NORTHWESTERN UNIVERSITY

Quantifying Clinical Relationships and Patient Care Activities to Predict Patient Outcomes: An Edge-Weighted Multilayer Network Approach

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Health Science Integrated Program

By

Ariel Evelyn Reitzel Chandler

EVANSTON, ILLINOIS

MARCH 2020

Abstract

The application of formal multilayer networks (MLNs)—networks which contain multiple types of relationships, data types, or other additional features of complexity and connectivity—has recently become popular in many fields. Areas of study ranging from social science to biology have utilized MLNs to explore, describe, and analyze interconnected complex systems. MLNs have demonstrated both flexibility and practicality in investigating high dimensional data due to the structural ability to integrate different types of related data into one mathematical model. An exemplar of such data are patient care activity logs, produced by clinician interaction with the electronic health record (EHR), which contains data on clinicians, patient encounters, and the care activities performed as part of treatment. This has inspired the development of an MLN model which analyzes patient care activity logs, with the aim of evaluating clinical processes by identifying when differences in clinical relationships are systematically predictive of patient outcomes.

This thesis presents an applied MLN methodology to answer the following question: During which care activities are groups of clinicians with highly successful relationships most likely to impact patient health outcomes? Evidence is presented in three papers that supports the hypothesis that the applied MLN methodology accurately identifies both 1) highly successful clinical relationships among providers and 2) the areas of care most associated with those relationships and outcomes. The first paper explores the necessity and effects of risk adjusting patient outcomes to ensure accurate evaluation of clinical relationships. The second paper describes the MLN network in further detail and applies the proportion of categorized clinical relationships (as measured by risk adjusted patient outcomes) as edge-weights representing the

connection between patient care activities performed on encounters, with the aim of identifying tasks where differences in relationships are linked to patient outcomes. This model structure is verified with simulated data and validated with treatment and outcomes data of intracranial hemorrhage (ICH) patients. In the final paper, the MLN model is further refined with the aim of increased clinical interpretability. Further evidence of method validity is presented from the examination of Computed tomography (CT) notes for the documentation of communication with other physicians. These methods present a new approach to leveraging EHR data by applying an MLN framework to investigate outcomes data with patient care activity logs. The evidence presented in this thesis supports the future utilization of these methods in targeted process improvement investigations and interventions.

Acknowledgment page

This work would not have been possible without the financial support of the Biomedical Data Driven Discovery (BD3) Training Program at Northwestern University, the Health Sciences Integrated Program (HSIP) and Dr. Naidech's lab (R01 NS 110779). I am especially indebted to my committee, Dr. Nicholas Soulakis, Dr. Andrew Naidech, and Dr. Denise Scholtens. Their time and guidance were invaluable to this work.

I would also like to thank Dr. Justin Starren, Dr. Neil Jordan and Dr. Lucy Bilaver for their support within the HSIP program for allowing me to explore my research interests at my own pace and direction.

Lastly, I would like to thank my family and friends for all their support and love during these last four years.

List of Abbreviations

Area under the receiver curve (AUC)

Computed tomography (CT)

Emergency Department (ED)

Enterprise data warehouse (EDW)

Intercranial hemorrhage (ICH)

Intensive care unit (ICU)

Modified Rankin Scale (mRS)

Multilayer network (MLN)

Northwestern University Brain Attack Registry (NUBAR)

Out of bag (OOB)

Random Forest (RF)

Shared Positive Outcomes Ratio (SPOR)

Table of Contents

List of Tables and Figures

1. Introduction

1.1. Background. The expansion of clinical informatics studies has occurred in conjunction with the recent ubiquity of electronic health records (EHRs). Clinical informatics is the science of how to use data, information and knowledge systems to improve human health and the delivery of healthcare services.[1] While the study of information and data in medicine has always been of interest, clinical informatics research was limited before widespread provider adoption of EHRs due to a scarcity of data. This changed with the passage of the 2009 HITECH act, which earmarked $35 billion in incentive payments for hospitals and physician groups to digitize their records and adopt EHRs.[2] By 2018, 96% of hospitals had met meaningful use standards,[3] the minimum government requirements for EHR technology and data interoperability.

Now, for better or worse, EHRs are at the heart of healthcare delivery—becoming the central repository for a patient's medical team to enter and share information during and after patient care. This has led to the growth of health and medical data—increasing at an annual rate of 48% since 2013.[4] This data includes the digitization of patient medical information like free text notes and lab results that used to reside in paper record. However, the increase is also due in part from the data generated from routine EHR use. Every new entry of clinical information into the EHR by a clinician (for example, creating or fulfilling an order, form submission, or entering a lab value) is stored in the enterprise data warehouse (EDW) as part of the patient's medical record. This trail of passively generated data can be gathered from the EDW and processed to create a granular timestamped log of all patient care activities performed in the treatment of the patient. This is typically referred to as an activity log, and in addition to logging all care performed, also includes the individual clinician performing the task. The increase in EHR use and these two types of data now readily available—the digitized patient record, and the activity log of patient care—has given clinical informatics a huge trove of data, once only accessible from paper records (or never recorded, in the case of activity logs).

The primary use of EHR data is for billing and clinical care; when these data are used for research it is typically termed secondary EHR analysis. Research methods utilizing these data have struggled to keep up—where there was once too little data, there is now too much—with concerns of it being of poor and incomplete quality. Initially, the first wave of secondary EHR research utilized digitized health records with an approach rooted in more traditional practices—clinical information extraction to improve clinical knowledge.[5,6] However, the limitation of inconsistent quality of digitized records were quickly apparent in clinical research, which require highly accurate information.[7] In recent years, clinical informatics researchers have brought their attention to the activity log generated from EHR use. There is potential in research utilizing activity logs, especially in the areas of operations and process-orientated studies. This data is structured, timestamped, and generated passively as a byproduct of care delivery, which circumvents many of the human-generated clinical data quality issues in digitized records. [8]

Of specific interest, is work utilizing activity logs from EHR data to characterize clinical workflows, often using process mining or other data mining methods. This work has included both the identification of care teams and care processes through a variety of techniques. Process mining, popularized in operations research, are methods that aim to discover, monitor and improve real processes from available event logs in an organization's IT system.[9] The frequent applications of process mining in healthcare has been to develop a process model, discovering the order of activities, understanding performance issues (e.g. time variation in models, bottle neck events), and the detection of process deviations compared to a pre-determined model.[10] Most of these case studies had a narrow scope, and provided little applicability outside of a specific problem and healthcare institution. The other common application of mining methods is the identification of care teams from activity logs, for example identifying the primary care team from patterns of patient chart access.[11] Other research categorized the type of clinician (nurse, physician, etc.) that should be performing an activity, and identified activities frequently performed by non-concordant clinicans.[12] Methods to identify care teams which frequently collaborated have been explored,[13] and patterns of care associated with length of stay in the hospital have been identified.[13,14]

Past research of process mining in healthcare has primarily been focused on the business case[15,16] and the technical aspects.[17,18] These data mining techniques are valuable in automating a process that would be difficult to do manually due to the large amount of data. However, they also result in complex and often uninterpretable results. To make this research applicable for a clinically oriented audience it is important to, first, direct the mining processes towards a meaningful result (e.g. patient health outcomes), and equally important, inputting human logic to filter out noise that may be contributing to overly complex results. This human input need neither be cumbersome or non-reproducible; rather it can be built into the structure of data.

Networks are an intriguing methodological approach to improve the interpretability and data filtering barriers present in much of the workflow mining informatics research. While traditional networks consisting of one relationship type and one or two types of data are not adequate for describing the complexity of EHR data representing hospital care delivery, a multilayer network (MLN) is more applicable. MLN are networks which contain multiple types of relationships, data types, or other additional features of complexity and connectivity. Creating a network from multidimensional data requires that data are defined in relation to other data dimensions. The requirement of defining dimensions by their relationships to one another is well suited to describing hospital care—a highly complex system where clinicians are connected to patients via the treatment they provide in the form of specific care activities, and clinicians are connected to one another through the shared goal of working towards improving the health of the patients they treat. The interconnectedness of the data representing hospital care can be leveraged when structured within a network model; relationships can be measured within the context of other related types of data. Instead of simply knowing who and what happened to a patient in the hospital, we can layer on additional knowledge of clinician relationships, as measured by shared patient success over the timeframe of their working relationship.

Evidence suggests that identity of clinicians and how they communicate with others impacts patient outcomes. Increases in collaborative communication are associated with improved patient outcomes.[19-21] Improved outcomes in patients with stroke are associated with effectiveness, task orientation, order and organization, and utility of quality information.[22] Relationships between clinicians may impact health outcomes, however, the specific patient care activities that may be most affected by individual clinicians have not been defined. Likely, this is due to the lack of measures for quantifying or monitoring these impacts of communication in hospitals, as well as limited initiatives that collect comprehensive and high-quality patient outcomes. Studying relationships between clinicians is labor-intensive, so research to date has necessarily been limited to a few health care team members within specific scenarios. Past studies primarily have focused on nurses and physicians, which may not reflect full care teams that can include clinicians who do not directly interact with patients, such as radiologists, or staff that do not make clinical decisions but are still involved with patient treatment. Furthermore, most studies collected data for less than 6 months, which may not be sufficient for capturing the complex dynamics of clinical relationships, especially due to the irregular schedules of many hospitalists.

Retrieving and analyzing activity log data from the EHR, combined with patient outcomes, would permit study of relationships between clinicians and patient outcomes on a larger scale than has heretofore been possible. All clinicians who contribute to caring for the patient are included in activity logs, and relationships could be measured over the course of many years due to ease of data collection. As previously discussed, network methods to analyze audit log data have already been established for analyzing hospital processes and relationships. However, these methods have yet to be widely applied using patient outcomes. This inspired the hypothesis that by applying complex network models to data which includes patient outcomes, the effect of both patient care activities and relationships between clinicians on patient outcomes could be measured.

1.2. Scope of Research. This thesis describes the rationale for developing a complex network model where activity log data is structured by relationships between data types and can be transformed across dimensions using patient outcomes to identifying influential relationships and patient care activities. Across three papers evidence is presented which supports the validity and contributions of this methodology.

The first paper, *Risk Adjusting a Healthcare Provider Collaboration Network,* explored the necessity and effects of risk adjusting the patient outcomes used to measure relationships between clinicians, relationally to other connections within a collaboration network. The second paper, *An Edge-Weighted Multilayer Network Model to Characterize Clinical Relationships and Quantify Patient Care Activities*, conceptually and mathematically describes an MLN model which structures and analyzes audit log data and patient outcomes. The end result is patient care activities weighted by the relationships of the clinicians which performed them. These weights can be used to identify patient care activities where systematic differences in collaboration is predictive of patient outcomes. This MLN model was verified using simulated audit log data, modified over various parameters, and validated using audit log and outcomes data of intercranial hemorrhage (ICH) patients. The third paper, *Specific Patient Care Activities Performed by Clinicians with Significant Relationships Predicts the Outcomes of Patients with Intracerebral Hemorrhage*, investigates the impact of modifying the MLN model to be more clinically applicable and interpretable, including activity clustering and limiting the number of predictive variables. The model is further validated by exploring if documentation of communication in CT notes is associated with relationship scores.

The work proposed here brings together three disassociated fields of study: (1) clinical research demonstrating the connection between medical teams and health outcomes, and workflow and health outcomes; (2) methodological work analyzing hospital workflow via data mining on EHR data, also known as process mining; and (3) complex network models. Patterns of care, care activities, hospital workflow, and transactional logs will be used interchangeably and refer to all the healthcare activities performed in the treatment of a patient by individual healthcare providers (for example: ordering or completing a test, taking vitals, or writing a note). Network methods and process mining have yet to be applied to hospital care with patient outcomes, with the exception of the Shared Positive Outcomes Ratio (SPOR)[23]. This previous work provides a valuable foundation for the further study of network models to quantify and identify clinical relationships, and patterns of care associated with high and low rates of positive patient outcomes. A detailed description of previous work involving the SPOR can be found in part two of this thesis. While network models have been more fully explored in other domains, such as physics or the social sciences, they are conceptually well suited to represent complex hospital systems because they can easily incorporate relationships into the model, and link together large amounts of information. By applying previously established methods in workflow analysis and MLN frameworks to activity log data and patient outcomes, we can quantify and identify healthcare providers and care activities that may be related to health outcomes.

1.3. Statement of Contribution. This work contributes to the body of informatics research by presenting a new approach to structuring and analyzing EHR data that quantifies patient care activities and the relationships between clinicians, which can be used to identify specific care activities where differences in relationships are predictive of differences in risk adjusted patient outcomes. In addition, evidence outlined in this thesis that suggests that this method addresses dimensionality reduction. This work presents a new approach to utilizing EHR data to study healthcare delivery in a hospital by adapting advanced methods that have been untried for this purpose. In particular the approach leverages multilayered networks to model complex medical processes. This MLN model has the following properties:

(1) Data dimensionality reduction. It will capture relationships that involve multiple interacting variables.

(2) Temporal and associative. The model will capture associations between variables across various transformations of time.

(3) Interpretability. Simplified results can be obtained by layering the multiple data types into one measure, and through stepwise data filtering.

In addition to the clinical informatics contributions, this is also an innovative approach to quality and hospital operations improvement: (1) This is a data mining approach, which makes few clinical assumptions about which clinicians or what activities are related to patient outcomes. (2) These methods are scalable and rely on passively collected data. (3) Results and model output are actionable. Identifying outlier provider relationships is important, but by incorporating these measures onto granular patient care activities investigations or interventions to improving communication or teamwork practices can be precisely applied.

