Modeling and Interpreting Patient Subgroups in Hospital Readmission: Visual Analytical Approach

Suresh K. Bhavnani, Ph.D.,[1,2§] Weibin Zhang, Ph.D.,[1] Shyam Visweswaran, M.D., Ph.D.,[3,4]
Mukaila Raji, M.D., M.S., F.A.C.P.,[5] Yong-Fang Kuo, Ph.D.[1]

[1]School of Public and Population Health, University of Texas Medical Branch, Galveston, TX, USA
[2]Institute for Translational Sciences, University of Texas Medical Branch, Galveston, TX, USA
[3]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
[4]Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA, USA
[5]Division of Geriatric Medicine, Department of Internal Medicine, University of Texas Medical Branch Galveston, TX, USA

[§]Corresponding author
Suresh K. Bhavnani, Ph.D., M.Arch., FAMIA
School of Public and Population Health
Institute for Translational Sciences
University of Texas Medical Branch
301 University Blvd
Galveston, TX, USA
email: subhavna@utmb.edu

ABSTRACT

Background

A primary goal of precision medicine is to identify patient subgroups and infer their underlying disease processes, with the aim of designing targeted interventions. However, few methods automatically identify both patient subgroups and their co-occurring characteristics simultaneously, measure their significance, and visualize the results. Such methods could enhance the interpretability of patient subgroups, and inform the design of classification and predictive models.

Objectives

To analyze patient subgroups in hospital readmitted patients using a three-step modeling approach. (1) *Visual analytical* modeling to automatically identify patient subgroups and their co-occurring comorbidities, and determine their statistical significance and clinical interpretability. (2) *Classification* modeling to classify patients into subgroups and measure its accuracy. (3) *Prediction* modeling to predict a patient's risk of readmission and compare its accuracy with and without patient subgroup information.

Methods

We extracted 2013-2014 Medicare data related to hospital readmission in three conditions: chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), and total hip/knee arthroplasty (THA/TKA). For each condition, we extracted cases defined as patients readmitted within 30 days of hospital discharge, and controls defined as patients not readmitted within 90 days of discharge, matched by age, gender, race, and Medicaid eligibility (n[COPD]=29,016, n[CHF]=51,550, n[THA/TKA]=16,498). These data were analyzed using: (1) bipartite networks to identify patient subgroups based on frequently co-occurring high-risk comorbidities; (2) multinomial logistic regression to classify patients into subgroups; and (3) hierarchical logistic regression to predict the risk of hospital readmission using subgroup membership, compared to standard logistic regression without subgroup membership.

Results

In each condition, the visual analytical model identified patient subgroups that were statistically significant (Q=0.17, 0.17, 0.31; *P*<.001, <.001, <.05), were significantly replicated (RI=0.92, 0.94, 0.89; *P*<.001, <.001, <.01), and were clinically meaningful to clinicians. (2) In each condition, the classification model had high accuracy in classifying patients into subgroups (mean accuracy=99.60%, 99.34%, 99.86%). (3) In two conditions (COPD, THA/TKA), the hierarchical prediction model had a small but statistically significant improvement in discriminating between the readmitted and not readmitted patients as measured by net reclassification improvement (NRI=.059, .11), but not as measured by the C-statistic or integrated discrimination improvement (IDI).

Conclusions

While the visual analytical models identified statistically and clinically significant patient subgroups, the results pinpoint the need to analyze subgroups at different levels of granularity for improving the interpretability of intra- and inter-cluster associations. The high accuracy of the classification models reflects the strong separation of the patient subgroups despite the size and density of the datasets. Finally, the small improvement in predictive accuracy suggests that comorbidities alone were not strong predictors for hospital readmission, and the need for more sophisticated subgroup modeling methods. Such advances could improve the interpretability and predictive accuracy of patient subgroup models for reducing the risk of hospital readmission and beyond.

INTRODUCTION

Background

A wide range of studies [1-9] on topics ranging from molecular to environmental determinants of health have shown that most humans tend to share a subset of characteristics (e.g., comorbidities, symptoms, genetic variants), forming distinct patient subgroups. A primary goal of precision medicine is to identify such patient subgroups and infer their underlying disease processes to design interventions targeted to those processes [2, 10]. For example, recent studies in complex diseases such as breast cancer [3, 4], asthma [5-7] and COVID-19 [11] have revealed patient subgroups, each with different underlying mechanisms precipitating the disease, and therefore each requiring different interventions.

A critical requirement for designing such interventions is the clinical interpretability of patient subgroups. Such interpretability requires clinicians to understand (a) how characteristics (e.g., comorbidities, symptoms, genetic variants) frequently and significantly co-occur across patients, and (b) the risk for adverse outcomes (e.g., mortality, hospital readmission) of patient subgroups that have those co-occurrences. An integration of the co-occurrence of characteristics, with the risk of outcomes in patient subgroups, is critical to infer the disease processes underlying each patient subgroup, and to design precision interventions targeted to those patient subgroups. However, few methods automatically identify both patient subgroups and their co-occurring characteristics simultaneously, which is important for measuring the risk for adverse outcomes and inferring their mechanisms. Such integrated methods could enhance the interpretability of patient subgroups by clinicians for designing interventions, and for informing the design of classification and predictive models that provide clinical decision support.

To address this need, we used a visual analytical method to identify and analyze patient subgroups in hospital readmitted patients. While we have previously demonstrated [12] the use of visual analytics to identify patient subgroups and their characteristics in hospital readmission, here we explore how the approach generalizes across three hospital readmission conditions and its use in classification and predictive modeling. This was done through an analytical framework for Modeling and Interpreting Patient Subgroups (MIPS) which used a three-step modeling approach: (1) *Visual analytical* modeling through bipartite networks to automatically identify patient subgroups and their co-occurring characteristics, and determine their statistical significance and clinical interpretability. (2) *Classification* modeling through multinomial logistic regression to classify patients into subgroups. (3) *Prediction* modeling through logistic regression with and without subgroup information to predict the risk of hospital readmission. Application of the MIPS analytical framework to three datasets helped pinpoint methodological and data limitations in our approach, which provided implications for improving the interpretability of patient subgroups in large and dense datasets, and for the design of clinical decision support systems to prevent adverse outcomes such as hospital readmissions.

Current Approaches for Identifying Patient Subgroups

A patient subgroup is a subset of patients drawn from a population (e.g., older adults) that share one or more characteristics (e.g., renal failure and diabetes). Patients have been divided into subgroups by using (a) investigator-selected variables such as race for developing hierarchical regression models [13], or assigning patients to different arms of a clinical trial, (b) existing classification systems such as the Medicare Severity-Diagnosis Related Group (MS-DRG) [14] to assign patients into a disease category for purposes of billing, and (c) computational methods such as classification [15-17] and clustering [5, 18] to discover patient subgroups from data (also referred to as *subtypes* or *phenotypes* depending on the condition and variables analyzed).

One of the simplest computational methods to identify patient subgroups is by enumerating conjunctions (identify all pairs, all triples, etc.) of variables such as comorbidities [19] that co-occur across patients and then examining the most prevalent subgroups. While such approaches are intuitive, they can lead to a combinatorial explosion (e.g., enumerating combinations of the 31 Elixhauser comorbidities would lead to 2^{31} or 2147483648 combinations), and most combinations do not incorporate the full range of comorbidities (e.g., the most frequent pair of comorbidities ignores other comorbidities that might exist in the profile of patients with that pair). Other approaches use *unipartite* clustering methods [17, 18] (e.g., clustering patients or comorbidities, but

not both together) such as k-means and hierarchical clustering; and dimensionality-reduction methods such as principal component analysis (PCA) [17] to identify *principal components* to define a reduced dimensionality plane on which patients or comorbidities are projected, which are then clustered using unipartite methods such as k-means (together referred to as spectral clustering).

However, because these methods are unipartite, there is no agreed-upon method to identify the patient subgroup defined by a cluster of characteristics, or vice-versa, which substantially reduces the interpretability of the results. Furthermore, such methods have well-known limitations, including (a) requiring a user-defined input for a similarity measure (e.g., Jaccard distance) to calculate the similarity between pairs of patients based on their profiles, or pairs of characteristics based on how their co-occurrence across patients, (b) requiring a user-defined input for the expected number of clusters, and (c) the absence of a quantitative measure to measure the quality of the clustering, critical for measuring the statistical significance of the clustering.

More recent *bipartite* network analytical methods [20] have attempted to address these limitations by automatically identifying biclusters [18, 21, 22] (e.g., clustering of patients and comorbidities simultaneously). A network consists of nodes and edges; nodes represent one or more types of entities (e.g., patients or comorbidities), and edges between the nodes represent a specific relationship between the entities. Figure 1A shows a unipartite network, where nodes are of the same type (often used to analyze co-occurrence of comorbidities [23]). In contrast, Figure 1B shows a bipartite network where nodes are of two types, and edges exist only between different types such as between patients (circles) and comorbidities (triangles). This approach uses bipartite modularity maximization [20, 24-26], a graph-theoretic approach to (a) quantitatively output the number, size, and statistical significance [18, 27] of biclusters (consisting of a patient subgroup and its most frequently co-occurring comorbidities), and (b) visualize those biclusters using layout algorithms [28, 29] to enable their clinical interpretation [11, 12, 30-36]. As shown in Fig. 1C, a bipartite visualization could enable clinicians to inspect the bicluster associations, infer potential mechanisms in each patient subgroup, and design targeted interventions. Our prior use of bipartite networks have enabled three types of discoveries related to subgroups: (1) *novel subtypes* (e.g., in asthma [33]); (2) *frequency of known subtypes* in a new condition (e.g., in COVID-19 [11]), and (3) *risk of subtypes* for adverse outcomes (e.g., in hip fracture hospital readmission [12]). Furthermore, the above subgroups could be used to train classifiers for classifying a new patient into a subgroup, and to build predictive models that leverage such patient subgroups to predict an outcome in a new patient.

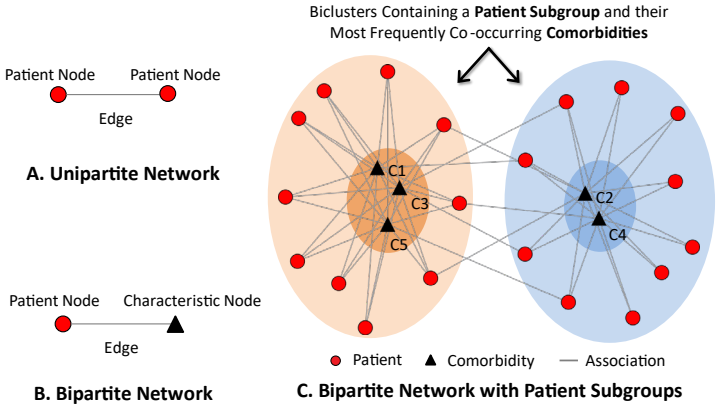

Fig. 1. Comparison between a unipartite (A) and a bipartite network representation (B), and how a bipartite network analysis can automatically identify biclusters containing patient subgroups and their most frequently co-occurring comorbidities (C).

Leveraging Patient Subgroups in Predictive Modeling

Patient subgroups are leveraged in predictive modeling using two common approaches [37] that trade-off simplicity with accuracy: (1) *Hierarchical Modeling* adds subgroup information (e.g., a subgroup membership variable specifying to which subgroup a patient belongs, predicted by a classifier) to a Standard Model without subgroup membership information to improve accuracy. However, while this approach is simple, it potentially trade-offs accuracy as the model's parameters (e.g., slope and intercept of a regression model) are fixed for all patients, regardless of subgroup membership. (2) *Subgroup-Specific Modeling* develops multiple models, one for each subgroup, allowing each model to have different model parameters, potentially improving accuracy. However, this improved accuracy trade-offs simplicity as the evaluation requires several additional steps: build multiple predictive models, predict the outcomes for each patient using the appropriate model (predicted by a

classifier), aggregate the accuracy of predictions across all patients, and compare it to the predictive accuracy of all patients generated from the Standard Model. Given this complexity, we used the simpler Hierarchical Modeling approach as a preliminary step for leveraging patient subgroups.

