 units were randomly selected. During the data collection period, 234 units in 63 hospitals attempted to submit patient days data. Two hundred and thirty units from 60 hospitals submitted data (including both complete and incomplete submissions). For analysis purpose, only units with data for at least 4 days out of the 7 selected data collection days were included. Our final multi-site patient census survey data consisted of 224 units from 58 hospitals. Two units were further excluded from the comparative analysis between the survey data and NDNQI data due to the lack of NDNQI patient days data in the survey month. **Table 3** summarizes the response rates for hospitals, units, and data collection days.

Table 3. Response rates of hospitals, units, and data collection days

	Enrolled	Any data submission	4 or more days of data submission	Response rate (%)
Hospitals	66	60	58	90
Units	282	230	224	82
Data collection days	1,974	1,573	1,552	80

Table 4 compares the units in the survey sample with all units in NDNQI member hospitals for which patient days data were submitted by patient days data collection methods, patient population, hospital bed size, teaching status, and magnet status. There were slightly more small (defined by bed size) non-teaching non-magnet hospitals in our survey sample than the NDNQI population, yet the largest difference was no more than 11%. The distributions of patient days data collection methods and patient population, the two most important factors related to the quality of patient days data, were very similar across the survey sample and the NDNQI population. Thus, despite some differences, we are confident our survey sample well represented the NDNQI population.

Table 4. Comparison of survey sample and NDNQI population by collection methods, patient population, and hospital characteristics

	Survey sample (n=224)		NDNQI population (n=15,368)	
	N	%	N	%
Collection method				
M1 - MC*	91	40.63	6,661	43.34
M2 - MC + PDs** from actual hours for SSPs***	90	40.18	5,103	33.21
M4 - PDs from act. hrs for inpatients and SSPs	12	5.36	1,391	9.05
M5 - PDs from multiple census reports	29	12.95	2,213	14.4
Patient population				
Adult inpatient	168	75.00	11,670	75.94
Neonatal inpatient	14	6.25	768	5.00
Pediatric inpatient	18	8.04	1,155	7.52
Psychiatric	9	4.02	1,139	7.41
Rehab inpatient	15	6.70	636	4.14
Hospital bed size				
< 100	37	16.52	1,332	8.67
100-199	74	33.04	3,289	21.40
200-299	30	13.39	3,359	21.86
300-399	28	12.50	2,451	15.95
400-499	20	8.93	1,692	11.01
>=500	35	15.63	3,245	21.12
Teaching status				
Academic Medical Center	30	13.39	3,286	21.38
Teaching	7	3	6	4

	2	2.14	,298	0.98
Non-teaching	1 22	5 4.46	5 ,784	3 7.64
Magnet status				
Non-Magnet	1 27	5 6.7	7 ,210	4 6.92
Applicant	4 0	1 7.86	3 ,005	1 9.55
Magnet	5 7	2 5.45	5 ,153	3 3.53

*Midnight census; **Patient days; ***Short stay patients

Reliability analysis

Table 4 displays the distribution of overall patient days and patient days by data collection methods in the survey sample and the matched NDNQI sample. Overall, the patient days data from our multi-site patient census survey and the matched NDNQI patient days data had almost the same distribution. When stratified by data collection methods, the largest difference (difference between means) was observed in Method 5, collecting patient day data from multiple census reports. The matched NDNQI data had an average of approximately 23 more days per month in June 2013 than the survey sample. The second largest difference existed when using method 4, collecting patient day data from actual hours for both inpatients and short stay patients, for which the matched NDNQI data had 21 days less per month on average. Using method 1, mid-night census only, and method 2, mid-night census and actual hours for short stay patients, both resulted in a 7-day difference but with reverse directions.

Table 4. Comparison of distributions of patient days between the survey sample and the matched NDNQI sample (n=213)*

	SURVEY		NDNQI	
	Mean (SD)	**Range**	**Mean (SD)**	**Range**
Overall	525.15 (289.23)	22.33-1672.42	526.74 (292.98)	33.00-1717.00
M1 - MC	551.65 (312.02)	22.33-1569.08	544.90 (313.31)	33.00-1591.00
M2 - MC + PDs from actual hours for SSPs	496.57 (253.57)	97.17-1106.75	503.69 (257.67)	92.00-1147.00
M4 - PDs from act. hrs for inpatients and SSPs	524.92 (415.39)	143.50-1672.42	504.26 (436.14)	142.00-1717.00
M5 - PDs from multiple census reports	534.12 (265.54)	52.67-1144.08	556.85 (265.24)	56.00-1148.00

To achieve a more accurate comparison, only units with 7-day data from the multi-site patient census survey were included.

The one-way random-effects ICC analyses identifying the overall and method-specific agreement of the patient days data between the survey sample and the matched NDNQI sample were conducted on units with at least 4 days of data submission (n=224) as well as on units submitted data on all 7 data collection days (n=213). No significant differences were observed in the results and here we only reported the results from the analysis including all units submitting data for 4 days or more. The results are displayed in **Table 6.** The reported ICC is the absolute-agreement ICC, which reflects the proportion of between units variance in total variance. In other words, the larger the ICC, the higher the agreement between the survey sample and the matched NDNQI sample.

Table 5. ICC of patient days data from the survey sample and the matched NDNQI sample

	N	**ICC**	**95 % C.I.**
Overall	224	0.95	0.94-0.96
M1 - MC	91	0.96	0.94-0.97
M2 - MC + PDs from actual hours for SSPs	90	0.94	0.91-0.96
M4 - PDs from act. hrs for inpatients and SSPs	12	0.96	0.88-0.99
M5 - PDs from multiple census reports	29	0.96	0.92-0.98

Our analyses indicated excellent agreement of the patient days data between the survey sample and the matched NDNQI sample. The overall ICC (1, 1) between the survey sample and the matched NDNQI sample was 0.95 (95%C.I.: 0.94-0.96), and the method-specific ICCs (1, 1) ranged from 0.94-0.96. All of them are above 0.8, which is usually considered as a cut-off point for excellent agreement (Forbes & Taunton, 1994;

Hughes & Anderson, 1994). Furthermore, given the representativeness of the survey sample compared with the NDNQI population (**Table 4**) and the ICCs (overall and method-specific), our study indicated a high reliability of the NDNQI patient days indicator.