2. The Shared Positive Outcomes Ratio

The shared positive outcomes ratio (SPOR) is a pairwise metric that quantifies the ratio of specific positive outcomes shared between two providers versus outcomes shared with other providers within a collaboration network.[23,24] The section outlines the previous work SPOR calculation and methodology, which is foundational to this dissertation. Subsequent references to the SPOR metric throughout this dissertation will refer to this section. The following outlines SPOR calculation, equations for risk adjusting, and categorizing clinician relationships with the SPOR edge-weight. Additional details can be found in the original papers. [23,24]

2.1. SPOR Calculation. The SPOR value, with risk adjusted outcomes for a provider pair j, j', is defined as *equation 1*:

$$SPOR_{j,j'} = \frac{\sum_{A_j \cap A_{j'}} \mathrm{r_i(y_i)} \Big/ \sum_{A_j \cup A_{j'}} \mathrm{r_i(y_i)}}{A_j \cap A_{j'} \Big/ A_j \cup A_{j'}} \tag{1}$$

Where A_j and $A_{j'}$ are the sets of patient encounters involving providers j and j', $\mathrm{r_i(y_i)}$ is the risk adjusted outcome, and y_i is the actual encounter positive/negative outcome (1 or 0). The denominator measures the frequency of encounters shared between two providers relative to the frequency of all encounters shared with either provider.

2.2. Risk Adjustment. The actual positive or negative outcome related to the patient encounter is denoted as y_i, where $y_i = 1$ indicates a positive event occurred and $y_i = 0$ indicates a negative event occurred. The risk adjusted outcome $r_i(y_i)$ incorporates the difference between the actual outcome y_i and the probability of positive outcome p_i is defined as *equation 2*:

$$r_i(y_i) = \frac{1+(y_i-p_i)}{2} \tag{2}$$

The risk adjusted outcome $r_i(y_i)$ will have a high value (close to 1) if the probability of a negative outcome is high (p_i is close to 0) but a positive outcome occurs $(y_i = 1)$. On the other hand, $r_i(y_i)$ will have a small value (close to 0) if the probability of a negative outcome is low (p_i is close to 1) but the patient experienced a negative outcome $(y_i = 0)$. In both cases, events that were less likely to occur based on risk-estimates, but in fact do occur, are more heavily weighted in the SPOR score. This method of risk adjusting outcomes results in rewarding unexpectedly good outcomes, penalizes unexpectedly bad outcomes, and gives smaller rewards or penalties for expected outcomes. The purpose of the risk adjustment is to account for the variability in demographics or disease severity of patients shared between providers.

2.3. Using the SPOR to Characterize Connections. Provider pairs which only share a few patients could have extreme or skewed SPOR values, which are not meaningful. To filter out these relationships, different threshold values for the number of encounters required to constitute a collaborative relationship will be tested. The standard deviation within the SPOR population is measured as the threshold value increases and will choose a point at which increasing the threshold does not meaningfully reduce the standard deviation. When the standard deviation plateaus as the threshold increases, this indicates that extreme values due to a sparse number of shared connections (i.e. unstable relationships) are no longer impacting the distribution of SPOR scores in the network.

To identify which provider pairs have statistically significant high or low SPOR scores, the simulated random networks method is employed.[25] Provider-patient networks structurally identical to the actual network are generated, and outcomes associated with patient outcome nodes are randomized and used to generate new SPOR scores for each new network. In this dissertation, 3000 randomly generated networks are generated, with each actual provider pair SPOR weight compared to the 3000 randomly generated SPORs. The p-value for each provider pair was calculated by percent of the SPOR calculations on the networks with randomly assigned node labels are evaluated, which exceeded the actual SPOR value. The connection between each providers pair is then classified by the percentage of generated SPORs which exceed the value of the actual SPOR score: ≤ 0.05 indicates "high scoring", ≥ 0.95 indicates "low scoring", and between < 95 to > 0.05 indicates "average". This is equivalent to running two one-sided significance tests on the collaboration network, with the 5% cut off corresponding to a 5% p-value of significance.

3. Risk Adjusting Health Care Provider Collaboration Networks

3.1. Introduction

Severity of illness and socio-demographic features of individual patients affect unplanned hospital readmissions.[26,27] The quality of hospital discharge care by medical teams is also a key determinant of readmission rates.[28,29] After controlling for key causes of variance in outcomes, patients' health status and socio-demographic factors, the effectiveness of teamwork can be more accurately measured by examining the rates of unplanned hospital readmissions between providers involved with the discharge process.

Risk adjustment has been an important component of hospital metrics that measure the quality of health care delivery.[30] Risk adjustment of these measures is critical so that entities providing care for higher risk patients are not unfairly penalized for poorer outcomes even in the face of high quality care.[31] Unplanned 30-day readmissions to the hospital has emerged as a leading measure tied to incentives,[32] and has become a common metric to quantify health outcomes in academic literature. This literature has mainly focused on how accurately the models can predict readmissions at a population or hospital level.[33] However, distilling this population-level risk down to the individual patient level, the level at which such knowledge of risk may be most impactful for processes of care, is relatively understudied. To our knowledge, current literature examining risk adjustment in individual patient outcomes used to measure health care provider performance is extremely limited, likely due to limited methodology for network analyses examining this performance in relation to patient outcomes.[34,35] Here, the focus is on understanding the impact of individual-level readmission risk adjustment when quantifying relationships.

Networks are frequently employed in the study of teams and collaboration because data in a network is structured around relationships. In a simple network, “nodes” or “vertices” are connected via “edges” or “links”. Networks where two types of nodes are present are bipartite networks. Special types of networks, termed collaboration networks, are derived from bipartite networks by directly connecting primary nodes, which are linked via the secondary node. In the health care context of this research, a collaboration network links providers based on shared patient encounters (the secondary node).[36] The connections between the single type of node, in this case the links between providers, can have quantities or weights based on characteristics from the secondary connections. Collaboration networks have been used across many contexts, most notably in scientometrics or the study of science. Scientists can be linked together in a collaboration network by shared authorship on a publication, with a possible edge weight of how many journals shared.[37] Health care providers who enter activity on the same patient encounter can be linked through the patient in collaboration networks using EHR administrative data. Not only can edges be drawn based on shared patient encounters, but also data about each encounter are easily stored and analyzed through weighted network methods.[38] As health care provider teamwork and care collaboration becomes a priority for U.S. medical organizations,[39,40] it is critical to understand the collaborative network among health care providers to improve quality of care in this area. If outcomes used in weighting the edges of this network merit risk adjustment, understanding the implications of that adjustment will allow us to more accurately interpret these complex measures of teamwork.

Quantifying the connections or edges between providers in a network, through their shared encounters, has been previously developed: The Shared Positive Outcome Ratio. [23,24] The

SPOR is a pairwise metric that quantifies the ratio of positive patient outcomes shared between two providers, versus all the positive patient outcomes when each of those two providers work with others in the network. This measure could be used in quality improvement processes, however, to be reliable the role and impact of risk adjustment must be better understood. Here, the SPOR edge weighting method to a specific organization of patient care activities as a use case: the patient discharge process of a cardiology unit. Patient unplanned hospital readmissions will be used as a shared outcome for providers who take part in discharge activities.

Readmissions may occur regardless of the quality of care provided. However, readmission has been shown to be more prevalent when there is a poor quality of care coordination, or missteps in the discharge planning and transitional care process.[41] Thus the quality of teamwork may plausibly impact readmission rates. The effects of risk adjusting outcomes on a patient-encounter level will aid in the understanding of how providers are scored using this methodology. The SPOR method calculates a ratio of positive outcomes between each provider pair and can only be compared to other scores within the same network, and therefore it is not obvious that risk adjusting would have a significant effect on this relational score. If there is a significant subset of provider pairs that disproportionately see patients that are at a higher or lower probability to readmit, then risk adjusting would be appropriate. Does risk adjusting mitigate penalizing providers-pair connections who share patients who are at more risk for readmission? If so, is this effect realized in one or both categorizes of high and low scoring provider connections? Answers to these questions will help to understand the role risk adjustment plays at the individual level, when outcomes are used as ratios instead of the traditional method of simply reporting risk adjusted rates.

3.2. Objectives

This study aims to measure the effect of introducing risk adjusted outcomes to the SPOR weighted edge in a collaboration network of health care providers. To achieve this objective, a risk adjusted model was developed, modeling of 30-day unplanned readmissions for our patient population at the individual level. Those SPOR weighted edges were compared to unadjusted scores using permutation testing. To determine the structural differences of risk adjustment the following was examined: (1) The concordance between models in categorizing provider-pairs as high or low scoring (2) the individual provider variation in SPOR scores. Less variation could indicate that the risk model is accounting for patient characteristic effects on readmission. (3) Demographics and co-morbidities of individual shared patients of high or low scoring providers. Risk adjustment theoretically should equalize the prevalence of disease or socio-demographic characteristics of patient populations that link high or low scoring provider pairs.

3.3. Methods

3.3.1. Study Settings. The cardiology unit at Northwestern Memorial Hospital (NMH) is an urban and academic medical facility consisting of 36 beds (24 inpatient and 12 observational). Data was retrospectively collected data from the EHR of patient encounters that resulted in admission to the hospital inpatient cardiology unit between December 2011 and February 2016, from the NHM EDW. After excluding patient encounters with implausible data due to overlapping hospital discharge dates, the final dataset contains 6671 inpatient admission encounters from 4771 patients.

3.3.2. Readmission Risk Model Development. In this study, the success of provider collaborations was measured using a binary patient outcome indicating whether a patient encounter had a subsequent unplanned 30-day readmission. The hypothesis was that by adjusting the readmission outcome based on how likely the patient was to be readmitted would provide a SPOR score that more accurately measured the effect of providers working together on patient outcomes. This risk adjustment is intended to correct for provider pairs who disproportionately treated patients who were more or less healthy than other providers in the unit.[39]

A Weibull model was used to calculate the probability of an unplanned 30-day readmission, which employs a continuous probability distribution for modeling survival time or time-to-event.[42] The event of interest was the time to readmission (in days), calculated as the difference between the date of hospital discharge and the following readmission date. If a patient's record included a date of death during the study's observation time, they were described as censored. For patients with multiple hospital stays or unplanned readmissions during the study period, only the first instance of each was used in the model. The Weibull model included the survival function of $S(t) = \exp\,(-\lambda t^m)$ where $\lambda = \exp(-\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \cdots),$ and *m* is the shape parameter or Weibull slope. The Weibull survival model was developed using the R package *survival*,[43]*SurvRegCensCov* [42] and *rms* packages.[44]

A full model was fitted, using the following variables available from patient encounters: age, gender, race, ethnicity, primary insurance, secondary insurance, total charges incurred (quintiles), discharge disposition, total length of stay at hospital, total length of stay in cardiology unit, number of total activities performed on patient during hospital stay (quintiles), if the patient had previously been to the ICU (binary, yes/no):, patient had previously been to the emergency department (binary, yes/no), and if patient had any the following comorbidity (binary, yes/no), myocardial infraction, heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, peptic ulcer disease, diabetes mellitus, metastatic solid tumor, liver disease, hypertension. Variable selection was performed by using fast backward variable selection with AIC, and ANOVA testing.[42] A patient's 30-day survival probability (p_i) was calculated as $S(t = 30) = \exp(-\lambda 30_m)$. This survival probability corresponds to the probability of no unplanned readmission event occurring within 30 days after discharge, since the SPOR score represents shared *positive* outcomes. After estimating p_i, the risk adjusted outcome was calculated for each patient encounter, $r_i(y_i) = \frac{1+(y_i-p_i)}{2}$. Details of this equation can be found in section *2.2.* SPOR risk adjustment.

3.3.3. Shared Positive Outcomes Ratio. The SPOR (equation 1) can be found in section *2.1.* SPOR. A provider collaboration network was built and the SPOR edge weight was calculated. Two providers were linked within the network if each logged activity into the EHR for the same patient encounter between February 1, 2015 and January 31, 2016. The activities were limited to 38 that have been previously verified as being related to the discharge process from the cardiology unit.[12] A two-sided permutation test estimated by Monte Carlo simulation was used to measure if there was a statistical difference between the SPOR edge calculated between provider pairs with the risk adjustment and without. The function *permTS* from R package perm was used,[45] and the p-value was estimated from 10000 Monte Carlo replications.

3.3.4. Using the SPOR Score to Characterize Collaboration. Providers were categorized as high or low scoring collaborators based on the SPOR scores calculated according to each model. Using the simulated random networks method,[25] as described in the Section 2.3 SPOR Introduction, SPOR weighted relationships were categorized as "High", "Low", or "Average". This was done for each network.

To understand the effects of risk adjusting on the SPOR scores of individual providers, the variance of SPOR edges of all providers in each model was examined. Less variance in edge scores for an individual provider indicated that the model was more accurately scoring the collaboration of that provider. The difference in SPOR score variance was measured for each individual provider as the variance of edges calculated from the unadjusted model subtracted from the variance of edges in the risk adjusted model. A positive difference indicated that there was less variation in the risk adjusted model. To better understand the differences in how high and low scoring provider pairs were characterized between models, encounter-level data containing patient demographics and disease characteristics were examined.

3.3.5. Ethical considerations. Northwestern University's Institutional Review Board approved this study with a waiver of patient's informed consent under IRB # STU00088968.

3.4. Results

3.4.1. Readmission Model. The dataset to model readmission risk included a total of 6671 patient encounters, with 624 encounters resulting in readmissions. The parameter values and standard errors of variables associated with predicting unplanned 30-day hospital readmission were calculated using a Weibull model (Table 1). Insurance, which was effect coded in the model, uses Medicare as the reference group. This model estimated p_i, closer to one indicating a higher likelihood of no readmission within 30 days of discharge. The mean of p_i was

.88, and the standard deviation was .04. These parameter variables are consistent with previous literature.[26,46-48]

Factor	Covariate Effect	Standard Error	Lower 0.95 Interval	Upper 0.95 Interval
age	0.16	0.09	-0.02	0.33
Total Length of Stay	-0.09	0.01	-0.11	-0.06
Myocardial Infarction (Yes)	-0.31	0.22	-0.75	0.12
Heart Failure (Yes)	-0.37	0.06	-0.48	-0.26
Hypertension (Yes)	-0.20	0.10	-0.40	6.30×10^{-4}
Transferred from ED (Yes)	0.27	0.06	0.16	0.38
Insurance (Commercial**:** Medicare)	0.32	0.08	0.16	0.47
Insurance (Medicaid**:** Medicare)	-5.94×10^{-3}	0.09	-0.18	0.16
Insurance (Self-pay: Medicare)	0.61	0.18	0.26	0.97
Discharged to Home without Assistance (No)	-0.31	0.07	-0.47	-0.19

Table 1. Weibull model covariates estimate the risk of unplanned 30-day readmission.