The Need for Automatic Identification of Patient Subgroups in Hospital Readmission

An estimated one in five elderly patients (over 2.3 million Americans) is readmitted to a hospital within 30-days after being discharged [38]. While many readmissions are unavoidable, an estimated 75% of readmissions are unplanned and mostly preventable [39], imposing a significant burden in terms of mortality, morbidity, and resource consumption. Across all conditions, unplanned readmissions cost almost $17 billion annually in the US [39], making them an ineffective use of costly resources, and therefore closely scrutinized as a marker for the poor quality of care by organizations such as the Centers for Medicare & Medicaid Services (CMS) [40].

To address this epidemic of hospital readmission, CMS sponsored the development of models to predict the patient-specific risk of readmission in specific index conditions such as chronic obstructive pulmonary disease (COPD) [41], congestive heart failure (CHF) [42], and hip/knee arthroplasty (THA/TKA) [43]. These models have two characteristics that are pertinent to the current study:

1. Inclusion of Comorbidities as Independent Variables. The independent variables (predictors) in the CMS models were prior comorbidities (as recorded in Medicare claims data), and demographics (age, gender, and race). The use of comorbidities was based on extensive literature showing the critical role comorbidities play in increasing the risk for adverse outcomes in older adults [38]. For example, almost two-thirds of older adults have two or more comorbid conditions, resulting in a heightened risk for adverse health outcomes such as hospital readmission and mortality [44]. Furthermore, multiple comorbidities often do not act independently, but rather interact with each other, resulting in processes that can precipitate readmission [45]. For example, due to the systemic nature of renal disease, a hip fracture patient with congestive heart failure and renal failure is at a higher risk of renal failure exacerbation, precipitating a hospital readmission, compared to one who only had renal failure [12]. To enable a head-to-head comparison with the CMS predictive models, we used the same independent variables for our predictive models.

2. Exclusion of Patient Subgroups. None of the CMS models used information related to patient subgroups. Therefore, while such models provide the risk of readmission for an individual patient, they do not leverage the existence of patient subgroups known to be present among patients with hospital readmission [12]. Such patient subgroups could be used in hierarchical regression models to potentially achieve higher predictive accuracy. Furthermore, while the primary focus of the CMS models was on predicting the risk of readmission of a patient, they provide little clinical guidance for the design of clinical interventions to address that risk. In contrast, if a patient belongs to a previously-identified patient subgroup with a comorbidity profile (often referred to as a phenotype), such information could be leveraged to classify patients into the best-fitting phenotype, and then to use that classification as a starting point to design clinical interventions targeted to the patient.

Here we demonstrate the development and use of an analytical framework for Modeling and Interpreting Patient Subgroups (MIPS) by using a three-step modeling approach: (1) bipartite networks to automatically identify subgroups of readmitted patients and their frequently co-occurring comorbidities, (2) classifiers to classify patients into a best-fitting subgroup, and (3) hierarchical predictive models which leverage the subgroup information to predict each patient's risk of readmission. This analytical framework was tested across three index conditions where readmission frequently occurs.

METHOD

Overview of Method

Fig. 2 provides a conceptual description of the data inputs and outputs from the three-step modeling in MIPS. As shown, the visual analytical model identifies the patient subgroups, and visualizes them through a network. The classification model predicts subgroup membership for cases and controls, and uses it to measure the risk of readmission within each subgroup based on its proportion of cases. This risk information is juxtaposed with

the visualization to enable clinicians interpret the readmitted patient subgroups. Finally, the predictive model uses the subgroup membership assignment of cases and controls to predict the readmission risk of a patient.

Data Description

Study population. We analyzed patients hospitalized for chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), and total hip/knee arthroplasty (THA/TKA). We selected these three index conditions because: (a) hospitalizations for each of these conditions are highly prevalent in older adults [38]; (b) hospitals report very high variations in their readmission rates [38]; and (c) there exist well-tested readmission prediction models for each of these conditions that did not consider patient subgroups [41-43, 46, 47].

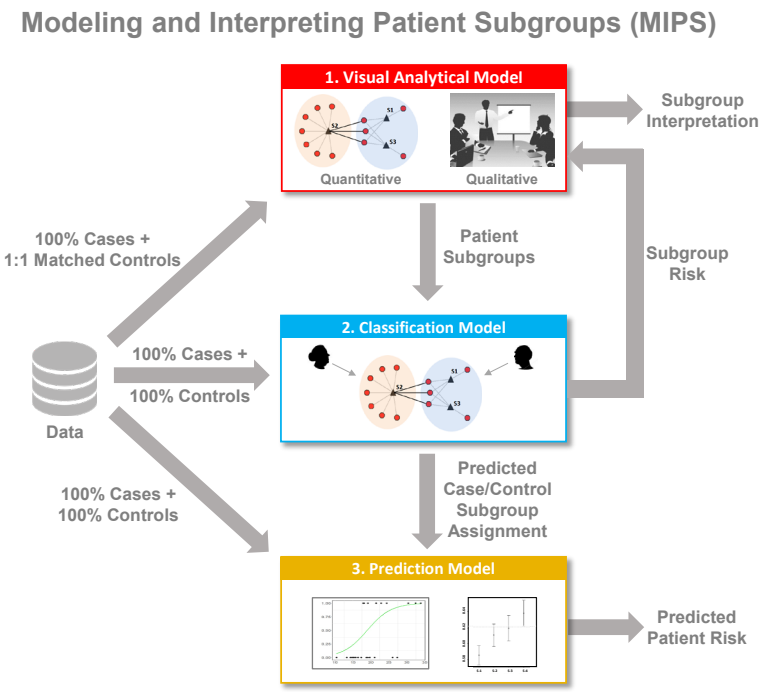

Fig. 2. Inputs and outputs for the three-step modeling in MIPS. The visual analytical model quantitatively identifies the patient subgroups, and visualizes them using a bipartite network. The classification model predicts subgroup membership of cases and controls in addition to the risk of each subgroup, which is juxtaposed with the visualization to enable clinicians to qualitatively interpret the readmission subgroups. The predictive model uses subgroup membership, comorbidities, and demographics to predict the risk of a new patient for being readmitted.

For each index condition, we used the same inclusion and exclusion criteria used to develop the CMS models, but with the most recent years (2013-2014) provided by Medicare when we started the project. We used 100% of the 30-day readmitted patients in 2013 and 2014 Medicare claims data, from which we extracted all patients that were admitted to an acute care hospital on or after July 2013-August 2014 with a principal diagnosis of the index condition, were 66 years of age or older, and were enrolled in both Medicare parts A and B fee-for-service plans in the 6 months before admission. Furthermore, we excluded patients who were transferred from other facilities, died during the hospitalization, or transferred to another acute care hospital. Similar to the CMS models, we selected the first admission for patients with multiple admissions during the study period, and did not use Medicare Part D (related to prescription medications).

Next, we extracted 100% controls who were not readmitted for at least 90 days since discharge. CMS uses this 90-day window of no re-admittance to ensure that the controls are substantially free of complications that result in readmission during this period [48, 49]. A small percentage (0.8%) of Medicare patients had “unknown race” for the Race attribute, so we grouped “unknown race” and “other race” and ensured that there was an equal number of them in the cases and control datasets. The low rate of missing data on race had too low a risk for bias to warrant a sensitivity analysis. Appendix-1 shows the detailed inclusion and exclusion criteria used to extract cases and controls for COPD, CHF, and THA/TKA, and the respective numbers of patients extracted at each step, in addition to the International Classification of Diseases, Ninth Version codes (ICD-9) codes for each of the three index conditions selected for analysis. Each modeling method used appropriate subsets of the above data described in the sections below.

Variables. The dependent variable (outcome) was whether a patient with an index admission (COPD, CHF, THA/TKA) had an unplanned readmission to an acute-care hospital within 30 days of discharge, as was recorded in the MEDPAR file (inpatient claims) in the Medicare database.

The independent variables included comorbidities, and patient demographics (age, gender, race). Comorbidities common in older adults were derived from three established comorbidity indices: Charlson Comorbidity Index (CCI) [50], Elixhauser Comorbidity Index (ECI) [51], and the Center for Medicare and Medicare Services Condition Categories (CMS-CC) used in the CMS readmission models [52] (the variables in the CMS models varied across the index conditions). As these indices had overlapping comorbidities, we derived a union of them, which was verified by the clinician stakeholders. They recommended that we also include the following additional variables as they were pertinent to each index condition: COPD (history of sleep apnea, mechanical ventilation); CHF (history of coronary artery bypass graft surgery); THA/TKA (congenital deformity of the hip joint, post-traumatic osteoarthritis). For each patient in our cohort, we extracted the above comorbidities and variables from the physicians, outpatient, and inpatient Medicare claims data in the 6 months before (to guard against miscoding), and on the day of the index admission.

Analytical and Evaluation Approach

Overview of the MIPS Framework. Table 1 provides a summary of the inputs and outputs of the three-step modeling approach in the MIPS framework, which was applied across the three index conditions.

Visual Analytical Modeling. The data used to build the visual analytical model consisted of 100% cases, and an equal number of 1:1 matched controls extracted by randomly selecting a control without replacement to match each case based on age, gender, race/ethnicity, and Medicaid eligibility [53]. The resulting dataset was divided randomly into a training (50%) and replication (50%) dataset (we use the term *replication* to avoid confusion with the term *validation* typically used in classification and prediction models). We used a bipartite network to

Model	Inputs	Outputs
1. Visual Analytical (Bipartite Network Analysis)	• **Training Dataset**: 50% random sample of 100% cases, and an equal number of 1:1 matched controls (used only for feature selection) • **Replication Dataset**: 50% random sample of 100% cases and equal number of 1:1 matched controls	• **Model Training** Feature Selection: Set of comorbidities univariably significant in both the training and replication datasets Biclustering: Modularity (degree of biclusteredness) and its significance, number of biclusters (subgroups), and their patient and comorbidity members in the training and replication datasets • **Model Replication** Comorbidity Co-Occurrence: Rand Index (degree of replication), and its significance to measure replicability of comorbidity co-occurrence • **Model Interpretation** Visualization: Layout of the bipartite network juxtaposed with risk of individual comorbidities and subgroups Clinical Significance: Interpretation by clinicians for face validity of patient subgroups based on comorbidity co-occurrence, leading to inference of mechanisms precipitating readmission, and interventions
2. Classification (Multinomial Logistic Regression)	• **Training Dataset**: Random sample of 75% cases, with bicluster membership • **Internal Validation Dataset**: Random sample of 25% of cases (with subgroup membership used to validate the model)	• **Model Training** Subgroup Membership: Probability of membership of each case to each subgroup (soft labels), with the highest used to determine subgroup membership (hard labels) • **Model Internal Validation** Internal Validation: Accuracy of classification model based on hard labels • **Model Application** Classification: Subgroup classification of 100% cases and 100% controls Subgroup Risk: Proportion of cases in each subgroup
3. Prediction (Binary Logistic Regression, and Hierarchical Binary Logistic Regression)	• **Training Dataset**: Random samples of 75% of 100% cases and controls, with subgroup membership • **Internal Validation Dataset:** Random sample of 25% of cases and controls (with case/control labels used to validate the model)	• **Model Training** Predicted Risk: Each patient's probability of being readmitted. • **Model Internal Validation** Internal Validation: C-statistic (discrimination), and calibration-in-the-large and calibration slope (calibration) • **Model Comparison** Accuracy: Net Reclassification Improvement (NRI) and Integrated Discrimination Improvement (IDI)

Table 1. Inputs used to train and replicate/validate the three models, and the analytical outputs they produced.

model the cases (30-day readmitted patients) and significant comorbidities in each index condition using the following steps:

A. *Model Training.* The training of the bicluster network model consisted of the following two steps:

I. *Feature Selection.* Given the large number of patients and comorbidities in the dataset, we used feature selection to identify comorbidities with the strongest signal and therefore interpretability for readmission using the following steps: (1) excluded comorbidities with prevalence less than 1% (as is commonly done in studies to reduce noise [54]); (2) selected significant comorbidities in the training dataset based on a 2-way interaction test using odds ratio (OR) with directionality, and correcting for multiple testing using Bonferroni, and (3) tested the surviving comorbidities for replication in the replication dataset, and selected those that were significant in both datasets. Appendix-2 shows the number of comorbidities, and variables that were included in the analysis for each of the three index conditions. The above feature selection generated a single set of significant and replicated comorbidities used for the following bipartite network analysis.