In addition to the ICC analyses, an examination of factors that may be related to the difference between the survey data and the NDNQI data was conducted. The study factor included data collection method, unit type, and degree of short stay patients (DSSP). In this analysis, the different between the two data was defined as the ratio of the NDNQI data to the survey data for each unit. The calculation of DSSP used information from a single item question in the survey asking whether there have been short stay patient(s) during that data collection day. The DSSP for each unit was the proportion of days with existence of short stay patient(s) out of the total data collection days (at least 4 days or more).

Figure 2 displays the distribution of the ratio of the NDNQI patient days to patient days in the survey sample. This overall difference and the method-specific differences are summarized in **Table 7**. Results from one-way variance analysis with Bonferroni correction for multiple comparisons confirmed that reporting methods had an influence on the accuracy of patient day data. No significant association was found between the NDNQI data to survey data ratio and the DSSP and between the the NDNQI data to survey data ratio and unit type.

Figure 2. Distribution of the NDNQI data to survey data ratios on patient days

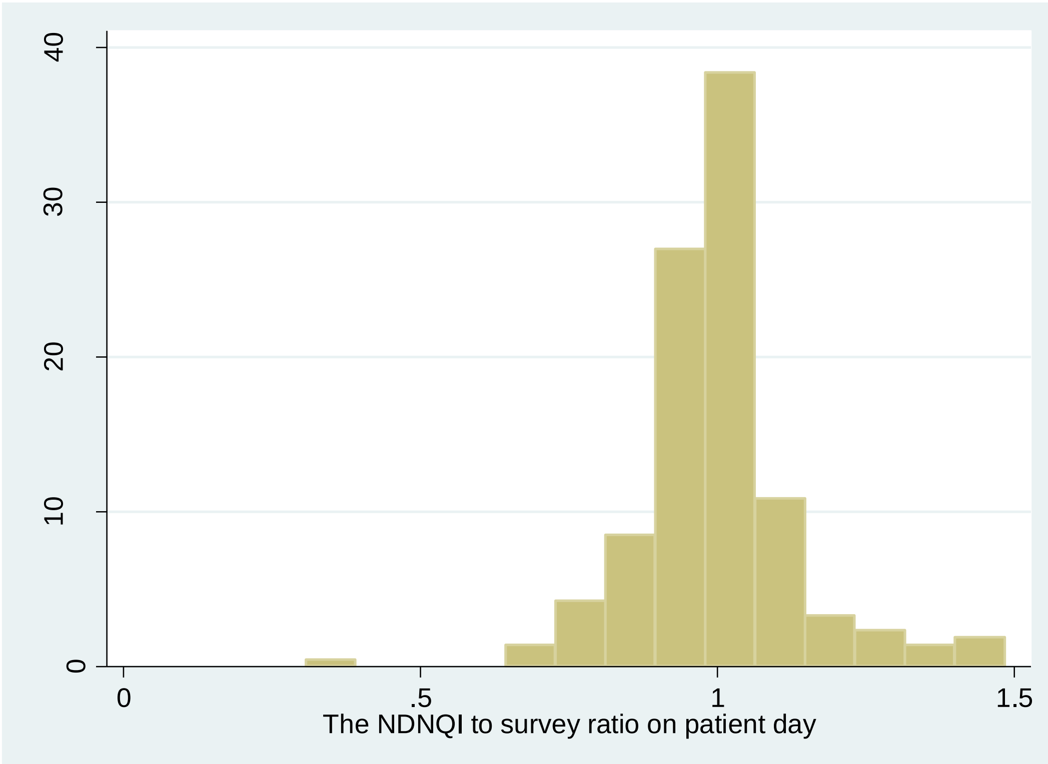


Table 7. Distribution of the NDNQI data to survey data ratios on patient days

		Mean (SD)	**Range**
Overall	19	1.01 (0.16)	0.31-1.74
M1 - MC	0	1.00 (0.13)	0.67-1.48
M2 - MC + PDs from actual hours for SSPs	9	1.00 (0.13)	0.74-1.48
M4 - PDs from actual hours for inpatients and SSPs	2	0.97 (0.26)	0.31-1.32
M5 - PDs from multiple census reports	8	1.11 (0.23)	0.88-1.74

Notes: MC: Midnight census; PDs: patient days; SSPs: short stay patients

Phase II: Site Coordinator Survey

Method

Site Coordinator Survey

In addition to the multi-site patient census survey, an online survey was designed to collection information from site coordinators (SCs) about their experience and perception of current patient days data collecting practice and other related issues in their hospitals (**Appendix B**). This information will assist NDNQI in verifying the reliability and validity of its patient days data received from member hospitals and provides informative evidence for future efforts in improving data collection. The survey was revised from the questionnaire used in the 2008-2009 NDNQI Patient Days Reliability Study Site Coordinator Survey. The face validity of this survey was assessed by NDNQI staff, including members from research team, analyst team, and liaison team.

SCs were chosen as the targeted population in this survey because of their vital role in ensuring high quality data collection and submission to NDNQI. SCs' responsibilities can be summarized in three areas. First, SCs are the primary point of contact between the NDNQI and member hospitals regarding any quarterly and annually NDNQI data collection and submission. Second, SCs are the coordinator in the process of unit enrollment/determination in NDNQI data submission and perform on-going unit monitoring and supervising of data collection and submission activities. Third, SCs are responsible for maintaining the security of his/her hospitals website by authorizing who is allowed access within his/her facility. The majority of the SCs are responsible for one hospital, however, some of them may coordinate two or more hospitals. For those who are in charge of more than one hospital, they were asked to choose the first hospital by alphabetical order of their hospitals' name to complete the survey.