3.4.2. Provider Collaboration Network. The network contained 651 unique providers, with 15,211 provider pair connections. The provider network that was limited to at least four shared encounters included 254 providers and 2359 provider pair edges. This is the network referred to in the remainder of this paper. The SPOR resulting from calculation using unadjusted readmission outcomes, y_i, and risk adjusted outcomes, $r_i(y_i)$ is in Table 2, and the SPOR distribution for each network is in Figure 1. A two-sided permutation test, estimated by Monte Carlo simulation, indicated that the mean of the SPOR values with risk adjusted outcomes were significantly different from SPOR values with unadjusted readmission as an outcome measure with a p-value of 0.0254.

	Min	25th percentile	Median	Mean	75th percentile	Max	Standard Deviation
Unadjusted SPORs (y_i)	0.29	0.90	1.00	0.99	1.11	1.29	.15
Risk adjusted SPORs $r_i(y_i)$	0.38	0.92	1.03	1.00	1.10	1.42	.13

Table 2. Summaries of SPOR edge-weights, calculated using unadjusted and risk adjusted outcomes.

For SPOR values with unadjusted outcomes, 118 (5.0%) provider pairs classified as high scoring, and 117 (4.9%) were classified as low scoring. In the risk adjusted network, 118 (5.0%) provider pairs were classified as high scoring, and 118 (5.0%) classified as low scoring. Of the 118 SPOR edges classified as high scoring in each network, 52 provider pairs were classified in both networks, and the remaining 66 were unique to each network. Of the 117 and 118 SPOR low scoring edges classified from the unadjusted and risk adjusted outcomes, respectively, 101 provider pairs were classified in both models, with the remaining 16 and 17 pairs being unique to

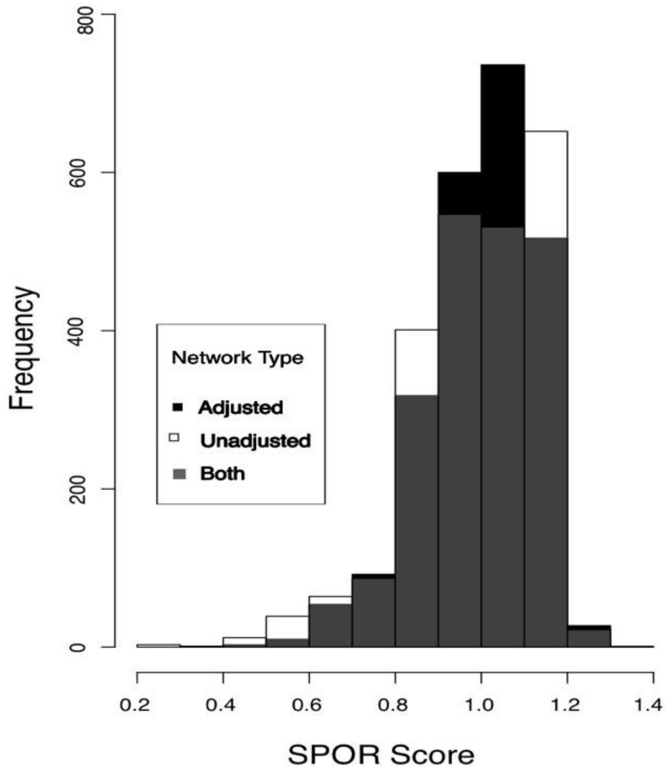

Figure 1. Distributions of SPOR edge-weights in unadjusted verse risk adjusted networks. The histogram represents the two networks' SPOR edge weight distributions overlaid on one another; the overlap is represented by the "both" category. Both networks have the same number of edge-weights.

each group. Examination of the difference in variance of SPOR scores for each individual provider revealed that the variance was reduced in the risk adjusted model (Figure 2). Values near zero indicated that the variance of SPOR scores between models were equal, while values above 0 indicated that there was less variation in the risk adjusted SPOR scores than the unadjusted scores.

To determine if risk adjusting affected the rate of disease and other relevant factors in high and low scoring provider pairs groups, the patient characteristics of shared encounters which linked these groups was examined (Table 3). All reported measures represent characteristics that tend to be associated with higher rates of unplanned readmissions, other than "discharged to home without assistance" and "commercial insurance", which are associated with lower rates. The unadjusted high scoring group had fewer patients with factors associated with readmission than the risk adjusted high scoring group. The risk adjusted high values, on average, have a larger percentage of patients with characteristics associated with unplanned readmission and the low scores have a smaller percentage of these patient characteristics versus the unadjusted scores.

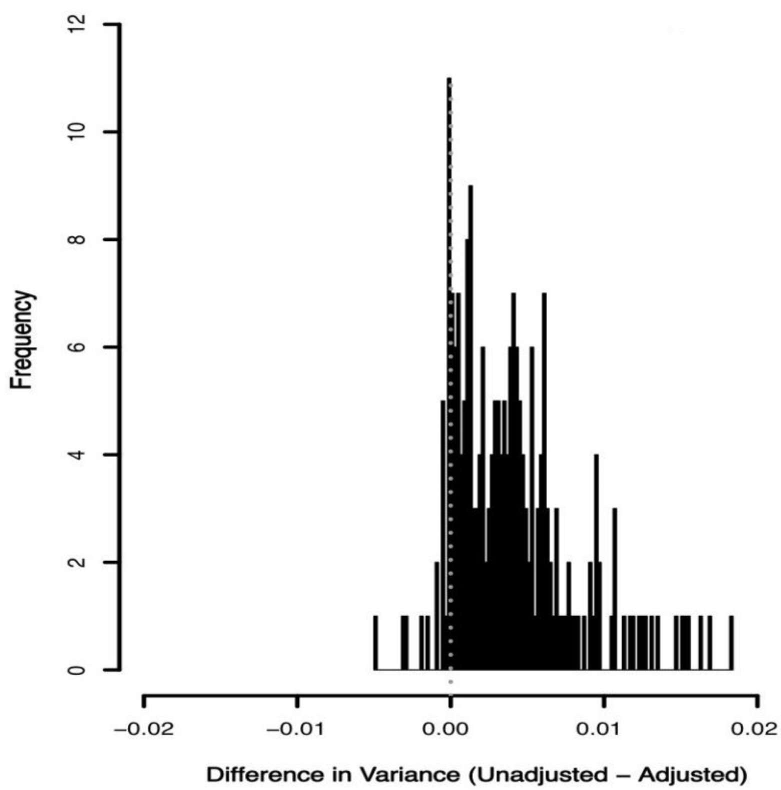

Figure 2. Difference in variance for individual SPOR edge-weights between unadjusted and risk adjusted networks. The variation difference was measured for each individual provider as the variance of scores calculated from the unadjusted model subtracted from the variance of the scores in the risk adjusted model. Dotted line indicates zero variance.

	Demographics of shared patients for providers with the following categorized relationships:			
	High SPOR risk adjusted	High SPOR unadjusted	Low SPOR risk adjusted	Low SPOR unadjusted
n=	209	293	326	324
Heart failure*	33%	29%	28%	31%
Myocardial infarction*	22%	20%	25%	26%
Peripheral vascular disease	21%	17%	23%	22%
Cerebrovascular disease	26%	20%	33%	35%
Chronic pulmonary disease	24%	25%	29%	27%
Diabetes mellitus	35%	38%	41%	42%
Hypertension*	72%	70%	73%	73%
Admitted from ED*	14%	17%	18%	17%
ICU before cardiology unit	58%	51%	40%	41%
Commercial insurance	24%	25%	19%	19%
Medicaid insurance	17%	15%	14%	15%
Medicare insurance	57%	59%	66%	65%
Self-pay /no insurance	1%	1%	0.1%	.1%
Discharged to home no assistance*	36%	46%	45%	43%

* indicates measures used in the risk adjusted positive patient outcome model.

Table 3. Demographics of unique patient encounters for provider in each network, compared across high and low SPOR categorized edge weights.

3.5. Discussion

In this paper, a methodology and test case are presented to accurately measure the care coordination in a specific workflow, through risk adjusting edge-weights in a collaboration network. This method has many hospital quality and patient safety applications, including identifying provider groups for testing improvement strategies, measuring new processes, and identifying structural issues or informal collaboration practices affecting patient outcomes. Upon examination, the SPOR scores calculated by each model are clearly different. However, this does not indicate that risk adjusting the outcome measures is necessarily a better approach. Traditional ways of testing our readmission model cannot be appropriated in this analysis since readmissions rates are being used in a more obscure manner, with the SPOR edge being calculated as different rates being compared in ratios. However, by exploring the different providers each model categorized as high or low scoring, the evidence strongly suggests that the risk adjustment of outcomes is a critical addition to the calculation of SPOR scores. From the distribution comparison and summary of the risk adjusted and unadjusted SPOR weights that the risk adjusted network is distributed more normally than the unadjusted network. More meaningful differences are also found in the tails of the distribution of SPOR scores, which matters because these are the values that represent high or low scoring provider pairs. The provider pairs being shifted away from the tail ends of the risk adjusted SPOR distribution—providers that otherwise may be scored higher or lower resulting from pre-existing risk in their patient population.

The practice of risk adjusting health outcomes tends to be driven by a desire to not penalize health care providers or organizations that care for sicker patients. By this reasoning, there should be less similarity between the risk adjusted and unadjusted low scoring SPOR groups, however, the opposite was observed. Nearly 86% of the low scoring provider pairs were categorized as such by both models. The two SPOR models overlapped in their categorization of high scoring provider pairs only 44% of the time. The provider pairs with shared patients that did not experience unplanned remissions, but had a high likelihood of such an event, highlighted the provider-pairs and individual providers who contributed most to positive outcomes.

After the introduction of risk adjusting SPOR scores a greater percentage of patients with characteristics associated with unplanned readmissions was observed in the high scoring group than the low scoring group. Even characteristics that were not directly associated with readmission in our model, such as cerebrovascular disease, increased in the patients which linked high scoring provider pairs. This over-correction was likely due to a large number of unexpectedly positive outcomes in these patients, whose risk adjusted probability indicated that they were likely to be readmitted to the hospital within 30-days. Therefore, this trend underscores the value of risk adjustment—providers are "rewarded" in their SPOR scores by treating patients who are more likely to have an unplanned readmission. While this conclusion does not provide evidence that the interaction during the discharge between high scoring provider pairs is directly responsible for this reduction in readmission, it does indicate that further research should be conducted to investigate the causes of this finding.

The risk adjustment also tended to either reduce the individual variance or have little effect on individuals' SPOR scores. Risk adjustment may not have had an effect due to the relational nature of the SPOR weighted edge—while individual providers may overall treat patients that are more or less likely than average to be readmitted, it is less likely that they share these patients in only one or a few of their relationships. However, many providers did have a reduction in variance, indicating that the risk adjusted model is more accurately calculating the individual collaboration scores of providers by accounting for variation in SPOR weights driven by patient characteristics associated with readmission.

3.5.1. Limitations. There are many data quality difficulties using EHR data in secondary data analysis.[7] Basic demographic data that may help to determine risk at a population level cannot be used in our study due to incompleteness of the EHR. For example, race and ethnicity has previously been shown to predict unplanned readmission rates for patients with heart failure and myocardial infarction.[26] However, these variables in our risk adjusted model; over 30% of race was coded in an unusable way (i.e. other, declined, unknown, unable to answer). Ten percent of ethnicity was coded in this way compared to 7% indicating that the patient identified as Hispanic or Latino, making these variables too unreliable for inclusion.

3.6. Conclusion

Risk adjusting unplanned readmission effects SPOR-weighted edges in a collaboration network, with the greatest difference seen in how high scoring SPOR provider pairs are classified. The risk adjusted model reduces the variance in SPOR weights of providers' edges. The risk adjustment model also adjusts shared patient characteristics that are associated with readmission in high and low scoring provider pairs more evenly. This indicates that the risk adjusted SPOR edges are measuring the impact of collaboration on readmissions, accounting for patients' risk of readmission. Based on these findings, rigorous risk adjusting should be considered when implementing the SPOR-weighted edge methodology.

4. An Edge-Weighted Multilayer Network Model to Characterize Clinical Relationships and Quantify Patient Care Activities: Model Description, Verification and Validation

4.1. Introduction

Multilayer networks (MLN) — networks with multiple dimensions, components, or features—are increasingly being applied to analyze complex systems, due to advances in network theory and availability of big data.[49,50] A MLN structure can be developed that describes dynamic and varying types of connections between heterogeneous data types, which compared to a simple network more accurately mirrors the interactions among the components of a real-life complex system. Once a network model is applied to data representing an interconnected system, dimensionality can be reduced by aggregating data types in relation to one another as an edge-weight, whose value represents the aggregated data. Patient care activity logs, produced as a data byproduct of treatment when clinicians interact with the EHR, is an exemplar of high dimensional heterogeneous data an interconnected complex system described. In hospitals, clinicians are connected to patients via treatments, and then connected to other clinicians through the shared treatment of patients. The interconnectedness of data representing patient care in healthcare systems is well suited to the application of an MLN framework that aims to measures both patient care activities and clinical relationships.

