BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Multiple chronic conditions at a major urban health system: a descriptive analysis of frequencies, costs and patterns

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029340
Article Type:	Research
Date Submitted by the Author:	22-Jan-2019
Complete List of Authors:	Majumdar, Usnish; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health Hunt, Christophe; Mount Sinai Health Partners Doupe, Patrick; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health Baum, Aaron; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health; Icahn School of Medicine at Mount Sinai, Health System Design and Global Health Heller, David; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health; Icahn School of Medicine at Mount Sinai, Health System Design and Global Health Kumar, Rashi; Healthfirst Futterman, Robert; Healthfirst Hajat, Cother; UAE University, Levine, Erica; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health Kishore, Sandeep; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health; Icahn School of Medicine at Mount Sinai, Health System Design and Global Health
Keywords:	HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PUBLIC HEALTH

SCHOLARONE™
Manuscripts

Multiple chronic conditions at a major urban health system: a descriptive analysis of frequencies, costs and patterns

Usnish Majumdar[1], Christophe Hunt[2], Patrick Doupe[1], Aaron Baum[1], David J. Heller[1], Rashi Kumar[3], Robert Futterman[3], Cother Hajat[4], Erica Levine[1], Sandeep P. Kishore[1]

1. Arnhold Institute for Global Health, Icahn School of Medicine at Mount Sinai, New York, New York
2. Mount Sinai Health Partners, Mount Sinai Health System, New York, New York
3. Healthfirst, New York, New York
4. United Arab Emirates University, Public Health Institute

Corresponding author:

Sandeep P. Kishore
1216 Fifth Avenue, Room 556
New York, New York 10129
Phone: 212-824-8957
Fax: 212 - 824 - 7995

Word Count (max 4500): 2383

Tables: 4

Figures: 3

Supplementary Material:
The code we used is available here: https://github.com/usnish/mcc_scripts

Funding:
This work was supported by Teva Pharmaceuticals for the Multiple Chronic Conditions Initiative with the Arnhold Institute for Global Health. Dr. Heller also reports support from the NIH Fogarty International Center (R21 TW010452-01).

For peer review only

Abstract

Objective

To (1) examine the burden of multiple chronic conditions (MCC) in an urban health system, and (2) propose a methodology to identify sub-populations of interest for both clinical and financial interventions.

Design: Retrospective cross-sectional study.

Setting: Mount Sinai Health System, set in all five boroughs of New York City.

Participants: 192,085 adult (18+) plan members of capitated Medicaid contracts between Healthfirst and Mount Sinai Health System in the years 2012-2014.

Methods

Adults were categorized as having 0,1, 2, 3, 4, 5+ chronic conditions from a list of 69 chronic conditions provided by the Agency for Healthcare Research & Quality (AHRQ). After summarizing the demographics, geography, and prevalence of MCC within this population, we then described groups of patients (clusters) using a novel methodology: We iteratively defined 26,495 potential clusters of patients by a pair of chronic conditions, a sex, and an age group, and then ranked them by 1) frequency, 2) cost and 3) ratios of observed to expected frequencies of co-occurring chronic conditions. Accordingly, we compiled pairs of conditions that occur more frequently together than otherwise expected.

Results

52.7% of the study population suffers from two or more chronic conditions. The most frequent dyad was hypertension and hyperlipidemia (18% age-adjusted) and the most frequent triad was diabetes, hypertension and hyperlipidemia (9% age-adjusted). Women aged 50-65 with hypertension and hyperlipidemia were the leading cost segment in the study population. Costs and prevalence of MCC increase with number of conditions and age. The most significant observed/expected ratio dyads were pulmonary disease and myocardial infarction.