II. *Biclustering.* We used bipartite networks on the training dataset to analyze heterogeneity in readmission using the following steps. (1) Removed all cases that did not have any comorbidities (as the modularity maximization algorithm will trivially put disconnected nodes into a separate cluster). (2) Represented the cases (30-day readmitted patients in the training dataset) and their significant and replicated comorbidities (selected in Step A) as a bipartite network. As shown in Fig. 1, the nodes represented cases or comorbidities, and edges represented which case had which comorbidity. (3) Used a bipartite modularity maximization algorithm to identify the number of biclusters, their boundaries, and degree of biclusteredness using modularity. Modularity is defined as the fraction of edges falling within a cluster, minus the expected fraction of such edges in a network of the same size with randomly assigned edges [20]. Modularity ranges from -0.5 to +1, with values >0 indicating biclustering that is higher than can be expected by chance. We used the bipartite version of modularity [55, 56] to find biclusters in the network. (4) Measured the significance of the bicluster modularity by comparing it to a distribution of the same quantity generated from 1000 random permutations of the network, by preserving the network size (number of nodes) and the network density (number of edges).

B. *Model Replication.* Repeated the above biclustering steps 1-4 to identify subgroups in the replication dataset, and compared the comorbidity co-occurrence in the training dataset, to that in the replication dataset using the Rand index (RI) [57]. RI measures the proportion of comorbidity pairs that co-occurred and did not co-occur in a cluster in the training and replication datasets (where 0=no inter-network cluster similarity, and 1=total inter-network cluster similarity). The significance of RI was measured by comparing it to a distribution of the same quantity generated from 1000 random permutations of the training and replication networks. All tests of statistical significance in Steps A and B were 2-sided.

C. *Model Interpretation.* The model interpretation consisted of the following steps:

I. *Visualization.* We used the following steps to visualize the network generated from the training dataset. (1) Used *Fruchterman-Reingold* (FR) [58], a force-directed algorithm to lay out the bipartite network. This layout algorithm pulls together nodes that are strongly connected, and pushes apart nodes that are not. This results in nodes with a similar pattern of connections to be placed close to each other in Euclidean space, and those that are dissimilar are pushed apart. (2) As the FR algorithm often cannot entirely separate clusters in large and dense networks, the network layout needs to be visually enhanced before it is interpretable by clinician stakeholders. Therefore, we used the *ExplodeLayout* algorithm [28, 29] to separate the biclusters to reduce their visual overlap. This algorithm preserves the distances of nodes within a bicluster, but increases the distance of nodes between clusters to improve interpretability. (3) Juxtaposed the risk of readmission with the network visualization (in response to a request from the clinical stakeholders). This was done by (a) displaying comorbidity labels with their univariable ORs for readmission (measured in Step A) ranked by their odds ratios (ORs) for each subgroup, and (b) measuring the readmission risk for each patient subgroup based on the full case-

control population (explained in more detail in the section on predictive modeling), and juxtaposing it with the respective subgroup.

II. *Clinical Interpretation.* We used the following steps to solicit clinical interpretations of the above bipartite network. (1) Recruited a pulmonologist specializing in COPD and hospital readmission to interpret the COPD results, and a geriatrician with expertise in treating older adults in CHF and THA/TKA to interpret the respective results. (2) Requested each clinician stakeholder to interpret the patient subgroups, their mechanisms, and potential interventions to reduce the risk for readmission.

Classification Modeling. As shown in the bipartite network example in Fig. 1, the biclusters identified through the modularity maximization algorithm contain patient subgroups and their *most frequently* co-occurring comorbidities with respect to other patients in the network. However, there are often many edges between biclusters, revealing that many patients within a bicluster have comorbidities that exist in other biclusters. As is true for most partitioning cluster methods, including modularity, membership of a new patient to each bicluster is therefore *probabilistic*. The classification of a patient into a cluster is therefore not defined by the *inclusion or exclusion of comorbidities* (e.g., hypertension and diabetes), but rather by the *probability* of being in a patient subgroup. Patients are therefore similar or different, not just in a handful of carefully-selected comorbidities while ignoring others, but based on *all* of their recorded comorbidities. This overall profile of patients reflects the reality of comorbid conditions.

To model the above complexity, we used multinomial logistic regression [17] to develop classification models in each index condition. This approach has the advantage of generating probabilities ("soft labels") for a patient to belong to each patient subgroup. The models were trained, internally validated, and then applied to generate information for the other two modeling methods, as described below:

A. *Model Training.* The data used to build the classification model consisted of the training dataset and subgroup membership from the visual analytical model. We trained a multinomial logistic regression model using the above data, with independent variables that included comorbidities identified through feature selection done for the visual analytical modeling. Accuracy of the trained model was measured by calculating the percentage of times the model correctly classifed the cases into the subgroups, using the highest predicted probability across the subgroups ("hard labels").

B. *Model Internal Validation.* To internally validate the classifier, we randomly split the above data into training (75%) and testing (25%) datasets, 1000 times. For each iteration, we trained a model using the training dataset, and measured its accuracy on the testing dataset. This was done by predicting the subgroup membership using the highest predicted probability among all the subgroups. The overall predicted accuracy was then estimated by calculating the mean accuracy across the 1000 models.

C. *Model Application.* Using the 100% cases, in addition to the 100% controls from July 2013-August 2014 (representing the entire Medicare population of each index condtion from those years), we generated the following two types of information for use in the other models. (1) Used the classifier trained in Step A above, to classify 100% cases and 100% controls into a subgroup. This information was used by the subsequent predictive modeling. (2) While the visual analytical model used the 1:1 matched controls for feature selection, this cohort did not represent the entire population. Therefore, to accurately measure the subgroup risk, we used the entire case-control population classified into the subgroups (as described in the above step), and measured the proportion of cases in each subgroup. Furthermore, as requested by the clinicians, we juxtaposed these subgroup risks next to the respective subgroups in the bipartite network visualization, to improve their interpretability.

Predictive Modeling. The data used to build the predictive models consisted of 100% cases and 100% controls, in addition to their subgroup membership generated from the above classification models. These data were randomly spilt into a training (75%) and validation (25%) dataset. The predictive models were trained, internally validated, and compared for predictive accuracy, as described below:

A. *Model Training.* We used the training dataset to train a Standard Model (binary logistic regression without subgroup membership similar to the CMS models), and a Hierarchical Model (binary logistic regression with subgroup membership), with 30-day unplanned readmission (yes vs. no) as the outcome. Independent variables for both models included comorbidities identified through the feature selection in each index condition (see Appendix-2), and demographics. The Hierarchical Model additionally included subgroup membership.

B. *Model Internal Validation.* We used the validation dataset to internally validate the models through the following two measures:

I. Discrimination (model's ability to distinguish readmitted patients from those not readmitted) was measured using the C-statistic, which is identical to the area under the receiver operating characteristic (ROC) curve. Model discrimination was examined using box plots to show the average risk prediction for patients with and without readmission.

II. Calibration (model's agreement of the predicted probabilities with the observed risk) was measured using calibration-in-the-large, and calibration slope, which was examined through a calibration plot showing the proportion of patients actually admitted, versus deciles of predicted probability of having readmission. Good calibration is when calibration-in-the-large is close to zero, and the calibration slope is close to one. Since the large sample size overpowered the study, we did not measure the calibration based on statistical significance (e.g., *P* values of the Hosmer-Lemeshow and calibration indices).

C. *Model Comparisons.* We used the chi-squared test to compare the C-statistic of the Standard Model to that of the Hierarchical Model. We also measured the C-statistic of the Standard Model applied to each subgroup separately. This enabled examination of how the Standard Model performed on patient subgroups to identify, for example, which subgroups underperformed when using the current Standard Model.

Because the above models used the feature selection step to select comorbidities for use as independent variables, they differed from those used in the published CMS models. Therefore, to perform a head-to-head comparison with the published CMS models, we additionally developed a logistic regression model using independent variables that were identical to the published CMS model (CMS Standard Model), which was compared to the same model that included subgroup membership (CMS Hierarchical Model). We used the chi-squared test to compare the C-statistic of the CMS Standard Model to that from the CMS Hierarchical Model, in addition to the following measures of model accuracy:

I. Net Reclassification Improvement (NRI) measured the proportion of patients whose predicted probability of readmission improved with reference to actual readmission status. We used two NRI statistics: (a) categorical NRI, which predicted readmission probabilities divided into 10 sequential categories ranging from 0-1, with improvement requiring a shift between categories; and (b) continuous NRI which is based on the proportions of patients with any improved predicted probability of readmission, regardless of the size of that improvement.

II. Integrated Discrimination Improvement (IDI) measured the difference in the average improvement in predicted risks between the CMS Standard Model and the CMS Hierarchical Model.

RESULTS

Data

Table 2 provides a summary of the number of cases and/or controls used to develop the three models in each condition.

Visual Analytical Modeling

The visual analytical modeling of readmitted patients in all three index conditions produced statistically and clinically significant patient subgroups and their most frequently co-occurring comorbidities, which were significantly replicated. Results from each condition are described below:

COPD. The inclusion and exclusion selection criteria (see Appendix-1) resulted in a training dataset (n=14,508 matched case/control pairs, of which 51 patient pairs with no dropped comorbidities), and a replication dataset (n=14,508 matched case/control pairs, of which 51 patient pairs with no dropped comorbidities), matched by age, sex, race, and Medicaid eligibility (a proxy for economic status). The feature selection method (see Appendix-2) used 45 unique comorbidities identified from a union of the three comorbidity indices, plus 2 condition-specific comorbidities. Of these, 3 were removed because of <1% prevalence. Of the remaining, 30 survived the significance and replication testing with Bonferroni correction. The visual analytical model used these surviving comorbidities (d=30), and cases consisting of CHF readmitted patients with at least one of those comorbidities (n=14,457). As shown in Fig. 3, the bipartite network analysis identified 4 biclusters, each

Model	Training	Replication/ Validation	Total
Visual Analytical*			
COPD (cases/controls)	14,508/14,508	14,508/14,508	29,016/29,016
CHF (cases/controls)	25,775/25,775	25,775/25,775	51,550/51,550
THA/TKA (cases/controls)	8,249/8,249	8,249/8,249	16,498/16,948
Classification			
COPD (cases)	10,842	3,615	14,457
CHF (cases)	19,254	6,418	25,672
THA/TKA (cases)	5,257	1,753	7,010
Prediction			
COPD (cases/controls)	21,692/117,839	7,334/39,176	29,026/157,015
CHF (cases/controls)	38,728/183,093	12,845/61,095	51,573/244,188
THA/TKA (cases/controls)	12,376/255,203	41,44/85,049	16,520/340,252

Table 2. Training and replication/validation datasets used to develop the three models in each of the three index conditions.

*The visual analytical models used 1:1 matched controls for the feature selection, and used only cases for the bipartite networks to analyze heterogeneity in readmission. The numbers shown for the visual analytical models are before removing patients with no comorbidities. The resulting cases-only datasets were used for the classification modelling as shown.