In late June and early July 2013 the online survey was constructed and tested using REDCap before public access by SCs. RedCap is developed by Vanderbilty University and is a secure, web-based application designed exclusively to support data collection for research purpose (Harris et al., 2009). Five hundred and sixty three SCs were randomly selected from a total of 1,492 SCs representing 1793 active NDNQI member hospitals that have unit(s) submitting patient days data in the past three years. An invitation letter along with detailed explanation of the research purpose and other related information was sent to those randomly selected SCs on July 10, 2013. SCs were asked to complete the online survey within a two-week survey period. A follow up email was sent out to remind SCs about survey deadline.

Data analysis

Descriptive statistics of the enrollment and response rates were first analyzed. It was followed by a comparison of hospital characteristics between the survey hospitals and the NDNQI member hospitals to determine the representativeness of this SC survey. Finally a detailed descriptive analysis of SCs' perception of different aspects of the patient days data collection practice in their hospitals.

Results

Sample

Two hundred and ninety three SCs who received an invitation of participation letter responded to the online survey, which resulted in a response rate of 52%. Among the 293 respondents, 279 SCs representing 279 hospitals completed the online survey, with 44% of them have been in the SC position for 3 years or more. In our analysis, only information from the 279 completed surveys was used.

One concern of our study sample was whether it was representative of the NDNQI member hospitals. To address this concern, a comparison of the survey respondents and the NDNQI population regarding their hospital characteristics, such as hospital type and bed size, was conducted. The results are presented in **Table 8**. The comparison analysis revealed minor differences between the NDNQI population and survey sample in term of hospital characteristics. For example, there were slightly fewer large hospitals with a bed size of 300 or more and more non-teaching hospitals in the survey population than the NDNQI population. However these differences are not expected to weaken the representativeness of this SC survey.

Table 8. Comparison of NDNQI population and SC survey by hospital characteristics

	NDNQI* (n=15,368)		**Survey (n=279)**	
	N	**%**	**N**	**%**
Hospital Type				
General	14,325	93.21	238	85.3
Pediatric	497	3.23	10	3.58
Critical Access	69	0.45	9	3.23
Rehabilitation	110	0.72	9	3.23
Phsychiatric	88	0.57	2	0.72
Other	279	1.82	20	7.12
Hospital Size				
0-24	139	0.90	5	1.79
25-49	290	1.89	20	7.17
50-74	34	2.2	19	6.8

75-99	3560	33.64	24	18.60
100-199	3,289	21.40	77	27.60
200-299	3,359	21.86	63	22.58
300-399	2,451	15.95	27	9.68
400-499	1,692	11.01	18	6.45
500 or more	3,245	21.12	26	9.32
Teaching Status				
Academic Medical Center	3,286	21.38	34	12.19
Teaching	6,298	40.98	114	40.86
Non-teaching	5,784	37.64	131	46.95
Magnet Status				
Yes	5,153	33.53	68	24.37
No	10,215	66.47	211	75.63

*NDNQI hospitals with unit(s) submitting patient day data in June 2013

Data collection method selection and the determining factor

In addition to unit managers and/or SCs, administrative and/or financial department played an important role in determining the method of collecting and reporting patient days (**Figure 3**). Approximately 34% of the survey respondents reported that the reporting method were decided by administrative department and/or financial department without their involvement; and another 33% of them indicated that the reporting method was a collaborative production of unit manager, SC, and administrative/financial department. The decision of which data collection method to use was usually based on data accuracy (43%) and the availability of patient days data sources (37%) (**Figure 4**).

Figure 3. Data collection method selection (%)