Conclusions

In this low-income, urban population, multiple chronic conditions are more prevalent than nationally, motivating further research and implementation efforts in this population. By identifying a number of potential target populations in a highly interpretable manner, this clustering methodology has utility for health services analysts.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Article summary: Strengths and limitations of this study

Strengths of the study:

- Large, robust dataset of patients with high prevalence of chronic disease
- New descriptive/analytic approach identifies unanticipated overlap of conditions
- Methodology applicable to other similar settings, including urban health systems

Weaknesses of the study:

- Cross-sectional data precludes causal analysis
- Use of cost claims data rather than clinical diagnosis
- Not necessarily representative of US population

Article summary (5 bullet points max)

- Retrospective data from capitated Medicaid contracts in an urban health system from 2012-2014 from 192,085 plan members were analyzed. Adults were categorized as having 0,1, 2, 3, 4, 5+ chronic conditions from a list of 69 chronic conditions provided by the Agency for Healthcare Research & Quality (AHRQ).
- We described groups of patients (clusters) using a novel methodology: We iteratively defined 26,495 potential clusters of patients as defined by a pair of chronic conditions, a sex, and an age group, and then ranked them by 1) frequency, 2) cost and 3) ratios of observed to expected frequencies of co-occurring chronic conditions.
- 52.7% of the study population suffers from two or more chronic conditions. The most frequent dyad was hypertension and hyperlipidemia (18% age-adjusted) and the most frequent triad was diabetes, hypertension and hyperlipidemia (9% age-adjusted). Women aged 50-65 with hypertension and hyperlipidemia were the leading cost segment in the study population.
- The most significant associations for observed/expected ratio dyads were pulmonary disease and myocardial infarction. In a low-income, urban population, multiple chronic conditions are more prevalent than nationally, motivating further research and implementation efforts in this population.
- By identifying a number of potential target populations in a highly interpretable manner, we argue that this clustering methodology has utility for health systems, financiers, and researchers working to address MCC. We provide a common methodology for targeting populations for financial and clinical intervention.

Data sharing statement

Data can be accessed by emailing the first author of the manuscript.

Patient and public involvement section

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

The study was a retrospective review using administrative claims data. The patients and public were not involved in this study.

Contributorship Statement

SPK conceived of the study. C Hajat advised on technical analysis. UM, C Hunt and PD completed analyses. AB, DJH, RK, RF and EL provided technical input to the manuscript. UM wrote the manuscript. SPK, DJH, C Hunt, C Hajat edited drafts of the manuscript.

Introduction

The management of multiple chronic conditions (MCC, here defined as the association of two or more chronic health conditions) constitutes a formidable clinical and financial challenge. An increasingly large proportion of the United States population suffers from MCC, including 42% of adults overall and 81% of those over the age of 65 years [1]. In the US, MCC patients account for more than 70% of all healthcare spending [2]. In patients over 65 years old, costs increase exponentially with each additional chronic condition, suggesting that there are additional costs associated with the complexity or inefficiency of care for MCC. [3–14].

Health systems have responded to these challenges with clinical and financial innovations. Clinical innovations include new models of care coordination, joint clinical guidelines for MCC patients and alternative delivery models which include bundling of services [14–18]. Financial innovations include value-based payments and bundled payment schemes. One growing form of value-based financial transformation involves capitation, where a fixed “budget” for each patient is agreed upon between the payer and the health system. Accordingly, the health system is incentivized to bring costs down while still maintaining a small margin of profit. It is in this context that a standard methodology to evaluate the potential interactions between conditions could be mutually beneficial. Importantly, risk adjustment generates appropriately large budgets for high-cost and complex patients, and by doing so accounts for changes in severity over time and incentivizes providing coverage to these high-cost individuals. Existing systems of risk adjustment employed by the Centers for Medicaid & Medicare Services (CMS) predict medical and pharmaceutical spending using demographics and diagnosis codes, and are employed in a standardized fashion for Medicare Advantage patients. State managed Medicaid plans can choose to employ any of many different risk adjustment models, some of which are based on the Medicare Advantage models [19].

Especially important in the setting of value-based payment schemes like capitation is the appropriate selection of sub-populations to receive clinical interventions. While increasingly popular nationally, measures targeting patients who are chronically hospitalized (sometimes known as “super-utilizers”) have demonstrated mixed cost savings, in part because of difficulties targeting patients who could benefit from interventions [18,20,21].