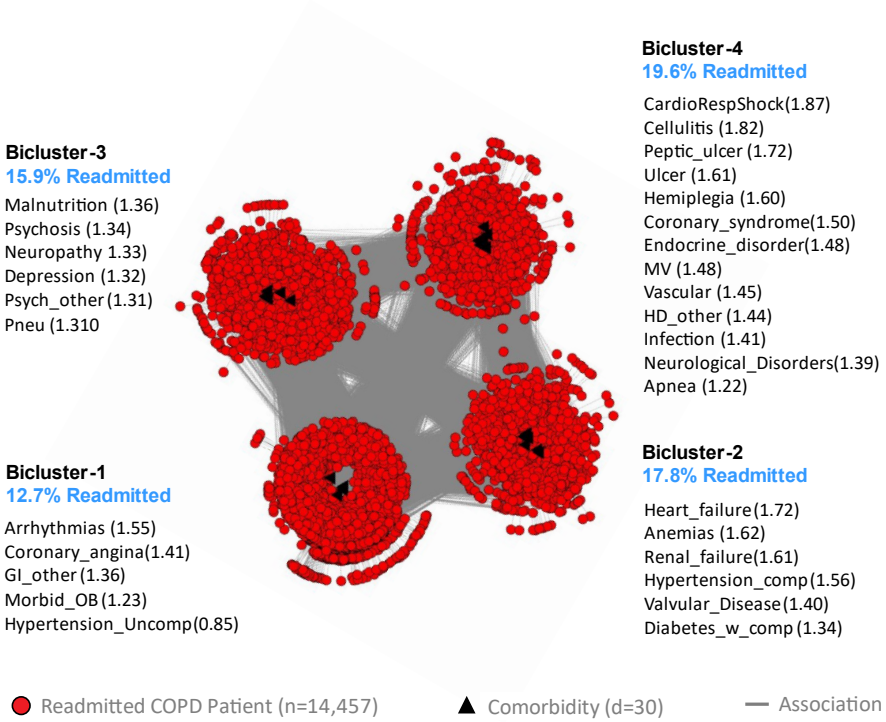

Fig. 3. The COPD visual analytical model showing four biclusters consisting of patient subgroups and their most frequently co-occurring comorbidities (whose labels are ranked by their univariable ORs, shown within parentheses), and their risk of readmission (shown in blue text).

Abbreviations: CardioRespShock, cardiorespiratory shock; COPD, chronic obstructive pulmonary disease; GI, gastrointestinal; Id, identifier; OB, obesity; Pneu, pneumonia; Psych, psychiatric; Uncomp, uncomplicated; HD_other, other and unspecified heart disease; MV, history of mechanical ventilation.

representing a subgroup of readmitted COPD patients and their most frequently co-occurring comorbidities. The biclustering had significant modularity (Q=0.17, z=7.3, *P*<.001), and significant replication (RI=0.92, z=11.62, *P*=<.001) of comorbidity co-occurrence. Furthermore, as requested by the clinician stakeholders, we juxtaposed a ranked list of comorbidities based on their ORs for readmission in each bicluster, in addition to the risk for each of the patient subgroups.

The pulmonologist inspected the visualization and noted that the readmission risk of the patient subgroups had a wide range (12.7% to 19.6%) with clinical (face) validity. Furthermore, the co-occurrence of comorbidities in each patient subgroup was clinically meaningful with interpretations for each subgroup. Subgroup-1 had a low disease burden with uncomplicated hypertension leading to the lowest risk (12.7%). This subgroup represented patients with early organ dysfunction and would benefit from using checklists such as regular monitoring of blood pressure in pre-discharge protocols to reduce the risk of readmission. Subgroup-3 had mainly psychosocial comorbidities, which could lead to aspiration precipitating pneumonia leading to an increased risk for readmission (15.9%). This subgroup would benefit from early consultation with specialists (e.g., psychiatrists, therapists, neurologists, and geriatricians) that had expertise in psycho-social comorbidities, with a focus on the early identification of aspiration risks and precautions. Subgroup-2 had diabetes with complications, renal failure and heart failure and therefore had higher disease burden leading to an increased risk for readmission (17.8%) compared to Subgroup-1. This subgroup had metabolic abnormalities with greater end-organ dysfunction and would therefore benefit from case management from advanced practice providers (e.g., nurse practitioners) with rigorous adherence to established guidelines to reduce the risk of readmission. Subgroup-4 had diseases with end-organ damage including gastro-intestinal disorders, and therefore had the highest disease burden and risk for readmission (19.6%). This subgroup would also benefit from case management with rigorous adherence to established guidelines to reduce the risk of readmission. Furthermore, as patients in this subgroup typically experience complications that could impair their ability to make medical decisions, they should be provided with early consultation with a palliative care team to ensure that care interventions align with patients' preferences and values.

CHF. The inclusion and exclusion selection criteria (see Appendix-1) resulted in a training dataset (n=25,775 matched case/control pairs, of which 103 patient pairs with no dropped comorbidities) and a replication dataset (n=25,775 matched case/control pairs, of which 104 patient pairs with no dropped comorbidities), matched by age, sex, race, and Medicaid eligibility (a proxy for economic status). The feature selection method (see Appendix-2) used 42 unique comorbidities identified from a union of the three comorbidity indices, plus 1 condition-specific comorbidity. Of these, 1 comorbidity was removed because of <1% prevalence. Of those remaining, 37 survived the significance and replication testing with Bonferroni correction. The visual analytical model (Fig. 4) used these surviving comorbidities (d=37), and cases consisting of CHF readmitted patients with at least one of those comorbidities (n=25,672). As shown in Fig. 4, the bipartite network analysis of the CHF cases identified 4 biclusters, each representing a subgroup of readmitted CHF patients and their most frequently co-occurring comorbidities. The analysis revealed that the biclustering had significant modularity (Q=0.17, z=8.69, *P*<.001), and significant replication (RI=0.94, z=17.66, *P*<.001) of comorbidity co-occurrence. Furthermore, as requested by the clinicians, we juxtaposed a ranked list of comorbidities based on their ORs for readmission in each bicluster, in addition to the risk for each of the patient subgroups.

The geriatrician inspected the visualization and noted that the readmission risk of the patient subgroups, ranging from 15.1% to 19.9%, was wide with clinical (face) validity. Furthermore, the co-occurrence of comorbidities in each patient subgroup was clinically meaningful. Subgroup-1 had chronic but stable conditions, and therefore had the lowest risk for readmission (15.1%). Subgroup-3 had mainly psychosocial comorbidities, but were not as clinically unstable or fragile compared to subgroups 2 and 4, and therefore had medium risk (16.6%). Subgroup-2 had severe chronic conditions, making them clinically fragile (with potential benefits from early palliative and hospice care referrals), and were therefore at high risk for readmission if non-palliative approaches were used (19.9%). Subgroup-4 had severe acute conditions which were also clinically unstable, associated with substantial disability and care debility, and therefore at high risk for readmission and recurrent intensive care unit (ICU) use (19.9%).

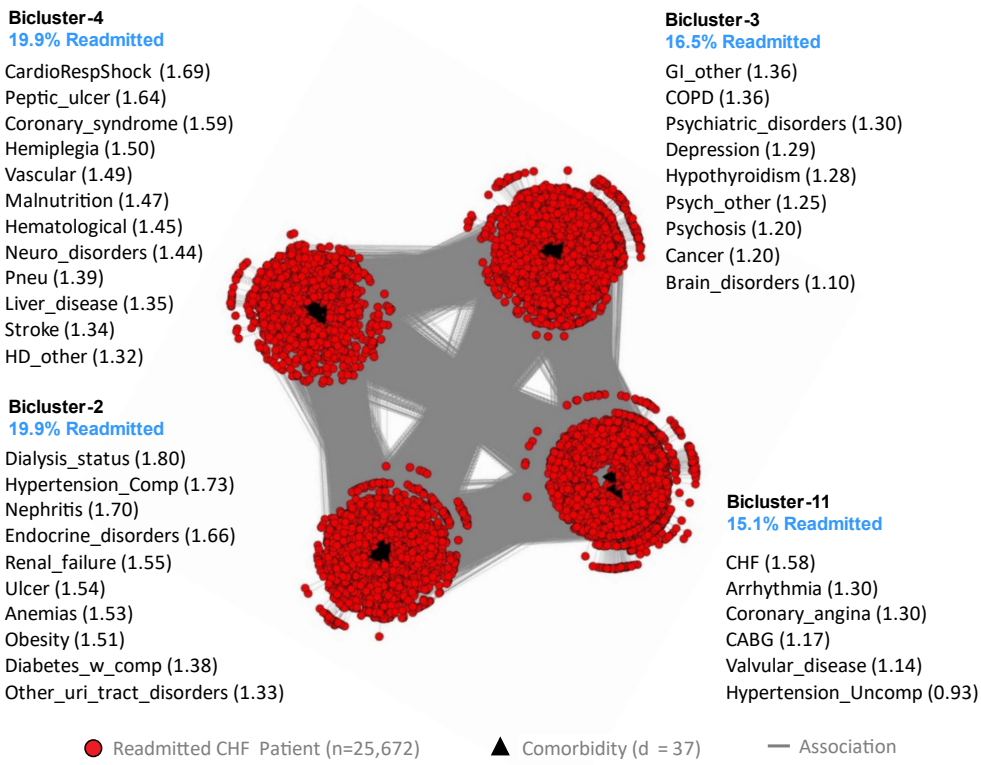

Fig. 4. The CHF visual analytical model showing four biclusters consisting of patient subgroups and their most frequently co-occurring comorbidities (whose labels are ranked by their univariable ORs, shown within parentheses), and their risk of readmission (shown in blue text).

Abbreviations: CABG, coronary artery bypass graft; CardioRespShock, cardiorespiratory shock; CHF, congestive heart failure; comp, complicated; COPD, chronic obstructive pulmonary disease; GI, gastrointestinal; Id, identifier; Neuro, neurologic; OB, obesity; Pneu, pneumonia; Psych, psychiatric; Uncomp, uncomplicated; uri, urinary; w_comp, with complications; HD_other, other and unspecified heart disease.

THA/TKA. The inclusion and exclusion selection criteria (see Appendix-1) resulted in a training dataset (n=8,249 matched case/control pairs, of which 1239 patient pairs with no dropped comorbidities) and a replication dataset (n=8,249 matched case/control pairs, of which 1264 patient pairs with no dropped comorbidities), matched by age, sex, race, and Medicaid eligibility (a proxy for economic status). The feature selection (see Appendix-2) used 39 unique comorbidities identified from the three comorbidity indices plus 2 condition-specific comorbidities. Of these, 11 comorbidities were removed because of <1% prevalence. Of the remaining, 11 survived the significance and replication testing with Bonferroni correction. The visual analytical model (Fig. 5) used these surviving comorbidities (d=11), and cases consisting of readmitted patients with at least one of those comorbidities (n=7,010).

As shown in Fig. 5, the bipartite network analysis of the THA/TKA cases identified 7 biclusters, each representing a subgroup of readmitted THA/TKA patients and their most frequently co-occurring comorbidities. The analysis revealed that the biclustering had significant modularity (Q=0.31, z=2.52, *P*=.011), and significant replication (RI=0.89, z=3.15, *P*=.002) of comorbidity co-occurrence. Furthermore, as requested by the clinician stakeholders, we juxtaposed a ranked list of comorbidities based on their ORs for readmission in each bicluster, in addition to the risk for each of the patient subgroups.

The geriatrician inspected the network and noted that TKA patients, in general, were healthier compared to THA patients, and therefore the network was difficult to interpret when the two index conditions were merged together. While our analysis was constrained because we were using the conditions as defined by CMS, these results nonetheless suggest that the interpretations did not suffer from a confirmation bias (manufactured interpretations to fit the results). However, he noted that the range of readmission risk had clinical (face) validity.