Previous research applying networks to healthcare system data has for the most part utilized simple networks, which contain only one or two types of data, with the aim of describing patterns in teams or patient care activities.[51] The approach of past work involved traditional network metrics (i.e. connectedness, centrality, etc.) and has not differentiated care patterns based on the outcomes of patients. To the best of our knowledge, there are no studies which explicitly apply a complex network framework to healthcare operations data, especially in the study of patient outcomes, clinical relationships and patient care activities. However, work foundational to this research has informally applied a complex network framework to measure and categorizing clinical relationships, based on the SPOR.[23,24] This method models relationships of clinicians from a log of EHR activity and patient outcomes, through the logical transformation of network layers in a manner, resulting in complexity reduction while retaining valuable information. This approach to measuring clinical relationships capitalized on the recent ubiquity of EHRs, where clinician-computer interaction generates a log of all patient care activities. This activity log represents both a granular and broad recording of treatment a patient received during a hospital stay. This log was utilized to connect clinicians to patients through patient care activities, construct medical teams through shared patients, and finally measure relationships between clinicians using the outcomes of those patients. The work in this paper builds on this past research, through the explicitly incorporation into an MLN framework, which uses the categorization of signification clinical relationship weight the care activities those clinicians performed.

Instead of associating patient outcomes with only one dimension of data—the care a patient received in the hospital—additional facets of information can be incorporated, like the clinical relationships of those performing the care activities. In this research, multiple interconnected data types are analyzed within an MLN framework, and the model is verified and validated using simulated and actual activity log data. This research validates the outlined MLN framework using the actual activity logs associated with the care of patients with intracranial hemorrhage (ICH), however, this methodology could be widely applied to logs of any patient population with associated outcomes data.

4.1.1. Overview of Multilayered Networks.

Networks can be applied to represent complex and dynamic systems, which can encompass many different types of nodes and edges, which need not stay static (i.e. time can be involved). These networks are called complex, multilayer, or multiplex networks. Multilayered network data models are highly logical and intuitive methodology of structuring data, and foundational mathematical concepts in graph theory and associated algorithms can be applied to complex networks illuminate structures and patterns, reduce complexity or transform the data through network aggregation.[52-54] While the complex or multilayered network framework is flexible and can be used and defined in a variety of ways, only the underlying network definitions that inform the scope of this will be defined here. Details in the application of this network structure to the data schema will be described in section 4.2.1.

A simple graph is defined as tuple $G=(V,E)$ where V the set of nodes and the edges are defined a tuple of nodes, $E=(u_1, u_2)$ given $u_1, u_2 \in V$. Complex graphs can represent this increased dimensionality with both aspects and layers. While both aspects and layers can be defined flexibly in relation to the data and problem scope, in the research outlined in this paper, the aspects of the network are defined as node types and layers are defined as a possible state of the system. A state is defined as a subset of aspects in the network or a newly defined relationship between the aspects present. One can imagine that each layer of a multilayer network represents one configuration of node groups, so that the entire network is different combinations of entity groups described by an ensemble of states. Every distinct combination of a network's aspect, α , and a layer, δ, is described as an elementary layer, $L = \{L_\delta\}_{\delta=1}^{\alpha}$, such that there is one set of elementary layers L_δ for each aspect α. The multilayer representation of the network $G=(V,E,L)$, where L is the set of elementary layers $L = \{L_\delta\}_{\delta=1}^{\alpha}$, V the set of nodes and E consists of triples (u_1, u_2, d), with $u_1, u_2 \in V$ and $d \in L$. Define $V_M \subseteq V \times L$ as the subset that contains only the node-layer combinations such that a node-layer tuple $(v, l) \in V_M$ if and only if v is present in layer L_δ^M. Let $E_M \subseteq V_M \times V_M$ be the subset of edges between node-layers. In the case of a weighted multilayer network, $G=(V,E,L,w)$, the edge is definition as $E=(u_1, u_2, d, w)$, where w is the weight on the link between nodes $u_1$and u_2 in the elementary layer d, given $u_1, u_2 \in V_M$ and $d \in L_\delta^M$.

To describe distinct data instances in a network containing multiple data types (i.e. unique keys in datasets are described by combinations of two or more variables), multiple nodes types need to be connected by a single edge. Hyperedges are edges that can connect any number of vertices (instead of being restrained to only two nodes). Hyperedges can be thought of as a subset of nodes within a network. Hyperedges are also referred to as an incidence set, which in the context of this work is the most applicable: each edge represents unique collection of nodes, which also correspond to unique observations or instances in the data. Therefore, in a weighted multilayer graph with hyperedges is defined as $G = (V, \mathcal{E}, L, w)$ a hypergraph with the vertex set V, L set of layers, and the hyperedge set $\mathcal{E} = (\{u_1, u_2, u_3\}, d, w)$.

4.2. Methods

This MLN framework allows for the incorporation data on patient outcomes, clinical relationships and patterns of care into a lower dimensional space, which should increase interpretability of model output and aid with variable selection. This model weights activities by clinicians' relationship scores, which represent a clinician's proportion of significant successful or unsuccessful relationships on an encounter. To confirm that the MLN model weights activities as expected, data is simulated to weight activities in section 4.2.2. Model verification. By weighting activities by relationship scores, specific care activities can be identified that are predictive of patient outcomes, which is described in section 4.2.3. Model validation. These identified activities represent areas of inpatient care where variations in clinical relationships are linked to with differences in outcomes.

4.2.1. MLN representation of Activity Logs. The aim of this approach is to conceptually and mathematically define the relationships between data types in hospital activity logs, as the network is transformed across these data types. The MLN framework is described by the relationships between network aspects (data types) in each network layer (new configuration of the network), and the relationship between network layers. Each network layer is constituted by a hyperedge set, defined by the network aspects present in the layer and in relation to adjacent layers. The hyperedge set is initially constructed from activity log data with four possible aspects: clinician, activities, time, patient (Table 4). The hyperedge set from this first layer is transformed through six stepwise adjacent layers (Table 5). The MLN model is conceptually defined in three parts: (1) initial construction of the hyperedge and node set from empirical data (activity logs); (2) edge-weighting relationships between clinicians with shared patient outcomes through aggregating the network; (3) edge-weighting patient care activity on encounters using clinician relationships by expanding the network.

Aspects in network layers:	**Aspect Definition**	**Associated attributes**
Clinician, *c*	An individual clinician	1.Position (i.e. nurse, resident)
Activity, *a*	A patient care activity	1. Type of activity (i.e. note, patient care) 2. Status of activity (i.e. ordered, performed)
Time, *t*	The moment when a patient care activity was performed	1. Date 2. Time
Patient, *p*	An individual patient encounter	1. Patient's outcome 28 days after discharge 2. Disease severity for risk adjustment 3. Demographic information (i.e. age, race)

Table 4. Network aspects, defined as different types of data, from the activity log

Lδ	Lδ definition	Hyperedge set, Ɛ, in each layer L_{δ}^{α}	Hyperedge weights
L1	Activity log layer	$\{L_1^c, L_1^a, L_1^t, L_1^p\}$	None
L2	Bipartite clinician-encounter layer	$\{L_c^2, L_p^2\}$	None
L3	Outcome weighted clinician layer	$\{L_c^3\}$	See *equation 1* in section *2.1*
L4	Relationship score weighted bipartite clinician-encounter layer	$\{L_4^{\mathrm{c}}, L_4^{\mathrm{p}}\}$	See *equation 3* in this section
L5	Relationship weighted clinician-activity-encounter layer	$\{L_5^{\mathrm{c}}, L_5^a, L_5^p\}$	See *equation 4* in this section
L6	Relationships weighted clinician -activity layer	$\{L_6^a, L_6^p\}$	See *equation 5* in this section

Table 5. MLN layers, hyperedge sets for each layer, and the relationships between layers.

The hyperedge and node set is initially defined in the first MLN layer from a patient care activity log. The activity log represents timestamped entries generated from the clinician-EHR interaction involved with patient care. Each log entry consisting of a unique instance across the four data types and each hyperedge in this first layer connects one of each node type and is defined by a log entry. These data types are: (1) a clinician, who performs (2) a care activity, at a certain (3) time and date, as part of the treatment of (4) a patient. In addition to these defined relationships, each node type has associated attribute data, which remain unchanged across all layers (Table 4). Job titles are attributed data for clinician node types —like nurse or resident. Activities have group types (i.e. patient care, notes, radiology). Associated with patient node types are attributes data on health outcomes, demographics, and disease severity.

Edge-weighting relationships between clinicians using shared patient outcomes is defined by two network layers, and is a MLN formalization of previously published work.[23,24,55] In the second network layer, the granular activities log network is aggregated to a bipartite network layer, where clinicians are connected directly to patients and adjacent to other clinicians. This represents the medical team for each patient encounter. The bipartite layer is then aggregated to a third network layer containing only clinicians, whose relationships are weighted by the SPOR metric, an aggregate measure of their shared patients' risk-adjusted outcomes. A detailed description of calculating, risk-adjusting and categorizing the SPOR can be found in the previously cited work, but the main value of this step is twofold. The SPOR characterized relationships requires the pair to exceed a minimum of shared patients, and therefore represents only clinicians who work together frequently. Additionally, each SPOR weighted relationship is measured relative to the network and therefore can be categorized into three groups, by comparing to randomly generated networks: two significant groups, the top and bottom 5% and then rest average (90% of connections). This labeling further simplifies the network by highlighting only the most important connections and represents curated and compressed information from the activity log.

The information about who has significant high or low levels of success when working together was distributed onto connections between other data aspects in the MLN model through three network layers. First, the categorized relationships between clinicians was transferred to weight individual clinicians on patient encounters on Layer 4. Through Layer 2, the bipartite layer of clinicians and patient encounters, clinicians were connected through individual encounters. We then determined how many of those connections have SPOR weights (indicating they are frequent collaborators) and which were significant (high and low) from Layer 4. The proportion of significant relationships was measured by the relationship score (Equation 1): the proportion of only high or low relationships on an encounter, or both can be incorporated by subtracting the number of high relationships from the low and dividing by the total number of SPOR categorized relationships. The relationship score, *w,* using both the high and low SPOR, of edge Ɛ =({c_i, p_i}, *d*, *w*), where $d \in L_4$, for clinician c_i and patient p_i, given $(c_i, p_i) \in \{L_2^c, L_2^p\}$, is defined as *equation 3:*

$$\mathcal{E}\left(\{L_4^c, L_4^p\}, w\right) = \frac{\sum_{(c_x,p_i)\in\{L_2^c,L_2^p\}} SPOR_{c_i,c_x}(high) - \sum_{(c_x,p_i)\in\{L_2^c,L_2^p\}} SPOR_{c_i,c_x}(low)}{\sum_{(c_x,p_i)\in\{L_2^c,L_2^p\}} SPOR_{c_i,c_x}} \quad (3)$$

How the relationship score is calculated reflects whether activities should be quantified only by high or low scoring relationships, or both should be included. These three methods of calculating the relationship score were tested, outlined in 4*.2.3.*

Model Validation. The relationship scores were then applied in layer 5 to each hyperedge, distinct by the activity, clinician, and patient encounter aspects. The relationship weighted activity, w, of hyperedge $\mathcal{E}=(\{c_i, a_i, p_i\}, d, w)$, where $d \in L_5$, for clinician c_i , activity a_i, and patient p_i, given $(c_i, a_i, p_i) \in \{L_1^c, L_1^a, L_1^p\}$, is defined as *equation 4:*

$$\mathcal{E}\left(\{L_5^c, L_5^a, L_5^p\}, w\right) = \frac{\mathcal{E}\left(\{L_4^c, L_4^p\}, w\right) \subseteq \left\{(c_i, a_i, p_i) \in \{L_4^c, L_4^a, L_4^p\}\right\}}{\sum \exists (c_i, a, p_i) \in \{L_1^c, L_1^a, L_1^p\}} \times \frac{\sum (c_i, a_i, p_i) \in \{L_1^c, L_1^a, L_1^p\}}{\sum (c_i, a_i) \in \{L_1^c, L_1^a\}} \tag{4}$$

The clinician's relationship score on that encounter was evenly distributed across the distinct activities they performed. The relationship score was distributed in this manner because if a clinician only performs few distinct activities on an encounter, compared to many, there is a greater likelihood that one of those activities is related to the relationship score. The relationship score for each of the clinician's distinct activities was adjusted by the proportion of the activity the clinician performed on that encounter, resulting in a distributed and adjusted relationship weight for each distinct activity a clinician performed on an encounter.

In Layer 6 the network's aspects were reduced to include only activities and encounters. The clinician-activity specific relationship weight was adjusted by the proportion performed by the clinician so the relationship weights of the same distinct activity could be summed for each encounter. The relationship weighted activity, w, of hyperedge $\mathcal{E}=(\{a_i, p_i\}, d, w)$, where $d \in L_6$, for activity a_i and patient p_i, given $(a_i, p_i) \in \{L_5^a, L_5^p\}$, is defined as *equation 5:*

$$\mathcal{E}\left(\left\{L_6^a, L_6^p\right\}, w\right) = \sum \mathcal{E}\left(L_5^a, L_5^p\}, w\right) \subseteq \left\{(c_i, a_i, p_i) \in \{L_5^c, L_5^a, L_5^p\}\right\} \tag{5}$$

This resulted in a total relationship weight for each distinct activity/per patient encounter. The final relationship weighted activity dataset contained a single value per encounter for each of the distinct activities in the activity log. The value of each relationship weighted activity either was the sum of all clinicians' activity scores for that activity, or zero. No value could indicate the activity was not performed, was only performed by clinicians with relationship scores of zero, or multiple clinicians performing the activity had relationship scores which cancelled out.