It remains difficult to compare and contrast the clinical and financial reforms enacted in different patient populations. While there exist numerous sophisticated

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

statistical methods for clustering populations of patients, such as random forests, single decision trees, k-means, and hierarchical cluster analysis, these methods suffer in their utility due to their limited interpretability, instability, and/or tendency for overfitting [22]. Rather than relying on complex statistical models, we propose a simple descriptive method that can be applied to any population for whom medical claims are available.

Prior studies of spending and MCC have focused on synergy in spending between conditions, or on a specific slice of a population or type of spending -- for example on inpatients or outpatient spending, or on those older than 65 [3–6,11–13]. Notably, literature on MCC patterns and trends among younger, lower socioeconomic status, and vulnerable populations remains scarce, despite their carrying a significant share of chronic disease burden and, accordingly, financial risk in value-based schemes [23]. Additionally, under global capitation both inpatient and outpatient costs must be considered together, as was done in this study.

In order to develop a methodology that would yield interpretable insights for both clinical interventions and financial incentives, we sought to first iteratively but simply generate many different sub-populations within the study population and then sort them via either clinically meaningful or financially relevant mechanisms. Clinical interventions can be developed from epidemiological information about which conditions are observed more frequently together than expected [24]. We theorized that observed/expected (independent) ratios would reveal groups of patients distinct from those based purely on frequency or cost. Combinations of chronic conditions could have shared risk factors (e.g. hypertension and diabetes), shared etiology (e.g. hypertension and congestive heart failure) or could be independent altogether (e.g. hypertension and arthritis). By contrast, financial interventions can be developed from cost information about which conditions and combinations of conditions occur in the most costly groups of patients. In practical terms, targeting the highest cost combinations of conditions (and therefore clusters of patients) could lead to proactive interventions to reduce avoidable or excess utilization.

Accordingly, in this manuscript we (1) develop a descriptive methodology to identify and describe unique clusters of MCC patients, and (2) apply the methodology in an urban health system using administrative claims data derived from a population of managed Medicaid patients at the Mount Sinai health system under global capitation -- a low-income, urban population unlike those previously studied. We also describe the general cost and geographic characteristics of this population, with the potential for use in future clustering applications.

Methods

Clustering

Clusters refer to groups of patients who meet certain disease criteria, demographic criteria, or both. For example, a cluster of patients would be defined by a dyad of diseases (i.e. hypertension and hyperlipidemia), an age range (ages 35-50 years), and sex (males). Such a cluster would consist of male patients aged 35-50 years with both hypertension and hyperlipidemia. As described, these clusters are not mutually exclusive (i.e. one

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

patient can belong to several clusters). We systematically investigated every possible cluster of patients defined by a combination of two chronic conditions (among 69), an age group (0-18, 18-35, 35-50, 50-65, 65+), and sex, yielding 26,495 potential clusters. For each of these clusters of patients, we computed a number of cluster characteristics by which to rank them: total cost attributable to cluster, average cost per person in cluster, and observed:expected ration of disease dyads in each cluster. The total cost attributable to the patients in each cluster was computed using claims provided by the payer. This calculation includes all costs for these patients, not just those attributable to the diseases defining the cluster. Clusters were also ranked by average cost per person per year of plan enrollment represented in the cluster. For each pair of diseases defining a cluster, an *observed:expected ratio* was computed by dividing the observed frequency of the pair of diseases in the study population by the expected frequency (multiplying together the individual frequencies of each disease in the pair). We chose a cutoff of 30 cluster members as the lower limit for understanding probable outcomes through a pilot program [25].

Chronic Conditions Lists

We completed a review of pre-existing approaches and opted to work with a defined list of 69 chronic condition categories from the Agency for Healthcare Research and Quality (AHRQ) [26–28]. This condition list was chosen because (1) it included the most expansive list developed by a consensus body of physicians, enabling us to detect uncommon combinations of conditions, and (2) it aligns with other federal multiple chronic condition projects.

