https://pubsonline.informs.org/series/educ

TUTORIALS IN OPERATIONS RESEARCH
 | ISBN 979-8-9882856-2-5
https://doi.org/10.1287/educ.2024.0279

Interventions for Patients with Complex Medical and Social Needs

Hari Balasubramanian,[a,*] Sindhoora Prakash,[a] Ali Jafari,[a] Arjun Mohan,[a] Chaitra Gopalappa[a]

[a] Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst, Massachusetts 01003

Contact: hbalasub@admin.umass.edu, https://orcid.org/0009-0002-8096-1954 (HB); sprakash@umass.edu, https://orcid.org/0009-0009-6930-2318 (SP); ajafari@umass.edu, https://orcid.org/0000-0002-4936-3684 (AJ); arjunmohan@umass.edu, https://orcid.org/0009-0000-7680-1371 (AM); chaitrag@umass.edu, https://orcid.org/0000-0001-8384-6041 (CG)

Abstract Patients with multiple chronic conditions and social needs represent a small percentage of the population but have a disproportionate impact on healthcare costs and utilization. Organizations around the United States have created programs, often referred to as complex care interventions, to improve the health and well-being of such patients and reduce avoidable hospital and emergency department use. In this tutorial, we focus on two emerging themes in the field: (1) identifying clinically meaningful subgroups in complex care populations through unsupervised learning methods and (2) describing the key operational features of interventions with an emphasis on staffing needs and the impact on patient outcomes. The material presented in this tutorial draws on the research of the Healthcare Operations Research Laboratory at the University of Massachusetts, Amherst, and its collaborating partners. To illustrate these themes and contextualize the details of complex care delivery, we use a range of patient-level examples, visualizations, descriptive summaries, case studies, and results from the clinical literature.

Funding: This research was funded by the U.S. National Science Foundation [Grants NSF CMMI 1254519 and NSF CMMI 2212237]. The views expressed in this paper are of the authors and not of the National Science Foundation.

Keywords multiple chronic conditions • complex care interventions • associations among chronic conditions • interpretable patient subgroups • staffing • stochastic process • randomized control trial

1. Introduction

Healthcare expenditures in the United States are disproportionately concentrated in a small percentage of the population. Five percent of the population accounts for one-half of annual expenditures, whereas 1% accounts for almost a quarter of annual expenditures (Cohen [14]). Many individuals among these highest-cost segments have multiple chronic conditions and experience higher hospital utilization rates, including avoidable hospitalizations, than the average.

Multiple chronic conditions affect all segments of the U.S. population, but for some individuals, the presence of other factors, such as homelessness, mental health conditions, substance abuse, poverty, the lack of employment or insurance, and the presence of disabilities, can

further complicate the care delivery process and add to the patients' vulnerabilities. This leads them to "experience combinations of medical, behavioral health, and social challenges that result in extreme patterns of healthcare utilization and cost" (Humowiecki et al. [27], p. 11). In the clinical and health services literature, several different terms are used to describe this heterogeneous subpopulation: multimorbid patients (Skou et al. [58]), patients with multiple chronic conditions, high-cost high-need patients (Blumenthal et al. [7]), patients with complex medical and social needs (Martinez et al. [36]), and frequent users (Finkelstein et al. [19]).

Healthcare organizations around the country have developed specialized programs to assist patients with medical and/or social complexity. The programs are often called complex care interventions. The goal of such interventions has been to improve the health and well-being of patients by reducing avoidable emergency department visits and inpatient stays, increasing access to outpatient primary and specialty care, and connecting patients to social and behavioral resources in the community. The target populations and mechanisms of intervention have varied considerably. Some interventions have been specifically designed for a particular demographic and disease. For instance, Naylor et al. [42] describe an intervention in which an advanced practice nurse assists heart failure patients over 65 years of age after a hospital discharge. Other interventions have focused on the intersection of social risk factors and health. The Boston Healthcare for the Homeless Program (O'Connell et al. [46]), which uses 600 medical, behavioral health, and social service providers to ensure that individuals and families experiencing homelessness receive comprehensive, high-quality healthcare, falls in this category. In other cases, interventions focus more generally on high-risk patients. The Veterans Affairs (VA), for example, provides intensive care management through a team of primary care providers, social workers, psychologists, nurses, and other support staff to patients with the highest risk of hospitalizations (Zulman et al. [72]).

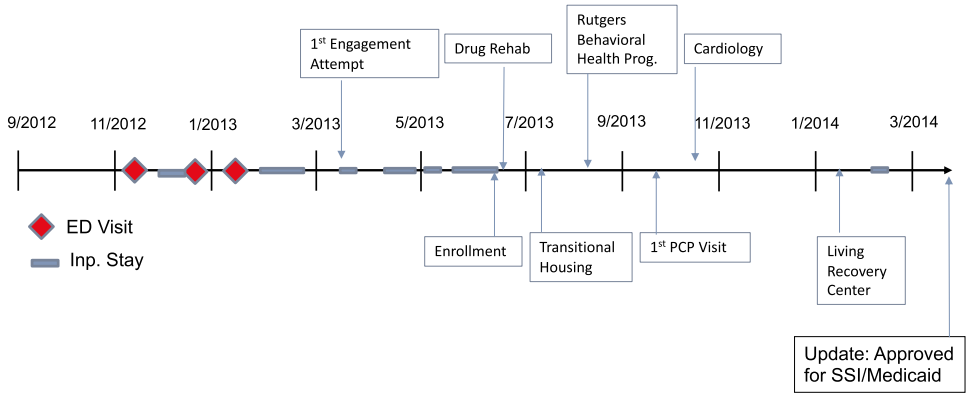
In fact, programs to assist individuals with complex medical and/or social needs exist in every community, town, city, or region. They can be led by small not-for-profit organizations, county public health departments, federally qualified health centers operating in medically underserved areas, social service organizations, or larger organizations such as hospitals, payers, and health systems. These efforts are not always visible within mainstream healthcare discussions, but they are vital nevertheless: they bring to light broader societal and population health concerns. Furthermore, although reductions in costs and decreases in avoidable hospitalizations are important criteria for complex care programs, among staff members who deliver care, the humanitarian impulse to genuinely connect with and help the most vulnerable individuals within a community is an equally important motivation. There's also a broad consensus that the help provided must be holistic and comprehensive; that is, the nature of the assistance is not only restricted to individual diseases but looks at all of an individual's medical needs, how they interact with each other, and also how social and behavioral factors might impede health and wellbeing.

1.1. Motivating Example

We now illustrate how a complex care intervention works with an actual patient example. The example is from the Camden Coalition, a New Jersey–based organization renowned in the field of complex care. The Coalition uses care teams consisting of community health workers, nurses, social workers, and clinical psychologists to assist patients. The Healthcare Operations Research laboratory has collaborated with the Camden Coalition since 2015. See Section 3.3 for a detailed case study of the operational features of their care model.

Figure 1 shows the nine-month pre- and post-intervention timeline for a 52-year-old man living in Camden, New Jersey. The care team enrolled the patient in their intervention in June 2013. Interviews and an assessment of the electronic health record revealed that the patient had three chronic conditions: hepatitis C, congestive heart failure, and diabetes. The care team also established that the patient did not have housing, was unemployed, had a substance use disorder, and had mental health conditions. This is a classic example of a patient with high

Figure 1. (Color online) Timeline of complex care interventions for a patient. The figure shows an example of a complex care intervention: the event sequence and timing for a 52-year-old man before and after being helped by a multidisciplinary care team. Healthcare expenditures in nine months prior to enrollment with care team totaled $112,583. Healthcare expenditures after enrollment with care team totaled $3,955. "Inp. Stay" refers to inpatient hospitalization.



medical and social complexity. In the nine months before enrollment with the care team, the patient had three emergency department (ED) visits, seven inpatient stays totaling 61 days, and $112,583 in hospital-based healthcare expenditures.

After enrollment, the care team worked closely with the patient, providing him with transitional housing and connecting him with a primary care provider (PCP), a cardiologist, a rehabilitation center for drug abuse, and a behavioral health program at Rutgers University. In the nine months after the intervention, the patient had only one three-day inpatient stay; the total hospital-based expenditures in this period were $3,955. The social worker in the care team facilitated a process by which the patient eventually received social security income and Medicaid healthcare insurance. The intervention required significant care team capacity, totaling 236 hours across the team. This included 139 hours from the social worker, 58 hours from the licensed practical nurse, 24 hours from the health coach, and 15 hours from the community health worker. Even if we assume that each member of the care team is paid $35 per hour and factor in the costs of care team hours, the total reduction in costs is more than $100,000. The care team's encounters with the patient are not shown on the postenrollment timeline in the figure, but if they were, they would present as a clustered series of points leading up to the principal highlights (such as the PCP visit).

It is important to not generalize based on a single example: Reductions in costs and utilization for many medically and socially complex patients are highly stochastic and vary significantly between individuals. Indeed, the Camden Coalition's randomized control trial (Finkelstein et al. [19]) demonstrates that hospitalizations and costs in individuals with similar medical and social complexity who did not receive the intervention can show a similar drop due to regression to the mean: "a statistical phenomenon that can make natural variation in repeated data make look like real change" (Barnett et al. [4], p. 1).

However, the example does illustrate the hypothesis underlying the interventions: that a care team can help patients (1) improve self-efficacy and well-being by enabling medical and social support, including access to primary and specialty care appointments, housing, employment, and health insurance and (2) reduce adverse and costly events such as emergency visits, medication-related complications, and hospitalizations. The example also illustrates that a truly holistic intervention is not limited to only medical issues, but spans other sectors, such as housing, employment, and legal assistance.

1.2. Focus and Organization of Tutorial

Intervention for individuals with complex health and social needs is a still emerging field and many open questions remain in both the clinical and operations research literature. This tutorial has two main parts.

1.2.1. Segmentation of Clinical Profiles. The first part discusses how to segment complex care populations into clinically meaningful subgroups. Almost all complex care programs provide care to heterogeneous groups of patients with multiple chronic conditions. Therefore, an important aspect of clinical segmentation is the identification of patterns in chronic condition co-occurrence. In Section 2, we introduce how the presence of multiple chronic conditions (MCCs) in patients can significantly increase the complexity of medical treatment. Next, in Section 2.1, we illustrate key patterns in the MCC prevalence using the Medical Expenditure Panel Survey and describe how chronic condition combinations observed in patients can exhibit a high degree of heterogeneity. To deliver personalized care to patients, clinicians, epidemiologists and administrators of complex care programs need a systematic, data-driven understanding of how chronic diseases that arise from different body/organ systems are associated with each other. Co-occurrence patterns can then be used to identify whether treatment, medications, and lifestyle changes for one disease are concordant or discordant with those of other diseases. For instance, Zulman et al. [71] point out that some disease subsets (such as high blood pressure, high cholesterol, and heart disease) are considered biologically linked and concordant: This explains why lifestyle changes related to diet and exercise often overlap for these conditions. Other disease subsets such as heart disease, depression, and asthma are considered discordant because they have very different treatment and management strategies.

To identify co-occurrence patterns and interpretable subgroups among complex care patients, we describe two unsupervised learning approaches: association rule mining (Section 2.2) and latent class analysis (LCA) (Section 2.3). Association rule mining is a data mining methodology that is widely used in the retail context to quantify groups of items that are frequently purchased together. In our case, we apply the method to identify groups of chronic diseases that frequently occur together. Association rule mining goes further, however; once the groups of co-occurring diseases have been identified, it quantifies pairwise associations/correlations between different subsets of diseases. This creates a comprehensive database of relationships that can be used by clinicians in complex care programs. Next, we focus on LCA, a method of clustering that is widely used in the clinical literature to infer subgroups. In both sections, we provide case studies, visualizations, and examples to illustrate insights.

1.2.2. Operational Details, Staffing Needs, and Impact on Patient Outcomes. In the second part of the tutorial (Section 3), we illustrate the operational features of complex care programs, a theme that has not received much attention in the literature. We illustrate how complex care interventions are essentially staffing interventions, where staff refers to some combination of physicians, nurses, community health workers, and social workers who assist patients through a series of recurring and stochastic service interactions. We summarize two examples of complex care programs (Section 3.2) and provide a detailed case study of our collaborating partner, the Camden Coalition (Section 3.3). We also emphasize how the Camden Coalition and other programs have tested the impact of their care delivery models on a variety of patient outcomes through randomized control trials (RCTs). Although staffing has been one of the most widely studied topic areas in healthcare operations research, few (if any) studies have demonstrated the impact on patient outcomes. RCT-based evaluations of complex care interventions therefore provide a unique opportunity to directly evaluate how service interactions between the care team and the patient impact outcomes. The Camden Coalition study also shows the benefits of rigorously collecting data related to patient-care

team encounters, and how such data can be used to (a) create staffing estimates for different members of the care team and (b) evaluate whether patients who engaged more with the care team had better outcomes. The case study provides a template that can be adapted by other complex care programs.

2. MCCs and Complexity of Medical Care

Chronic diseases are defined as conditions that last a year or more, require ongoing medical attention, and/or limit the activities of daily living (Buttorff et al. [9]). Examples of highly prevalent chronic conditions include high blood pressure, diabetes, depression, inflammatory joint disorders, and asthma. MCCs refer to the presence of two or more chronic conditions in an individual. In the United States for example, the number of individuals with zero, one, two, three, four, and five or more chronic conditions was estimated to be 40%, 18%, 13%, 9%, 7% and 12% respectively in a RAND report. Although 40% of the population has no apparent chronic conditions and contributes only 10% of total healthcare costs, 12% of the U.S. population has five or more chronic conditions and this segment accounts for 41% of total healthcare costs (Buttorff et al. [9]).

The RAND report on the prevalence of multiple chronic conditions is based on the Medical Expenditure Panel Survey (MEPS) from 2014. MEPS [37] is a set of large-scale, nationally representative surveys taken annually of families and individuals, their medical providers, their insurance companies, and their employers. Approximately 30,000 individuals are surveyed each year. MEPS is also a rich source of patient-level data and the basis of our analysis in this section. Medical encounters for each surveyed individual, the type of specialist or other care provider seen, the associated diagnoses code, and the exact month and year in which the encounter happened are anonymized and released to the public each year. By merging various data files, longitudinal information about all health events for each surveyed individual can be assembled for two years.