4.2.2. Model Verification. The purpose of model verification was to ensure that the model performs as intended.[56] To verify the MLN model, activity log data were simulated by modifying the parameters of the data corresponding with terms in *Equation 2,* which weighted activities using relationship scores. These include: the number of high or low scoring relationships a clinician has (*relationship score*), the number of distinct activities performed on a patient encounter, and the percentage of a distinct activity performed by a clinician (*activity count* and *activity percentage*, respectively). Each parameter's rationale, upper and lower limits and expected trends on the weights of activities are described under section 4.2.2.1. Simulation Parameters. The simulation parameters' upper and lower limits are also succinctly described in Table 6, which includes each variable name, description, ranges to be simulated, and expected effects on weighting activities. A detailed description of simulating patient care activity logs is outlined in section 4.2.2.2. Simulation Methodology.

4.2.2.1. Simulation Parameters. The relationship score is the most critical parameter in how patient care activities are weighted, with the other two activity parameters modifying the application. The relationship scores of clinicians in the network are limited by the total number of significantly categorized relationships; 10% of relationships are categorized as high and low, accumulatively. Therefore, the relationship score is directly related to how clustered the significant relationships are among individuals. On the lower limit, significant relationships are evenly distributed among all clinicians and most care activities would be weighted by relationship scores close to zero. This would lead to all activity weights clustered around zero. This should hold true regardless of the activity count and activity percentage parameters. On the high limit, approximately 10% of clinicians would have 100% of significant SPOR connections. The activities of these clinicians would have an outsized influence, leading to an uneven distribution with some activities' relationship weight close to -1 or 1, though most activities would have no relationship weight. Due to the set limit of high and low SPOR categorized relationships in a dataset, the relationship score parameter is adjusted by modifying both the percent of clinicians with significant relationships and the number of significant connections each of those clinicians possess.

Parameter name	**Variable Description**	**Modification Description**	**Expected effect on activity weight**
Relationship score	A measure of how clustered or spread out significant (i.e. high or how categorized SPOR connections) are among clinicians.	**Lower Limit:** Evenly distribute high and low categorized relationships across all clinicians in the network. **Higher Limit:** Cluster high and low categorized relationships among approximately 10% of clinicians.	As the value changes from evenly distributed to maximally clustered, the variation in relationship weighted activities should increase.
Activity count	A measure of how many other distinct activities a clinician performed (per encounter).	**Lower Limit:** 100, which indicates that the clinician performed 100 distinct activities on the encounter. **Higher Limit:**1, which indicates that the activity was only activity performed by a single clinician.	As the activity count changes from the lower limit (many activities) to the upper limit, the variation in relationship weighted activities should increase.
Activity percentage	The percent of an activity (per encounter) performed by one clinician.	**Lower Limit:** 10%, which indicates that the clinician only performed 10% of a distinct activity **Higher Limit:** 100% (1), which indicates only one clinician performed the activity on a patient encounter.	As the activity percentage increases (i.e. a clinician does a greater percentage of an activity on an encounter) the variation in relationship weighted activities should increase.

Table 6. Parameters to be modified to verify the MLN model weights activities with relationship scores as intended.

The activity count is one of two activity patterns that modify the application of the relationship score to the edge-weight of an activity. Both activity patterns in this study will be investigated by generating individual activities that represent various ranges of the parameter. The activity count represents how many distinct activities a clinician is performing on an encounter and is important because the value indicates how weight should be assigned to any single activity performed by a clinician. If a clinician with a high relationship score performs many activities on an encounter, each activity receives a proportional piece of the weight. However, if a clinician only performs one activity, it is very clear that the one activity is very important and receives all the weight. The lower range of the activity count parameter in data simulations is 100, where all clinicians performing that activity, are also performing 99 other activities on an encounter. This would greater minimize the relationship scores applied to activities, and most would be weighted near zero. This upper limit is when a clinician only does one activity, which would lead to a large variance of relationship weighted activity values.

The activity percentage is the other parameter in which relationship weighted activities are modified. This parameter considers what percent of a distinct activity was performed by an individual clinician on an encounter. This variable is important because it allows for an aggregate of relationship edge values across instances of activities of an encounter. The minimum value of this parameter will be 10%, which indicates that a clinician only performed a tenth of the activity on an encounter. This could indicate the clinician only performed the activity once out of a total of ten times, or ten times of a total of 100 times. While this parameter could be infinitely small, for simplicity sake, this will be the minimum value. The maximum value is one clinician performing all instances of the activity on the encounter, or 100% of an activity.

4.2.2.2. Simulation Methodology. Randomized values for each of the three parameters was simulated based on the activity log described in section 4.2.3.1. Data Sources. Each parameter was simulated for three separate distributions: an upper and low limit, and randomized. Three datasets containing the three distributions of relationship scores were generated, containing 27 simulated weighted activity distributions. Each simulated weighted activity distribution represented one possible combination of the three modified parameters (L=low limit, H =high limit, R=randomized) of activity count and activity percentage (LL, LR, RL, RR, LH, HL, HR, RH, HH). In total, 27 simulated activities were generated across all possible combinations of the three parameters, in all possible combinations. The relationship weighted activities were assessed by the increase or decrease in variance, since increased variance demonstrates increased relationship weights were applied to activities.

4.2.3. Model Validation. Activity log and patient outcome data from patients with ICH were used to validate the model. The activity log data was processed in six different ways, across three parameters, (described in section 4.2.3.2 and outlined in Table 7) to test if quantifying patient care activities by relationship scores accurately predicted patient outcomes in Random Forest (RF) modeling.

4.2.3.1. Data Sources. We utilized the Northwestern University Brain Attack Registry (NUBAR) [57] as a source of patient outcomes and risk-adjustment data to validate the MLN model. NUBAR is a prospective registry of patients with intracerebral hemorrhage (ICH, bleeding into brain tissue). Patients were diagnosed with ICH by computed tomography (CT), interpreted by a board-certified neurologist. All patients with ICH were admitted to a Neuro/Spine ICU in the context of a certified stroke center that predated the existence of

NUBAR by several years, and a high-intensity model of intensive care (e.g., all patients were attended to daily by an intensive care team). All patients or a legally authorized representative provided written informed consent for the use of EHR, with the exception of patients who were permanently comatose, died, or who were not consented and had no available legal representative, in which case the Institutional Review Board granted a waiver from informed consent.

Only patients with ICH scores of 0 to 2 were included in analysis predicting outcomes. Patients with ICH scores of 3 or greater were highly likely (96.5%) to have a poor outcome (disabled or death) at follow-up, indicating patient care activities had little to no impact these patient outcomes. Data on patient outcomes, we well as the data used for risk adjustment (ICH scores) were obtained from the NUBAR registry. Transaction logs of patient care activity for patient encounters that were included in the NUBAR registry were retrospectively collected data from the hospital's EDW. A neurologist coded the patient care activities deemed non-essential (e.g., "discontinue suction tube") or not performed by humans (e.g., an automatic alert from a laboratory computer that an order was completed). Those not performed by humans were excluded from models, non-essential activities were also excluded from the activity log in the four models indicated.

4.2.3.2 Data Preprocessing. Data was processed in 3 different manners to test the impact on accuracy of predicted patient outcomes: (1) excluding expert coded non-essential activities; (2) using activity data unweighted by relationship scores (count of an activity on each encounter, and whether an activity occurred or not on an encounter); (3) calculating relationship scores using either significantly high or low categorized relationships, or both (see relationship score description in section 4.2.1.). The baseline model was the relationship weighted activity model, where the relationship score was calculated using both high and low categorized relationships and the activities coded as non-essential were excluded. This was compared to five other models. The first was a relationship weighted activity model where activities coded as non-essential were included, and the relationship score included both high and low scores. All other models excluded activities coded as non-essential. The second and third models were relationship weighted activities models where the relationship score was calculated using only high or low categorized relationships, respectively. The fourth and five comparative models were activities unweighted by relationship scores, processed in two different manners. In the activity count model, the frequency at which a distinct activity occurred was counted for each encounter. In the binary activity model, whether an activity occurred or not on an encounter was coded (yes=1, no=0).

4.2.3.3. Predicting Outcomes. To test if the outcomes of patient encounters could be accurately predicted using relationship score weighted patient care activities in the MLN model, outcomes were predicted (positive outcome = 1) in random forest (RF) models, using the six datasets outlined in section 4.2.3.2 and Table 7. There were three objectives of predicting patient outcomes across these data. First, we wanted to know if the labor-intensive step of an expert coding essential activities is a necessary data cleaning step in future applications. Second, by comparing relationship weighted and unweighted activities the additional predictive power of aggregating multiple dimensions of data could be quantified. Additionally, the top predictors could be used for further clinical investigation of specific care activities, given the model is accurate.

An RF algorithm was chosen for this study for a number of reasons. First, the methodology allows for processing high-dimensional non-parametric data, while not over fitting the model to the data. Also, RF returns a measure of the importance for each predictor variable, that makes results more interpretable. Lastly, a RF model handles collinear data, by "spreading" the variable importance across all the variables (instead of discarding collinear variables), which is important since many activities are likely to be performed in tandem. [58]

The RF model in this study was fitted using 10-fold cross validation to minimize error and variable section maximized AUC (area under the ROC curve), using a backward elimination process based on the initial ranking of the variables.[59] To compare the performance between models the following was reported: AUC, the out of bag (OOB) error rate, sensitivity, and specificity, number of variables in the model.

4.3. Results

The distributions of the 27 simulated relationship weighted activities for each combination of the three modified parameters can be found in Figure 3. The results of the six models predicting patient outcomes across different quantifications of patients care activities can be found in Table 7.

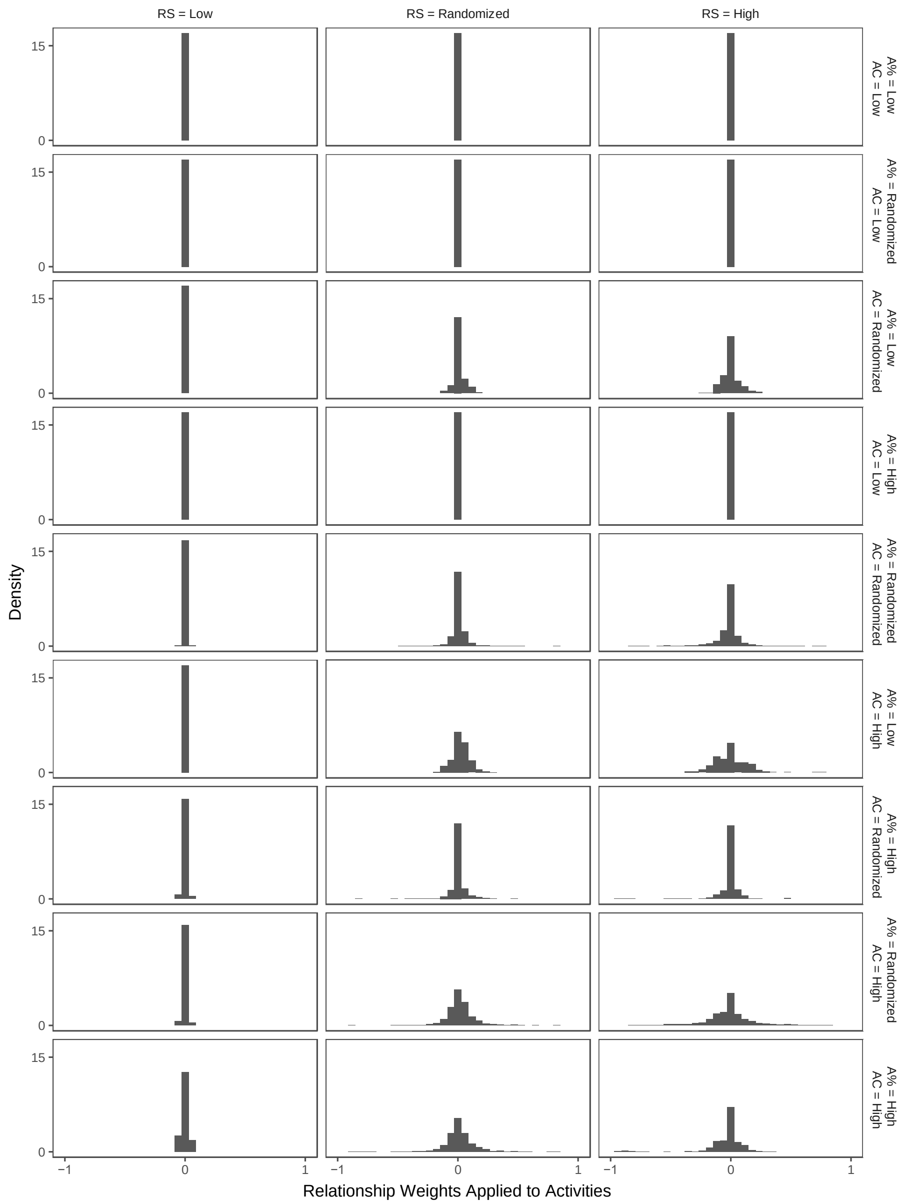

Figure 3. Distribution results of relationship weighted activities across 27 simulation. Each activity was modified by relationship score (RS), activity count (AC), and activity percentage (A%).

Activity variables	Activity weight	Relationship score	Number of distinct activities in dataset	AUC	OOB Error	Sensitivity	Specificity	Number of predictors in final model
Excluded non-essential	Relationship score	High and low	644	0.945	0.053	0.930	0.960	203
All activities	Relationship score	High and low	3139	0.947	0.053	0.944	0.950	1605
Excluded non-essential	Relationship score	High only	619	0.886	0.109	0.848	0.925	149
Excluded non-essential	Relationship score	Low only	620	0.906	0.095	0.910	0.900	119
Excluded non-essential	Count	None	838	0.763	0.233	0.728	0.799	43
Excluded non-essential	Binary	None	838	0.786	0.207	0.736	0.836	110

Table 7. Comparison of models predicting patient outcomes across models. Differences in models, (excluding non-essential activities, differences in calculating activity weight and relationship scores) can be found in Section 4.2.3.1. Data Sources.