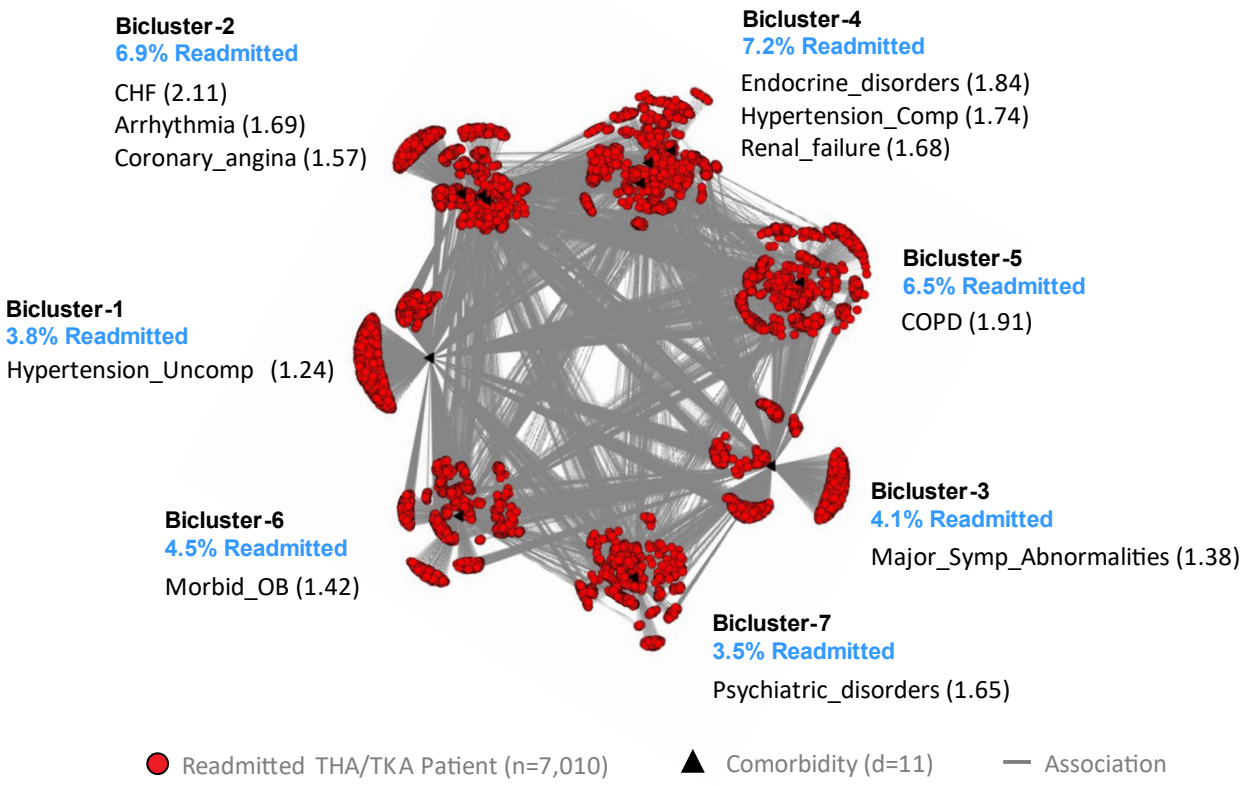

Fig. 5. The THA/TKA visual analytical model showing four biclusters consisting of patient subgroups and their most frequently co-occurring comorbidities (whose labels are ranked by their univariable ORs, shown within parentheses), and their risk for readmission (shown in blue text).

Abbreviations: CHF, congestive heart failure; comp, complicated; COPD, chronic obstructive pulmonary disease; Id, identifier; OB, obesity; Symp, symptom; THA/TKA, total hip/knee arthroplasty; Uncomp, uncomplicated.

Furthermore, subgroups 2, 4, and 5 had more severe comorbidities related to lung, heart, and kidney, and therefore had a higher risk for readmission compared to subgroups 1, 6, and 7 that had less severe comorbidities with a lower risk for readmission. In addition, subgroups 2, 5, 6 and 7 would benefit from chronic care case management from advanced practice providers (e.g., nurse practitioners). Finally, subgroups 2 and 5 could benefit from using well-established guidelines for CHF and COPD, subgroup 7 would benefit from mental health care and management of psycho-social comorbidities, and subgroup 6 would benefit from care for obesity and metabolic disease management.

Classification Modeling

The classification model used multinomial logistic regression in each index condition (see Appendix-3 for the model coefficients) to predict the membership of patients using subgroups (identified from the above visual analytical models). The results revealed that in each index condition, the classification model had high accuracy in classifying all the cases in the full dataset (training dataset used in the visual analytical modeling). Similarly, the internal validation results using a 75%-25% split of the above dataset also had high classification accuracy (Table 3 with classification accuracy divided into quantiles). We report both results for each index condition:

COPD. The model correctly predicted subgroup membership for 99.90% of the cases (14443/14457) in the full dataset. Furthermore**,** as shown in Table 3, the internal validation results revealed that the percentage of COPD cases correctly assigned to a subgroup in the testing dataset, ranged from 99.10% to 100.00%, with a median (Q.50) of 99.60%, and with 95% being in the range from 99.30% to 99.80%.

CHF. The model correctly predicted subgroup membership for 99.20% of the cases (25476/25672) in the full dataset. Furthermore**,** as shown in Table 3, the internal validation results revealed that the percentage of CHF

Models	Quantiles					Summary			
	Q .025	Q .25	Q .50	Q .75	Q .975	Min	Max	Mean	SD
COPD									
Training (n=10842)	99.90	100.00	100.00	100.00	100.00	99.70	100.00	100.00	0.02
Testing (n=3615)	99.30	99.40	99.60	99.60	99.80	99.10	100.00	99.60	0.15
CHF									
Training (n=19254)	99.40	99.50	99.60	99.60	99.80	99.00	99.90	99.57	0.11
Testing (n=6418)	99.00	99.30	99.30	99.40	99.60	98.70	99.70	99.34	0.15
THA/TKA									
Training (n=5257)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	0.00
Testing (n=1753)	99.70	99.80	99.90	99.90	100.00	99.40	100.00	99.86	0.09

Table 3. Internal validation results showing the percentage of COPD, CHF, and THA/TKA patients correctly-assigned to a subgroup by the classification models in each condition.

cases correctly assigned to a subgroup in the testing dataset, ranged from 98.70% to 99.70%, with a median (Q.50) of 99.30%, and with 95% being in the range from 99.00% to 99.60%.

THA/TKA. The model correctly predicted subgroup membership 100.00% of the cases (7010/7010) in the full dataset. Furthermore**,** as shown in Table 3, the internal validation results revealed that the percentage of CHF cases correctly assigned to a subgroup in the testing dataset, ranged from 99.40% to 100.00%, with a median (Q.50) of 99.90%, and with 95% being in the range from 99.70% to 100.00%.

Application of the Classification Model to Generate Information for Other Models. The above classification model was used to classify 100% cases and 100% controls for use in the prediction model (described below). Furthermore, the proportion of cases and controls classified into each subgroup was used to calculate the risk of readmission for each subgroup (see Appendix 3). As this subgroup risk information was requested by the clinicians to improve interpretability of the visual analytical model, the values were juxtaposed next to the respective subgroups in the bipartite network visualizations (see blue text in Fig. 3-5).

Prediction Modeling

For each of the three index conditions, we developed two binary logistic regression models to predict readmission, with comorbidities in addition to sex, age, and race: (1) **Standard Model** representing all patients without subgroup membership, similar to the CMS models; and (2) **Hierarchical Model** with an additional variable that adjusted for subgroup membership.

COPD. The inclusion and exclusion selection criteria (see Appendix-1) resulted in a cohort of 186,041 patients (29,026 cases and 157,015 controls). As shown in Fig. 6A, the Standard Model had a C-statistic of 0.624 (95% CI: 0.617-0.631) which was not significantly (*P*=.8578) different from the Hierarchical Model that had a C-statistic of 0.625 (95% CI: 0.618-0.632). The calibration plots revealed that both models had a slope close to 1, and an intercept close to 0 (see Appendix-4).

As shown in Fig. 6B, the Standard Model was used to measure the predictive accuracy of patients in each subgroup separately. The results showed that Subgroup-1 had a lower C-statistic compared to Subgroup-3 and Subgroup-4. While the C-statistics in Fig. 6A and Fig. 6B cannot be compared as they are

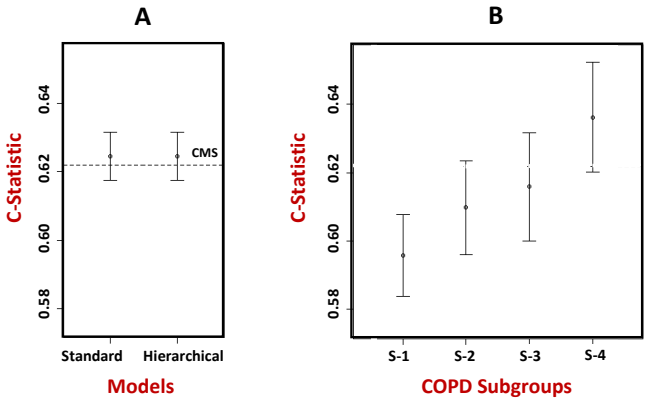

Fig. 6. **(A)** Predictive accuracy of the Standard Model compared to the Hierarchical model in COPD, as measured by the C-Statistic. The C-statistic for the CMS published model is shown as a dotted line. **(B)** Predictive accuracy of the Standard Model when applied separately to patients classified to each subgroup. S-1 has lower accuracy compared to S-3 and S-4. (C-statistics in A and B cannot be compared as they are based on models from different populations).

based on models developed from different populations, these results reveal that the current CMS readmission model for CHF might be underperforming for one COPD patient subgroup, pinpointing which one might benefit by a Subgroup-Specific Model.

CHF. The inclusion and exclusion selection criteria (see Appendix-1) resulted in a cohort of 295,761 patients (51,573 cases and 244,188 controls). As shown in Fig. 7A, the Standard Model had a C-statistic of 0.600 (95% CI: 0.595-0.605), which was not significantly different (*P*=.2864) from the Hierarchical Model that also had a C-statistic of 0.600 (95%CI: 0.595-0.605). The calibration plots revealed that all models had a slope close to 1, and an intercept close to 0 (see Appendix-4).

As shown in Fig. 7B, the Standard Model was used to measure the predictive accuracy of patients in each subgroup separately. The results showed that Subgroup-1 had a lower C-statistic compared to Subgroup-4. While the C-statistics in Fig. 7A and Fig. 7B cannot be compared as they are based on models developed from different populations, but similar to the results in COPD, these results reveal that the current CMS readmission model for CHF might be underperforming for one CHF patient subgroup, pinpointing which one might benefit by a Subgroup-Specific Model.

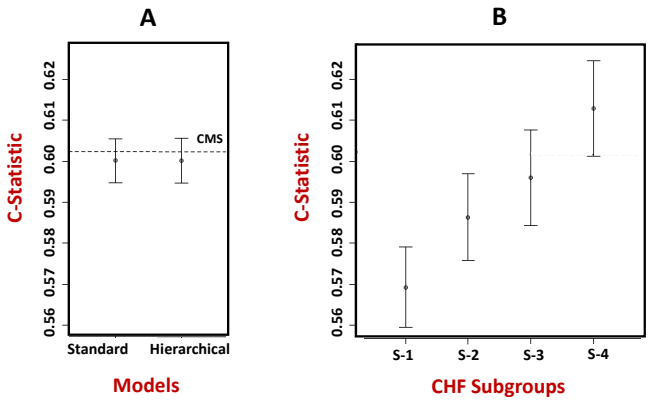

Fig. 7. **(A)** Predictive accuracy of the Standard Model compared to the Hierarchical model in CHF as measured by the C-Statistic. The C-statistic for the CMS published model is shown as a dotted line. **(B)** Predictive accuracy of the Standard Model when applied separately to patients classified to each subgroup. S-1 has lower accuracy compared to S-3 and S-4. (C-statistics in A and B cannot be compared as they are based on models from different populations).

THA/TKA. The application of the inclusion and exclusion selection criteria (see Appendix-1) resulted in a cohort of 356,772 patients (16,520 cases and 340,252 controls). As shown in Fig. 8A, the Standard Model had a C-statistic of 0.638 (95% CI: 0.629-0.646), which was not significantly different (*P*=.6817) from the Hierarchical Model that had a C-statistic of 0.638 (95% CI: 0.629-0.647). The calibration plots (see Appendix-4) revealed that both models had a slope close to 1, and an intercept close to 0 (see Appendix-4).