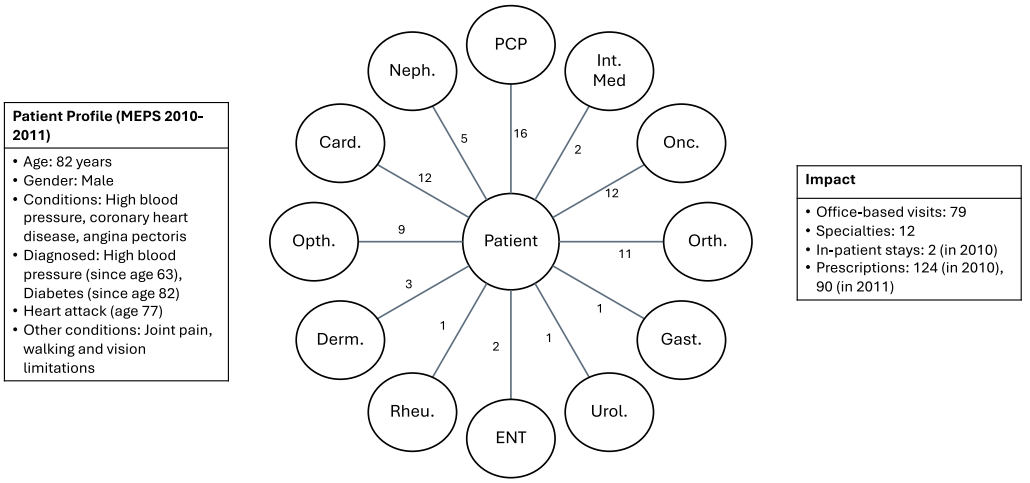
Example. Consider the following patient example MEPS 2010 and 2011. In Figure 2, we present details of the total outpatient visits for a single 82-year-old male patient with high blood pressure, coronary heart disease, angina pectoris, joint pain, walking, and vision limitations. The patient had 79 doctor's office visits spread over 12 different specialties in the 2-year survey period. In addition, the patient also had two inpatient stays and refilled more than 200 prescriptions.

Most MCC patients, even younger patients in the 40–65 age group, have a similar utilization profiles, and it is easy to see why they have an outsize impact on healthcare utilization and costs. The example also shows how challenging this two-year period must have been for the patient and his caregivers. The burden of scheduling appointments, arranging transportation, checking insurance paperwork and payments, ensuring the right medications are being taken at the right time, and reconciling potentially conflicting advice received from the different specialists: All such responsibility falls on the patient and the patient's caregivers.

Patients with MCC also pose several challenges for healthcare providers. The presence of MCCs leads to visits to a large number of specialists because multiple organ/body systems need to be monitored. Furthermore, MCC patients are often on more than five unique medications, and the risk of interactions between medications is high.

Medical complexity, in turn, leads to operational challenges. Physicians must spend more time communicating with each other, the patient, and the patient's family, by email, phone, and messaging systems, to ensure that moving parts of a patient's care are not out of sync. They must also spend greater time evaluating laboratory tests and other diagnostic information in the patient's electronic health record. Thus, all clinicians involved must budget a greater nonvisit or non–face-to-face capacity for MCC patients. The burden of coordinating care most often falls on the patient's PCP (Rossi and Balasubramanian [54], Press [49]). MCC patients also have higher visit rates, which in turn implies shorter intervals between

Figure 2. (Color online) MCC patient profile: an Example.



Notes. The outer circles stand for the type of physicians visited: PCP, primary care physician; Int. Med., internal medicine; Onc., oncology (cancer specialist); Orth., orthopaedics (musculoskeletal specialist); Gast., gastroenterology; Urol., urology; ENT, ear-nose-throat, also known as Otorhinolaryngology; Rheu., rheumatology (specialist for musculoskeletal diseases and autoimmune conditions); Derm., dermatology; Opth., ophthalmology; Card., cardiology; Neph., nephrology (kidney specialist). The numbers on the edges indicate the number of visits to the respective specialty in the 2 years.

appointments. This means that each provider must have sufficient appointment capacity to accommodate MCC patients at short notice.

Such complexities in care delivery are precisely the motivation for setting up a specialized team dedicated to assisting MCC patients and coordinating their care. This eases the burden on patients and their caregivers and makes it easier for all clinicians involved in the patient's care.

2.1. Heterogeneity in Chronic Condition Combinations

We now turn to another complicating factor in complex care delivery: the high levels of heterogeneity observed in MCC combinations. To understand what this means, consider the chronic condition combinations of three different patients from MEPS. Conditions for each patient are separated by ";".

Patient A (5 conditions): {Lipid Metabolism Disorders; Anxiety; Depression; Non-Traumatic Joint Disorders; Spondylosis/Intervertebral disc disorders/other back issues}

Patient B (5 conditions): {Diabetes; Lipid Metabolism Disorders; High Blood Pressure; Coronary Atherosclerosis/Heart Disease; Connective Tissue Disease}

Patient C (6 conditions): {Diabetes; Lipid Metabolism Disorders; High Blood Pressure; Chronic Obstructive Pulmonary Disease (COPD) and Bronchiectasis; Coronary Atherosclerosis/Heart Disease; Spondylosis/Intervertebral disc disorders/other back issues}

To create the combination for each patient, we first identified all diagnosis codes for the patient in a year related to different visit types (outpatient, inpatient, emergency room, home health visits, etc.). Diagnosis codes in MEPS are based on the International Classification of Diseases (ICD), maintained and revised by the World Health Organization (WHO). Each diagnosis code for the patient is then mapped to a higher-level grouping, the U.S. Agency of Healthcare Research and Quality's Clinical Classification Software (CCS) codes. These are clinically meaningful aggregates of ICD diagnosis codes. (ICD codes could be directly used to

create combinations, but the diagnosis descriptions are too detailed for meaningful patterns to emerge.)

Names of some of the conditions, such as diabetes, high blood pressure, anxiety, and depression, are well known. Others are less familiar to a general audience. Lipid metabolism disorders refers to abnormal cholesterol levels; spondylosis/intervertebral disc disorders/other back issues to a specific subcollection of related musculoskeletal spine problems; Nontraumatic joint disorders to a different subcollection of musculoskeletal problems; coronary atherosclerosis to a subcollection of heart disease diagnoses; chronic obstructive pulmonary disease (COPD) and bronchiectasis to a subcollection of lung/respiratory diagnoses.

We notice the patients share certain subsets of conditions. Patients B and C, for example, share a subset of four conditions: diabetes; lipid metabolism disorders; high blood pressure; and coronary atherosclerosis/heart disease. However, the combination for each of the patients is unique in that the full set of conditions for a patient is not repeated in another. Unique combinations lead to unique clinical profiles, requiring the patient's treatment plan to be personalized.

In a four-year extract of the MEPS data set (2016–2019), we found 34,880 unique chronic condition combinations involving two to seven diseases. Let K denote the number of chronic conditions in a combination and let $\mathcal{N}_K$ denote the set of all unique combinations of size K. Let $\mathcal{N}_{K,l} \subseteq \mathcal{N}_K$ denote the set of combinations of size K that appeared in l different years where $l = 1, 2, ...4$. Note that $|\mathcal{N}_K| = \sum_{l=1}^{4} |\mathcal{N}_{K,l}|$. Combinations that appeared in all four years have a high degree of consistency (they are also relatively more prevalent), whereas combinations that appear in just one of the five years are rare and often appear in a single patient.

In Table 1, we show the $|\mathcal{N}_K|$ and $|\mathcal{N}_{K,l}|$ for $K = 2, 3, \ldots, 7$. For instance, we found 7,842 unique combinations of size 4 in MEPS 2016–2019. Of these, 7,350 (more than 93%) combinations appeared in only one of the four years (i.e. they were rare enough to be observed only in one year); 401 appeared in two of the four years; 63 appeared in three of the four years; and 28 appeared consistently in all four years.

An increase in K aggravates this pattern sharply: the vast majority of combinations with five, six, and seven conditions rarely repeat beyond a single year. This is because each combination has a very low prevalence. Yet, paradoxically, there are such a large number of such combinations that they add up to a high combined prevalence in the population: Recall from the introduction of Section 2 that 19% of the U.S. population, that is, around 60 million individuals, had four or more conditions in 2014. Rezaee et al. [51] aptly call this phenomenon "the high prevalence of low prevalence chronic disease combinations." This pattern is not just restricted to MEPS; it has also been found in Medicaid data (Sorace et al. [62]).

In the next section, we show how association rule mining can be used to systematically (a) identify the most frequent subsets of diseases found in chronic condition combinations; and (b) establish pairwise relationships between subsets.

Table 1. Cardinality of sets $\mathcal{N}_K, \mathcal{N}_{K,l}$ for $K = 2, 3, \ldots, 7$ based on MEPS 2016–2019.

K	$\lvert\mathcal{N}_K\rvert$	$\lvert\mathcal{N}_{K,1}\rvert$	$\lvert\mathcal{N}_{K,2}\rvert$	$\lvert\mathcal{N}_{K,3}\rvert$	$\lvert\mathcal{N}_{K,4}\rvert$
2	3,496	1,792	750	471	483
3	7,929	6,592	977	238	122
4	7,842	7,350	401	63	28
5	6,538	6,355	170	13	0
6	5,079	5,025	53	1	0
7	3,996	3,963	33	0	0

2.2. Identifying Chronic Disease Co-Occurrence Patterns Through Association Rule Mining

Our presentation of association rule mining adapts the notation used in Hastie et al. [25]. Suppose we have $j = 1, 2, \ldots, p$ diseases, and the binary variable Z_j denotes whether a patient has disease j ($Z_j = 1$) or not ($Z_j = 0$). The goal is to find subsets of the p variables that frequently take on a value of one in a data set. Suppose that we have $i = 1, 2, \ldots, N$ patients in the data set. Each variable Z_j is assigned two values: $z_{i,j} = 1$ if the jth disease is present in patient i, $z_{i,j} = 0$ otherwise. Subsets of diseases that jointly take on the value of one represent diseases that frequently co-occur together: These are called frequent itemsets. In the retail/grocery store setting, variables that take on a value of one represent items that have been purchased together; hence, the method described later is often called market basket analysis.

More formally, we seek to find an itemset of diseases $\mathcal{K} \subset 1, 2, \ldots, p$ that co-occur with "high" probability. This probability can be written as $\Pr[\cap_{k\in\mathcal{K}}(Z_k = 1)]$. In the context of the data set, the estimated value of this probability is simply the fraction of the N patients in the data who have all the diseases in $\mathcal{K}$:

$$\hat{\Pr}\left[\bigcap_{k\in\mathcal{K}}(Z_k = 1)\right] = \frac{1}{N}\left[\sum_{i=1}^{N}\prod_{k\in\mathcal{K}} z_{ik}\right]. \tag{1}$$

We call this estimated probability the "support" or "prevalence" and denote it by $T(\mathcal{K})$. A patient i for whom $\prod_{k\in\mathcal{K}} z_{ik} = 1$ has all the diseases in $\mathcal{K}$. In association rule mining, we seek to identify all disease itemsets or subsets (we use these terms interchangeably) that have a support higher than some predetermined value t. Formally, we seek all subsets $\mathcal{K}_l$ such that $\{\mathcal{K}_l | T(\mathcal{K}_l) > t\}$. In other words, there are 2^p disease subsets in total, of which we are interested only in those that appear in the least $t * N$ patients of the data set. The value of t must be carefully calibrated. It cannot be too large because there are a large number of low-prevalence disease subsets. Many of these low-prevalence subsets would not be included if t is large. However, because of the combinatorial structure of the frequent itemsets problem, small values of t make the problem computationally challenging.

The "Apriori" algorithm (Agrawal and Srikant [1]) efficiently computes all the itemsets for a given value of t. It uses the following principle to eliminate subsets: Any item set L that comprises a subset of the items in $\mathcal{K}$ must have support greater than or equal to that of $\mathcal{K}$, indicated by $L \subseteq \mathcal{K} \Rightarrow T(L) \geq T(\mathcal{K})$.

We provide a brief overview of how the Apriori algorithm works. In the first pass through the data, the algorithm computes the support for all single itemsets; that is, it computes how frequently individual diseases occur. Diseases that have support below t are eliminated. In the next pass, it computes support of all disease pairs from the diseases that survived the first pass; itemsets of size two with support below t are dropped. Each successive traversal only considers itemsets formed by combining those that survived previous passes. To generate all frequent itemsets of size $|\mathcal{K}| = m$, we need to only consider ancestral items of size $m - 1$ or less that cross the threshold t. The algorithm keeps traversing through the data until all candidate itemsets from the preceding pass have support below t.

Sample Results. We applied the Apriori algorithm to the data set of $n = 28{,}512$ individuals surveyed in MEPS 2019. Ideally, the data set should be segmented based on age, race, geographic location, and other demographic characteristics. However, in the interest of keeping the analysis concise, we applied it to the entire data set. Furthermore, segmenting the population reduces the number of patients within each segment, making it more difficult to identify frequent itemsets. Large health systems that care for millions of patients (e.g., Veteran Affairs Health System) are better suited for Apriori analysis that considers demographic adjustments.

The input to the algorithm was a binary matrix with each row corresponding to a patient and a column to 1 of 274 Clinical Classification Software (CCS) codes. For instance, CCS 49 indicates diabetes; CCS 98 indicates high blood pressure (essential hypertension); CCS 204 indicates nontraumatic joint disorders; and so on. We set the minimum support threshold $t = 0.0003$, which implies that all itemsets identified in the Apriori algorithm appear in at least $0.0003 * 28{,}512 \approx 10$ patients. This yields hundreds of dyads, triads, quartets, and quintets; we show some examples of these frequent subsets/itemsets here. For each subset, we provide the name of the CCS code and its number; codes are separated by a ";". We also provide support for each itemset.