Data Set and Inclusion Criteria

We used claims data from patients operating under a capitated contract between Mount Sinai Health System and Healthfirst, the largest managed care organization for federal Medicaid funds in New York State. These data include all medical claims from 2012 to 2014 including 6,676,867 claims for 213,091 plan members. This period represents the first full year of claims following the start of the Mount Sinai-Healthfirst contract to the last year when claims were made with the International Classification of Diseases version 9 (ICD-9). Costs represent paid amounts, not charged amounts.

We used the Agency for Healthcare Quality and Research (AHRQ) Healthcare Cost and Utilization Project (HCUP) mapping of 4,427 ICD-9 codes to 69 clinically-relevant chronic condition categories. We omitted 2015 data because ICD-10 codes were used inconsistently alongside ICD-9 codes, and the HCUP mapping of ICD-10 codes to chronic condition categories is incomplete. We performed a complete case analysis and excluded participants with missing age or gender. The study was approved through Institutional Review Board at the Icahn School of Medicine at Mount Sinai.

Variables

We studied age, gender, location, chronic condition codes, number of chronic conditions, and total cost of care during the member's plan enrollment. Multiple chronic conditions were studied as dyads and triads. The analysis of different combinations of cluster criteria was limited by processing power and computational cost.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Statistical Analyses

The observed frequency of each cluster was age-adjusted using the New York State age distribution. Clusters were segmented by gender. Estimates were calculated for clusters defined by chronic condition codes, gender, age, total costs of care. Chi-squared tests were used to analyze differences in frequency between cluster groupings. We defined statistical significance as a two-sided $p < 0.05$. Claims were aggregated by patient-year via SQL, and subsequent cleaning, analysis, and plotting was performed with R and Python (code available in Supplementary Information).

Results

Prevalence of MCC by selected characteristics

52.7% of the study population (49.1% in women, 44.1% in men) suffers from two or more chronic conditions, as compared to 42% nationally. **Table 1** displays demographic data of the sample (n=192,285 patients). Median age was 26 years (25th percentile = 9.0; 75th percentile = 46.0), and 58% (112,141) were female. We identified the most prevalent combinations of two and three chronic conditions. Each identified dyad or triad represents the prevalence of patients with that combination of chronic conditions, including those that also have additional conditions (for example, a patient with hypertension, hyperlipidemia, and diabetes, would still be counted within both the hypertension & hyperlipidemia and hyperlipidemia & diabetes dyads).

These overlapping clusters of patients, ranked by age-adjusted frequency are reported in **Table 2.** Of these, 20,675 clusters contained at least one patient with the largest cluster containing an average of 4,329 plan members per year. The most common dyads were hypertension and hyperlipidemia (18%) and the most common triad was hypertension, hyperlipidemia and diabetes (9%).

Healthcare expenditures

Figure 1 shows healthcare expenditure among patients with different numbers of chronic conditions. Patients with missing demographic data have been excluded (12.3% of all patients). Costs increase by over 40% with each additional condition, as does the patient-to-patient variance in yearly cost.

Clusters by Age, Sex, Costs

Supplementary Table 1 indicates the top clusters and characteristics by chronic conditions by age and gender using the classification outlined in the Methods section. The lists are presented by top 10 highest frequency (3A), top 10 dyads with the highest costs and at least an average of 30 members per year (3B) and by top 10 dyads with the highest cost and at least an average of 1,000 members per year.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

One important point is that by amending the minimum number of patients in the cluster, there are effects on the kinds of diseases represented. For example, when the minimum is 30 members the highest cost segment was males, age 35-50, with “Anemia and other non-cancer hematological disorder” & “conduction disorder or cardiac dysrhythmia” and when the minimum was 1,000 members it was females, 50-65, with “Hypertension & Coronary atherosclerosis”. In general, these smaller clusters (>30 members) tended to be higher in average individual cost but lower in total cost than the larger clusters (>1,000 members).