As shown in Fig. 8B, the Standard Model was used to measure the predictive accuracy of patients in each subgroup separately. The results showed that Subgroup-1 had a lower C-statistic compared to Subgroup-4. Again, while the C-statistics in Fig. 8A and Fig. 8B cannot be compared as they are based on models developed from different populations, similar to the results in COPD, these results reveal that the current CMS readmission model for THA/TKA might be underperforming for 4 patient subgroups, pinpointing which ones might benefit by Subgroup-Specific Models.

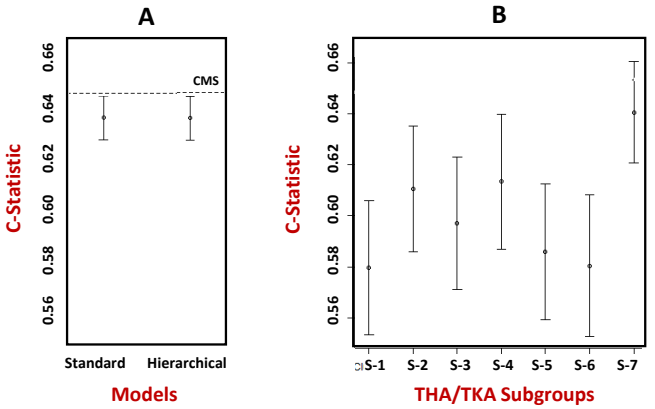

Fig. 8. **(A)** Predictive accuracy of the Standard Model compared to the Hierarchical model in THA/TKA as measured by the C-Statistic. The C-statistic for the CMS published model is shown as a dotted line. **(B)** Predictive accuracy of the Standard Model when applied separately to patients classified to each subgroup. S-1 has lower accuracy compared to S-7. (C-statistics in A and B cannot be compared as they are based on models developed from different populations).

CMS Standard Model vs. CMS Hierarchical Model. Unlike the CMS published models, the above models used only the comorbidities that survived feature selection. Therefore, to perform a head-to-head comparison with the published CMS models, we also developed a CMS Standard Model (using the same variables from the published CMS model), and

Model	NRI		IDI
	Categorical (95% Interval)	Continuous (95% Interval)	IDI (95% Interval)
COPD	.023 (.012, .034)*	.059 (.034, .083)*	.0002 (-.0004, .0008)
CHF	-.010 (-.016, -.0004)*	-.038 (-.057, -.019)*	-.0006(-.0009,- .0003)*
THA/TKA	.022 (.012, .032)*	.111 (.080, .142)*	-.003(-.002,- .003)*

Table 4. Comparison of the CMS Standard Model with the CMS Hierarchical Model across the three index conditions based on NRI and IDI (* = significant at the .05 level).

compared it to the corresponding CMS Hierarchical Model (with an additional variable for subgroup membership) in each condition. Similar to the models in Fig. 6-8, there were no significant differences in the C-statistics between the two modeling approaches in any condition (see Appendix-4). However, as shown in Table 4, the CMS Hierarchical Model for COPD had significantly higher NRI, but not significantly higher NDI compared to the CMS Standard Model; the CMS Hierarchical Model for CHF had a significantly lower NRI and IDI compared to the CMS Standard Model, and the CMS Hierarchical Model for THA/TKA had a significantly higher NDI and IDI compared to the CMS Standard Model. Furthermore, similar to the results in 6B-8B, when the CMS Standard Model was used to predict readmission separately in subgroups within each index condition, it identified subgroups that underperformed, pinpointing which ones might benefit by a Subgroup-Specific Model (See Appendix-4). In summary, the comparisons between the CMS Standard Models and the respective CMS Hierarchical Models showed that in two conditions (COPD and THA/TKA), there was a small but statistically significant improvement in discriminating between the readmitted and not readmitted patients as measured by NRI, but not as measured by the C-statistic or IDI, and that a subgroup in each index condition might be underperforming when using the CMS Standard Model.

DISCUSSION

Overview

Our overall approach of using the MIPS framework for identifying patient subgroups through visual analytics, and using those subgroups to build classification and prediction models, revealed strengths and limitations for each modeling approach, and for our data source. This examination led to insights for developing future clinical decision support systems, and a methodological framework for improving the clinical interpretability of subgroup modeling results.

Strengths and Limitations of Modeling Methods and Data Source

Visual Analytical Modeling. The results revealed three strengths of the visual analytical modeling: (1) the use of bipartite networks to simultaneously model patients and comorbidities, enabled the automatic identification of patient-comorbidity biclusters, and the integrated analysis of co-occurrence and risk; (2) the use of a bipartite modularity maximization algorithm to identify the biclusters enabled the measurement of the strength of the biclustering, critical for gauging its significance; and (3) the use of a graph representation enabled the results to be visualized through a network. Furthermore, the request from the domain experts to juxtapose the risk of each subgroup with their visualizations appeared to be driven by a need to reduce working memory loads (from having to remember that information spread over different outputs), which could have enhanced their ability to match bicluster patterns with chunks (previously-learned patterns of information) stored in long-term memory. The resulting visualizations enabled them to recognize subtypes based on co-occurring comorbidities in each subgroup, reason about the processes that precipitate readmission based on the risk of each subtype relative to the other subtypes, and propose interventions that were targeted to those subtypes and their risks. Finally, the fact that the geriatrician could not fully interpret the THA/TKA network because it mixed two fairly different conditions, suggests that the clinical interpretations were not the result of a *confirmation bias* (interpretations leaning towards fitting the results).

However, the results also revealed two limitations: (1) while modularity is estimated using a closed-form equation (formula), no closed-form equation exists to estimate the modularity variance, which is necessary to measure its significance. To estimate modularity variance, we therefore used a permutation test by generating

1000 random permutations of the data, and then compared the modularity generated from the real data to the mean modularity generated from the permuted data. Given the size of our datasets (ranging from 7K-25K patients), this computationally-expensive test took approximately 7 days to complete, despite the use of a dedicated server with multiple cores; and (2) while bicluster modularity was successful in identifying significant and meaningful patient-comorbidity biclusters, the visualizations themselves were extremely dense, and therefore potentially concealed patterns within and between the subgroups. Future research should explore a closed-form equation to estimate modularity variance, with the goal of accelerating the estimation of modularity significance, and more powerful analytical and visualization methods to reveal intra- and inter-cluster associations in large and dense networks.

Classification Modeling. The results revealed two strengths of the classification modeling: (1) the use of a simple multinomial classifier was adequate to predict with high accuracy to which subgroup a patient belonged; (2) because the model produced membership probabilities for each patient for each subgroup, the model captured the dense inter-cluster edges observed in the network visualization; and (3) the coefficients of the trained classifier could be inspected by an analyst making it more transparent (relative to most deep-learning classifiers which tend to be a black box).

However, because we dichotomized the classification probabilities into a single subgroup membership, our approach did not fully leverage the membership probabilities for modeling and visual interpretation. For example, some patients have high classification probabilities (representing strong membership) to a single subgroup (as shown by patients in the outer periphery of the biclusters with edges only within their bicluster), whereas others have equal probabilities to all subgroups (as shown in the inner periphery of the biclusters with edges going to multiple clusters). Future research should explore incorporating the probability of subgroup membership into the design of hierarchical models to improve predictive accuracy, and visualization methods to help clinicians interpret patients with different profiles of membership strength, with the goal of designing patient-specific interventions.

Predictive Modeling. The results revealed two strengths of the predictive modeling: (1) the use of the Standard Model to measure predictive accuracy across the subgroups helped to pinpoint which subgroups tend to have lower predictive accuracy compared to the rest, and therefore which of them could benefit from a more complex but accurate subgroup-specific model; and (2) despite the use of a simple Hierarchical Model with a dichotomized membership label for each patient, the predictive CMS models detected significant differences in the prediction accuracy as measured by NRI in two of the conditions, when compared to the CMS Standard Models. However, the results also revealed that the differences in predictive accuracy as measured by the C-statistic and NDI were small, suggesting that comorbidities on their own were potentially insufficient for accurately predicting readmission. Future research should explore the use of electronic health records, and multiple subgroup-specific models targeted to each subgroup, to improve the predictive accuracy of the models.

Data Source. The Medicare claims data had four key strengths: (1) scale of the datasets which enabled subgroup identification with sufficient statistical power; (2) spread of the data collected from across the US which enabled generalizability of the results; (3) data about older adults which enabled examination of subgroups in an underrepresented segment of the US population; and (4) data used by CMS to build predictive readmission models, which enabled a head-to-head comparison with the hierarchical modeling approach.

However, the data had two critical limitations. (1) As we compared our models with the CMS models, we had to use the same definition for controls (90 days with no readmission) that had been used, which introduced a selection bias that exaggerates the separation between cases and controls. Similarly, by excluding patients who died, this exclusion criterion potentially biased the results towards healthier patients. (2) Administrative data have known limitations such as the lack of comorbidity severity and test results, which could strongly impact the accuracy of predictive models. Future research should consider the use of national-level electronic health record (EHR) data such as those being assembled by the National COVID Cohort Collaborative (N3C) [59], and the TriNetX [60] initiatives, which could overcome the above limitations by providing laboratory values and comorbidity severity, but could also introduce new as yet unknown limitations.

Implications for Clinical Decision-Support Systems that Leverage Patient Subgroups

While the focus of this project was to develop and evaluate the MIPS framework, its application to three index conditions coupled with extensive discussions with clinicians led to insights for designing a future clinical decision support system. Such a system could integrate outputs from all three models. As we have shown, the visual analytical model automatically identified and visualized the patient subgroups, which enabled the clinicians to comprehend the co-occurrence and risk information in the visualization, reason about the processes that lead to readmission in each subgroup, and design targeted interventions. The classification model leveraged the observation that many patients have comorbidities in other biclusters (shown by a large number of edges between biclusters), and accordingly generated a membership probability of a patient belonging to each bicluster, from which the highest was chosen for bicluster membership. Finally, the predictive model predicts the risk for readmission for a patient, by using in the future the most accurate model designed for the bicluster to which the patient belongs.

The outputs from the above models could be integrated into a clinical decision support system to provide recommendations for a specific patient using the following algorithm: (1) use the classifier to generate the membership probability (MP) of a new patient belonging to each subgroup; (2) multiply the MP in each subgroup with the patient's risk (R) for readmission provided by the predictive model for that subgroup, to generate an importance score [IS = f(MP) X g(R)] for the respective intervention; (3) rank the subgroups and their respective interventions using IS; and (4) use the ranking to display in descending order, the subgroup comorbidity profiles along with their respective potential mechanisms, recommended treatments, and the respective IS. Such model-based information, displayed through a user-friendly interface, could enable a clinician to rapidly scan the ranked list to (a) determine why a specific patient's profile fits into one or more subgroups, (b) review the potential mechanisms and interventions ranked by their importance, and (c) use the combined information to design a treatment that is customized for the real-world context of the patient. Consequently, such a clinical decision support system could not only provide a quantitative ranking of membership to different subgroups, and the importance score for the associated interventions, but also enable the clinician to understand the rationale underlying those recommendations, making the system interpretable and explainable. Comparative evaluation of such a system to standard care could determine its clinical efficacy.

Implications for Analytical Granularity to Enhance the Interpretability of Patient Subgroups

While the visual analytical model enabled the clinicians to interpret the patient subgroups, they were unable to interpret the associations within and between the subgroups due to the large number of nodes in each bicluster and the dense edges between them. Several network filtering methods [61, 62] have been developed to "thin out" such dense networks such as by dropping or bundling nodes and edges based on user-defined criteria, to improve visual interpretation. However, such filtering could bias the results, or modify the clusters resulting from the reduced data.