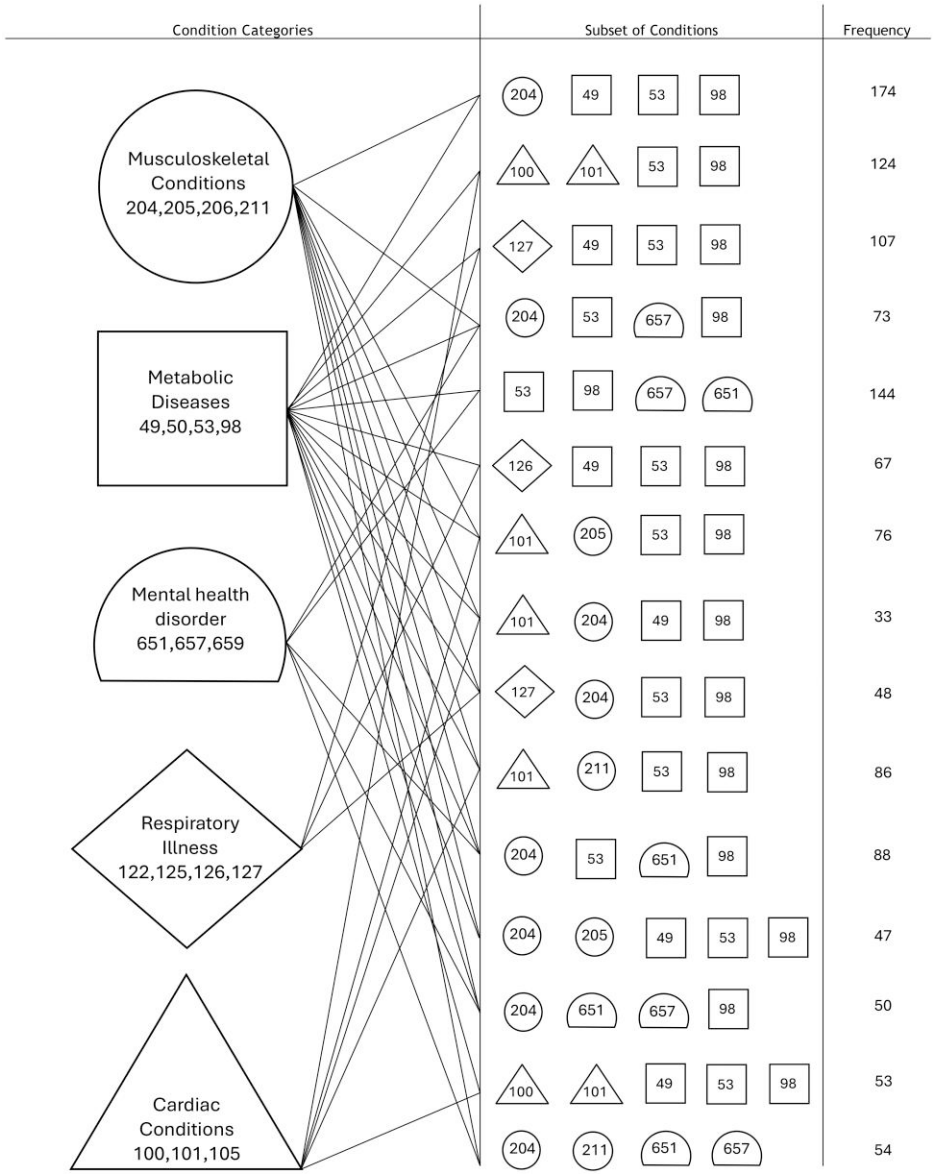
- {Lipid Metabolism Disorder (53); Essential Hypertension (98)}, Support: 0.098
- {Anxiety Disorders (651); Mood Disorders (657)}, Support: 0.032
- {Non-Traumatic Joint Disorders (204); Spondylosis/Intervertebral disc disorders/other back issues (205)}, Support: 0.017
- {Lipid Metabolism Disorder (53); Essential Hypertension (98); Diabetes (49)}, Support: 0.036
- {Lipid Metabolism Disorder (53); Diabetes (49); Coronary Atherosclerosis/Heart Disease (101)}, Support: 0.006
- {Lipid Metabolism Disorders (53); Essential Hypertension (98); Coronary Atherosclerosis/Heart Disease (101); Chronic Obstructive Pulmonary Disease and Bronchiectasis (127)}, Support: 0.002
- {Lipid Metabolism Disorder (53); Essential Hypertension (98); Diabetes (49); Non-Traumatic Joint Disorders (204); Spondylosis/Intervertebral disc disorders/other back issues (205)}, Support: 0.001

As subset sizes get larger, we begin to see diseases from different body systems appear together. For instance, in the last subset, metabolic diseases such as lipid metabolism disorder (abnormal cholesterol) diabetes, essential hypertension (high blood pressure), and diabetes combine with musculoskeletal conditions such as joint disorders and spondylosis/invertebral disc disorders/back issues. In the second to last subset, we see metabolic (abnormal cholesterol, high blood pressure), heart (coronary atherosclerosis), and lung-related conditions (chronic obstructive pulmonary disease and bronchiectasis).

We visualize this pattern in more detail in Figure 3. The middle panel in the figure shows a selection of subsets involving four or five condition codes. The right panel shows the frequency, that is, the number of patients where the subset was present. The left panel shows five clinically meaningful disease groupings of condition codes: musculoskeletal; metabolic; mental health; respiratory; and cardiac. For instance, the cardiac grouping collects condition codes 100 (acute myocardial infarction),101 (coronary atherosclerosis and heart disease), and 105 (conduction disorders), all of which are heart related. Each disease grouping is marked by a different shape so that the groupings present in a subset can be easily identified. We observe that all itemsets contain diseases from at least two different groupings.

As discussed earlier, diseases within a grouping can often be concordant; for instance, lifestyle changes concerning diet and exercise are similar for diabetes, high blood pressure, and high cholesterol. However, this may not be true when diseases belong to different groups. For instance, exercise is recommended for cardiovascular disease, but the presence of osteoarthritis or chronic obstructive pulmonary disease (COPD) may make exercise difficult (Zulman et al. [71]). In such cases, the diseases are discordant. Diseases that map to different body systems may also increase the risk that a particular medication for one disease may have compelling contraindication (i.e., negative outcome) for another. Thus, frequent itemset analysis provides a systematic method for identifying the most prevalent patterns of chronic condition co-occurrence, which in turn can be used to guide further research on medication, lifestyle changes, and treatment options.

Figure 3. Frequencies (row count) of selected subsets with four and five conditions. Itemsets shown here contain disease codes from multiple body/organ systems, indicated by different shapes.



2.2.1. Quantifying Relationships Between Frequent Itemsets. We next discuss how pairs of frequent itemsets can be analyzed for further insights. We first define an antecedent itemset A and a consequent itemset B. A and B do not share diseases in common. The association rule between A and B is denoted as follows:

$$A \Rightarrow B. \tag{2}$$

We now illustrate three quantities, support, confidence, and lift, related to this association rule. The support of the rule $T(A \Rightarrow B)$ is the fraction of patients in the data set that have all the conditions in the union of the antecedent and consequent. The confidence $C(A \Rightarrow B)$ is defined by Equation (3).

$$C(A \Rightarrow B) = \frac{T(A \Rightarrow B)}{T(A)} \tag{3}$$

This is simply an approximation of the probability of B given A, denoted as $\Pr(B|A)$. Here, $\Pr(A)$ represents the probability of item set A occurring in a basket, shorthand for $\Pr(\prod_{k\in A} Z_k = 1)$. The "lift" of the rule is given in Equation (4).

$$L(A \Rightarrow B) = \frac{C(A \Rightarrow B)}{T(B)} = \frac{T(A \Rightarrow B)}{T(A)T(B)} \tag{4}$$

This is a ratio of the joint occurrence of A and B (numerator) to the product of individual occurrences (denominator). If A and B are independent of each other, the ratio will be close to one. If A and B are positively correlated with each other, the ratio will be higher than one. If they are negatively correlated, the ratio will be between zero and one.

The objective of this analysis is to generate association rules (2) that exhibit high levels of both support and confidence (Equation (3)). The Apriori algorithm is used to identify all item sets with significant support, determined by the support threshold t. Now an additional confidence threshold c is established, and any rules formed from pairs of itemsets (2) that surpass this confidence threshold are identified and reported. More formally, we now generate a set of rules between all pairs of mutually exclusive itemsets that meet the following two criteria:

$$T(A \Rightarrow B) > t \quad \text{and} \quad C(A \Rightarrow B) > c. \tag{5}$$

Alternatively, instead of confidence, we can generate rules that cross an additional lift threshold l:

$$T(A \Rightarrow B) > t \quad \text{and} \quad L(A \Rightarrow B) > l. \tag{6}$$

Sample Results. To illustrate $T(A \Rightarrow B)$, $C(A \Rightarrow B)$, and $L(A \Rightarrow B)$, we use one antecedent, A: {Diabetes, High Blood Pressure, Cholesterol}, and the three consequents:

- B_1: {Chronic obstructive pulmonary disease; Asthma}
- B_2: {Joint Disorders; Spondylosis, invertebral disc disorders, other back problems}
- B_3: {Anxiety Disorders; Mood Disorders}

The antecedent is the most prevalent triad of chronic conditions in the United States; the conditions in the triad belong to the group of metabolic diseases. The three consequents are pairs of chronic conditions chosen from respiratory, musculoskeletal, and mental health conditions respectively. We deliberately chose antecedents and consequents from disparate body systems to see whether any noteworthy associations were indicated by confidence or lift. We used the same MEPS 2019 data for this analysis. Our binary data set comprises n = 28,512 patients (rows) and 274 Clinical Classification Codes (columns).

There are $n(A)$ = 1,050 patients in the data set with the three diseases that constitute the Antecedent group (A). Therefore the support of the antecedent is

$$T(A) = \frac{n(A)}{N} = 0.036 \qquad \text{Support} = 0.036.$$

Association Rules for Antecedent A and Consequent B_1:

$$\begin{bmatrix}\text{Diabetes}\\ \text{High blood pressure}\\ \text{Cholesterol}\end{bmatrix} \Downarrow \begin{bmatrix}\text{Chronic obstructive pulmonary disease and bronchiectasis}\\ \text{Asthma}\end{bmatrix}$$

Support = 0.0015 (43 patients), Confidence = 4.1%, Lift = 4.23

Association Rules for Antecedent A and Consequent B_2:

$$\begin{bmatrix}\text{Diabetes}\\ \text{High blood pressure}\\ \text{Cholesterol}\end{bmatrix} \Downarrow \begin{bmatrix}\text{Other non-traumatic joint disorders}\\ \text{Spondylosis; intervertebral disc disorders; other back problems}\end{bmatrix}$$

Support = 0.0016 (47 patients), Confidence = 4.5%, Lift = 2.57

Association Rules for Antecedent A and Consequent B_3:

$$\begin{bmatrix}\text{Diabetes}\\ \text{High blood pressure}\\ \text{Cholesterol}\end{bmatrix} \Downarrow \begin{bmatrix}\text{Anxiety disorders}\\ \text{Mood disorders}\end{bmatrix}$$

Support = 0.0021 (62 patients), Confidence = 5.9%, Lift = 1.8

Although confidence values differ only slightly across the three comparisons, the lift values are greater than one for all three comparisons. In particular, the lift of the metabolic diseases in the antecedent with the two respiratory diseases (COPD and asthma) in consequent B_1 is quite high at 4.23.

These results should be treated with caution because the analysis uses a single year of MEPS data and factors such as age, gender, and socioeconomic indicators are currently not included in the analysis. Nevertheless, the previous examples illustrate how association rule mining can be used to create a comprehensive database consisting of thousands of pairwise associations quantified with support, confidence, and lift. The database can systematically identify disease subsets with high pairwise confidence and lift values. If a clinician sees a patient with a particular subset of diseases, confidence and lift could be used to determine the risk of developing other disease subsets and thereby formulate prevention strategies. Subsets that have high pairwise confidence or lift can be prioritized for (i) resource planning and coordination among the relevant specialties; (ii) the creation of guidelines for medication choices (in particular to avoid contraindications), treatment options, and lifestyle changes; and (iii) to postulate physiological connections between diseases that belong to different organs/body systems.

2.3. Identifying Interpretable Subgroups Through LCA

LCA is a statistical method used to identify distinct subgroups, known as latent classes, within populations. These classes are characterized by observable traits known as indicator variables, for example, the age group of a patient and whether they have certain

comorbidities. The assumption behind latent variable models is that the observed distribution of the indicators results from a finite latent mixture of underlying distributions. These models are used to identify solutions that can accurately represent the latent variables, within which the indicators follow the same distribution (Sinha et al. [57]).

The indicator and latent variables can be either continuous or categorical. In the healthcare context, continuous variables could be the number of ED visits, the number of days of inpatient hospitalization in a given time period, the patient's age, their comorbidity index (such as Charlson Comorbidity Index; Charlson et al. [13]), healthcare utilization/total expenses, and so on. Some of these variables, such as ED visits, take on discrete values but can be modeled as continuous variables. Categorical indicator variables could be whether they have certain chronic conditions, their mobility level (good, needs some support, needs wheelchair), their self-rating of health (fair, good, not good, poor), and so on. Depending on the types of indicator and latent variables in our data set, different latent variable models are applicable as described in Skrondal and Rabe-Hesketh [59]. For instance, when both variables are continuous, models such as the common factor model, structural equation model, linear mixed model, and covariate measurement error model are appropriate. If the indicator variable is continuous and the latent variable is categorical, the latent profile model is applicable. In contrast, when the indicator variable is categorical and the latent variable is continuous, the latent trait model (also known as item-response theory) is suitable. Finally, when both variables are categorical, the latent class model is the recommended approach. In this tutorial, our primary focus is on exploring the LCA algorithm. However, by delving into this algorithm, we aim to establish a foundational understanding that will serve as a springboard for other related algorithms as well.

2.3.1. LCA Algorithm. LCA uses the expectation-maximization (EM) algorithm to determine the underlying latent classes. The process begins by randomly assigning model probabilities, encompassing both the overall probability of observing each class in the given population and the probability of a specific response given the latent class probability. This entails treating the randomized class membership as an observed variable. During the expectation (E) step, posterior probabilities of class membership are calculated. This involves determining, for a given pattern of indicator values, the probability that the pattern belongs to each class. These probabilities are mutually exclusive for each pattern and sum to one. The maximization (M) step entails the assignment of classes based on the posterior probabilities derived in the E-step. Subsequently, the model probabilities (or model parameters) are recalculated. The E- and M-steps are alternatively repeated until the change in the log-likelihood value stabilizes, indicating convergence (Vermunt [64]).

The algorithm for LCA can be further explained using an illustrative example adapted from Vermunt et al. [65]. Let us consider three chronic conditions, anxiety, depression, and substance-related disorders, which are typically prevalent among patients with mental or behavioral illnesses. Let C_1, C_2, and C_3 refer to the binary indicator variables representing whether a patient has anxiety, depression, and substance abuse, respectively. The vector notations $\mathbf{C}$ and $\mathbf{c}$ are used to refer to a complete response pattern. Furthermore, let X be the underlying latent variable and N the number of latent classes; each latent class (LC) is enumerated by the index x, $x = 1, 2, \ldots, N$. In this example, the goal is to identify subgroups of patients exhibiting varying levels of mental illness diagnoses.