Table 3 shows all clusters segmented by age (5 categories) and gender (male/female). This table indicates dyads of chronic conditions organized by observed/expected ratios. This data reveal a different relationship of chronic conditions to one another than the frequency and cost tables. By selecting clusters of patients with at least 30 included, we demonstrate relationships between unexpected diseases in small yet high-cost groups of patients. For example, paralysis and immunity disorders occur at 15.63 times the expected rate, accounting for an average yearly cost of $86,182. By selecting clusters of patients with at least 1,000, we demonstrate relationships that are more commonly observed (and more frequently expected), such as between peripheral atherosclerosis and coronary atherosclerosis, or between anxiety disorders and bipolar disorder.

Age, Spatial distribution and rising risk for patients with multiple chronic conditions

Figure 2 shows frequency of multiple chronic conditions as a function of age across the 5 counties in New York City. Significant disparities are observed between boroughs. A 50% prevalence of MCC is seen at age 30-34 in the Bronx, a historically lower-income borough of the city, whereas in Brooklyn at in the same 30-34 age-group, the prevalence is only 34%.

Discussion

In this paper, we argue that this simple descriptive clustering methodology has utility for resource planning, care coordination, and care delivery. This methodology would be especially useful in the context of public and private benefits schemes focused on low-income populations.

We find that 52.7% of our population suffers from two or more chronic conditions as compared to 42% nationally, motivating efforts to build MCC interventions and tools in the Medicaid population [2]. Using an established list of conditions, we found that total costs increase with each condition added, consistent with findings from other research groups [29–36]. We also found that the most frequent dyad of co-occurring chronic conditions was hypertension and hyperlipidemia (18% age-adjusted) and the most frequent triad was diabetes, hypertension and hyperlipidemia (9% age-adjusted), each in turn more frequent in our study population than nationally (13.6%, as estimated from NHANES in 2010, and 6.3%, from NHANES in 2012) [37,38]. This is a striking finding, considering that the NHANES cohort includes a larger proportion of older adults than our study. As NHANES includes fixed sample-size targets and weighting to generate a national sample of households that is representative of the US adult population, the median age at the time of

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

these studies was 37.2, significantly older than the median of 26 in our dataset. This age discrepancy could be due to two reasons: (1) As adults who are dual-eligible for Medicaid and Medicare are often re-directed to managed Medicare contracts, our study population under-represents adults over 65. (2) Studies of chronic conditions in adults using NHANES tend to utilize a minimum age of 20, as people aged 19 or younger are categorized as 'youth'; compared to the age cutoff of 17 or younger in our study population [35, 36].

Women aged 50-65 with hypertension and hyperlipidemia were the leading cost segment in the health system for dyads. Overall, women age 50-65 and hypertension, osteoarthritis, hyperlipidemia were the leading triad in terms of prevalence and cost. The most significant observed/expected ratio dyads were pulmonary disease and myocardial infarction. We provided various approaches to grouping these chronic conditions in service of broader research objectives to identify conditions that drive multiplicative, rather than additive, health or cost burdens.

The O/E approach provides a clinically oriented view of examining which conditions occur disproportionately together. For example, we find that in our study population, anemia, pulmonary heart disease, congestive heart failure and conduction disorders occur together more frequently than expected. We also observe that patients' costs balloon when they have these conditions. This would suggest an area where healthcare systems need to focus – screening, dedicated counseling, resources and research dollars. For instance, by targeting patients with conditions like anemia and pulmonary heart disease that do not appear to be physiologically related, care managers can minimize fractures in care. If taken together with our finding that MCC burden differs by locale (**Figure 2)** health systems should elect to co-locate specialty clinics, share clinical teams, and develop joint management protocols for these conditions. While these kinds of innovations have been prototyped around episodic procedural care, such as knee and hip replacements, they have yet to be adopted in managing MCC [17,39,40]. Meanwhile, patients with multiple chronic conditions are already requesting these changes [41]. Importantly, this approach yields specific chronic disease targets beyond the most frequent conditions.

Conditions like anemia and pulmonary heart disease are not currently considered among the interaction terms included in existing CMS models (which focus instead on predicting indicators of severe disease like sepsis, pulmonary embolism, or seizure disorders), but may be more locally appropriate measures of disease severity or spending in this population. Further validation would be required of these novel disease interactions in a larger or different sample population.