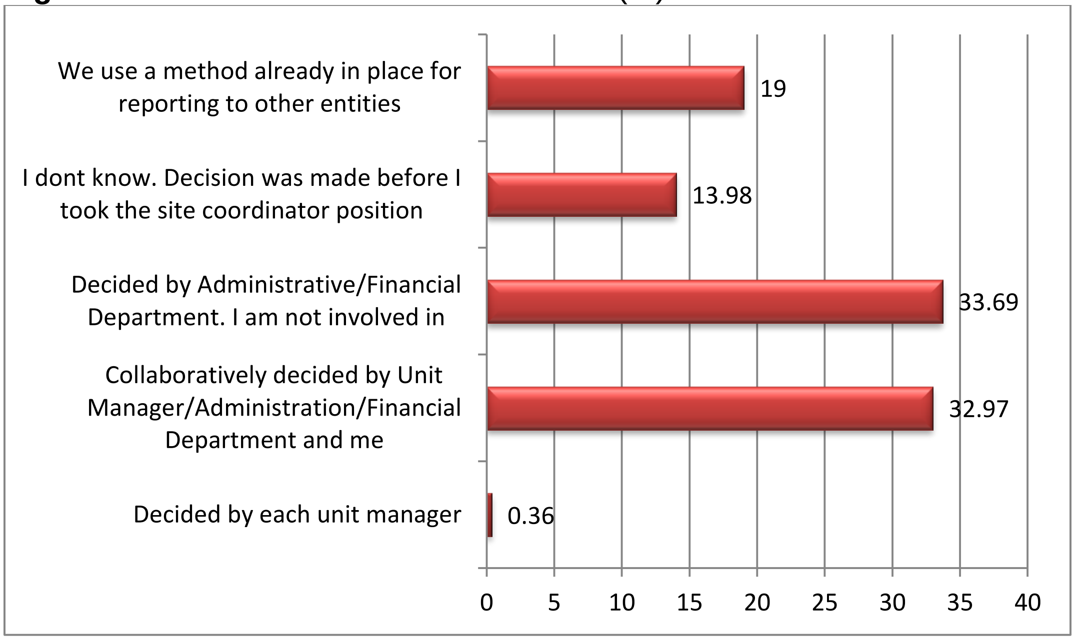
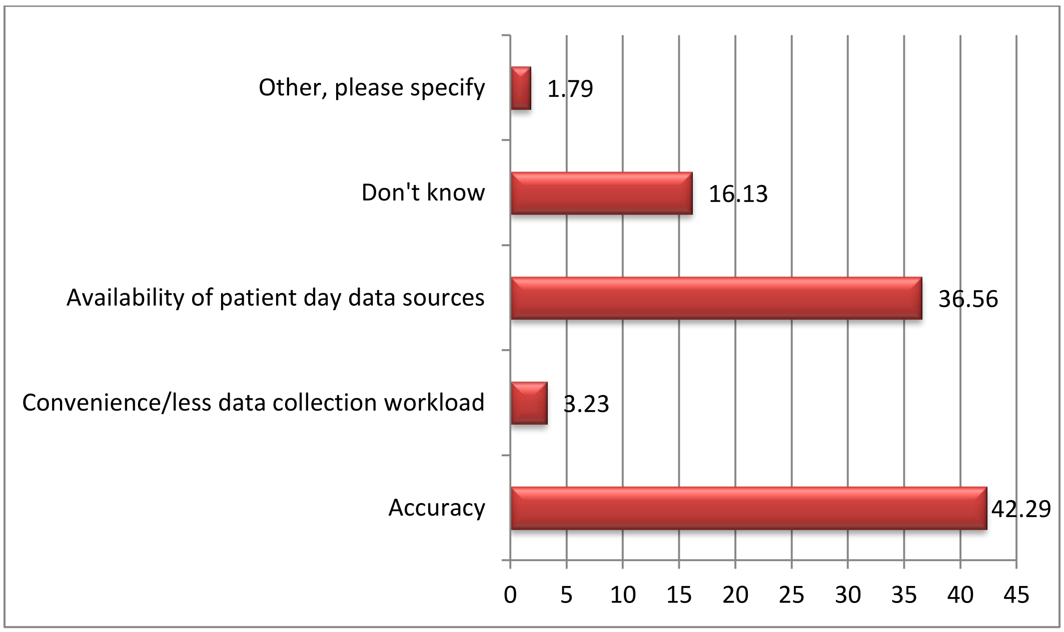


Figure 4. Most important reason in data collection method selection (%)



Appropriateness of data collection method and accuracy of patient days data

SCs were asked in the survey about their perceptions of the appropriateness of their data collection method and the accuracy of their submitted patient days data by units with and without short stay patients respectively. Because previous NDNQI patient days reliability study suggested that the presence and proportion of short stay patients could affect data accuracy. SCs indicated that the patient day collection method used in units without short stay patients were very appropriate (42%) or appropriate (45%), and in units with short stay patients were very appropriate (34%) or appropriate (42%) (**Figure 5**). In term of the accuracy of the submitted patient days data, 50% of the SCs reported it as excellent and 37% reported it as good in units without short stay patients; and 38% indicated it as excellent and 38% indicated it as good in units with short stay patients (**Figure 6**). Regardless of the presence of short stay patients or not, over 10% of the SCs were not sure about the appropriateness of method and the accuracy of submitted data.

Utilization of NDNQI Guidelines and data verification approaches

We asked SCs questions about the use of NDNQI Guidelines for Data Collection manual or online tutorial and approaches for data verification as indirect indicators of the overall quality of patient days data. As shown in **Figure 7**, roughly 37% of the SCs reported that they referred to the NDNQI Guideline or online tutorial for data collection at least once a quarter.

Figure 5. Appropriateness of data collection method by units with/without short stay patients (%)