The fundamental concept behind an LC model is that the probability of obtaining a response pattern $\mathbf{c}$, $P(\mathbf{C} = \mathbf{c})$, is the weighted average of the N class-specific probabilities $P(C = c \mid X = x)$, that is

$$P(\mathbf{C} = \mathbf{c}) = \sum_{x=1}^{N} P(X = x) P(\mathbf{C} = \mathbf{c} \mid X = x), \tag{7}$$

Table 2. Prevalence of different response patterns in the data.

						Expanded frequencies	
C_1	C_2	C_3	Frequency	$P(X=1\|\mathbf{C}=\mathbf{c})$	$P(X=2\|\mathbf{C}=\mathbf{c})$	$X=1$	$X=2$
0	0	0	696	0.998	0.002	694	2
0	0	1	68	0.929	0.071	63	5
0	1	0	275	0.876	0.124	241	34
0	1	1	130	0.168	0.832	109	21
1	0	0	34	0.848	0.152	29	5
1	0	1	19	0.138	0.862	3	16
1	1	0	125	0.080	0.920	10	115
1	1	1	366	0.002	0.998	1	365

Source. Adapted from Vermunt and Magidson [65].

where $P(X=x)$ is the proportion of patients belonging to the LC x. The assumption of local independence, that is, mutual independence of the indicator variables within each LC, is formulated as

$$P(\mathbf{C}=\mathbf{c}\,|\,X=x)=\prod_{i=1}^{3}P(C_i=c_i\,|\,X=x). \tag{8}$$

Once the conditional response probabilities $P(C_i=y_i\,|\,X=x)$ have been estimated, observing the variability among these probabilities across the LCs reveals the distinct features of each class. Combining Equations (1) and (2), the model for $P(\mathbf{C}=\mathbf{c})$ can be written as

$$P(\mathbf{C}=\mathbf{c})=\sum_{x=1}^{N}P(X=x)\prod_{i=1}^{3}P(C_i=c_i\,|\,X=x). \tag{9}$$

The data given in Table 2 results in the model described in Table 3.

The two classes identified have 62% and 38% of the patients. Because the first class has higher probabilities of having the response 0 (patient does not have the chronic condition) for the three indicator variables compared with the second class, the LCs can be named "no serious mental health issues" and "serious mental health issues," respectively.

The posterior class membership probability, that is, the probability of belonging to class x given a certain response pattern $\mathbf{c}$, can be used to classify patients into the appropriate LC. This can be calculated using Bayes rule as

$$P(X=x\,|\,\mathbf{C}=\mathbf{c})=\frac{P(X=x)P(\mathbf{C}=\mathbf{c}\,|\,X=x)}{P(\mathbf{C}=\mathbf{c})}. \tag{10}$$

The class membership probabilities $P(X=x)$ reported in Table 3 are based on modal assignment, which means each patient is assigned to the LC with the highest $P(X=x\,|\,\mathbf{C}=\mathbf{c})$ value. This is the most commonly used classification rule in LCA and other clustering methods. The

Table 3. Probabilities for two latent classes.

	$X=1$ (no serious mental health issues)	$X=2$ (serious mental health issues)
$P(X=x)$	0.62	0.38
$P(C_1=0\,\|\,X=x)$	0.96	0.23
$P(C_2=0\,\|\,X=x)$	0.74	0.04
$P(C_3=0\,\|\,X=x)$	0.92	0.24

Source. Adapted from Vermunt and Magidson [65].

average posterior membership probabilities within each class are indicators of the effectiveness of the latent variable in segregating patients into distinct subgroups. In this example, the average posterior membership probabilities are 0.913 for the "no serious mental health issues" class and 0.903 for the "serious mental health issues" class, indicating a high likelihood that patients truly belong to their respective groups. In other words, patients within each class exhibit similar characteristics or features that align with the classification.

Algorithm 1 provides a more comprehensive explanation of the iterative procedure mentioned previously. If there are I unique response patterns in the data and f_i is the frequency of each pattern i, the log-likelihood can be evaluated for each iteration as follows:

$$ln\,\mathcal{L} = \sum_{i=1}^{I} f_i ln P(\mathbf{C} = \mathbf{c}_i), \tag{11}$$

$$P(\mathbf{C} = \mathbf{c}_i) = \sum_{j=1}^{N} P(X = x_j) P(\mathbf{C} = \mathbf{c}_i | X = x_j). \tag{12}$$

The algorithm is run multiple times with different random initializations to avoid finding just the local optimum. Various tools and software, including Mplus (Muthén and Muthén [41]), the poLCA package in R (Linzer and Lewis [34]), and the Stepmix package in Python (Morin et al. [38]), are used to implement LCA. These packages typically include customizable EM optimization parameters, often with predefined default values. For instance, in Stepmix, parameters such as the maximum number of EM iterations, the tolerance for stopping the optimization (default 10^{-3}), and the number of different initializations to try are provided (Morin et al. [38], Lacourse et al. [33]).

Algorithm 1 (LCA Algorithm)

1: Initialize the current model probabilities in Table 3 with random values.
2: Calculate the conditional response probabilities using Equation (8) and the complete response probabilities using Equation (9), based on the current model probabilities.
3: Evaluate the log-likelihood based on the newly found pattern probabilities $P(\mathbf{C} = \mathbf{c})$, using Equations (11) and (12).
4: [E-Step] Calculate the class membership probabilities using Equation (10) and create an expanded frequency table based on these probabilities for each pattern (i.e., the new class membership frequencies for each pattern).
5: [M-Step] Recalculate the model probabilities in Table 3 based on values in the expanded frequency table.
6: Repeat steps 2–5 until the change in log-likelihood stabilizes to zero or some acceptable minimum threshold.

2.3.2. LCA in Practice. While using LCA in practice, it is imperative to account for certain key considerations.

1. **Multicollinearity**: Mixture modeling assumes "local independence" within latent classes, implying that observed variables are independent within each class. However, the degree of acceptable correlation between variables and its impact on model fit is uncertain. Sensitivity analyses, such as excluding highly correlated variables or allowing them to be correlated in the model, are recommended to assess the effects on class composition and model fit statistics, ensuring robustness and accuracy in LCA (Sinha et al. [57]).

2. **Sample Size**: Prior studies indicate that sample sizes between 300 and 1,000 are sufficient for most of the commonly used fit indices in mixture models. However, for simpler LCA models featuring only a few well-separated classes, as few as 30 samples might be

sufficient. Conversely, it is recommended to have larger sample sizes for complex LCA models with a large number of indicators and classes, to ensure accurate parameter estimation and identification of true classes (Nylund-Gibson and Choi [45]). It is also important to check if the sample size within each latent class is big enough to spot meaningful differences in the predetermined outcome measures, such as clinical outcomes (Sinha et al. [57]). A constraint such that each latent class should comprise at least 10% of the patient population could also be applied, as smaller class sizes may not be substantial for facilitating tailored interventions (Smeet et al. [60]).

3. **Number of Classes**: The process of selecting the most appropriate LCA model involves fitting models ranging from 1 to N classes, where N is the largest reasonable number of latent classes that might be present in the data, it is subjective to each data set. Fit statistics such as Akaike information criteria (AIC, ideally minimized), Bayesian information criteria (BIC, ideally minimized), log likelihood (ideally maximized), entropy (ideally maximized), and average latent class posterior probability (ideally maximized) can be plotted against the number of classes to select the best model. However, it must be noted that the statistical criteria should complement the interpretability of the model. A model exhibiting superior statistical metrics holds limited utility if it lacks clinical coherence (Weller et al. [67]).

4. **Item-Response Probabilities**: The probability that an indicator variable takes on a particular value when the latent class membership is known, is defined as $P(C_i = c_i | X = x)$ (the values in Table 3). These probabilities represent the strength of the relationship between the observed indicator variables and the latent variable and are used to interpret and define the latent classes. In a good LCA fit, we should observe 2 properties: homogeneity and latent class separation (Bray [8]). Homogeneity means that all the item-response probabilities are either close to zero or close to one; that is, there is either a strong negative or strong positive relationship. Latent class separation refers to the degree to which the latent classes can be distinguished from each other. If all item-response probabilities are homogeneous, but have the same strength, there is no separation between variables. Table 4 shows what good and bad fits might look like for the example discussed in Section 2.3.1.

In the healthcare domain, the indicator variable selection depends on the research question and is generally an iterative process involving preliminary data collection, analysis, model fit assessments, and feedback from clinicians. The LCA study by Davis et al. used 53 hierarchical chronic condition categories and identified seven clinically distinctive subgroups: "end-stage renal disease," "cardiopulmonary conditions," "diabetes with multiple comorbidities," "acute illness superimposed on chronic conditions," "conditions requiring highly specialized care," "neurologic and catastrophic conditions," and "patients with few comorbidities" (Davis et al. [16]). A study by Smith et al. [61] aimed to characterize the impact of multimorbidity on individuals experiencing homelessness by identifying distinct groups based on medical, psychiatric, and substance use disorder profiles and comparing clinical outcomes across these groups. Indicators were extracted from EHR for 497 adults referred to the Durham Homeless Care Transitions (DHCT) program over a period of four years. Five distinct groups were identified

Table 4. Homogeneity and latent class separation.

	High homogeneity + latent class separation (good fit)		High homogeneity + no latent class separation (bad fit)	
	$X = 1$	$X = 2$	$X = 1$	$X = 2$
$P(C_1 = 0 \mid X = x)$	0.84	0.13	0.84	0.95
$P(C_2 = 0 \mid X = x)$	0.96	0.22	0.96	0.89
$P(C_3 = 0 \mid X = x)$	0.01	0.89	0.01	0.05

using LCA: "low morbidity" (referent)," "high comorbidity," "high tri-morbidity," "high alcohol use," and "high medical illness." LCA has also been used to categorize the top 10% of care users or patients with multimorbidity with above-average care utilization, based on characteristics more geared toward their socio-demographic conditions in a study by Smeets et al. [60]. A cohort of 12,602 patients was divided into four classes distinguished by dominant characteristics such as age, household position, and source of income. The indicator variables used in the final model can be a subset of the originally chosen variables as well. The original LCA model included 41 indicator variables, of which 32 were discarded in the final model due to low statistical relevance (i.e., the item-response probabilities were not distinguishable among the different classes). The final model used only nine variables to identify the same latent classes in a more streamlined manner, which made it easier to interpret and analyze.

Once the latent classes have been identified using the indicator variables, associations and patterns can be determined by analyzing exogenous variables (i.e., data that were not used in the LC model). For example, in the study by Davis et al. [16], survival rates ranged from 43% to 88% across the identified subgroups (Table 5). The difference in survival rates marked a clear difference in the healthcare needs of each subgroup. Likewise, in the research conducted by Smith et al. [61], individuals categorized under the high medical illness classification exhibited higher mortality rates within 12 months of being referred to the DHCT program, in contrast to those classified under the low morbidity category. Additionally, both the high comorbidity and high tri-morbidity groups demonstrated elevated risks of drug overdose within the same timeframe following referral compared with the referent group: low morbidity.

In contrast to these two studies, Smeeths et al. [60] conducted a post hoc analysis on the prevalence of chronic conditions within latent classes, which were identified using demographic, biomedical, and socioeconomic characteristics in combination with pharmaceutical costs. The four latent classes they identified are as follows: "older adults living with partner," "older adults living alone," "middle-aged, employed adults with family," and "middle-aged adults with social welfare dependency" (Figure 4). They discovered that diabetes was the most prevalent across all classes (30.5%–43.4%). The second most prevalent condition varied between osteoarthritis in classes with older adults (21.7%–23.8%), asthma in the middle-aged and employed group (25.3%), and mood disorders in the group with middle-aged adults with social welfare dependency (23.1%). Moreover, although there was an increase in the utilization of general practitioner (GP) care among the older adult classes over the follow-up period, it remained fairly consistent among middle-aged classes.

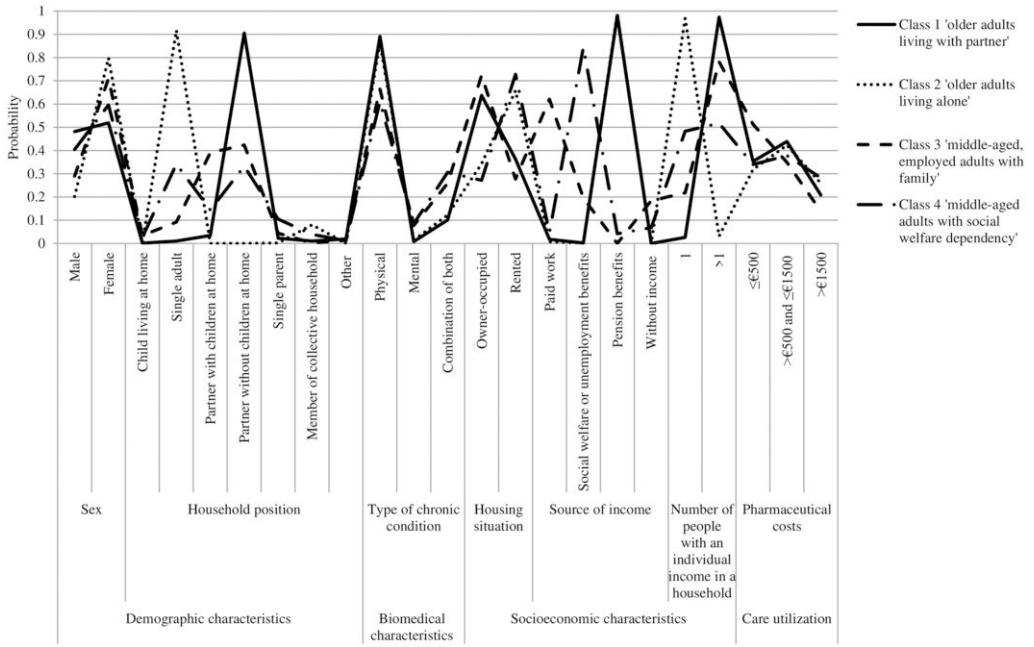
The findings presented in these studies carry significant implications; they showcase how the integration of social, behavioral, and medical data can yield a detailed understanding of high-risk patient subgroups, hence opening up possibilities for targeted interventions and appropriate care planning strategies (Rinehart et al. [52]).