At the same time, the sorting of clusters by highest cost and frequency provides a simple view of groups where minor interventions could result in larger-scale cost-savings, particularly for health systems facing value-based financing schemes. Addressing the top clusters of patients with bundled financial incentives could supplement the clinical innovations described above. Indeed, recent analyses of the Medicare Shared Savings plan have found that a significant proportion of savings were derived from incremental cost interventions that applied to large swathes of the insured population [42].

The limitations of our proposed approach include the following: (1) the use of health insurance claims itself limits the epidemiologic utility of the analyses. Claims are effectively

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

billing receipts and therefore have limited reliability in reporting disease states [43]. Additionally, we did not control for variations in coding by center or physician. We plan to integrate these claims data with EMR data going forward in order to retrieve higher quality epidemiological insights (2) Our analysis is limited by the study period. Data from 2012-2014 is likely not recent enough to enact present-day interventions in a health system -- this is largely because the mapping of ICD-10 codes to chronic condition categories has not been finalized, with some remaining discontinuities between ICD-9 and ICD-10-based classifications, limiting our ability to use data from 2015 onwards. We plan to include more recent data once the mapping is completed, as well as prototype this methodology using the CMS Chronic Condition Warehouse algorithm, which functions with ICD-10 codes but includes fewer conditions (27 rather than 69). [44]. Additionally, we did not examine epidemiologic trends through time, as a period as short as 3 years is not long enough to elucidate relationships between diseases that share etiology (i.e. hypertension, stroke). (3) The generalizability of our analysis is limited by the geospatial distribution of patients in the study population -- because provider attribution is accomplished regionally, there is an enrollment bias towards patients who live near Mount Sinai practices. Accordingly, this study population of managed Medicaid patients is not necessarily representative of the Medicaid or U.S. population at-large, or the fee-for-service Medicaid population served by Mount Sinai. (4) We did not include pharmacy claims in our analysis, which will result in an underestimation of spending. This underestimation is most significant regarding conditions that require expensive medications (i.e. high-cost injectables for HIV and hepatitis C). However, we also note that risk adjustment methodologies employed by Medicaid Advantage and State Medicaid programs tend to predict spending on pharmaceuticals separate from other costs. [19] (5) Lastly, a significant portion (12.3%) of our study population was excluded on account of missing demographic data, introducing some bias into which clusters of patients were highlighted. Any more pragmatic application of this methodology would also require an approach to patients with missing data.

Taken together, these analyses have implications for health systems, financiers, and researchers working to address MCC, and provide a common methodology for targeting populations for financial and clinical intervention. Most notably, this tool yields a simple, transparent methodology for selecting coherent, clearly-defined populations of patients for intervention, and can be applied to any commercial claims dataset. With application in the right contexts, this methodology could help improve the selection strategy of super-utilizer clinics and other clinical innovations, yielding further advancements in our health systems' management of chronic conditions. Ultimately, however, more research is needed to evaluate this methodology's utility in business scenarios, and applicability to different sizes and kinds of patient populations.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

References

1 Buttorff C, Ruder T, Bauman M. Multiple Chronic Conditions in the United States. RAND Corporation 2017. doi:10.7249/tl221

2 Gerteis J, Izrael D, Deitz D, *et al.* Multiple Chronic Conditions Chartbook. Rockville, MD: Agency for Healthcare Research and Quality; 2014. https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/prevention-chronic-care/decision/mcc/mccchartbook.pdf

3 Crystal S, Johnson RW, Harman J, *et al.* Out-of-pocket health care costs among older Americans. *J Gerontol B Psychol Sci Soc Sci* 2000;**55**:S51–62.

4 Fishman P, Von Korff M, Lozano P, *et al.* Chronic care costs in managed care. *Health Aff* 1997;**16**:239–47.

5 Hoffman C. Persons with chronic conditions. Their prevalence and costs. *JAMA* 1996;**276**:1473–9.

6 Moxey ED, O'Connor JP, Novielli KD, *et al.* Prescription drug use in the elderly: a descriptive analysis. *Health Care Financ Rev* 2003;**24**:127–41.