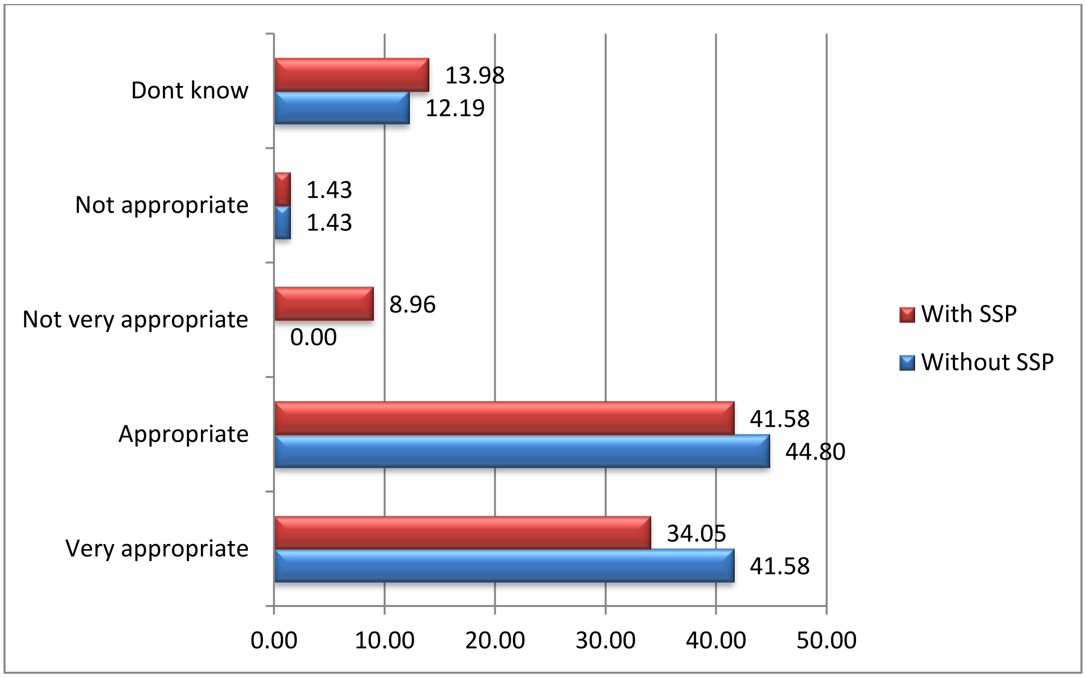
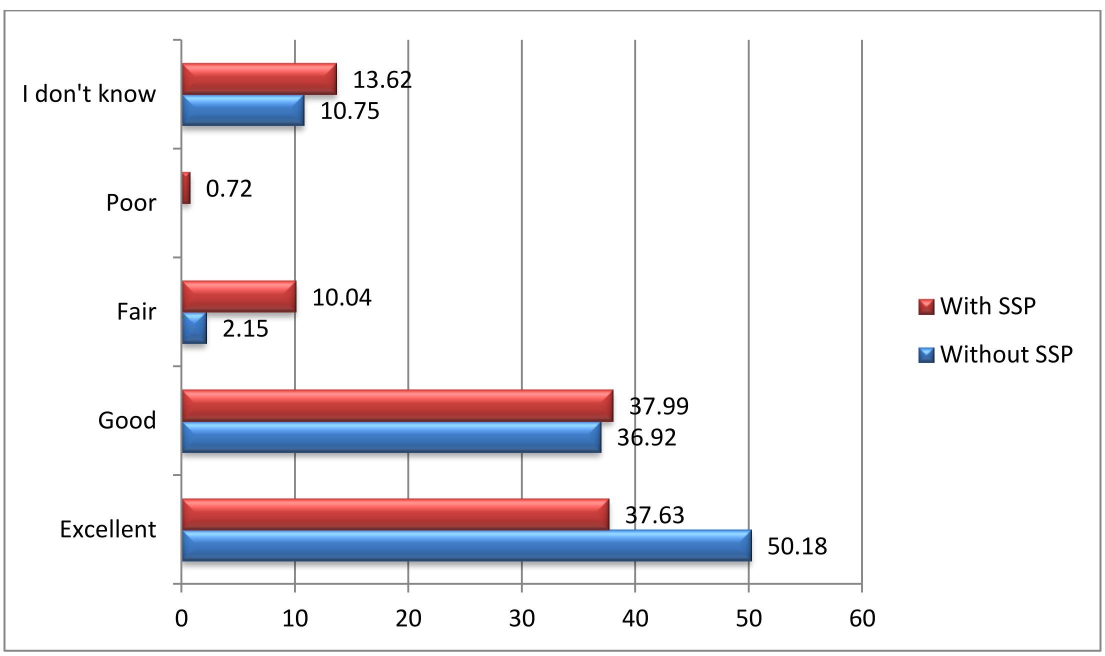


Figure 6. Accuracy of patient days data by units with/without short stay patients (%)



In addition to the NDNQI Guidelines and online tutorials, we also asked SCs about other data verification approaches. **Figure 8** shows the number of verification approaches used and **Figure 9** presents how frequently these approaches were used. Approximately three in four of the SCs reported that their hospitals used at least one method to verify patient days data before submission, and about 38% of the SCs indicated their hospitals used 2-4 methods for data verification. Among the listed four data verification methods, “comparing each unit’s value to earlier quarters’ data” was most frequently employed (56%), followed by “comparing to values used in other reports” (43%, **Figure 9**). There were 23% of the SCs reported that they did not use other verification and submitted data as received.

Figure 7. Utilization of NDNQI Guidelines or online tutorial for data collection (%)

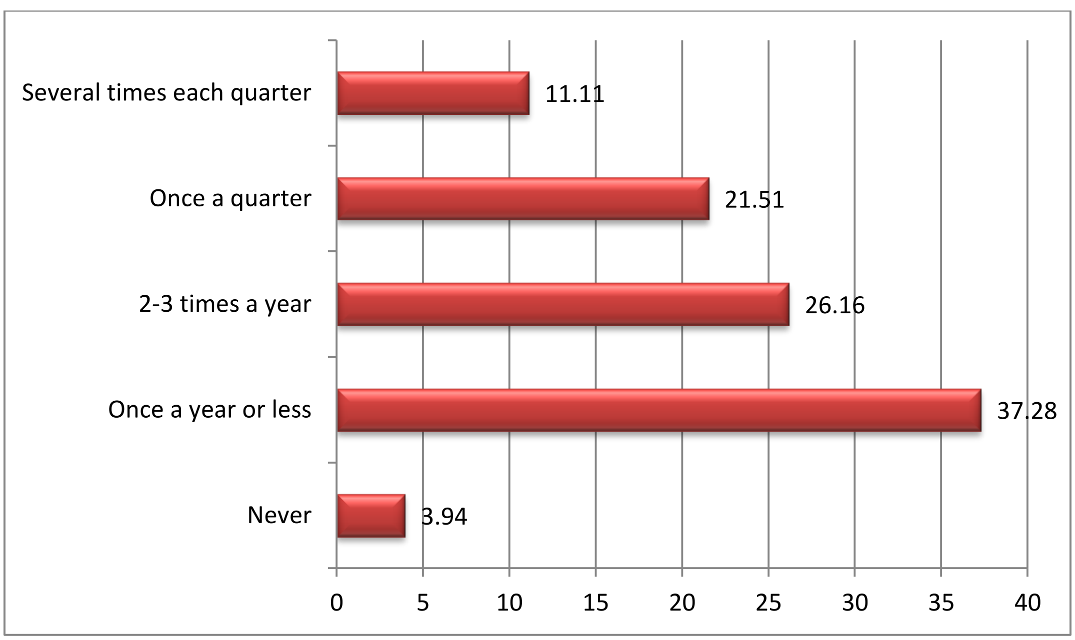


Figure 8. Number of data verification methods used (%)