Table 5 Survival rates by group.

	Percentage surviving through 2011	Percentage surviving through 2014
All Top 1%	88	69
Class 1: End Stage Renal Disease	91	68
Class 2: Neurologic and Catastrophic	90	73
Class 3: Cardio-Vascular and Pulmonary	87	66
Class 4: Diabetes with Multiple Comorbid Conditions	80	46
Class 5: Highly Specialized Treatments	81	62
Class 6: Acute Exacerbation of Chronic Conditions	72	43
Class 7: Few Comorbidities	96	88

Source. Data from Davis et al. [16].

Figure 4. Item-response probabilities of each indicator variable the latent classes, in a Dutch primary care group.



Source. Republished from Smeets et al. [60].

3. Complex Care Interventions: Operational Details, Staffing Needs, and Impact on Patient Outcomes

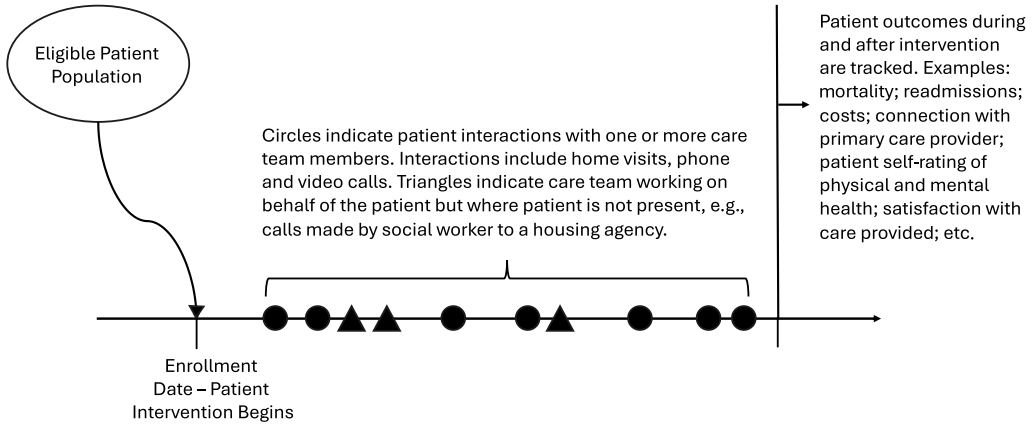
As described in Section 1, organizations around the country have developed programs or interventions for patients with complex medical and social needs. In this second part of the tutorial, we turn our attention to the operational and staffing features of such interventions, and the evidence of their effectiveness in improving patient outcomes. The section is organized as follows. After describing the features of complex care programs (Section 3.1), we summarize two examples (Section 3.2). We then turn to our primary case study: the complex care program of our collaborating partner, the Camden Coalition (Section 3.3). We provide patient-level examples of intervention progression, quantify how they vary between patients, and present an algorithm to adequately staff care teams. We conclude the section with the Camden Coalition's RCT-based evaluation of their interventions.

3.1. Features of Complex Care Interventions

Complex care programs across the United States vary considerably in their target population, staff used, duration of the intervention, and the outcomes evaluated (see Chang et al. [12] for the most recent systematic review and meta-analysis of programs). However, the programs do share some key features which we encapsulate in Figure 5. We describe these features in detail here.

3.1.1. Eligibility. All complex care programs begin with defining the pool of eligible patients. Eligibility depends on the scope of the intervention. The most common eligibility criteria include patients with two or more conditions; patients in the top percentiles of annual medical expenditures; patients with a high number of ED visits or hospitalizations or unique

Figure 5. Key features of complex care interventions.



medications; or some combination of these. Patients are typically identified based on data in electronic health records and claims data and are approached by one or more care team members at a primary care appointment or during an inpatient hospitalization.

For specialized programs such as the Boston Healthcare for the Homeless (O'Connell et al. [46]), chronically homeless individuals in the Greater Boston area living in shelters would be the eligible population. For the Transitions Clinic (Shavit et al. [56]), a consortium of 45 primary care programs active in 14 states and Puerto Rico that provides community health worker support for patients returning to the community from incarceration, the eligible population consists of individuals older than 50 with multiple chronic conditions who are about to be released from prison.

3.1.2. Care Team Composition and Intervention Goals. The intervention begins when the patient consents to be part of the program. Programs vary widely in the types of staff members used. Interventions are often provided in teams. Teams can consist of physicians, nurses, community health workers, social workers, pharmacists etc. Even in cases where the intervention is led by a single member, for instance, community health workers, the staff member still works with a medical team.

A typical doctor's visit is generally focused on a specific disease, treatment, procedure, or resolving a set of symptoms; it often addresses only a small portion of the patient's full set of medical needs. Patient-care team encounters in complex care, in contrast, are designed to consider how all medical conditions, associated medications, and treatments of the patient interact and affect well-being. Interventions typically begin with the care team having detailed conversations and interviews with the patient to get a holistic, comprehensive picture of the patient's current situation. A set of patient-centered goals define the scope of the intervention. The care team tries to facilitate better access to primary, specialty, and mental health care; some members of the care team even accompany the patients to their appointments. The care team's activities also touch on aspects of the patient's personal life that are barriers to health and well-being. They include addressing the lack of family member or peer support, housing instability and homelessness, and challenges with alcohol and substance abuse.

Such in-depth conversations and related action plans require care team members to be empathetic and build trust: a slow process that can take multiple attempts and may not always be successful. The role of community health workers, who often have the same

socioeconomic background as the patient and in many cases have gone through the same health and social challenges, is crucial for patient engagement. The intervention ends when the patient's immediate goals have been met and they can go about their daily life with minimal support from the care team.

3.1.3. Intervention Progression as a Stochastic Process. From an operational point of view, the patient and the care team have recurring interactions, shown in Figure 5, creating a longitudinal event history. This event history is a realization of an underlying stochastic process. Encounters include interactions in which the patient is present (physically or virtually) and activities that the care team carries out on behalf of the patient but where the patient is not present (such as a nurse in the care team calling the patient's cardiologist, or a social worker calling a landlord to secure apartment housing for the patient).

The process is stochastic because there is uncertainty in the types of encounters (home visit, accompanying the patient to a specialty visit, phone call, etc.), the time spent in the encounter, and the number of days between encounters. The precise sequence of encounters varies from patient to patient because each patient has a different set of needs. The duration of the intervention varies significantly between programs, some last for two weeks, whereas others last for six months or longer. The total time spent by the care team across all encounters from the start to end of the intervention is often termed as the dosage of the intervention. Dosage too can vary significantly between patients, as we will demonstrate in Section 3.3.

3.1.4. Intervention Outcomes. Complex care programs have tracked a wide range of patient outcomes such as mortality, hospital and ED visit rates, number of days spent in hospital, and expenditures. Intermediate measures such as the number of outpatient primary and specialty care visits and whether patients received a primary care appointment after hospital discharge are also of interest because these are indirect indicators that the patient's chronic conditions are being managed. In one study (Kangovi et al. [29]), clinical measures such as systolic blood pressure, blood sugar levels, and the number of cigarettes smoked per day were tracked. Finally, more qualitative metrics such as patient satisfaction with care and self-rating of physical and mental health are also collected.

The most common time frames for tracking outcomes are 3-, 6-, and 12-month periods before and after the start of the intervention. However, simply conducting a pre-post analysis is methodologically problematic, especially for utilization measures such as the number of hospitalizations and ED visits. Patients are enrolled in complex care programs precisely because their hospitalizations and ED visits have reached the highest observed percentiles in, say, a preceding six-month period. In the following six-month period, these same measures are more likely to exhibit regression toward the mean, that is, they are more likely to drop naturally, with or without an intervention. This may give the false impression that an improvement has occurred due to the intervention. The correct approach, therefore, is to include a control group, that is, track outcomes in similar time periods for a group of eligible patients who did not receive a complex care intervention.

RCTs are the most unbiased methods to create intervention and control groups. Results of several complex care RCTs have been published in recent years. However, RCTs are logistically challenging and costly to organize. Furthermore, complex care RCTs involve a small percentage (1%–5%) of the population, which results in low trial recruitment rates. Most trials have a few hundred patients in total across intervention and control arms. Because of these challenges, most complex care programs are unable to evaluate their outcomes through RCTs. An alternative to RCTs is to identify a control group retrospectively by matching covariates through methods such as propensity score matching (Caliendo and Kopeinig [10]).

3.2. Examples of Complex Care Interventions

Although our primary case study is from the Camden Coalition (Section 3.3), to give a sense of the diversity of approaches to complex care, we summarize two additional examples of complex care programs.

3.2.1. Kaiser Permanente Mid-Atlantic States Complex Care Program. In 2017–2018, the managed care provider Kaiser Permanente started a new Complex Care Program (CCP) in the mid-Atlantic states (Roblin et al. [53]). A team consisting of a primary care physician and a nurse assisted a special panel of patients with advanced clinical disease and a history of recent hospitalizations. A typical primary care physician has a panel of 1,800 patients, whereas a Kaiser physician dedicated to serving complex care patients was assigned a panel of 200 patients. This allowed the physician to spend more time on in-person visits (45 minutes as opposed to the typical 15- or 20-minute visit) in addition to nontraditional visits (visits to the patient's home or nursing home and video calls). The physician and nurse team also spent a significant amount of time providing personalized care: reviewing patients' medical records, reconciling medications prescribed by multiple specialists, and addressing social needs such as financial challenges in making medical payments and lack of transportation.

A major finding was that patients who were assigned to complex care panels had lower mortality rates (17.2%) than propensity score–matched control patients (26%) six months after enrollment in CCP. However, hospital readmission rates six months after enrollment were not statistically different between CCP patients and matched control patients.

3.2.2. University of Pennsylvania Community Health Worker Program. In this program (Kangovi et al. [29]), a community health worker (CHW) assisted patients in a high-poverty region of Philadelphia. The patients had two or more of the following chronic diseases: hypertension, diabetes, obesity, and tobacco dependence. The program was titled Individualized Management of Patient-Centered Targets (IMPaCT), and enrollment of patients happened during a primary care visit. The CHW and the patient worked together to develop "patient-driven action plans." For instance, if the patient's goal was to address obesity by losing weight through better nutrition, the CHW might accompany the patient to a food pantry where affordable fresh produce was available. If the patient wanted more information about diabetes, the CHW would guide them to a diabetes educator at the local YMCA. The CHW interacted with each of the patients for at least six months from the enrollment date via phone calls, texts, and in-person visits and facilitated patient support groups. The average time a CHW spent with a patient, that is, the average dosage, was 38.4 hours.

The CHW-led intervention was tested in a randomized control trial involving 302 participants (150 in intervention and 152 in control). All patients in the trial collaboratively set a disease management goal with their primary care physician; however, the intervention groups received additional support from a CHW (see previous examples) to assist with the goal. The intervention group saw statistically significant drops in clinical outcomes such as blood sugar levels, weight loss, and the number of cigarettes smoked per day six months after the intervention. For instance, blood sugar levels, as measured by glycosylated hemoglobin (known widely as the HbA1c level) dropped by 0.4 points in the intervention group compared with the preintervention baseline; there was no change in the control group. Self-rated mental health also improved in the intervention group. 23% of the patients in the intervention group were hospitalized one year after trial enrollment while the same rate in the control group was 31%. Not all measures improved in the intervention group. For instance, systolic blood pressure (SBP) was reduced by 11.2 points in the control group (a clinically significant reduction) compared with the pre-intervention baseline; in the intervention group, it was reduced only by 1.8 points.

Other evaluations of the IMPaCT program (Kangovi et al. [30], Kangovi et al. [31]) have yielded generally positive results, and the program has been adopted widely around the country by state and local governments, insurance providers, and the VA.

3.3. Case Study: Camden Coalition

The Camden Coalition is a multidisciplinary nonprofit organization that works to improve care for patients with complex health and social needs in Camden, New Jersey, and nationally. We showed an example in Section 1 of how the Coalition uses multidisciplinary care teams to assist a patient with complex medical and social needs.

Founded by a family physician named Jeffrey Brenner in 2002, the effort grew into a city-wide initiative (hence the term Coalition) that brought together "hospitals, primary care, social service providers, and community representatives" in Camden, New Jersey (Noonan and Craig [44]). The Coalition's goal is to empower patients with the skills and support they need to avoid preventable hospital use and improve their well-being. In 2007, the Coalition piloted its first care team intervention. Since then, it has worked with thousands of people from the Camden region struggling with chronic health issues, addiction, mental health challenges, poverty, unemployment, housing instability, child welfare issues, and criminal justice involvement. Their work has attracted nationwide attention and has been featured in *The New Yorker* (Gawande [22]) and *PBS Frontline* (Gawande [21]).

The intervention developed by the Coalition is called the Core Model. In what follows, we describe and quantify key operational features of the Core Model; see Martinez et al. [36] and Koker et al. [32] for further details. To enroll patients, the care team first identifies patients currently admitted in Camden area hospitals with a history of two or more hospitalizations in the previous six months. Among this set, patients are considered eligible for intervention if they satisfy two or more of the following criteria:

- Two or more chronic conditions
- Polypharmacy, as defined by five or more medications
- A mental health diagnosis
- Homelessness
- Active substance use
- Lack of social support
- Difficulty accessing services (e.g., limited physical mobility, language barrier, lack of transportation, etc.)

The care team approaches the patient in the inpatient setting. The intervention begins once the patient consents to be part of the program. The patient is assigned to a community health worker (CHW) and licensed practice nurse (LPN) pair who lead the intervention. Other team members who assist the pair include a social worker (SW), a registered nurse (RN), and a clinical psychologist (CP). Patients and the care team collaboratively determine domains in which the patient needs assistance. The Coalition lists 16 such domains; examples include medication and medication supplies, transportation support, housing, relationships with primary care providers, legal assistance, addiction support, etc.

Care team members aim to meet with each patient in their home within five days of discharge to conduct a review of the patient's current medications. They also aim to schedule a primary care appointment for the patient within seven days of discharge. Nurses in the care team (LPN or RN) accompany the patient to the PCP appointment. The care team continues to assist the patient with clinical and social coordination in the ensuing weeks. When the team determines that the patient's clinical and social goals have been met and can go about their daily lives with minimal support, the patient is said to "graduate" from the intervention. Otherwise, the intervention is considered "incomplete." Incomplete

interventions can happen when a patient cannot be reached despite multiple attempts by the care team, is no longer interested in further services, has moved out from the Camden region, or has passed away. In some cases, incomplete patients may re-enroll again after a hospitalization.