7 Rice DP, LaPlante MP. Medical expenditures for disability and disabling comorbidity. *Am J Public Health* 1992;**82**:739–41.

8 Schneider KM, O'Donnell BE, Dean D. Prevalence of multiple chronic conditions in the United States' Medicare population. *Health Qual Life Outcomes* 2009;**7**:82.

9 Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med* 2002;**162**:2269–76.

10 Lehnert T, Heider D, Leicht H, *et al.* Review: health care utilization and costs of elderly persons with multiple chronic conditions. *Med Care Res Rev* 2011;**68**:387–420.

11 Cortaredona S, Ventelou B. The extra cost of comorbidity: multiple illnesses and the economic burden of non-communicable diseases. *BMC Med* 2017;**15**:216.

12 Brilleman SL, Purdy S, Salisbury C, *et al.* Implications of comorbidity for primary care costs in the UK: a retrospective observational study. *Br J Gen Pract* 2013;**63**:e274–82.

13 He Z, Bian J, Carretta HJ, *et al.* Prevalence of Multiple Chronic Conditions Among Older Adults in Florida and the United States: Comparative Analysis of the OneFlorida Data Trust and National Inpatient Sample. *J Med Internet Res* 2018;**20**:e137.

14 Hajat C, Stein E. The global burden of multiple chronic conditions: A narrative review.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Prev Med Rep 2018;**12**:284–93.

15 Hajat C, Kishore SP. The case for a global focus on multiple chronic conditions. *BMJ Glob Health* 2018;**3**:e000874.

16 Hajat C, Stein E, Yach D. Multiple chronic conditions: the global state. https:// lnkd. in/ etb9eAJ

17 Keswani A, Koenig KM, Bozic KJ. Value-based Healthcare: Part 1-Designing and Implementing Integrated Practice Units for the Management of Musculoskeletal Disease. *Clin Orthop Relat Res* 2016;**474**:2100–3.

18 Berkowitz SA, Parashuram S, Rowan K, *et al.* Association of a Care Coordination Model With Health Care Costs and Utilization: The Johns Hopkins Community Health Partnership (J-CHiP). *JAMA Netw Open* 2018;**1**:e184273–e184273.

19 March 31, 2016, HHS-Operated Risk Adjustment Methodology Meeting Discussion Paper. Centers for Medicare & Medicaid Services, Center for Consumer Information & Insurance Oversight https://www.cms.gov/CCIIO/Resources/Forms-Reports-and-Other-Resources/Downloads/RA-March-31-White-Paper-032416.pdf

20 Kanzaria HK, Hoffman JR. Hot-Spotters Aren't 'The Problem'...But They Are Emblematic of the Failure of U.S. Healthcare. *J Gen Intern Med* 2017;**32**:6–8.

21 Lee NS, Whitman N, Vakharia N, *et al.* High-Cost Patients: Hot-Spotters Don't Explain the Half of It. *J Gen Intern Med* 2017;**32**:28–34.

22 Breiman L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Stat Sci* 2001;**16**:199–231.

23 Shaw KM, Theis KA, Self-Brown S, *et al.* Chronic Disease Disparities by County Economic Status and Metropolitan Classification, Behavioral Risk Factor Surveillance System, 2013. *Prev Chronic Dis* 2016;**13**:E119.

24 Schäfer I, Kaduszkiewicz H, Wagner H-O, *et al.* Reducing complexity: a visualisation of multimorbidity by combining disease clusters and triads. *BMC Public Health* 2014;**14**:1285.

25 Hertzog MA. Considerations in determining sample size for pilot studies. *Res Nurs Health* 2008;**31**:180–91.

26 Harrison C, Britt H, Miller G, *et al.* Examining different measures of multimorbidity, using a large prospective cross-sectional study in Australian general practice. *BMJ Open* 2014;**4**:e004694.