4.4. Discussion

In this study a multilayer network model was proposed, which aggregated data across multiple dimensions by weighting patient care activities by the significant relationships of the clinicians. The relationship weighted activities were highly predictive of patient outcomes, and the predict activities in the model direct one towards areas in treatment where differences in relationships are most related to patient outcomes.

The model verification demonstrated that the equation that weighted patient care activities with relationship scores, applied weights as expected. We saw the most variance in highly clustered relationship scores (high limit), and most of the relationship weights zero or extremely clustered around zero when all clinicians had a similar number of significant relationships. We also observed if the activity count is low, the activities will not receive any weight, regardless of the value of the relationship score or activity percentage. Interestingly, when the activity percentage was high, across all activity count values, a greater variance in activity weights occurred.

The results of model validation using activity logs from ICH patients provided evidence that the MLN model outlined in this paper accurately can identify when relationships between clinicians are most predictive of patient outcomes. Model validation also indicated that both high and low categorized relationships contain unique and specific information regarding when clinical relationships are most predictive of outcomes. Including both in weighting patient care increased the accuracy of the model and reduced the OOB error. The validation process also provided evidence that the MLN framework proposed could be applied to data that has not been pre-processed with disease-specific domain knowledge to filter irrelevant people or care. Hand coded variables versus unfiltered variables had a very similar model results, so this step may not be required and future analyses. The model automatically filtering out unimportant or uncommon practices, which lead strikingly similar model results between the model using unprocessed activity data, verses clinician filtered activities. This is important because coding relevant variables is burdensome and time-consuming.

To better understand the importance of activities as they relate to real life care, steps should be taken to further reduce complexity in the activities, especially regarding variables that are likely colinear. For example, in the relationship weighted model, many predictors were conceptually linked—such as writing and verifying a note. It is unclear that verifying a note is somehow predictive of outcomes in an unrelated manner to writing the note itself. By combining linked predictive variables, areas of hospital workflow where differences in clinician relationships are associated with patient outcomes can be more easily identified. Classical network metrics (such as centrality and density) are being developed for MLNs, as well as newer network-based metrics that describe interactions between layers, node types and time. These could all be applied to this baseline MLN model to gain actionable insight into how differences in clinical relationships may be impacting patient outcomes.

4.4.3. Limitations. There are several limitations to this work. First, data used to validate the network model was from one medical center and the population was only one disease type. This method of analyzing activity log data will need further validation across different populations and disease types. It is unclear what the level of heterogenicity of data would be incompatible with a multilayer network model; could accurate results be obtained from activity log data containing various diseases, illnesses or injuries, and if so, could different hospital units be analyzed together as well? While presumably each treatment workflow is highly associated with both the patient's condition and the hospital unit they are being treated in, hypothetically new aspects or layers in the model could be introduced to structure an analysis which accurately accounts for these new dimensions.

The other primary limitation is that the activity log is only an artifact of patient care. This method is useful in generating hypothesizes that should be further investigated by directing examining the real-life care associated with the record of care. Acute inpatient critical care is a complex and siloed practice, where practitioners often don't see many of the long-term effects of their care on the patient. This makes it more difficult to critically examine how individuals' interpretations or implementations of standardized practices may improve or harm patients.

Therefore, these methods, while not showing cause and effect, are valuable for analyzing an environment has become too complex for traditional methods to measure both care and care practitioners may be impacting patient outcomes.

4.5. Conclusion

By structuring activity logs as MLNs, care activities can be quantified by the categorized relationships of the clinicians performing those activities identify patient care that may be associated with successful or unsuccessful relationships. This network framework was successfully verified and validated. The MLN outlined in this study can be applied to study other inpatient populations, to identify areas of hospital care where differences in teamwork related practices may be impacting patient outcomes.

5. Identifying Patient Care Activities that Predict ICH Outcomes

5.1. Introduction

Clinicians (e.g., physicians, nurses, respiratory therapists) are presumed to be interchangeable, especially in highly standardized and evidence-based fields. Health care systems compensate for variability in the actions and performance of health care team members, helping to standardize care, minimize unnecessary practice variation, and maximize patient outcomes. Evidence-based guidelines emphasize routine tasks in patient care, such as diagnostic tests, elements of the history, and common interventions. Health care team members with different roles are typically not considered as individuals, i.e., any physician is expected to order aspirin for a patient with acute vascular disease, any pharmacist is expected to approve it, and any nurse is expected to administer it to the patient. The effect of any particular patient care activity is presumed to be consistent, regardless of who performs it.

Evidence suggests that identity of clinicians impacts patient outcomes. Increases in collaborative communication are associated with improved patient outcomes.[19-21] Improved outcomes in patients with stroke are associated with effectiveness, task orientation, order and organization, and utility of quality information.[22] Relationships between clinicians may impact health outcomes, however, the specific patient care activities that may be most affected by individual clinicians have not been defined.

Studying relationships between clinicians is labor-intensive, so research to date has necessarily been limited to a few health care team members, in a few specific scenarios, for brief epochs of time. Teamwork training,[60] increased compliance with clinical guidelines,[61] and simulation labs are all strategies to improve patient care and patient outcomes by improving how health care team members work together. Beyond a few examples, however, how to specifically target patient care activities performed by specific clinicians to improve patient outcomes is not clear. While valuable, in-person qualitative methods are difficult to scale to large groups, as dozens of clinicians routinely participate in patient care. Retrieving and analyzing data on which clinicians perform patient care activities, combined with patient outcomes, would permit study of relationships between clinicians and patient outcomes on a larger scale than has heretofore been possible. The hypothesis was tested that the effect of patient care activities on patient outcomes would be modified by measures of the relationship between clinicians.

5.2. Methods

5.2.1. Study Population. The NUBAR data repository was utilized in this research, which is a prospective registry of patients with ICH. [57] Patients were diagnosed with ICH by CT, interpreted by a board-certified neurologist. All patients with ICH were admitted to a Neuro/Spine ICU in the context of a certified stroke center that predated the existence of NUBAR by several years, and a high-intensity model of intensive care (e.g., all patients were attended to daily by an intensive care team). All patients or a legally authorized representative provided written informed consent for the use of EHR, with the exception of patients who were permanently comatose, died, or who were not consented and had no available legal representative, in which case the Institutional Review Board granted a waiver from informed consent.

Only patients with ICH scores of 0 to 2 were included in analysis predicting outcomes. Patients with ICH scores of 3 or greater were highly likely (96.5%) to have a poor outcome (disabled or death) at follow-up, indicating patient care activities had little to no impact these patient outcomes.

5.2.2. Study Outcomes. Patient outcome was defined by the modified Rankin Scale (mRS), an ordinal score from 0 (no symptoms) to 6 (dead). The mRS was prospectively collected by the NUBAR registry, assessed at 28 days after hospital discharge using a validated questionnaire, as previously described. The mRS outcome was dichotomized as "good" (mRS 0-3, independent for ambulation), or "poor" (mRS 4-6, disabled or dead).[62]

Patient outcomes were used to define "successful" patient care because the mRS is the defined outcome for pivotal clinical trials of ICH. Thus, activities and relationship that increase the likelihood of a good outcome at follow-up are of interest. Put another way, our intention was to determine if specific clinicians performing specific patient care activities increased the accuracy of predicting patient outcomes after adjusting for severity of injury (see section 5.2.5. Weighting Patient Care Activities).

5.2.3. Data Sources. Data on patient outcomes, as well as the data used for risk adjustment (ICH scores) were obtained from the NUBAR registry. Transaction logs of patient care activity for patient encounters that were included in the NUBAR registry were retrospectively collected data from the EDW. This data included a timestamped log of all activities performed in a patient's EHR for all ED and inpatient care, and the individual clinician who performed the activity (e.g., Dr X ordered metoprolol at 02/03/2007 20:24, Nurse Y performed a neurological assessment note at 12/26/2016 07:17). A neurologist coded the patient care activities deemed non-essential (e.g., "discontinue suction tube") or not performed by humans (e.g., an automatic alert from a laboratory computer that an order was completed). These were excluded from the activity log.

5.2.4. Clustering Activities. Patient care activities do not always occur independently from one another. Often, a single real-life task will require multiple EHR interactions, resulting in several corresponding activity log entries. For example, a physician may write, sign, and verify the same progress note, resulting in three log entries. Through a two-step process, the activities which were frequency performed in sequence were clustered. First, the frequency (or event relation) between pairs of activities was calculated, and then applied to weight the edges between activities. Then this edge-weighted activity network was clustered to identify communities of activities.

Using previously described methods,[17] the frequency at which activities were performed together was calculated. The frequency, or event relation, between any two unique activities, e_j and e_i, across all subsequences of activities on a patient encounter can be measured using *equation 6:*

$$\text{Event Relation}(\mathrm{e}_j, \mathrm{e}_i) = \Sigma \frac{1}{\left(\mathrm{P}(e_j) - \mathrm{P}(e_i)\right)^2}, \text{where: } \left(0 < \mathrm{p}(e_j) - \mathrm{p}(e_i) \leq \alpha\right) \qquad (6)$$

The event relation measures the positional relation between e_j and e_i, where $p(e_i)$ is the position of the event e_i in a sequence. The first event in each subsequence is in position one, and every subsequent activity is sequentially order by timestamp. Using alpha, provides flexibility in the distance between events; if the events must be directly after each other the $\alpha = 1$ and the event distance is 1. If $\alpha = 4$, the activities up to 4 positions away will be counted, the distance of fourth event being 1/16. In this study an alpha of 5 was used. The frequency was calculated at which two patient care activities of the same category (i.e. notes or pharmacy) and were performed one after the other by the same clinician on each patient encounter.

These activities were then connected in a network, with the link between activities weighted by the event relation. Using a fast-greedy clustering algorithm,[63] activities in the network were clustered into communities by the frequency weight between each pair and the total connectivity of a group. For example, if activity pairs (A, B), (B, C), and (A, C) were connected with a high frequency weight, then A, B, C would be clustered. This resulted in the identification of activity clusters, which represent activities of the same category which are frequently performed in close temporal proximity. These activity clusters then replaced corresponding individual activities in the data log.

5.2.5. Weighting Patient Care Activities with the Relationships Scores. To measure patient care activities using the relationships between clinicians on patient encounters, a relationship score was developed using a previously described method of categorizing clinical relationships, the SPOR. [23,24,55] The SPOR is a metric designed to identify clinician pairs, whose shared patients experienced statistically significant high or low rate of good outcomes compared to shared patients with other clinicians (after risk-adjustment). Clinicians get paired by sharing a certain number of patients in common and are connected to patients by performing patient care activities on their encounter (as recorded in activity logs). The SPOR weighted relationships are distributed around a mean of one, since most clinicians have a similar rate of positive outcomes across all pairings (SPOR ~1). The greater the SPOR is from one, the more variant the rate of shared positive outcomes. The top and bottom 5% of SPOR weighted relationships are tested to determine if the rate of patient outcomes could occur due to chance, by comparing with randomly generated data. If a relationship's rate of positive outcomes likely did not occur by chance (<0.05), it is categorized as high (if positive) or low (if negative). Detailed methodological descriptions of SPOR calculation, risk-adjustment, and categorization is included in section 2 Shared Positive Outcomes Ratio.

In this study, the SPOR was calculated for every pair of clinicians who shared at least 12 patients in common (of any ICH score), and shared positive outcomes were risk-adjusted for disease severity using patient ICH scores. Clinician relationships were categorized into three groups by SPOR values—high (~5% of pairings), low (~5%), and average (~90)—resulting in a list of categorized pairings of clinicians that frequently share patients.

The list of categorized clinician relationships and patients shared by each relationship, was used to calculate a clinician's relationship score on every patient encounter through the number of high, low, and average relationships connected to the clinician via the encounter. The relationship score equals to the number of high relationships subtracted from low, divided by the number of total SPOR relationships. The relationship score assesses the long-term outcome success for each clinician when working with a certain combination of team members on a patient encounter. If a clinician has more high than low relationships on an encounter, the relationship score is positive; more low than high relationships results in a negative relationship score. Due to 90% of relationships pairs being categorized as average, most clinician only average relationships on encounters, resulting in most relationship scores equaling 0.

The relationship score was then applied to each distinct patient care activity a clinician performed on an encounter. The clinician's relationship score on that encounter was evenly distributed across the distinct activities they performed. The relationship score gets distributed in this manner because if a clinician only performs few distinct activities on an encounter, compared to many, there is a greater likelihood that one of those activities is related to the relationship score. The relationship score for each of the clinician's distinct activities gets adjusted by the proportion of the activity the clinician performed on that encounter, resulting in a distributed and adjusted relationship weight for each distinct activity a clinician performed on an encounter. The clinician-activity specific relationship weight was adjusted by the proportion performed by the clinician so the relationship weights of the same distinct activity could be summed for each encounter, resulting in a total relationship weight for each distinct activity/per patient encounter. The final relationship weighted activity dataset contained a single value per encounter for each of the 644 distinct activities in the activity log. The value of each relationship weighted activity either is the sum of all clinicians' activity scores for that activity or zero. Zero indicates the activity was not performed, was only performed by clinicians with relationship scores of zero, or multiple clinicians performing the activity had relationship scores which cancelled out.

5.2.6. Statistical Procedures. Patient encounter outcomes (good outcome = 1) were predicted and compared across four RF models for patients with low to moderate ICH severity at diagnosis (ICH score 0-2). The baseline model predicted patient outcomes using a patient's ICH score, which is currently used to predict patient outcomes after ICH.[62,64,65] Two models predicted patient outcomes from patient care activities, unweighted by clinician relationship scores. These included the activity count, the number of times an activity was performed on an encounter and the occurrence of an activity on an encounter (1=occurrence, 0= did not occur). The last model predicted outcomes using the relationship weighted activities, described above. The RF models in this study were fitted and adjusted using k-fold cross validation techniques to minimize OOB error and AUC. The predictive variables in each model were those which maximized model AUC, while minimizing the number of predictors by accounting for the relative incremental increase in AUC per variable. AUC, OOB error, sensitivity, specificity, and number of predictors were reported for each model.