3.3.1. Analyzing Intervention Progression. As conceptualized earlier, the intervention leads to a series of stochastic recurring encounters between the patient and the care team. The Coalition collects very detailed operational data on these encounters.

Figure 6 shows a visualization of intervention events for three patients from the start of the intervention (day 0) to their graduation from the intervention. Shapes indicate the type of care team-patient encounter and color indicates the type of staff member involved. Stacking indicates that multiple staff types can be involved in the same encounter. Home visits, for instance, are often conducted by a pair of staff members. Importantly, the Camden Coalition also collects the amount of time spent by each staff type in each encounter. Home visits and accompanying patients to specialty and PCP visits can take more than an hour, whereas phone calls can be as short as a few minutes. The sequence of encounters and time spent in each encounter allows us to reconstruct the precise sequence of encounters for each patient and infer how the duration of intervention and dosage (amount of care team spent) varies among its patients.

Let S be the set of all staff types and s refer to a particular staff type. At the Camden Coalition, $s \in \{chw, lpn, rn, sw, cp\}$. Let E denote the set of all encounter types (home visit, accompanying patient to the PCP, phone call, etc.), and $e \in E$ a particular type of encounter. Let $z^e_{i,d,s}$ denote the number of hours spent by staff type s with (or on behalf of) patient i in encounter type e on day d of the intervention. Here d can range from enrollment day $d = 0$ to week $d = \mu_i$, when the intervention concludes and an outcome O_i for the patient $\{incomplete, graduated\}$ is determined. The cumulative intervention dosage for patient i on day k, $0 \le k \le \mu_i$, across all care team staff members is given by

$$Z_{i,k} = \sum_{d=0}^{k} \sum_{s \in S} \sum_{e \in E} z^e_{i,d,s}. \tag{13}$$

When $k = \mu_i$, $Z_{i,k}$ denotes the total dosage for patient i from the start of the intervention to the outcome date. Figure 7 shows $Z_{i,d}$ (y axis) for $d = 0, 1, \ldots, \mu_i$ for one patient. Steeper parts

Figure 6. (Color online) Examples of intervention timelines.

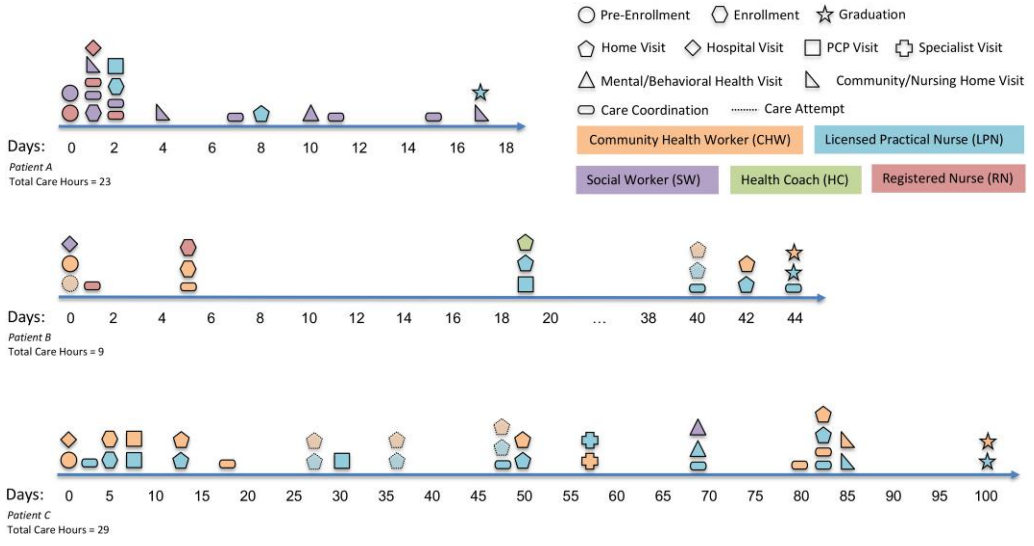
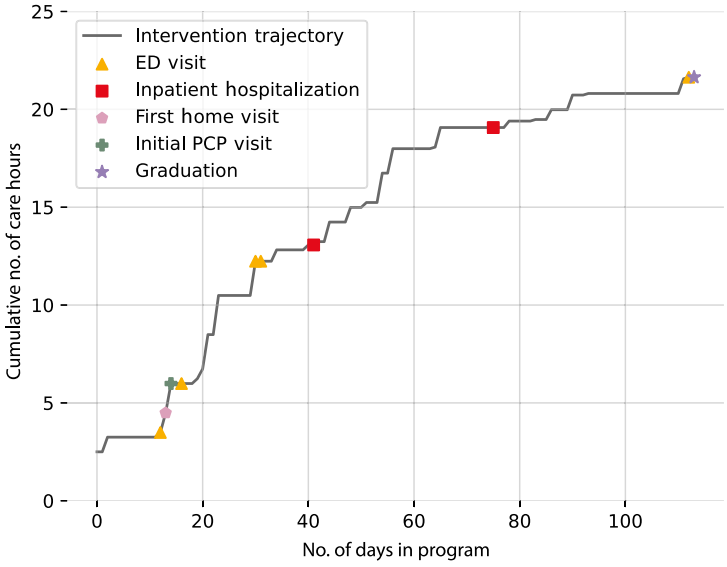


Figure 7. (Color online) Intervention trajectory of a specific patient.

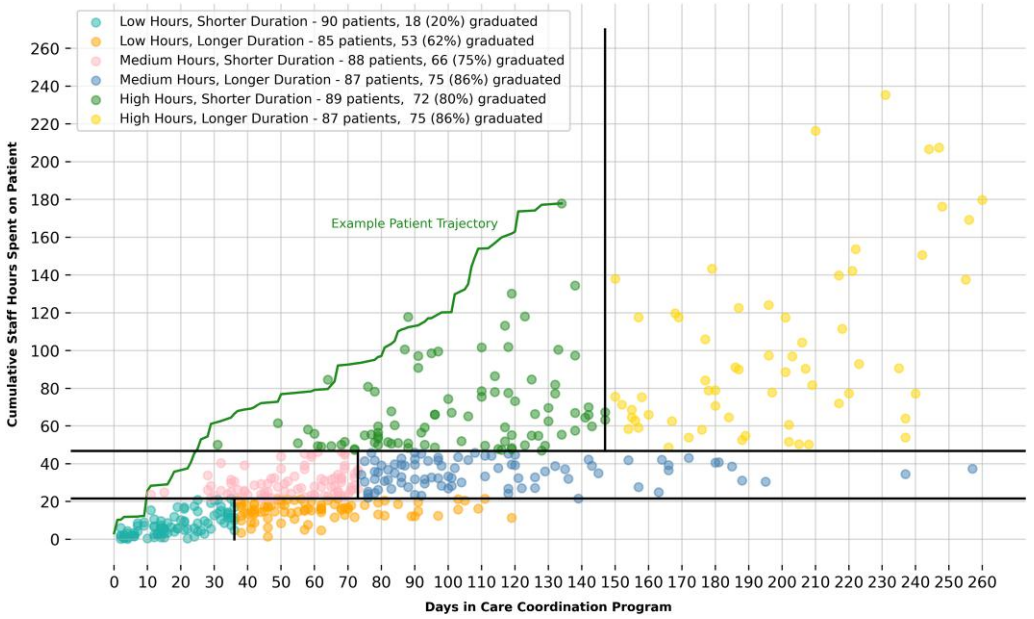


of the trajectory indicate more care team effort in a short period while plateaus indicate no care team effort. The trajectory also illustrates important events during intervention such as the first home visit, the initial PCP visit, ED visits and inpatient hospitalizations.

The Coalition enrolled 531 patients between 2012 and 2015. In Figure 8, we show a scatter plot (μ_i, Z_{i,μ_i}) for these patients. The plot tell us how the interventions vary in two measures: the length of the intervention (μ_i shown on the x axis) and the total dosage (Z_{i,μ_i} on the y axis). To cluster patients in the two-dimensional space, we first divided Z_{i,μ_i} into three groups. Each group was further subdivided into two subgroups based on μ_i. This gives us six interpretable clusters with approximately equal numbers of patients. We also report the percentage of patients who graduated in each cluster. Patients in the lowest left cluster ("low hours, lower duration"), for instance, had the shortest intervention durations (less than 40 days) and the lowest intervention dosage (less than 20 hours); this cluster also had the lowest graduation rates (18.2%). In contrast, patients in the top right cluster ("high hours, higher duration"), had the longest intervention durations (greater than 150 days) and the highest intervention dosage (greater than 45 hours); graduation rates in this cluster were the highest (86.5%) suggesting that care team was able to address patient goals in most cases. We also show the full trajectory $Z_{i,k}$ for $k = 1, 2, \ldots, \mu_i$ for one patient whose final endpoint is in the top left cluster. Trajectories in this cluster are some of these steepest, requiring a large amount of care team effort in a short period of time.

These results demonstrate how intervention duration and dosage can vary significantly even among patients with complex medical and social needs. Indeed, the Core Model was originally planned as a 90-day intervention but results in Figure 8 suggest that interventions can be far shorter or longer than 90 days. Similarly, intervention dosage also varies significantly: The 90th percentile of intervention dosage was three times the median dosage. Uncertainty in intervention duration and dosage complicates staffing decisions. Ideally, medical or social covariates could be used to predict these quantities. However, complex care programs treat a small number of patients which makes it difficult to build accurate predictive models. Additionally, although we identified characteristics associated with the top quintile of intervention dosage through a random forest model (Martinez et al. [36]), such as count of mental health conditions, homelessness and housing needs, inadequate nutrition, and lack of

Figure 8. (Color online) Variation in patient trajectories.



Source. Adapted from Martinez et al. [36].

relationships with medical providers, these models still have high variability when it comes to predicting staff hours needed each week.

3.4. Nonstationary Features

Another important feature of the interventions is that they are front-loaded and nonstationary; that is, intervention intensity is higher in the first two weeks after enrollment and reduces with time. Table 6 shows that approximately a fifth of the 24,249 total hours spent by the care team on the 531 patients occurred in the first 12 days after enrollment and 40% of the total effort happened within 33 days of enrollment. This front-loaded pattern reflects the need for the care team to develop strong relationships with their patients through home visits, accompanying the patient to primary care appointments, and taking all the needed steps to avoid readmissions within 30 days of discharge (recall that the patients are enrolled during an inpatient hospitalization). The workload of the nurses (RN and LPN) is high in the first two weeks of the intervention as clinical needs after hospital discharge take precedence. As patients graduate, relatively smaller numbers are active in later stages (active patients are those that do not have a graduated or incomplete outcome yet). However, we see that 166 of the 531 patients were still active four months after the start of the intervention, and 20% of the care team's effort was spent on these patients for over a year (days 188–575). A deeper

Table 6. Distribution of (approximate) deciles of care team hours by time period since intervention enrollment.

Fifth of postenrollment effort	First	Second	Third	Fourth	Fifth	All
Delivery period (days)	0–12	13–33	34–62	63–117	118–575	0–575
Patients active	531–509	509–457	457–347	347–166	166–1	531–1
Total care team time (h)	4,813	4,797	4,864	4,919	4,856	24,249
Total care team time (percent of total)	19.8%	19.8%	20.1%	20.3%	20%	100%

Source. Adapted from Martinez et al. [36].

analysis reveals that most of these patients needed higher engagement of the social worker's time for activities such as assistance in obtaining housing. These results show that different staff types are needed at different stages of the intervention.

3.5. Staffing Algorithm

An important concern in complex care is how multidisciplinary care teams should be staffed for a given rate of patient enrollment. Specifically, how many hours of each staff type (community health worker, nurse, social worker, etc.) would be needed for, say, an enrollment of three patients per week, to ensure that patients receive timely assistance? Staffing decisions are challenging because, as our preceding analyses demonstrate, intervention durations are highly unpredictable, the progression of the intervention is nonstationary, and the workload of different staff members peaks at different stages of the intervention. Analytical queueing network models are unable to capture these features. Even a discrete event simulation of the nonstationary intervention process is challenging since many parameters need to be estimated from a data set with a small number of patients.

To overcome these challenges we use a simple sampling-based simulation algorithm (Koker et al. [32]). The algorithm uses the patient's complete sequence of encounters from enrollment to graduation; see Figure 6 for examples of encounter sequences. The advantage of using a patient's chronologically sequenced record is that it implicitly captures the nonstationary dynamics of the intervention and differences in staff involvement. By sampling in such a way that different combinations of patients are active at any time, we can calculate estimates of staff hours needed for a given patient enrollment rate. We briefly describe the algorithm next.

Consider a simulation time horizon of $t = 1, 2, \ldots, T$ days, and an enrollment rate that follows a Poisson distribution with a mean of λ patients per day. On each day, starting with $t=1$, we sample from this Poisson distribution to determine the number of enrollments (distributions other than Poisson can also be used in the framework). Let n_t denote the number of enrollments on day t. We then randomly sample n_t patient indices without replacement from our total set of 531 patients and assign them to be enrolled on day t. The event history of each sampled patient is assumed to unfold in the exact sequence observed in the data from day t onward, with day t being the first day of the intervention, $t + 1$ being the second day, and so on, until $t + \mu_i$ when the patient is assigned an outcome and graduates from the intervention. The simulation runs until no more patient indices remain to be sampled (recall that indices are sampled without replacement). If a patient is enrolled once every two days, that is, with a rate of $\lambda = \frac{1}{2}$ per day (observed at the Coalition), then the simulation will run on average for $531 \times 2 = 1{,}062$ days, which is a long enough period to estimate key metrics. We choose sampling without replacement because the encounter histories of patients with longer interventions and larger intervention dosages can bias the results if picked repeatedly.