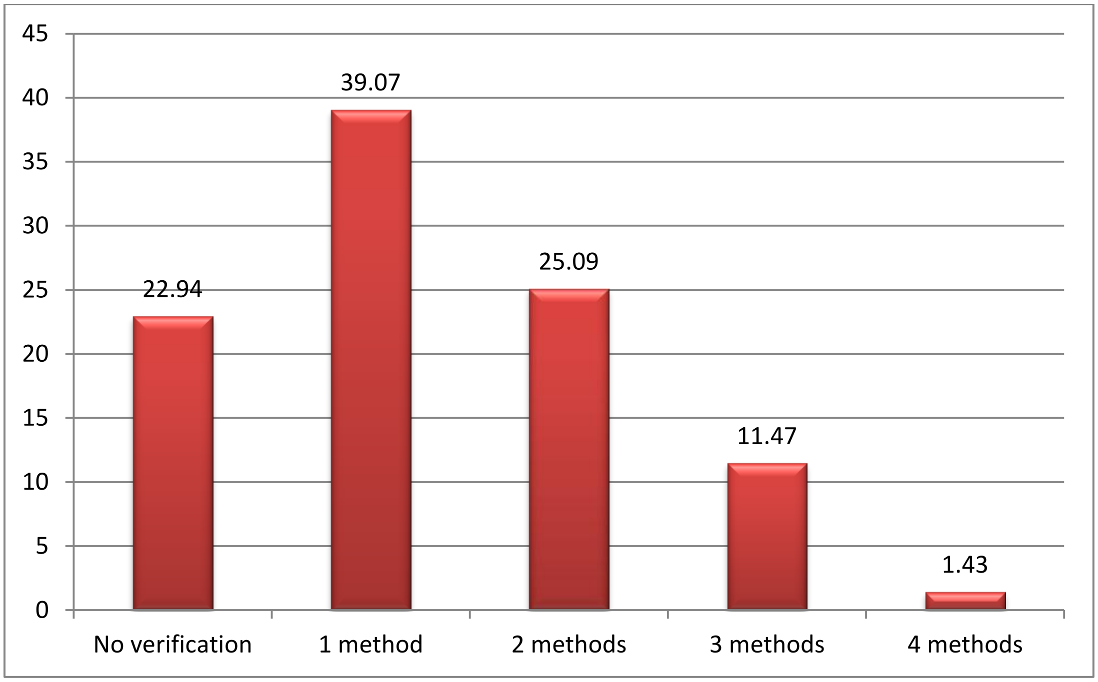
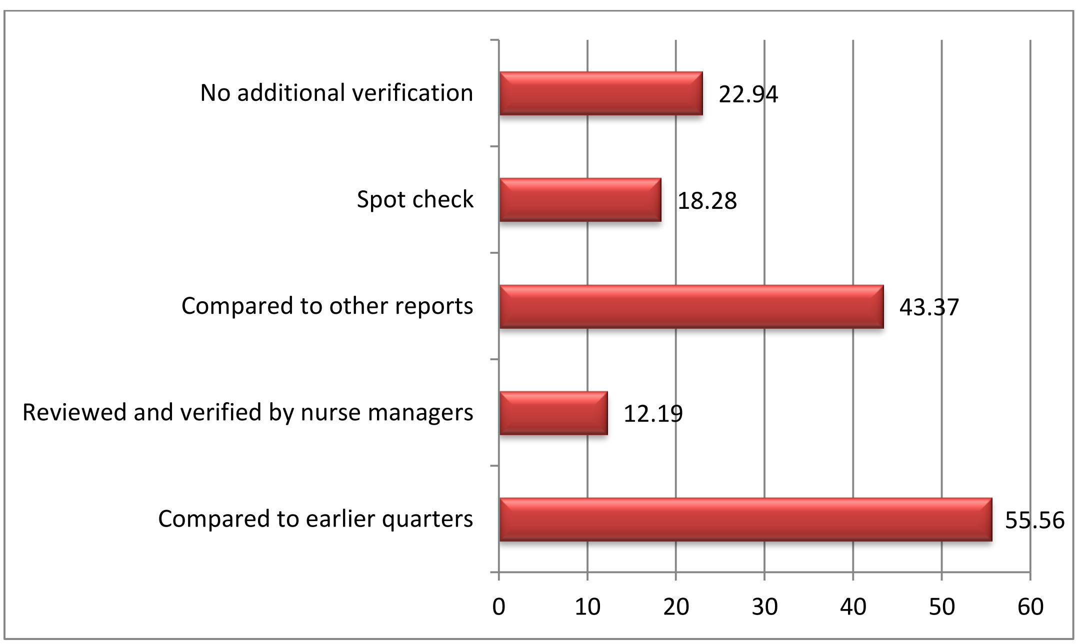


Figure 9. Utilization of different data verification methods (%)