An alternate approach that preserves the full dataset leverages the notion of analytical granularity, where the data is progressively analyzed at different levels. For example, we have analyzed COVID-19 patients [11] at the cohort, subgroup, and patient levels, and we are currently using the same approach to examine symptom co-occurrence and risk at each level in Long COVID patients. Our preliminary results suggest that analyzing data at different levels of granularity enables clinicians to progressively interpret patterns such as within and between subgroups, in addition to guiding the systematic development of new algorithms. For example, at the subgroup level, we have designed an algorithm that identifies which patient subgroups have a significantly higher probability for having characteristics that are clustered in another subgroup, providing critical information to clinicians about how to design interventions for such overlapping subgroups; at the patient level, we have identified patients that are outliers to their subgroups based on their pattern of characteristics inside and outside their subgroup. Such patient outliers could be flagged to examine if they need individualized interventions versus those recommended for the rest of their subgroup. Such analytical granularity could therefore inform the design of interventions by clinicians, in addition to the design of decision support systems that provide targeted and interpretable recommendations to physicians, who can then customize them to fit the real-world context of a patient.

Implications of the MIPS Framework for Precision Medicine

While we have demonstrated the application of the MIPS framework across multiple readmission conditions, its architecture has three properties that should enable its generalizability across other medical conditions. First, as shown in Fig. 2, the framework is *modular* with explicit inputs and outputs, enabling the use of other methods at each of the three modeling steps. For example, the framework could use other biclustering (e.g., Non-negative Matrix Factorization [63]), classification (e.g., deep learning [64]), and prediction methods (e.g., subgroup-specific modeling [17]). Second, the framework is *extensible,* enabling an elaboration of the methods at each modeling step to improve the analysis and interpretation of subgroups. For example, as discussed above, the analytical granularity at the cohort, subgroup, and patient levels could improve the interpretability of subgroups in large and dense datasets. Third, the framework is *integrative* as it systematically combines the strengths of machine learning, statistical, and precision medicine approaches. For example, the visual analytical modeling leverages search algorithms to discover co-occurrence in large datasets; the classification and prediction modeling leverages probability theory to measure the risk of co-occurrence patterns; and clinicians leverage medical knowledge and human cognition to interpret patterns of co-occurrence and risk for designing precision-medicine interventions. Such integration of different models and their interpretation operationalizes *team-centered informatics* [65] designed to facilitate data scientists, biostatisticians, and clinicians in multidisciplinary translational teams [66] to work more effectively across disciplinary boundaries, with the goal of designing interventions for precision medicine. Our current research tests the generality of the MIPS framework in other conditions such as Long COVID and Post-Stroke Depression, with the goal of designing and evaluating precision medicine interventions targeted to patient subgroups.

CONCLUSIONS

Although a primary goal of precision medicine is to identify patient subgroups and to design targeted interventions, few methods automatically identify both patient subgroups and their co-occurring characteristics simultaneously, measure their significance, and visualize the results. Here we demonstrated the use of the MIPS framework, which used a three-step approach to model and interpret patient subgroups. A visual analytical method automatically identified statistically significant and replicated patient subgroups and their frequently co-occurring comorbidities. Next, a multinomial logistic regression classifier had high accuracy in correctly classifying patients into the patient subgroups identified by the visual analytical model. Finally, despite using a simple hierarchical logistic regression model to incorporate subgroup information, the predictive models had a statistically significant improvement in discriminating between the readmitted and not readmitted patients in two of the three readmission conditions, and further analysis pinpointed for which patient subgroups the current CMS model might be underperforming. Finally, by integrating the co-occurrence and risk patterns in a visualization, the MIPS framework enabled clinicians to interpret the patient subgroups, reason about mechanisms precipitating hospital readmission, and design targeted interventions.

However, evaluation of the methods across three readmission index conditions also helped to identify limitations of the models and the data. The visual analytical model was too dense to enable the clinicians to interpret the associations within and between the subgroups, and the absence of a closed-form equation to measure modularity variance required a computationally-expensive process to measure the significance of the biclustering. Furthermore, the small improvement in predictive accuracy suggested that comorbidities on their own were insufficient for predicting hospital readmission.

By leveraging the modular and extensible nature of the MIPs framework, future research should address the above limitations by developing more powerful algorithms which analyze subgroups at different levels of granularity to improve the interpretability of intra- and inter-cluster associations, and the evaluation of subgroup-specific models to predict outcomes. Furthermore, EHR data made available through national-level data initiatives such as N3C and TriNetX now provide access to critical variables including laboratory results and comorbidity severity, which should lead to higher predictive power for predicting adverse outcomes. Finally, extensive discussions with clinicians provided implications for the design of future decision support systems, which could integrate outputs from the three models to provide for a specific patient, predicted subgroup memberships, ranked interventions, along with associated subgroup profiles and mechanisms. Such

interpretable and explainable systems could enable clinicians to use patient subgroup information for informing the design of precision medicine interventions, with the goal of reducing adverse outcomes such as unplanned hospital readmissions and beyond.

AKNOWLEDGEMENTS

We thank Tianlong Chen, Clark Andersen, Yu-Li Lin, and Emmanuel Santillana for performing the analyses on this project. This study was supported in part by the Patient-Centered Outcomes Research Institute (ME-1511-33194), the Clinical and Translational Science Award (UL1 TR001439) from the National Center for Advancing Translational Sciences at the National Institutes of Health, and by the National Library of Medicine (R01 LM012095) at the National Institutes of Health. The content is solely the responsibility of the authors, and does not necessarily represent the official views of the Patient-Centered Outcomes Research Institute, or the National Institutes of Health. The Medicare data were analyzed using a CMS data-use agreement (CMS DUA RSCH-2017-51404).

REFERENCES

1. McClellan J, King M-C. Genetic Heterogeneity in Human Disease. Cell. 141(2):210-7. doi: 10.1016/j.cell.2010.03.032.
2. Waldman SA, Terzic A. Therapeutic targeting: a crucible for individualized medicine. Clinical Pharmacology & Therapeutics. 2008;83(5):651–4.
3. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clinical cancer research: an official journal of the American Association for Cancer Research. 2005;11(16):5678-85. Epub 2005/08/24. doi: 10.1158/1078-0432.ccr-04-2421. PubMed PMID: 16115903.
4. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(19):10869-74. Epub 2001/09/13. doi: 10.1073/pnas.191367098. PubMed PMID: 11553815; PubMed Central PMCID: PMCPmc58566.
5. Fitzpatrick AM, Teague WG, Meyers DA, Peters SP, Li X, Li H, et al. Heterogeneity of severe asthma in childhood: confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. The Journal of allergy and clinical immunology. 2011;127(2):382-9.e1-13. Epub 2011/01/05. doi: 10.1016/j.jaci.2010.11.015. PubMed PMID: 21195471; PubMed Central PMCID: PMCPmc3060668.
6. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster analysis and clinical asthma phenotypes. American journal of respiratory and critical care medicine. 2008;178(3):218-24. Epub 2008/05/16. doi: 10.1164/rccm.200711-1754OC. PubMed PMID: 18480428; PubMed Central PMCID: PMCPmc3992366.
7. Lotvall J, Akdis CA, Bacharier LB, Bjermer L, Casale TB, Custovic A, et al. Asthma endotypes: a new approach to classification of disease entities within the asthma syndrome. The Journal of allergy and clinical immunology. 2011;127(2):355-60. Epub 2011/02/02. doi: 10.1016/j.jaci.2010.11.037. PubMed PMID: 21281866.
8. Nair P, Pizzichini MMM, Kjarsgaard M, Inman MD, Efthimiadis A, Pizzichini E, et al. Mepolizumab for Prednisone-Dependent Asthma with Sputum Eosinophilia. New England Journal of Medicine. 2009;360(10):985-93. doi: doi:10.1056/NEJMoa0805435. PubMed PMID: 19264687.
9. Ortega HG, Liu MC, Pavord ID, Brusselle GG, FitzGerald JM, Chetta A, et al. Mepolizumab Treatment in Patients with Severe Eosinophilic Asthma. New England Journal of Medicine. 2014;371(13):1198-207. doi: 10.1056/NEJMoa1403290.
10. Collins FS, Varmus H. A new initiative on precision medicine. The New England journal of medicine. 2015;372(9):793-5. Epub 2015/01/31. doi: 10.1056/NEJMp1500523. PubMed PMID: 25635347.
11. Bhavnani SK, Kummerfeld E, Zhang W, Kuo Y-F, Garg N, Visweswaran S, et al. Heterogeneity in COVID-19 Patients at Multiple Levels of Granularity: From Biclusters to Clinical Interventions. Proceedings of the American Medical Informatics Association Summits. 2021:112-21. doi: PMID: 34457125.
12. Bhavnani SK, Dang B, Penton R, Visweswaran S, Bassler KE, Chen T, et al. How High-Risk Comorbidities Co-Occur in Readmitted Patients With Hip Fracture: Big Data Visual Analytical Approach. JMIR Med Inform. 2020;8(10):e13567. doi: 10.2196/13567.
13. Lacy ME, Wellenius GA, Carnethon MR, Loucks EB, Carson AP, Luo X, et al. Racial Differences in the Performance of Existing Risk Prediction Models for Incident Type 2 Diabetes: The CARDIA Study. Diabetes care. 2015. Epub 2015/12/03. doi: 10.2337/dc15-0509. PubMed PMID: 26628420.
14. Baker JJ. Medicare payment system for hospital inpatients: diagnosis-related groups. Journal of health care finance. 2002;28(3):1-13. Epub 2002/06/25. PubMed PMID: 12079147.
15. Lipkovich I, Dmitrienko A, Denne J, Enas G. Subgroup identification based on differential effect search--a recursive partitioning method for establishing response to treatment in patient subpopulations. Statistics in medicine. 2011;30(21):2601-21. Epub 2011/07/26. doi: 10.1002/sim.4289. PubMed PMID: 21786278.
16. Kehl V, Ulm K. Responder identification in clinical trials with censored data. Comput Stat Data Anal. 2006;50(5):1338-55. doi: 10.1016/j.csda.2004.11.015.

17. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York, NY, USA: Springer New York Inc.; 2001.
18. Abu-jamous B, Fa R, Nandi AK. Integrative Cluster Analysis in Bioinformatics. Chichester, West Sussex, United Kingdom: John Wiley & Sons, Ltd.; 2015.
19. Lochner KA, Cox CS. Prevalence of multiple chronic conditions among Medicare beneficiaries, United States, 2010. Preventing chronic disease. 2013;10:E61. Epub 2013/04/27. doi: 10.5888/pcd10.120137. PubMed PMID: 23618541; PubMed Central PMCID: PMCPmc3652723.
20. Newman MEJ. Networks: An Introduction. Oxford, United Kingdom: Oxford University Press; 2010.
21. Shabalin AA, Weigman VJ, Perou CM, Nobel AB. Finding large average submatrices in high dimensional data. 2009:985-1012. doi: 10.1214/09-AOAS239.
22. Odibat O, Reddy CK. Efficient Mining of Discriminative Co-clusters from Gene Expression Data. Knowledge and information systems. 2014;41(3):667-96. Epub 2015/02/03. doi: 10.1007/s10115-013-0684-0. PubMed PMID: 25642010; PubMed Central PMCID: PMCPmc4308820.
23. Folino F, Pizzuti C, Ventura M. A comorbidity network approach to predict disease risk. Proceedings of the First international conference on Information technology in bio- and medical informatics; Bilbao, Spain. 1885260: Springer-Verlag; 2010. p. 102-9.
24. Newman MEJ. Modularity and community structure in networks. Proceedings of the National Academy of Sciences. 2006;103(23):8577-82. doi: 10.1073/pnas.0601602103.
25. Newman MEJ. Fast algorithm for detecting community structure in networks. Physical Review E. 2004;69(6):066133.
26. Trevino III S, Nyberg A, Del Genio CI, Bassler KE. Fast and accurate determination of modularity and its effect size. J Stat Mech. 2015;P02003.
27. Chauhan R, Ravi J, Datta P, Chen T, Schnappinger D, Bassler KE, et al. Reconstruction and topological features of the sigma factor regulatory network of Mycobacterium tuberculosis. In Review.
28. Dang B, Chen T, Bassler KE, Bhavnani SK. ExplodeLayout: Enhancing the Comprehension of Large and Dense Networks. AMIA Jt Summits Transl Sci Proc 2016.
29. Bhavnani SK, Chen T, Ayyaswamy A, Visweswaran S, Bellala G, Rohit D, et al. Enabling Comprehension of Patient Subgroups and Characteristics in Large Bipartite Networks: Implications for Precision Medicine. Proceedings of AMIA Joint Summits on Translational Science. 2017:21-9. Epub 2017/08/18. PubMed PMID: 28815099; PubMed Central PMCID: PMCPMC5543384.
30. Bhavnani SK, Eichinger F, Martini S, Saxman P, Jagadish HV, Kretzler M. Network analysis of genes regulated in renal diseases: implications for a molecular-based classification. BMC bioinformatics. 2009;10 Suppl 9:S3. Epub 2009/09/26. doi: 10.1186/1471-2105-10-s9-s3. PubMed PMID: 19761573; PubMed Central PMCID: PMCPMC2745690.
31. Bhavnani SK, Bellala G, Ganesan A, Krishna R, Saxman P, Scott C, et al. The nested structure of cancer symptoms. Implications for analyzing co-occurrence and managing symptoms. Methods of information in medicine. 2010;49(6):581-91. Epub 2010/11/19. doi: 10.3414/me09-01-0083. PubMed PMID: 21085743; PubMed Central PMCID: PMCPMC3647463.
32. Bhavnani SK, Ganesan A, Hall T, Maslowski E, Eichinger F, Martini S, et al. Discovering hidden relationships between renal diseases and regulated genes through 3D network visualizations. BMC research notes. 2010;3:296. Epub 2010/11/13. doi: 10.1186/1756-0500-3-296. PubMed PMID: 21070623; PubMed Central PMCID: PMCPMC3001742.
33. Bhavnani SK, Victor S, Calhoun WJ, Busse WW, Bleecker E, Castro M, et al. How cytokines co-occur across asthma patients: from bipartite network analysis to a molecular-based classification. Journal of biomedical informatics. 2011;44 Suppl 1:S24-30. Epub 2011/10/12. doi: 10.1016/j.jbi.2011.09.006. PubMed PMID: 21986291; PubMed Central PMCID: PMCPMC3277832.
34. Bhavnani SK, Bellala G, Victor S, Bassler KE, Visweswaran S. The role of complementary bipartite visual analytical representations in the analysis of SNPs: a case study in ancestral informative markers. Journal of the American Medical Informatics Association: JAMIA. 2012;19(e1):e5-e12. Epub 2012/06/22. doi: 10.1136/amiajnl-2011-000745. PubMed PMID: 22718038; PubMed Central PMCID: PMCPMC3392853.

35. Bhavnani SK, Dang B, Bellala G, Divekar R, Visweswaran S, Brasier AR, et al. Unlocking proteomic heterogeneity in complex diseases through visual analytics. Proteomics. 2015;15(8):1405-18. Epub 2015/02/17. doi: 10.1002/pmic.201400451. PubMed PMID: 25684269; PubMed Central PMCID: PMCPMC4471338.
36. Bhavnani SK, Dang B, Kilaru V, Caro M, Visweswaran S, Saade G, et al. Methylation differences reveal heterogeneity in preterm pathophysiology: results from bipartite network analyses. Journal of perinatal medicine. 2018;46(5):509-21. Epub 2017/07/01. doi: 10.1515/jpm-2017-0126. PubMed PMID: 28665803; PubMed Central PMCID: PMCPMC5971156.
37. Raudenbush SWBAS. Hierarchical linear models: applications and data analysis methods. Thousand Oaks: Sage Publications; 2002.
38. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. N Engl J Med. 2009;360(14):1418-28. doi: doi:10.1056/NEJMsa0803563. PubMed PMID: 19339721.
39. Report to Congress: Promoting Greater Efficiency in Medicare. Washington D.C.: MedPac (Medical Payment Advisory Commission); 2007.
40. Ashton CM, Del Junco DJ, Souchek J, Wray NP, Mansyur CL. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. Medical care. 1997;35(10):1044-59. Epub 1997/10/24. PubMed PMID: 9338530.
41. Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation. 2015 Condition-specific measures updates and specifications report: Hospital-level 30-day risk-standardized readmission measures on acute myocardial infarction, heart failure, pneumonia, chronic obstructive pulmonary disease, and stoke. A report prepared for the Centers for Medicare & Medicaid Services (CMS) 2015 [April 20, 2015]. Available from: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html.
42. Keenan PS, Normand SL, Lin Z, Drye EE, Bhat KR, Ross JS, 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 failure. Circ Cardiovasc Qual Outcomes. 2008;1(1):29-37. Epub 2008/09/01. doi: 10.1161/circoutcomes.108.802686. PubMed PMID: 20031785.
43. Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation. 2015 Procedure-specific readmission measures updates and specifications report: Elective primary total hip arthroplasty and/or total knee arthroplasty, and isolated coronary artery bypass graft surgery. A report prepared for the Centers for Medicare & Medicaid Services (CMS) 2015 [April 20, 2015]. Available from: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html.
44. Lochner KA, Cox CS. Prevalence of Multiple Chronic Conditions Among Medicare Beneficiaries, United States, 2010. Prev Chronic Dis. 2013;10:E61. doi: 10.5888/pcd10.120137.
45. Hillege HL, Girbes AR, de Kam PJ, Boomsma F, de Zeeuw D, Charlesworth A, et al. Renal function, neurohormonal activation, and survival in patients with chronic heart failure. Circulation. 2000;102(2):203-10. Epub 2000/07/13. PubMed PMID: 10889132.
46. Sharif R, Parekh TM, Pierson KS, Kuo YF, Sharma G. Predictors of early readmission among patients 40 to 64 years of age hospitalized for chronic obstructive pulmonary disease. Ann Am Thorac Soc. 2014;11(5):685-94. Epub 2014/05/03. doi: 10.1513/AnnalsATS.201310-358OC. PubMed PMID: 24784958; PubMed Central PMCID: PMCPmc4225809.
47. Grosso LM, Curtis JP, Lin Z, Geary LL, Vellanky S, Oladele C, et al. Hospital-level 30-Day All-Cause Risk-Standardized Readmission Rate Following Elective Primary Total Hip Arthroplasty (THA) And/Or Total Knee Arthroplasty (TKA). Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation, 2012.
48. Medicare Payment Advisory Commission (MedPAC). Report to the Congress: Medicare and the Health Care Delivery System. Chapter 3. Approaches to Bundle Payment for Post-Acute Care Washington, DC2013 [updated June]. 57-88]. Available from: http://www.medpac.gov/documents/reports/jun13_ch03.pdf?sfvrsn=0.

49. Evaluation YNHHSCCfOR. Procedure-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measures. A report prepared for the Centers for Medicare & Medicaid Services (CMS) 2017 [June 1, 2017]. Available from: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html.
50. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-83. PubMed PMID: 3558716.
51. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Medical care. 1998;36(1):8-27. PubMed PMID: 9431328.
52. Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation. 2017 Condition-specific measures updates and specifications report: Hospital-level 30-day risk-standardized readmission measures on acute myocardial infarction, heart failure, pneumonia, chronic obstructive pulmonary disease, and stoke. A report prepared for the Centers for Medicare & Medicaid Services (CMS) 2017 [June 1, 2017]. Available from: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html.
53. SAS. Using SAS® to Perform Individual Matching in Design of Case-Control Studies 2010 [cited 2020 May, 5th]. Available from: https://support.sas.com/resources/papers/proceedings10/061-2010.pdf.
54. Islam MM, Valderas JM, Yen L, Dawda P, Jowsey T, McRae IS. Multimorbidity and comorbidity of chronic diseases among the senior Australians: prevalence and patterns. PloS one. 2014;9(1):e83783. Epub 2014/01/15. doi: 10.1371/journal.pone.0083783. PubMed PMID: 24421905; PubMed Central PMCID: PMCPmc3885451.
55. Treviño S, Nyberg A, Del Genio CI, Bassler KE. Fast and accurate determination of modularity and its effect size. Journal of Statistical Mechanics: Theory and Experiment. 2015;2015(2):P02003. doi: 10.1088/1742-5468/2015/02/p02003.
56. Chauhan R, Ravi J, Datta P, Chen T, Schnappinger D, Bassler KE, et al. Reconstruction and topological characterization of the sigma factor regulatory network of Mycobacterium tuberculosis. Nature communications. 2016;7:11062. Epub 2016/04/01. doi: 10.1038/ncomms11062. PubMed PMID: 27029515; PubMed Central PMCID: PMCPMC4821874.
57. Rand WM. Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association. 1971;66(336):846-50. doi: 10.2307/2284239.
58. Fruchterman T, Reingold E. Graph Drawing by Force-Directed Placement. Software – Practice & Experience. 1991;21(11):1129–64.
59. Bennett TD, Moffitt RA, Hajagos JG, Amor B, Anand A, Bissell MM, et al. The National COVID Cohort Collaborative: Clinical Characterization and Early Severity Prediction. medRxiv. 2021. Epub 2021/01/21. doi: 10.1101/2021.01.12.21249511. PubMed PMID: 33469592; PubMed Central PMCID: PMCPMC7814838 Amin Manna, and Nabeel Qureshi: employee of Palantir Technologies; Brian T. Garibaldi: Member of the FDA Pulmonary-Allergy Drugs Advisory Committee (PADAC); Matvey B. Palchuk: employee of TriNetX; Kristin Kostka: employee of IQVIA Inc.; Julie A. McMurry: and Melissa A. Haendel Cofounders of Pryzm Health; Chris P. Austin and Ken R. Gersing, employees of the National Institutes of Health. No conflicts of interest reported for all other authors.
60. Topaloglu U, Palchuk MB. Using a Federated Network of Real-World Data to Optimize Clinical Trials Operations. JCO clinical cancer informatics. 2018;2:1-10. Epub 2019/01/18. doi: 10.1200/cci.17.00067. PubMed PMID: 30652541; PubMed Central PMCID: PMCPMC6816049.
61. Dogrusoz U, Karacelik A, Safarli I, Balci H, Dervishi L, Siper MC. Efficient methods and readily customizable libraries for managing complexity of large networks. PloS one. 2018;13(5):e0197238. Epub 2018/05/31. doi: 10.1371/journal.pone.0197238. PubMed PMID: 29813080; PubMed Central PMCID: PMCPMC5973603 the following competing interests: I.S., H.B., and L.D. were supported through Google Summer of Code for implementing some of the algorithms in this work as part of open source software projects. Others have no competing interests. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

62. Wu J, Zhu F, Liu X, Yu H. An Information-Theoretic Framework for Evaluating Edge Bundling Visualization. Entropy (Basel, Switzerland). 2018;20(9). Epub 2018/08/21. doi: 10.3390/e20090625. PubMed PMID: 33265714; PubMed Central PMCID: PMCPMC7513140.
63. Dhillon IS, Sra S. Generalized nonnegative matrix approximations with Bregman divergences. Proceedings of the 18th International Conference on Neural Information Processing Systems; Vancouver, British Columbia, Canada. 2976284: MIT Press; 2005. p. 283-90.
64. Dilsizian ME, Siegel EL. Machine Meets Biology: a Primer on Artificial Intelligence in Cardiology and Cardiac Imaging. Current cardiology reports. 2018;20(12):139. Epub 2018/10/20. doi: 10.1007/s11886-018-1074-8. PubMed PMID: 30334108.
65. Bhavnani SK, Visweswaran S, Divekar R, Brasier AR. Towards Team-Centered Informatics: Accelerating Innovation in Multidisciplinary Scientific Teams Through Visual Analytics. The Journal of Applied Behavioral Science. 2018:0021886318794606. doi: 10.1177/0021886318794606.
66. Wooten KC, Calhoun WJ, Bhavnani S, Rose RM, Ameredes B, Brasier AR. Evolution of Multidisciplinary Translational Teams (MTTs): Insights for Accelerating Translational Innovations. Clinical and translational science. 2015;8(5):542-52. Epub 2015/03/25. doi: 10.1111/cts.12266. PubMed PMID: 25801998; PubMed Central PMCID: PMCPmc4575623.