5.2.7. Textual Analysis of CT Interpretations. After the results from the above analysis were obtained, a substantial number of clinical notes were associated with differences in relationship scores. Further validation of the results was performed by analyzing the free text of the first CT imaging note available for each patient, for several reasons: (1) CT notes were only predictive in the weighted activity model; (2) CT imaging is one of the first activities performed on all encounters because it is required for ICH diagnosis; (3) The free text content of radiology notes are typically more reliable and structurally predictable then other clinical notes, which are notorious for errors and obfuscation from their copy and paste construction.[66]

CT notes were affirmatively coded (1) when containing explicit verbal communication of imaging results from the radiologist to another physician within the free text, which included both clinicians' names, the date/time and mode of communication (phone or in person). The relationship score was dichotomized if the clinician had a positive value (indicating the clinician has more high scoring relationships than low on an encounter), or not (both a zero and negative relationship score included). A chi-square test for significance was performed by comparing documented communication and positive relationship scores.

On 12 patient encounters, the clinicians writing the first CT note did not have any SPOR categorized relationships, indicating they did not frequently participate in the network (i.e. did not share at least 12 patients with any other clinician). These 12 notes were excluded from textual analysis.

5.3. Results

Demographics of the cohort are shown in Table 8. The mean patient age and race/ethnicity were typical of patients with ICH. The final dataset contained 284 inpatient encounters, with 644 distinct patient care activities (e.g., nursing assessment, progress note) performed a total of 548,250 times.

	Encounter outcome = 0	**Encounter outcome = 1**
N=	159	125
Age (mean±SD)	65.95±13.67	61.68±13.65
Gender (Female %)	42.77%	45.60%
Race (%)		
White	47%	40%
Black	31%	34%
Asian	2%%	2%
Other	8%	14%
Unknown	12%	10%
Ethnicity (Hispanic %)	7%	13%
ICH score (median± IQR)	1.31 ± 0.72	0.624 ±0.69
Modified Rankin Score (median± IQR)	4.91 ±0.77	1.82± 1.08
Pneumonia (%)	0%	23.90%
Vent Free Days (mean± SD)	7.84±5.43	12.88±2.80
LOS (mean± SD)	19.01 ±19.35	7.85±5.45
DVT (%)	10.14%	3.50%

Table 8. Demographics of patients with low to moderate ICH

Details of patient care activities, grouped by outcome, are shown in Table 9. Patients with poor outcome at follow-up had a greater number of activities performed, treated by more clinicians, and had more distinct patient care activities during their hospital admission. This is consistent with the longer length of stay and increased complications seen in Table 8.

	Encounter outcome = 0		Encounter outcome = 1	
	Mean±SD	**Range (min – max)**	**Mean±SS**	**Range (min – max)**
Total Number of activities	35,577±45,089	213- 298,453	6,104±7750	183 – 48,551
Distinct activities	259±83.4	60-451	193±51.90	93-380
Clinicians per encounter	101±50.1	18-229	54.9 ±26.4	22-170

Table 9. Patterns of patient care activities from the treatment of patients with ICH.

Model performance is reported in Table 10 for the relationship weighted activity, activity count, and the activity occurrence models. Each model had progressively improved prediction of patient outcome above the baseline of ICH Score (severity of injury). The weighted activity model was a more accurate across all measures. The types of activities that were predictive variables in these three models are reported as percentages in Table 11. Full list of predictive variables in each model can be found in Table 12.

The first CT note performed was searched for documentation of communication within the free text. Clinicians with positive relationship scores documented communication (e.g., “this was discussed with Dr. X by telephone”) in 57% (85/149) of CT notes, compared to only 35.8% (44/123) notes of clinicians with zero or negative relationship scores (P = 0.001).

Model	AUC	OOB error	Sensitivity	Specificity	# Variables
Weighted	0.93	0.07	.93	.93	12
Count	0.79	0.21	.77	0.81	12
Occurrence	0.76	0.24	0.74	0.78	12
ICH Score	.67	0.31	0.50	0.85	1

Table 10. Model results predicting outcomes of patients with ICH, using relationship weighted activities (weighted), the count of each activity per encounter (count), whether an activity occurred on an encounter (occurrence), and patients’ disease severity as measured by ICH score.

Model	Number of Activities	Notes	Pharmacy	Intake & Output	Labs	Patient Care	Radiology
Full Dataset	644	23%	33%	1.5%	8%	21%	13%
Weighted	12	50%	8%	17%	0	17%	8%
Count	12	17%	33%	25%	8%	8%	8%
Occurrence	12	0	42%	25%	25%	0	8%

Table 11. Patient care activities predictor by type (%).

Predictive variables in models		Order of predictive variables' relative importance in the model:		
Activity type	**Activity**	**Weighted activity model**	**Count of activity model**	**Occurrence of activity model**
Notes	Pharmacy Medication Reconciliation	1		
	Care Team Form	9		
	Critical Care Progress Note	3	4	
	Braden Scale Form	5		
	CT Brain WO Contrast	7		
	Performed Notes: Neurological Assessment, PC Neurological Assessment, Patient Assessment, Patient Response to Medication, Med Surg Restraints	6	2	
Patient care	Complete Orders: Fall Risk Assessment, Glucose Fingerstick, Line Initiation and Care, MD to RN Communication, Pain Assessment, Patient Assessment	10		
	Complete Orders: Evaluate Patient For Moisture, MPET ICU (Assessment), Neurological Checks, Nutrition Assessment, Order Reconciliation, Pain Assessment, Patient Assessment (4 hours), Turn Reposition Patient	2	1	
Intake/ Output	Oral (PO) Fluid Intake	11	3	1
	Urine Description	8	9	9
	Tube Feeding Fluid, Tube Flushes,		8	5
Radiology	Modify Orders CT Brain WO Contrast	4		
	Activate Orders: Transcranial Doppler Examination, X Ray Chest AP Portable, X Ray Abdomen AP for Tube Placement		10	3
Pharmacy	Order: chlorhexidine topical, Heparin, ocular lubricant		12	4
	Order chlorhexidine topical			12
	Order vancomycin		11	2

	Discontinue vancomycin			8
	ciprofloxacin			10
	Complete Orders: Free Water Flush, heparin, levetiracetam, ocular lubricant		7	
	Complete Orders: amlodipine, docusate, famotidine, hydralazine, metoprolol, phenytoin Order: potassium chloride	12	6	
Laboratory	Order: Partial Thromboplastin Time, Prothrombin Time INR		5	
	Order Sodium Level			11
	Discontinue Sodium Level			7
	Order Urine Culture			6

Table 12. Patient care activities predictors in each model by variable importance ranking.

5.4. Discussion

This study hypothesized that patient outcomes would be associated with patient care activities weighted by the relationship scores of the performing clinicians. This hypothesis was tested in a well-described cohort of patients with mild to moderate ICH, because severe ICH is usually fatal regardless of medical intervention. After restricting the clinicians to those who worked together frequently and adjusting for severity of injury, it was found that specific patient care activities, weighted by relationship scores of clinicians, were highly predictive of patient outcomes. These results suggest that both the specific patient care activities and who performs them have additional impacts on patient outcomes after ICH.

Prediction of patient outcomes improved after accounting for patient care activities, and, subsequently, weighting patient care activities by relationship scores between clinicians. Severity of injury, measured here by the ICH score, is highly predictive of patient outcomes after ICH and was used as a baseline to predict patient outcomes. Predicting patient outcomes using only patient care activities (the number of times each type of activity was performed, or if it occurred, on an encounter), improved prediction over ICH scores. This is intuitive, for example, activities related to respiratory ventilation or infections indicate the patient is more likely to have a poor outcome. However, these activities were not instructive of hospital care which could be improved, rather they simply indicated a treatment response to patient's worsening condition. However, the prediction of patient outcomes improved after adjusting for relationship scores and further, different variables were predictive in this new model. In particular, clinical notes constituted half the predictive variables in the relationship weighted model, compared to zero note predictors in the activity occurrence model. The two notes predictive in the count model (critical care progress notes and neuro ICU assessments) were likely predictive due to increased LOS, which could be due to an underlying hospital complication.

It was also hypothesized that the activities identified in the relationship weighted model were predictive due to differences in teamwork or communication related practices involved with performing the activity. Indeed, radiologists with positive relationship scores were more likely to document a conversation about communicating results in the first CT scan interpretation. This evidence links the network measurement with independently documented clinical communication. Overall, accounting for severity, patient care activities, and who performed patient care activities was most predictive of patient outcome.

These results may have implications for improving patient care. Once severity of injury is considered, specific patient care activities may be targeted to improve patient outcomes, especially related to information flow or communication practices. These data suggest that specific patient care activities, such as notes for medication reconciliation, neurosurgery and critical care teams, and specific bedside nursing assessments may have outsized importance. Further research might codify how improved relationships between clinicians during specific care activities leads to differences in patient outcomes. For example, communication between radiologists and bedside clinicians might be a particular opportunity to improve communication, leading to improved patient outcomes.

There are several potential limitations to this work. The inpatient treatment of only one disease was examined. This work only examined one hospital, however, clinicians who participate in patient care after ICH are standardized. The relationship scores of clinicians may be measuring factors that are not entirely clear, and further work could illuminate additional clinical actions associated with positive or negative relationship scores. In addition, the activity log representing clinician and EHR interaction may not capture every patient care activity that is potentially important to patient outcome (e.g., handwashing). In addition, the patient care activity log may be overly granular or not contain enough detail in all cases to directly connect the digital record to practices in real life. Strengths of this work include the use of a large, well-characterized patient cohort with prospective documentation of severity of injury and patient outcomes after hospital discharge.

5.5. Conclusion

Care activities, weighted by clinicians' relationship scores, were excellent predictors of outcomes in patients with ICH of low-to-moderate severity, compared to ICH scores and unweighted activities. CT notes were examined after being identified as a predictor in the relationship weighted model. CT notes written by radiologists with positive relationships scores were more likely to include documentation that imaging results were communicated to others via the phone or in person. Findings suggest that identifying and targeting areas of care frequently performed by clinicians with significant relationship scores could improve patient outcomes.

6. Conclusion

This dissertation contains three articles that present evidence in support of one overarching thesis: a multilayer network model approach can be used to measure both clinical relationships and patient care activities, and such measurement can be used to accurately predict patient outcomes and identify the points during treatment when clinical relationships are most related to of outcomes. This was accomplished by first quantifying and categorizing clinical relationships with risk-adjusted patient outcomes. These scores were then used to weight patient care activities. The product of this MLN model was the aggregation of multiple dimensions of data to identify when during treatment clinical relationships were most predictive of outcomes. Several steps were taken to determine the underlying theoretical assumptions and accuracy of the MLN model outlined in this dissertation. First, the effects and necessity of risk-adjusting the outcomes of patients used to characterize relationships between clinicians was examined. Evidence from this investigation supported the assumption that risk-adjusting is an important step in accurately measuring relationships. Second, the MLN model and the activity-weighting methodology were verified using simulated activity logs and validated using real patient data. This produced several important findings, including a better understanding of how the model applied and modified the relationship scores to activities, the value of using both the high- and low-categorized relationships, and how it may be unnecessary to have a domain expert pre-code non-essential patient care activities in future research. Lastly, a clustering algorithm was applied and a variable selection equation to determine the clinical utility of this methodology to a quality improvement application. A small number of relationship-weighted activities were found to be excellent predictors of patient outcomes. In addition, further clinical evidence linked to relationship scores demonstrations the utility in hospital quality improvement. Brain CT scans, weighted by relationship scores, were one of the predictors in the model and one of the first patient care activities performed on all patients as part of ICH diagnosis. It was found that in the first CT scan, radiologists who had a positive relationship score were more likely to document a conversation about communicating results. This evidence links our teamwork measurement with independently documented clinical communication.

Taken together, the evidence produced in this dissertation demonstrates how an MLN approach to measuring both clinical relationships and patient care activities through edge-weighted network aggregation can help identify specific care activities where variations in clinical relationships may have an effect on patient outcomes. In this research, the treatment of patients with ICH consists of thousands of distinct activities performed by clinicians within the complex system of a hospital. It is not realistic to expect any clinician or medical director to know which among these thousands of care activities is linked to patient outcomes 28 days after discharge. This dissertation demonstrated that this methodology can highlight areas of hospital workflow where differences in relationships are linked to patient outcomes. In this particular study, 12 distinct activities were identified that occur to nearly all patients, regardless of their post-discharge outcome. Notes and patient care activities were both the most predictive and prevalent variables in the MLN relationship-weighted model, which suggests that the model is identifying activities related to communication and/or teamwork practices. For example, if the radiologist who performs a patient's first brain CT note has a positive relationship scores on the encounter, they were more likely to document communicating the results to other physicians.

However, it is still unclear how widely these findings will apply to other populations or identify other clinical actions related to communication or teamwork practices.

There are several potential limitations to this work. The inpatient treatment of only one disease was examined. Further, the treatment was only examined at one hospital, though ICH care practices are standardized. The relationship scores of clinicians may be measuring factors that are not entirely clear, and further work could illuminate additional clinical actions associated with positive or negative relationship scores. The activity log representing clinician-EHR interaction may not capture every patient care activity that is potentially important to patient outcome (e.g., handwashing). In addition, the patient care activity log may be overly granular or not contain enough detail in all cases to directly connect the digital record to practices in real life. Strengths of this work include the use of a large, well-characterized patient cohort with prospective documentation of severity of injury and patient outcomes after hospital discharge.