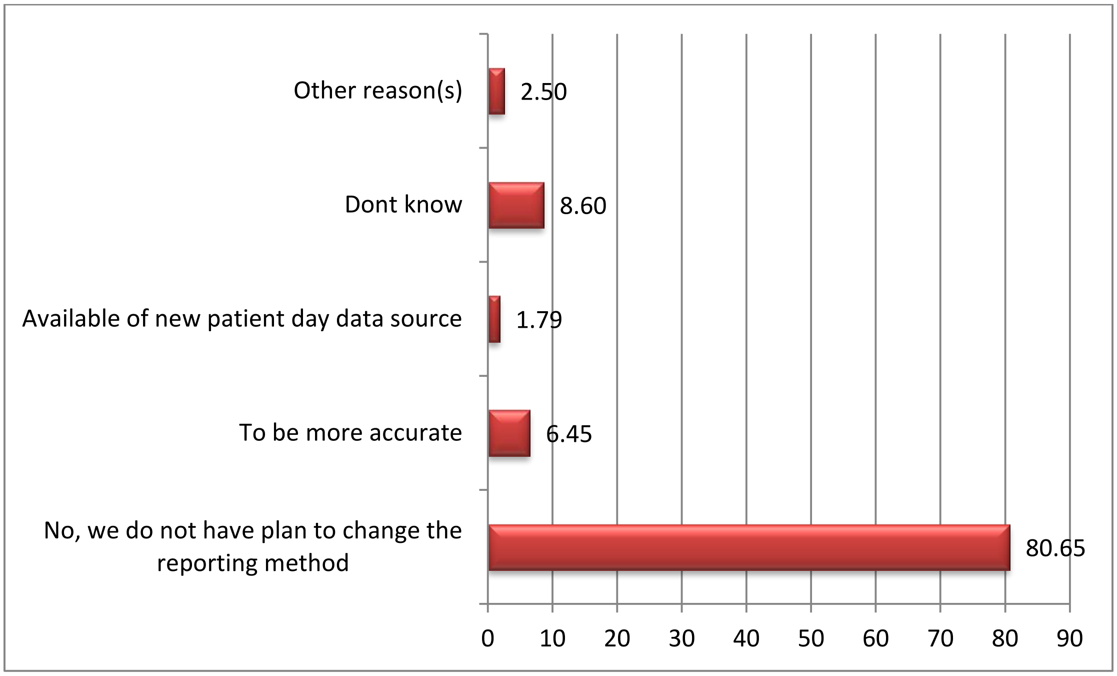


Other issues related to patient days data collection

When asked about future plans to change patient days data collection method within the next six months, about 81% of the SCs expressed that they did not plan to change the method (**Figure 10**). Among those who indicated a plan of changing the methods (11% or 30 SCs), the reasons for change varied. These reasons included "to be more accurate", "available of new patient day data sources," "changing units with sd and cc beds", and "increase of short stay or observational patients on units."

SCs were also asked to report how their hospitals track actual patient hours from short stay patients, "medical records (electronic, paper, or both)," "administrative/financial department data," "others, please specify" or "do not know." Over half of the SCs (57%) indicated that the administrative/financial department provided hours from short stay patients for NDNQI data submission. Approximately 29% reported they used medical records. Among those who reported "other," the answer they gave varied and included "other patient hour tracking programs/systems, such as patient classification system," "tracked by quality data personnel," "specific ADT information," "bed control program that counts bed at noon and midnight," "use 4 data points in 24 hours to capture patients in a bed," "time tracked by nursing supervisors," and "do not submit short stay hours or not applicable,"

Figure 10. Plan to change data collectin method within the 6 months (%)



Part IV: Conclusion

The purpose of this study is to examine the reliability of the NDNQI quarterly patient days data collected via four different methods by comparing it to a dataset collected through a uniquely designed multi-site patient census study. Our analysis indicates excellent reliability in general as well as method-specific. All the identified ICCs (1, 1) were above 0.9 (**Table 6**) and none of their low boundary of 95% CI is below 0.88 (Lee, Koh, & Ong, 1989), which indicates no revision is required at the current moment NDNQI patient day data collection methods.

This study also confirmed that using different data collection methods can result in different levels of data accuracy. This finding is as expected and consistent with other literature (Beswick, Hill, & Anderson, 2010; Simon et al., 2011). However, it should be noted that this difference only exists in Method 5 (patient days from multiple census reports). Furthermore, results from the linear regression model indicated that reported patient days using Method 5 were likely to be overestimated.

We did not find a relationship between the DSSP and the difference between survey data and the NDNQI data (measured as the ratio of the NDNQI data to the multi-site patient census survey data). This should not be simply considered as DSSP does not influence patient days data collection method and accuracy. Our analyses showed that units with higher proportions of short stay patients were more likely to select the reporting methods tracking actual hours of the short stay patients. It should be noted that although the calculated DSSP partially reflected the situation of short stay patients on study units, it did not fully depict the picture of the proportion of patient days attributable to short stay patients. SCs also expressed in the survey that they felt more confident in both the appropriateness of the data collection method in use and the accuracy of the collected data on units without short stay patients compared to units with short stay patients (**Figure 5 and Figure 6**). These reasons may also explain that why no relationship was detected between unit types and the data accuracy.

Data accuracy is the main driving factor in selecting patient days data collection method; however method selection and data accuracy is also limited due to the availability of patient days data sources. Theoretically, units using Method 1 (midnight census) should be only those without short stay patients. Our analyses showed that 44% of the units with Method 1 (midnight census) have some short stay patients. Findings from the SC survey supported this conclusion as well (**Figure 4**).

Our study suggested that the NDNQI should pay close attention to the content in the NDNQI data collection guideline and/or online tutorials and maybe provide several optional approaches for units to validating data before submission, in order to help units collect accurate patient day data (**Figure 7, Figure 8 and Figure 9**). SCs expressed the frequent use of the NDNQI guidelines and online tutorial during data collection. Other information such as how short stay patients can influence patient days measurement (and associated factors like total nursing care hours per patient day or falls) and how to conduct pilot studies to collect patient days for short stay patients could also be added into the NDNQI data collection guidelines.

Midnight census is the most frequently used method in collecting patient days data. Yet, it is the method least likely to result in accurate patient day data at the presence of short stay patients (Beswick et al., 2010). One of the approaches that can better capture the existence of short stay patients without a dramatic increase in workload is the midnight/noon census method. There are units that have used this method for patient days data collection according to the SC survey. Moreover, evidence has suggested that collecting the counts of patient admissions and discharges along with the collection of patient days data can better reflect the care workload for nurses (Beswick et al., 2010).

Given the increasing use of the Electronic Health Records (EHRs) in health care settings, units should be encouraged to use EHRs which can provides more accurate information regarding patients' stay on a unit. The SC survey showed that

EHRs is not used (29%) or only partially used (32%) in reporting patient days data, although 93% of the SCs claimed that their hospitals had EHRs. This also requires a more active role of the SCs coordinating data collection activities between the units and department in charge of the EHRs.