By Little's Law (Little and Graves [35]), we know that the simulation reaches a steady state when the mean number of active patients in the program is $\lambda \times \mu$ where μ is the mean intervention duration. The precise day that the simulation reaches steady state can be determined either visually or through more formal moving average approaches such as Welch's method [66]. Consider any day k when the simulation has reached steady state. Let W_k denote the patients that are active in the intervention on day k and let $\tau_{i,k}$ be the number of days that patient $i \in W_k$ has been active on simulation day k. Then, the total demand for staff type s on day k is given by

$$D_{s,k} = \sum_{i \in W_k} z_{i,\tau_{i,k},s}. \tag{14}$$

Here $z_{i,\tau_{i,k},s}$ denotes the time spent by staff type s on patient i on day $\tau_{i,k}$ of the intervention. After recording $D_{s,k}$ for all steady-state days across multiple replications of the simulation, we can estimate the distribution of workload for staff type s. This method of estimating

demand is called offered load analysis where we measure the capacity needed as if there were no limits on its availability. See Whitt et al. [68] for more details. The distribution of the offered load (estimated using $D_{s,k}$ values) can be used to determine staffing levels to meet performance targets (e.g., ensure that delays for complex care patients do not exceed a predetermined target). Offered load analysis is particularly helpful when there when time-varying dynamics and when queues have a network structure [68].

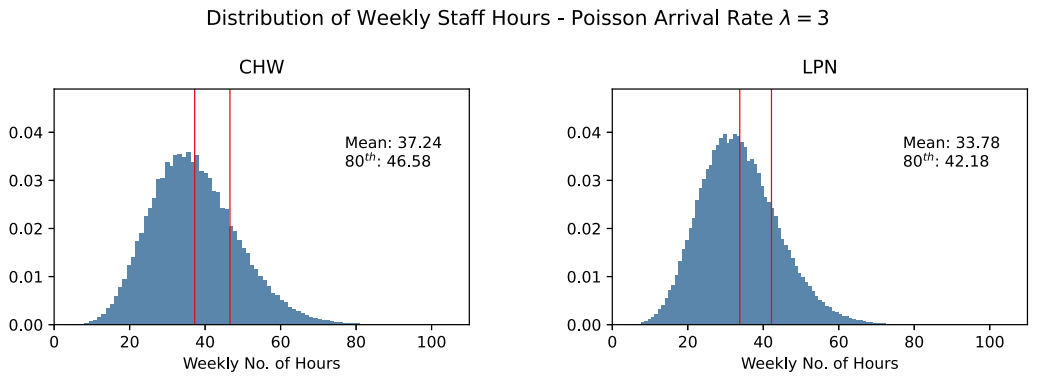
In Koker et al. [32], offered load demand estimates were obtained at the level of each week. Figure 9 shows the histograms of the demand in hours for CHWs and LPNs for a mean enrollment rate of three per week. We see that the CHW and the LPN in the care team will spend on average 37.24 and 33.78 hours, respectively, on behalf of active patients. However, the 80th percentiles of the two distributions suggest that overtime hours beyond the 40-hour workweek will be necessary for both CHW and LPN, especially so for the CHW. These results can be used to make decisions on whether part-time support staff need to be recruited for a given patient enrollment rate. Several other staffing results including joint offered load analysis (because multiple staff are involved in an encounter, staff workloads can often be correlated), and nonstationary arrival rates have been presented in Koker et al. [32].

3.6. Randomized Control Trial of Camden Core Model Interventions

The Camden Coalition conducted a multiyear randomized control trial (2014–2017) in partnership with the Poverty Action Laboratory (J-PAL) at the Massachusetts Institute of Technology. The goal of the RCT was to test whether their interventions were impacting inpatient hospitalizations. Patients who consented to be part of the trial were randomly assigned to intervention and control groups. Results of the trial, published in January 2020 in the *New England Journal of Medicine* (Finkelstein et al. [19]), revealed that the six-month readmission rate (a binary yes/no variable for each patient) in the intervention ($n = 393$) and control ($n = 389$) groups was 62.3% and 61.7%, respectively, six months after assignment. There was no statistical difference. Other measures such as the average number of readmissions, number of days in hospital, and hospital charges were also not statistically different between intervention and control.

Although this null result was a disappointment to many complex care practitioners, it is instructive to revisit the assumptions of an RCT. The RCT assumes that all patients underwent the same intervention when in reality the trial took place over a four-year span where the intervention changed significantly (different staffing arrangements, new service offerings for patients with housing instability, etc.). The RCT thus studies only the average impact of an aggregated measure for a diverse patient population recruited in four years. Furthermore, complex care interventions do not deal with a single disease or administer a single drug or

Figure 9. (Color online) Distribution of weekly staff hours.



Source. Adapted from Koker et al. [32]).

vaccine whose effect can be isolated; they deal with a mix of chronic conditions and social determinants, resulting in a great deal of variability, as revealed by our analysis of intervention dosage and duration. Each patient's clinical profile, when combined with their social needs, makes them essentially unique, and an exact match in medical and social covariates between an intervention and control patient is quite rare.

Two recent analyses that revisited the Camden RCT data set have revealed some intriguing insights. The first demonstrates that readmission rates were statistically lower for patients who engaged more with the care team, and the second found that the RCT was successful in other outcomes. We summarize the two studies here.

3.6.1. RCT Outcomes by Variation in Patient Engagement. Yang et al. [70] looked at variation in patient engagement as an explanatory variable for intervention outcomes. As our intervention analysis demonstrated, many patients received an incomplete status because they did not respond despite multiple attempts by the care team. Yang et al. [70] defined an intervention patient to be "engaged" if at least two of the following three criteria were satisfied: (1) the patient received at least three intervention hours during the first two weeks of enrollment; (2) the patient had contact with staff at least once per week for four of the initial six weeks; and (3) the patient was retained in the program for 60 days (half of the average treatment length) or graduated within that time frame.

Next, they identified medical and social covariates that predicted the binary engagement variable using a gradient-boosting machine learning model. They found that patients with the highest probability of engagement were less likely to have been arrested and less likely to have substance abuse disorder or alcohol-specific hospital diagnoses. The model was used to predict the probability of engagement for each intervention and control patient. Finally, the study compared intervention and control patients with progressively higher probabilities of engagement. Between intervention and control patients who were in the top decile of engagement probability, there were statistically different outcomes when it came to readmission rates (66% readmission rate in control versus 53% in intervention) and average number of hospital visits (1.57 for intervention versus 1.22 for control) six months after enrollment.

3.6.2. Other RCT Outcomes: Access to Primary Care, Specialty Care, and Medical Equipment. Finkelstein et al. [20] looked at whether Camden's Core Model intervention improved access to primary and specialty care in the immediate weeks, months, and a year following enrollment. Reducing hospitalizations can be challenging in complex care patients; however, improved access to primary and specialty care gives the patient a strong preventive foundation and tools for long-term chronic care management. Using Medicaid claims data, the study found that among control patients, only 18.93% received a primary care appointment 14 days after being discharged from the hospital (recall that patients were enrolled in the trial after a hospitalization). Among intervention patients, 33.1% received a primary care appointment. Similar statistically different differences in both the percent of individuals as well as average number of visits were observed at six months and even a year after enrollment. Intervention patients were more likely to receive durable medical equipment (e.g., wheelchairs, oxygen equipment) than the control group (43.06% versus 29.29%).

Thus, a more nuanced picture of the effectiveness of the Camden Core Model has emerged since the aggregate RCT results with the null result were first published in January 2020. These new results suggest that the Core Model did indeed improve outcomes in certain groups of patients and in providing patients access to primary and specialty care.

There's also a subtle link between staffing and RCT outcomes that we wish to explore in future research. The RCT assumes the delivery of intervention to one patient does not impact the intervention for another. This is certainly true in a vaccine or drug trial. However, because complex care is a staffing intervention, a patient's intervention has subtle but important dependencies on the interventions of other patients. This is because the care team's time is a

resource shared by all concurrently active patients. If the care team is short of capacity, among currently active patients a particular subset may be prioritized over others, thus delaying the intervention process for other patients and possibly affecting outcomes. Similarly, patients who enroll when care team workloads are at their peak may have less than optimal outcomes compared with those who enroll when care team workloads are low. Thus, care team availability is key to intervention success and therefore RCT outcomes. The same principle applies also to other resources critical to intervention success such as access to primary and specialty care appointments.

4. Conclusions, Discussion, and Future Research

In summary, we focused on two different themes related to interventions for patients with complex medical and social needs. In the first theme, we illustrated unsupervised learning approaches to segment heterogeneous patient populations into meaningful subgroups that share common groups of diseases and other characteristics. In the second theme, we considered operational features of complex care interventions, in particular, the question of staffing care teams and quantifying the impact on patient outcomes. The methodologies illustrated, association rule mining, latent class analysis, descriptive analyses of patient-care team encounter sequences, and sampling patient sequences to create staffing estimates, can be used by clinicians and administrators of complex care programs.

The tutorial, however, has discussed only a fraction of a challenging domain. There are numerous other questions under each theme and opportunities for research at the intersection of traditional OR/MS-based methods and AI/ML. We highlight some topics here.

MCCs and Care Needs: Although our work focused on the first step of quantifying associations between diseases and disease groups, there is a pressing need to develop a comprehensive understanding of causal mechanisms. Mechanisms could be biological, that is, one disease can compromise an individual's clinical risk for another, or they could be behavioral, for example, lack of exercise and poor diet could be common risk factors for diabetes and cardiovascular diseases. Furthermore, poor diet and exercise could be driven by social determinants such as access and affordability to physical activity and healthy foods.

The biggest challenge in this field is the heterogeneity in disease combinations and the low prevalence of each combination, as illustrated in Section 2. This makes it difficult to estimate the joint prevalence of health, social, and behavioral conditions. Even the largest data sets, such as those available at U.S. government providers such as Veterans Affairs, Medicare, and Medicaid, can become sparse once a population is segmented by age, race, geographic location, income, and education levels. There is a need for systematic methods to estimate joint prevalence distributions that extrapolate beyond marginal and pairwise associations observed in such data sets. Examples of such methods include Markov random fields or undirected probabilistic graphical models solved using maximum entropy or iterative proportional fitting (Murphy [40], Fienberg [18], Bishop et al. [6], Gopalappa and Khoshegbhal [24]), Copula-based methods (Nelson [43], Geenens [23]), and maximum entropy optimization (Phillips et al. [47]). When it comes to establishing causal mechanisms, Bayesian methods, probabilistic graphical methods, and graph neural networks will be most relevant.

A different way to address the sparse data problem is to integrate insights from diverse data sources and published findings. The difficulty here is that features relevant for the analyses may reside in data sets that have their own data collection design and further take different modality, for example, longitudinal medical records, expert inputs, and insights obtained from meta-analyses of smaller controlled studies presented in the literature.

Complex Care Interventions: The Camden Coalition case study in Section 3 illustrates the importance of collecting data on medical and social covariates and granular details of intervention progression. The section also illustrates the importance of control groups to infer causality, either through RCTs or other methods, to ensure that the problem of regression to

the mean is properly accounted for. RCTs are among the best methods for testing causality between interventions and outcomes. Although they have been widely used in medical treatment and pharmaceutical interventions, their use in testing innovations in healthcare delivery (such as complex care interventions) is still in the early stages (Alsan and Finkelstein [3]). The biggest barrier is that RCTs for complex care have low recruitment rates and are logistically challenging and costly for organizations to implement. However, they do allow for a richer understanding of intervention impact and the development of further hypotheses. In particular, they facilitate the identification of subgroups of intervention patients based on medical and social characteristics that might have performed better or worse than the same subgroups of control patients (see Bertsimas et al. [5] for an operations research/management science (OR/MS) perspective). Given the heterogeneity of the patient population, such subgroup analyses are a vital direction of future research because they could lead to trials to test more personalized interventions.

Furthermore, the literature in the social and behavioral sciences domain is rich in studies that jointly analyze social and health support programs (Courtin et al. [15], Akinyemiju [2], Hill-Briggs et al. [26], Powell-Wiley et al. [48], Remes et al. [50]). Each study may focus on a narrower set of issues, such as effectiveness of depression treatment and medication adherence for a chronic condition from one study (Świątoniowska-Lonc et al. [63]) and a separate study on effectiveness of social support for housing instability in patients with the chronic condition (Mosley-Johnson et al. [39]). For a person with that chronic condition with both depression and housing instability, a combined intervention may be relevant but if there is a causation between housing instability and depression the fraction of the population that needs both interventions may be lower. Similarly, if the evidence from the published literature on other conditions and social determinants is collected, this would yield a large number of intervention combinations for different patient subgroups. The methods of OR/MS and artificial intellligence can help guide the portfolio of candidate interventions that are likely to have the greatest impact; these, then, could be evaluated through randomized controlled trials.

Another direction of future work in this space relates to modeling the longitudinal progression of disease and intervention actions in time. While traditional intervention analysis techniques have used simulation or Markov processes to model disease stage progressions (Denton [17]), this approach may become challenging as the number of conditions and interventions increase, and further development of such a model will be reliant on an understanding of causal mechanisms. Alternate approaches could focus on use of machine learning, such as generative adversarial networks and transformer models (Shankar et al. [55]), for prediction by training of longitudinal data (Hwang et al. [28], Yang et al. [69], Cascarano et al. [11]). Although these may not address causality directly, they can be combined with expert opinion to develop appropriate preventive decision support.

References

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. J. B. Bocca, M. Jarke, and C. Zaniolo, eds. *Proceedings of the 20th International Conference of Very Large Data Bases*, volume 1215. Morgan Kaufmann Publishers Inc., San Francisco, 487–499, 1994.

[2] F. T. Akinyemiju. Socio-economic and health access determinants of breast and cervical cancer screening in low-income countries: Analysis of the world health survey. *PLoS One* 7(11):e48834, 2012.

[3] M. Alsan and A. N. Finkelstein. Beyond causality: Additional benefits of randomized controlled trials for improving healthcare delivery. *Milbank Quarterly* 99(4):864–881, 2021.

[4] G. A. Barnett, J. C. Van Der Pols, and A. J. Dobson. Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology* 34(1):215–220, 2005.

[5] D. Bertsimas, N. Korolko, and M. A. Weinstein. Identifying exceptional responders in randomized trials: An optimization approach. *Informs Journal on Optimization* 1(3):187–199, 2019.

[6] M. Y. Bishop, S. E. Fienberg, and P. W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. Springer Science & Business Media, Boston, 2007.

[7] D. Blumenthal, B. Chernof, T. Fulmer, J. Lumpkin, and J. Selberg. Caring for high-need, high-cost patients-an urgent priority. *New England Journal of Medicine* 375(10):909–911, 2016.

[8] B. Bray. 1&1 webinar on latent class analysis (LCA) with Bethany Bray. Accessed July 17, 2024, https://www.youtube.com/watch?v=BGdvoEzLCYA, 2019.

[9] C. Buttorff, T. Ruder, and M. Bauman. *Multiple Chronic Conditions in the United States*. Rand Corporation, Santa Monica, CA, 2017.