These limitations present the opportunity for further investigation. There are three important next steps to continue to validate that this approach to studying hospital healthcare delivery processes is accurately identifying care activities when clinical relationships are most associated with patient outcomes. First, the clinical practices and EHR artifacts (i.e. note text) associated with the patient care activities identified by this MLN model should be more systematically investigated for differences in communication and teamwork practices that are related to the edge-weighted metrics outlined in this dissertation. In addition, these methods should be applied to other patient populations. For example, the same clinicians who treated the patients with ICH in this study also treated patients with subarachnoid hemorrhage with similar workflows in the same hospital unit. If the same significant clinical relationships and patient care activities were independently identified, it would greatly strengthen the validity of this network approach. Third, this network model could be used to predict patient outcomes of patients not included in this study. This validation approach would demonstrate the real-life application of this method.

Assuming this approach is effective, further methodology could be applied to this foundational model to better understand workflow in other ways. More complex methods of activity clustering could be employed to understand with greater detail the time component of relationship-weighted activities. For example, are the relationships of the clinician performing the first critical care progress note more predictive of outcomes than the last note? These applications would need to adjust for confounders, such as longer lengths of stay associated with poor outcomes. Other additions to this MLN framework could also include measuring relationship scores between more than two clinicians, or aggregating relationship-weighted activities across all encounters for each clinician to measure if the variance of relationship weights is greater for certain activities. In addition to other approaches to aggregation and edge-weighting, classical network metrics (such as centrality and density) are being applied to MLNs, as well as newer network-based metrics that describe interactions between layers, node types and time. These could all be applied to this baseline MLN model to gain actionable insight into how differences in hospital operations processes may be impacting patient outcomes.

As a whole, this thesis presents evidence that a multilayer network approach to quantifying clinical relationships and patient care activities can be used to predict patient outcomes, and these predictions have the potential to be used to implement hospital quality improvement initiatives. The specific areas of care that are most likely to improve patient outcomes can be identified and targeted through investigating the interconnected and dynamic system of hospital acute care through a complex network framework.

References

1. What is Informatics? AMIA. https://www.amia.org/fact-sheets/what-informatics. Published 2011. Accessed.

2. Thune J, Alexander L, Roberts P, Burr R, Enzi M. Where Is HITECH's $35 Billion Dollar Investment Going? In. Vol 2019: Health Affairs; 2015.

3. 2019 Medicare Electronic Health Record (EHR) Incentive Program Payment Adjustment Fact Sheet for Hospitals. CMS. https://www.cms.gov/newsroom/fact-sheets/2019-medicare-electronic-health-record-ehr-incentive-program-payment-adjustment-fact-sheet-hospitals. Published 2019. Accessed.

4. *Harnessing the Power of Data in Health.* Stanford University School of Medicine;2017.

5. Dean BB, Lam J, Natoli JL, Butler Q, Aguilar D, Nordyke RJ. Review: Use of Electronic Medical Records for Health Outcomes Research:A Literature Review. *Medical Care Research and Review.* 2009;66(6):611-638.

6. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics.* 2012;13(6):395-405.

7. Safran C, Bloomrosen M, Hammond WE, et al. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association.* 2007;14(1):1-9.

8. Meystre S, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann C. Clinical data reuse or secondary use: current status and potential future progress. *Yearbook of medical informatics.* 2017;26(01):38-52.

9. Mans RS, van der Aalst WM, Vanwersch RJ. Healthcare Processes. In: *Process Mining in Healthcare.* Springer; 2015:11-15.

10. Rojas E, Munoz-Gama J, Sepúlveda M, Capurro D. Process mining in healthcare: A literature review. *Journal of Biomedical Informatics.* 2016;61:224-236.

11. Vawdrey DK, Wilcox LG, Collins S, et al. Awareness of the Care Team in Electronic Health Records. *Appl Clin Inform.* 2011;2(4):395-405.

12. Kricke GS, Carson MB, Lee YJ, et al. Leveraging electronic health record documentation for Failure Mode and Effects Analysis team identification. *Journal of the American Medical Informatics Association.* 2016;24(2):288-294.

13. Chen Y, Lorenzi NM, Sandberg WS, Malin BA, Wolgast K. Identifying collaborative care teams through electronic medical record utilization patterns. *Journal of the American Medical Informatics Association.* 2016;24(e1):e111-e120.

14. Chen Y, Malin BA, Patel MB, McNaughton CD. Interaction patterns of trauma providers are associated with length of stay. *Journal of the American Medical Informatics Association.* 2018;25(7):790-799.

15. Poelmans J, Dedene G, Verheyden G, Van der Mussele H, Viaene S, Peters E. Combining Business Process and Data Discovery Techniques for Analyzing and Improving Integrated Care Pathways. 2010; Berlin, Heidelberg.

16. Rebuge Á, Ferreira DR. Business process analysis in healthcare environments: A methodology based on process mining. *Information systems.* 2012;37(2):99-116.

17. Chen Y, Xie W, Gunter CA, et al. Inferring Clinical Workflow Efficiency via Electronic Medical Record Utilization. *AMIA Annual Symposium proceedings AMIA Symposium.* 2015;2015:416-425.

18. Yoo S, Cho M, Kim E, et al. Assessment of hospital processes using a process mining technique: Outpatient process analysis at a tertiary hospital. *International journal of medical informatics.* 2016;88:34-43.

19. Baggs JG, Ryan SA, Phelps CE, Richeson JF, Johnson JE. The association between interdisciplinary collaboration and patient outcomes in a medical intensive care unit. *Heart & lung : the journal of critical care.* 1992;21(1):18-24.

20. Baggs JGP, RN, Schmitt MHP, RN, FAAN;, Mushlin AIM, ScM; , et al. Association between nurse-physician collaboration and patient outcomes in three intensive care units. *Critical Care Medicine.* 1999;Volume 27(9): pp 1991-1998.

21. Wheelan SA, Burchill CN, Tilin F. The link between teamwork and patients' outcomes in intensive care units. *Am J Crit Care.* 2003;12(6):527-534.

22. Strasser DC, Falconer JA, Herrin JS, Bowen SE, Stevens AB, Uomoto J. Team functioning and patient outcomes in stroke rehabilitation. *Archives of Physical Medicine and Rehabilitation.* 2005;86(3):403-409.

23. Carson MB, Scholtens DM, Frailey CN, et al. Characterizing Teamwork in Cardiovascular Care Outcomes: A Network Analytics Approach. *Circulation Cardiovascular quality and outcomes.* 2016;9(6):670-678.

24. Carson MB, Scholtens DM, Frailey CN, Gravenor SJ, Kricke GE, Soulakis ND. An Outcome-Weighted Network Model for Characterizing Collaboration. *PloS one.* 2016;11(10):e0163861.

25. Dunn AG, Westbrook JI. Interpreting social network metrics in healthcare organisations: A review and guide to validating small networks. *Social science & medicine.* 2011;72(7):1064-1068.

26. Damiani G, Salvatori E, Silvestrini G, et al. Influence of socioeconomic factors on hospital readmissions for heart failure and acute myocardial infarction in patients 65 years and older: evidence from a systematic review. *Clin Interv Aging.* 2015;10:237-245.

27. Donzé J, Aujesky D, Williams D, Schnipper JL. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA internal medicine.* 2013;173(8):632-638.

28. Ashton CM, Kuykendall DH, Johnson ML, Wray NP, Wu L. The association between the quality of inpatient care and early readmission. *Ann Intern Med.* 1995;122(6):415-421.

29. Naylor MD, Brooten DA, Campbell RL, Maislin G, McCauley KM, Schwartz JS. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. *J Am Geriatr Soc.* 2004;52(5):675-684.

30. DesHarnais SI, McMahon Jr LF, Wroblewski RT, Hogan AJ. Measuring hospital performance: The development and validation of risk-adjusted indexes of mortality, readmissions, and complications. *Medical Care.* 1990:1127-1141.

31. Keenan PS, Normand S-LT, Lin Z, et al. An Administrative Claims Measure Suitable for Profiling Hospital Performance on the Basis of 30-Day All-Cause Readmission Rates Among Patients With Heart FailureCLINICAL PERSPECTIVE. *Circulation: Cardiovascular Quality and Outcomes.* 2008;1(1):29-37.

32. Medicare Cf, Services M. measure information about the 30-day all-cause hospital readmission measure, calculated for the value-based payment modifier program. *Baltimore: Centers for Medicare & Medicaid Services.* 2015.

33. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *Jama.* 2011;306(15):1688-1698.

34. Bae S-H, Nikolaev A, Seo JY, Castner J. Health care provider social network analysis: a systematic review. *Nursing outlook.* 2015;63(5):566-584.

35. Sabot K, Wickremasinghe D, Blanchet K, Avan B, Schellenberg J. Use of social network analysis methods to study professional advice and performance among healthcare providers: a systematic review. *Systematic reviews.* 2017;6(1):208.

36. Meltzer D, Chung J, Khalili P, et al. Exploring the use of social network methods in designing healthcare quality improvement teams. *Social science & medicine.* 2010;71(6):1119-1130.

37. Newman ME. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences.* 2001;98(2):404-409.

38. Robinson I, Webber J, Eifrem E. *Graph databases: new opportunities for connected data.* " O'Reilly Media, Inc."; 2015.

39. Sehgal N. Annual Perspective 2014: Handoffs and Transitions. AHRQ Patient Safety Network. http://psnet.ahrq.gov/perspective.aspx?perspectiveID=170. Published 2015. Accessed.

40. Medicare Coordinated Care Demonstration. The Centers for Medicare & Medicaid Services. https://innovation.cms.gov/initiatives/Medicare-Coordinated-Care. Published 2017. Accessed.

41. Pope GC, Kautter J, Ellis RP, et al. Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health care financing review.* 2004;25(4):119.

42. Zhang Z. Parametric regression model for survival data: Weibull regression model as an example. *Annals of Translational Medicine.* 2016;4(24):484.

43. *A Package for Survival Analysis in S* [computer program]. Version 2.382015.

44. *Regression Modeling Strategies* [computer program]. Version 5.1-22018.

45. Fay MP, Shaw PA. Exact and asymptotic weighted logrank tests for interval censored data: the interval R package. *Journal of statistical software.*36(2).

46. O'Connor M, Murtaugh CM, Shah S, et al. Patient Characteristics Predicting Readmission Among Individuals Hospitalized for Heart Failure. *Medical care research and review : MCRR.* 2016;73(1):3-40.

47. Khawaja FJ, Shah ND, Lennon RJ, et al. Factors associated with 30-day readmission rates after percutaneous coronary intervention. *Arch Intern Med.* 2012;172(2):112-117.

48. Saunders ND, Nichols SD, Antiporda MA, et al. Examination of unplanned 30-day readmissions to a comprehensive cancer hospital. *Journal of oncology practice.* 2015;11(2):e177-e181.

49. Interdonato R, Atzmueller M, Gaito S, Kanawati R, Largeron C, Sala A. Feature-rich networks: going beyond complex network topologies. *Applied Network Science.* 2019;4(1):4.

50. Aleta A, Moreno Y. Multilayer Networks in a Nutshell. *Annual Review of Condensed Matter Physics.* 2019;10:45-62.

51. Brunson JC, Laubenbacher RC. Applications of network analysis to routinely collected health care data: a systematic review. *Journal of the American Medical Informatics Association.* 2017;25(2):210-221.

52. Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. *Journal of Complex Networks.* 2014;2(3):203-271.

53. Barrat A, Barthelemy M, Vespignani A. *Dynamical processes on complex networks.* Cambridge university press; 2008.

54. Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. *Proceedings of the national academy of sciences.* 2004;101(11):3747-3752.

55. Chandler AE, Mutharasan RK, Amelia L, Carson MB, Scholtens DM, Soulakis ND. Risk Adjusting Health Care Provider Collaboration Networks. *Methods Inf Med.* 2019;58(2-03):71-78.

56. John S. Carson I. Verification validation: model verification and validation. Proceedings of the 34th conference on Winter simulation: exploring new frontiers; 2002; San Diego, California.

57. Naidech AM, Garg RK, Liebling S, et al. Anticonvulsant use and outcomes after intracerebral hemorrhage. *Stroke.* 2009;40(12):3810-3815.

58. Cutler DR, Edwards Jr TC, Beard KH, et al. Random forests for classification in ecology. *Ecology.* 2007;88(11):2783-2792.

59. Calle ML, Urrea V, Boulesteix A-L, Malats N. AUC-RF: a new strategy for genomic profiling with random forest. *Human heredity.* 2011;72(2):121-132.

60. Strasser DC, Falconer JA, Stevens AB, et al. Team Training and Stroke Rehabilitation Outcomes: A Cluster Randomized Trial. *Archives of Physical Medicine and Rehabilitation.* 2008;89(1):10-15.

61. Middleton S, McElduff P, Ward J, et al. Implementation of evidence-based treatment protocols to manage fever, hyperglycaemia, and swallowing dysfunction in acute stroke (QASC): a cluster randomised controlled trial. *The Lancet.* 2011;378(9804):1699-1706.

62. Cheung RT, Zou LY. Use of the original, modified, or new intracerebral hemorrhage score to predict mortality and morbidity after intracerebral hemorrhage. *Stroke.* 2003;34(7):1717-1722.

63. Sousa FBd, Zhao L. Evaluating and Comparing the IGraph Community Detection Algorithms. Paper presented at: 2014 Brazilian Conference on Intelligent Systems; 18-22 Oct. 2014, 2014.

64. Hemphill JC, 3rd, Bonovich DC, Besmertis L, Manley GT, Johnston SC. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. *Stroke.* 2001;32(4):891-897.

65. Clarke JL, Johnston SC, Farrant M, Bernstein R, Tong D, Hemphill JC. External validation of the ICH score. *Neurocritical Care.* 2004;1(1):53-60.

66. Hirschtick RE. Copy-and-Paste. *JAMA.* 2006;295(20):2335-2336.