This study has a couple of limitations. First, we were not able to obtain the exact admission and discharge time of each patient on participating units. However, as illustrated in **Figure 7** and described in on page 6, collecting patient counts every two hours does not substantially increase workload for nurses while generate patient days data that is very similar as using admission and discharge time. Another limitation is that the multi-site patient census survey did not collect information on the proportion of patient days from short stay patients. However, the proportion of short stay patients is not the primary aim of this study; and we were able to create a variable that partially reflects the existence of short stay patients by asking whether there were short stay patient(s) on each data collection day.

References

Beswick, Susan, Hill, Pamela D, & Anderson, Mary Ann. (2010). Comparison of nurse workload approaches. *Journal of Nursing Management, 18*(5), 592-598.

Carayon, Pascale, Wetterneck, Tosha B, Cartmill, Randi, Blosky, Mary Ann, Brown, Roger, Kim, Robert, . . . Wood, Kenneth E. (2013). Characterising the complexity of medication safety using a human factors approach: an observational study in two intensive care units. *BMJ quality & safety*, bmjqs-2013-001828.

Chen, L. F., Sexton, D. J., Kaye, K. S., & Anderson, D. J. (2009). Patient-days: a better measure of incidence of occupational bloodborne exposures. *Am J Infect Control, 37*(7), 534-540. doi: 10.1016/j.ajic.2009.02.004

Flynn, Linda, Liang, Yulan, Dickson, Geri L, Xie, Minge, & Suh, Dong-Churl. (2012). Nurses' practice environments, error interception practices, and inpatient medication errors. *Journal of Nursing Scholarship, 44*(2), 180-186.

Forbes, S, & Taunton, RL. (1994). Reliability of aggregated organizational data: An evaluation of five empirical indices. *Journal of Nursing Measurement, 2*(1), 37-48.

Giraudeau, B, & Mary, JY. (2001). Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Statistics in medicine, 20*(21), 3205-3214.

Harris, Paul A, Taylor, Robert, Thielke, Robert, Payne, Jonathon, Gonzalez, Nathaniel, & Conde, Jose G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics, 42*(2), 377-381.

He, Jianghua, Dunton, Nancy, & Staggs, Vincent. (2012). Unit-level time trends in inpatient fall rates of US hospitals. *Medical Care, 50*(9), 801-807.

Hughes, L.C., & Anderson, R.A. (1994). Issues regarding aggregation of data in nursing systems research. *Journal of Nursing Measurement, 2*(1), 79-101.

Kalisch, Beatrice, Friese, Christopher R, Choi, Seung Hee, & Rochman, Monica. (2011). Hospital nurse staffing: choice of measure matters. *Medical Care, 49*(8), 775.

Kalisch, Beatrice J, Tschannen, Dana, & Lee, Kyung Hee. (2012). Missed nursing care, staffing, and patient falls. *Journal of Nursing Care Quality, 27*(1), 6-12.

Lake, Eileen T, Shang, Jingjing, Klaus, Susan, & Dunton, Nancy E. (2010). Patient falls: association with hospital Magnet status and nursing unit staffing. *Research in nursing & health, 33*(5), 413-425.

Lee, James, Koh, David, & Ong, CN. (1989). Statistical evaluation of agreement between two methods for measuring a quantitative variable. *Computers in biology and medicine, 19*(1), 61-70.

Lerner, Nancy B. (2013). The Relationship Between Nursing Staff Levels, Skill Mix, and Deficiencies in Maryland Nursing Homes. *The health care manager, 32*(2), 123-128.

Li, Yu-Fang, Wong, Edwin S, Sales, Anne E, Sharp, Nancy D, Needleman, Jack, Maciejewski, Matthew L, . . . Liu, Chuan-Fen. (2011). Nurse staffing and patient care costs in acute inpatient nursing units. *Medical Care, 49*(8), 708-715.

Montalvo, I. (2007). The National Database of Nursing Quality Indicators (NDNQI). The Online Journal of Issues in Nursing, 12(3),

Nedved, Patricia, Chaudhry, Rabhea, Pilipczuk, Dina, & Shah, Shital. (2012). Impact of the Unit-Based Patient Safety Officer. *Journal of Nursing Administration, 42*(9), 431-434.

Simon, M., Yankovskyy, E., Klaus, S., Gajewski, B., & Dunton, N. (2011). Midnight census revisited: Reliability of patient day measurements in US hospital units. *Int J Nurs Stud, 48*(1), 56-61. doi: 10.1016/j.ijnurstu.2010.07.002

Simon, M., Yankovskyy, Y., & Dunton, N. (2010). Solving the mystery of patient days and the midnight census. *Nurs Manage, 41*(2), 12-14. doi: 10.1097/01.numa.0000368560.92652.66

Staggs, Vincent S. (2013). Nurse staffing, RN mix, and assault rates on psychiatric units. *Research in nursing & health, 36*(1), 26-37.

Staggs, Vincent S, & He, Jianghua. (2013). Recent Trends in Hospital Nurse Staffing in the United States. *The Journal of nursing administration*.

Twigg, D., Duffield, C., Bremner, A., Rapley, P., & Finn, J. (2011). The impact of the nursing hours per patient day (NHPPD) staffing method on patient outcomes: a retrospective analysis of patient and staffing data. *Int J Nurs Stud, 48*(5), 540-548. doi: 10.1016/j.ijnurstu.2010.07.013

Van den Heede, Koen, Sermeus, Walter, Diya, Luwis, Clarke, Sean P, Lesaffre, Emmanuel, Vleugels, Arthur, & Aiken, Linda H. (2009). Nurse staffing and patient outcomes in Belgian acute hospitals: cross-sectional analysis of administrative data. *International Journal of Nursing Studies, 46*(7), 928-939.

Wilson, Sally, Bremner, Alexandra, Hauck, Yvonne, & Finn, Judith. (2011). The effect of nurse staffing on clinical outcomes of children in hospital: a systematic review. *International Journal of Evidence-Based Healthcare, 9*(2), 97-121.