[10] M. Caliendo and S. Kopeinig. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22(1):31–72, 2008.

[11] A. Cascarano, J. Mur-Petit, J. Hernandez-Gonzalez, M. Camacho, E. Nina de Toro, P. Gkontra, M. Chadeau-Hyam, et al. Machine and deep learning for longitudinal biomedical data: A review of methods and applications. *Artificial Intelligence Review* 56(suppl 2):1711–1771, 2023.

[12] E. Chang, R. Ali, J. Seibert, and D. N. Berkman. Interventions to improve outcomes for high-need, high-cost patients: A systematic review and meta-analysis. *Journal of General Internal Medicine* 38(1):185–194, 2023.

[13] E. M. Charlson, P. Pompei, K. L. Ales, and C. R. MacKenzie. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases* 40(5):373–383, 1987.

[14] S. Cohen. The concentration of health care expenditures and related expenses for costly medical conditions, 2012. Statistical Brief #455. Agency for Healthcare Research and Quality, Rockville, MD. Accessed July 17, 2024, http://www.meps.ahrq.gov/mepsweb/data_les/publications/st455/stat455.pdf, 2014.

[15] E. Courtin, S. Kim, S. Song, W. Yu, and P. Muennig. Can social policies improve health? A systematic review and meta-analysis of 38 randomized trials. *Milbank Quarterly* 98(2):297–371, 2020.

[16] C. A. Davis, E. Shen, N. R. Shah, B. A. Glenn, N. Ponce, D. Telesca, M. K. Gould, et al. Segmentation of high-cost adults in an integrated healthcare system based on empirical clustering of acute and chronic conditions. *Journal of General Internal Medicine* 33:2171–2179, 2018.

[17] T. B. Denton. Optimization of sequential decision making for chronic diseases: From data to decisions. E. Gel, L. Ntaimo, and D. Shier, eds. *Recent Advances in Optimization and Modeling of Contemporary Problems*. INFORMS, Catonsville, MD, 316–348, 2018.

[18] E. S. Fienberg. An iterative procedure for estimation in contingency tables. *Annals of Mathematical Statistics* 41(3):907–917, 1970.

[19] A. Finkelstein, A. Zhou, S. Taubman, and J. Doyle. Healthcare hotspotting—A randomized, controlled trial. *New England Journal of Medicine* 382(2):152–162, 2020.

[20] A. Finkelstein, C. J. Cantor, J. Gubb, M. Koller, A. Truchil, R. A. Zhou, and J. Doyle. The Camden coalition care management program improved intermediate care coordination: A randomized controlled trial: Study examines the Camden coalition care management program results. *Health Affairs* 43(1):131–139, 2024.

[21] A. Gawande. Frontline—Doctor hotspot (full report)—pbs. Accessed July 17, 2024, https://www.youtube.com/watch?v=0 DiwTjeF5AU, 2011.

[22] A. Gawande. The hot spotters. *The New Yorker* 86(45):40–51, 2011.

[23] G. Geenens. Copula modeling for discrete random vectors. *Dependence Modeling* 8(1):417–440, 2020.

[24] C. Gopalappa and A. Khoshegbhal. Mechanistic modeling of social conditions into disease predictions for public health intervention-analyses: Application to HIV. Accessed July 17, 2024, https://doi.org/10.1101/2023.03.01.23286591, 2023.

[25] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, volume 2. Springer, Berlin, 2009.

[26] F. Hill-Briggs. N. E. Adler, S. A. Berkowitz, M. H. Chin, T. L. Gary-Webb, A. Navas-Acien, P. L. Thornton, et al. Social determinants of health and diabetes: A scientific review. *Diabetes Care* 44(1):258, 2021.

[27] M. Humowiecki, T. Kuruna, R. Sax, M. Hawthorne, A. Hamblin, S. Turner, K. Mate, C. Sevin, and K. Cullen. Blueprint for complex care. Accessed July 17, 2024, https://camdenhealth.org/wp-content/uploads/2019/03/Blueprint-for-Complex-Care_UPDATED_030119-2.pdf, 2019.

[28] U. Hwang, S. Choi, H.-B. Lee, and S. Yoon. Adversarial training for disease prediction from electronic health records with missing data. Preprint, submitted May 22, https://arxiv.org/abs/1711.04126, 2018.

[29] S. Kangovi, N. Mitra, D. Grande, H. Huo, A. R. Smith, and J. A. Long. Community health worker support for disadvantaged patients with multiple chronic diseases: A randomized clinical trial. *American Journal of Public Health* 107(10):1660–1667, 2017.

[30] S. Kangovi, N. Mitra, D. Grande, L. M. White, S. McCollum, J. Sellman, R. P. Shannon, et al. Patient-centered community health worker intervention to improve posthospital outcomes: A randomized clinical trial. *JAMA Internal Medicine* 174(4):535–543, 2014.

[31] S. Kangovi, N. Mitra, L. Norton, R. Harte, X. Zhao, T. Carter, D. Grande, et al. Effect of community health worker support on clinical outcomes of low-income patients across primary care facilities: A randomized clinical trial. *JAMA Internal Medicine* 178(12):1635–1643, 2018.

[32] E. Koker, H. Balasubramanian, R. Castonguay, A. Bottali, and A. Truchil. Estimating the workload of a multi-disciplinary care team using patient-level encounter histories. *Health Systems* 1–21, 2023.

[33] É. Lacourse, R. de la Sablonnière, C. É. Giguère, S. Morin, R. Legault, F. Laliberté, and Z. Bakk. Interface to 'Python' Package 'StepMix'. Accessed July 17, https://cran.r-project.org/web/packages/stepmixr/stepmixr.pdf, 2024.

[34]. D. A. Linzer and J. B. Lewis. poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software* 42:1–29, 2011.

[35] J. D. C. Little and S. C. Graves. Little's law. D. Chhajed, T. J. Lowe, and F. Hillier, eds. *Building Intuition: Insights from Basic Operations Management Models and Principles*. Springer Science + Business Media, LLC, New York, 81–100, 2008.

[36] Z. Martinez, E. Koker, A. Truchil, and H. Balasubramanian. Time and effort in care coordination for patients with complex health and social needs: Lessons from a community-based intervention. *Journal of Interprofessional Education & Practice* 15:142–148, 2019.

[37] MEPS-AHRQ. Medical expenditure panel survey. Accessed July 17, 2024, https://meps.ahrq.gov/mepsweb/, 2023.

[38] S. Morin, R. Legault, F. Laliberté, Z. Bakk, C.-É. Giguère, R. de la Sablonnière, and É. Lacourse. Stepmix: A python package for pseudo-likelihood estimation of generalized mixture models with external variables. Preprint, submitted June 17, https://arxiv.org/abs/2304.03853, 2024.

[39] E. Mosley-Johnson, R. J. Walker, M. Thakkar, J. A. Campbell, L. Hawks, S. Pyzyk, and L. E. Egede. Relationship between housing insecurity, diabetes processes of care, and self-care behaviors. *BMC Health Services Research* 22:1–8, 2022.

[40] P. K. Murphy. *Machine Learning: A Probabilistic Perspective.* MIT Press, Cambridge, MA, 2012.

[41] B. Muthén and L. Muthén. Mplus. *Handbook of Item Response Theory.* Chapman and Hall/CRC, Boca Raton, 507–518, 2017.

[42] D. M. Naylor, D. A. Brooten, R. L. Campbell, G. Maislin, K. M. McCauley, and J. S. Schwartz. Transitional care of older adults hospitalized with heart failure: A randomized, controlled trial. *Journal of the American Geriatrics Society* 52(5):675–684, 2004.

[43] B. R. Nelsen. *An Introduction to Copulas.* Springer, Berlin, 2006.

[44] K. Noonan and K. Craig. From siloed system to ecosystem: The evolution of the camden coalition's complex care model. Accessed July 17, 2024, https://camdenhealth.org/wp-content/uploads/2023/03/core-model-print-10.18.19.pdf, 2023.

[45] K. Nylund-Gibson and A. Y. Choi. Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science* 4(4):440, 2018.

[46] J. J. O'Connell, S. C. Oppenheimer, C. M. Judge, R. L. Taube, B. B. Blanchfield, S. E. Swain, and H. K. Koh. The Boston healthcare for the homeless program: A public health framework. *American Journal of Public Health* 100(8):1400–1408, 2010.

[47] J. S. Phillips, R. P. Anderson, and R. E. Schapire. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190(3–4):231–259, 2006.

[48] M. T. Powell-Wiley, Y. Baumer, F. O. Baah, A. S. Baez, N. Farmer, C. T. Mahlobo, M. A. Pita, et al. Social determinants of cardiovascular disease. *Circulation Research* 130(5):782–799, 2022.

[49] M. J. Press. Instant replay: A quarterback's view of care coordination. *New England Journal of Medicine* 371:489–491, 2014.

[50] O. Remes, J. F. Mendes, and P. Templeton. Biological, psychological, and social determinants of depression: A review of recent literature. *Brain Sciences* 11(12):1633, 2021.

[51] E. M. Rezaee, L. LeRoy, A. White, E. Oppenheim, K. Carlson, and M. Wasserman. Understanding the high prevalence of low-prevalence chronic disease combinations: Databases and methods for research. Working paper, ABT Associates, Cambridge, MA, 2013.

[52] J. D. Rinehart, C. Oronce, M. J. Durfee, K. W. Ranby, H. A. Batal, R. Hanratty, J. Vogel, et al. Identifying subgroups of adult superutilizers in an urban safety-net system using latent class analysis: Implications for clinical practice. *Medical Care* 56(1):e1–e9, 2018.

[53] W. D. Roblin, J. E. Segel, R. J. McCarthy, and N. Mendiratta. Comparative effectiveness of a complex care program for high-cost/high-need patients: A retrospective cohort study. *Journal of General Internal Medicine* 36:2021–2029, 2021.

[54] M. C. Rossi and H. Balasubramanian. Panel size, office visits, and care coordination events: A new workload estimation methodology based on patient longitudinal event histories. *MDM Policy & Practice* 3(2):2381468318787188, 2018.

[55] V. Shankar, E. Yousefi, A. Manashty, D. Blair, and D. Teegapuram. Clinical-gan: Trajectory forecasting of clinical events using transformer and generative adversarial networks. *Artificial Intelligence in Medicine* 138:102507, 2023.

[56] S. Shavit, N. B. Jenerius, A. Aminawung, S. Greenberg, T. Berthold, A. Fishman, H. S. Busch, and E. A. Wang. Transitions clinic network: Challenges and lessons in primary care for people released from prison. *Health Affairs* 36(6):1006–1015, 2017.

[57] P. Sinha, S. C. Calfee, and K. L. Delucchi. Practitioner's guide to latent class analysis: Methodological considerations and common pitfalls. *Critical Care Medicine* 49(1):e63–e79, 2021.

[58] T. S. Skou, F. S. Mair, M. Fortin, B. Guthrie, B. P. Nunes, J. J. Miranda, C. M. Boyd, S. Pati, S. Mtenga, and S. M. Smith. Multimorbidity. *Nature Reviews Disease Primers* 8(1):48, 2022.

[59] A. Skrondal and S. Rabe-Hesketh. Latent variable modelling: A survey. *Scandinavian Journal of Statistics* 34(4):712–745, 2007.

[60] G. M. R. Smeets, A. M. J. Elissen, M. E. A. L. Kroese, N. Hameleers, and D. Ruwaard. Identifying subgroups of high-need, high-cost, chronically ill patients in primary care: A latent class analysis. *PLoS One* 15(1):e0228103, 2020.

[61] M. C. Smith, J. Feigal, R. Sloane, and D. J. Biederman. Differences in clinical outcomes of adults referred to a homeless transitional care program based on multimorbid health profiles: A latent class analysis. *Frontiers in Psychiatry* 12:780366, 2021.

[62] J. Sorace, H.-H. Wong, C. Worrall, J. Kelman, S. Saneinejad, and T. MaCurdy. The complexity of disease combinations in the Medicare population. *Population Health Management* 14(4): 161–166, 2011.

[63] N. Świątoniowska-Lonc, W. Tański, J. Polański, B. Jankowska-Polańska, and G. Mazur. Psychosocial determinants of treatment adherence in patients with type 2 diabetes: A review. *Diabetes, Metabolic Syndrome and Obesity* 14:2701–2715, 2021.

[64] K. J. Vermunt. The EM algorithm for latent class analysis. Accessed July 17, 2024, https://www.youtube.com/watch?v=btUaJd35hYE, 2021.

[65] K. J. Vermunt and J. Magidson. Latent class analysis. M. S. Lewis-Beck, A. Bryman, and T. F. Liao, eds. *The Sage Encyclopedia of Social Sciences Research Methods*, volume 2. SAGE Publications, Inc, London, 549–553, 2004.

[66] D. P. Welch. A graphical approach to the initial transient problem in steady state simulations. *Proceedings of the 10th IMACS World Congress on Systems, Simulation, and Scientific Computation*. North-Holland Publishing Co., Amsterdam, 219–221, 1982.

[67] E. B. Weller, N. K. Bowen, and S. J. Faubert. Latent class analysis: A guide to best practice. *Journal of Black Psychology* 46(4):287–311, 2020.

[68] W. Whitt. OM Forum: Offered load analysis for staffing. *Manufacturing & Service Operations Management* 15(2):166–169, 2013.

[69] Z. Yang, A. Mitra, W. Liu, D. Berlowitz, and H. Yu. Transformehr: Transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records. *Nature Communications* 14(1):7857, 2023.

[70] Q. Yang, D. Wiest, A. C. Davis, A. Truchil, and J. L. Adams. Hospital readmissions by variation in engagement in the healthcare hotspotting trial: A secondary analysis of a randomized clinical trial. *JAMA Network Open* 6(9):e2332715, 2023.

[71] M. D. Zulman, S. M. Asch, S. B. Martins, E. A. Kerr, B. B. Hoffman, and M. K. Goldstein. Quality of care for patients with multiple chronic conditions: The role of comorbidity interrelatedness. *Journal of General Internal Medicine* 29(3):529–537, 2014.

[72] M. D. Zulman, C. P. Chee, S. C. Ezeji-Okoye, J. G. Shaw, T. H. Holmes, J. S. Kahn, and S. M. Asch. Effect of an intensive outpatient program to augment primary care for high-need veterans affairs patients: A randomized clinical trial. *JAMA Internal Medicine* 177(2):166–175, 2017.