27 Fortin M, Stewart M, Poitras M-E, *et al.* A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Ann Fam Med* 2012;**10**:142–51.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

28 Le Reste JY, Nabbe P, Manceau B, *et al.* The European General Practice Research Network presents a comprehensive definition of multimorbidity in family medicine and long term care, following a systematic review of relevant literature. *J Am Med Dir Assoc* 2013;**14**:319–25.

29 Crystal S, Johnson RW, Harman J, *et al.* Out-of-pocket health care costs among older Americans. *J Gerontol B Psychol Sci Soc Sci* 2000;**55**:S51–62.

30 Fishman P, Von Korff M, Lozano P, *et al.* Chronic care costs in managed care. *Health Aff* 1997;**16**:239–47.

31 Hoffman C. Persons with chronic conditions. Their prevalence and costs. *JAMA* 1996;**276**:1473–9.

32 Moxey ED, O'Connor JP, Novielli KD, *et al.* Prescription drug use in the elderly: a descriptive analysis. *Health Care Financ Rev* 2003;**24**:127–41.

33 Rice DP, LaPlante MP. Medical expenditures for disability and disabling comorbidity. *Am J Public Health* 1992;**82**:739–41.

34 Schneider KM, O'Donnell BE, Dean D. Prevalence of multiple chronic conditions in the United States' Medicare population. *Health Qual Life Outcomes* 2009;**7**:82.

35 Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med* 2002;**162**:2269–76.

36 Lehnert T, Heider D, Leicht H, *et al.* Review: health care utilization and costs of elderly persons with multiple chronic conditions. *Med Care Res Rev* 2011;**68**:387–420.

37 Song Y, Liu X, Zhu X, *et al.* Increasing trend of diabetes combined with hypertension or hypercholesterolemia: NHANES data analysis 1999-2012. *Sci Rep* 2016;**6**:36093.

38 Egan BM, Li J, Qanungo S, *et al.* Blood pressure and cholesterol control in hypertensive hypercholesterolemic patients: national health and nutrition examination surveys 1988-2010. *Circulation* 2013;**128**:29–41.

39 Bleich SN, Sherrod C, Chiang A, *et al.* Systematic Review of Programs Treating High-Need and High-Cost People With Multiple Chronic Diseases or Disabilities in the United States, 2008-2014. *Prev Chronic Dis* 2015;**12**:E197.

40 Bandara S, Lynch G, Cooke C, *et al.* Using Care Bundles to Improve Surgical Outcomes and Reduce Variation in Care for Fragility Hip Fracture Patients. *Geriatr Orthop Surg Rehabil* 2017;**8**:104–8.

41 Bayliss EA, Edwards AE, Steiner JF, *et al.* Processes of care desired by elderly patients with multimorbidities. *Fam Pract* 2008;**25**:287–93.

42 McWilliams JM, Chernew ME, Landon BE. Medicare ACO Program Savings Not Tied

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

To Preventable Hospitalizations Or Concentrated Among High-Risk Patients. *Health Aff* 2017;**36**:2085–93.

43 Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997;**127**:666–74.

44 HCUP CCS. Healthcare Cost and Utilization Project (HCUP). 2017.www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp (accessed 27 Oct 2017).

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 1: Demographics of Medicaid patients at Mount Sinai Health System in Healthfirst capitated contracts.

	0 Chronic Condition(s)			1 Chronic Condition(s)			2 Chronic Condition(s)			3 Chronic Condition(s)			4 Chronic Condition(s)			5 Chronic Condition(s)			
	2012	2013	2014	2012	2013	2014	2012	2013	2014	2012	2013	2014	2012	2013	2014	2012	2013	2014	p
n	30,246	38,653	42,699	30,902	38,728	40,518	21,679	28,200	28,860	15,533	20,044	20,087	11,646	15,183	15,423	38,667	51,636	56,775	
Age (mean (sd))	16.09 (14.63)	16.77 (14.80)	17.7 (15.56)	20.41 (17.26)	21.49 (17.71)	23.01 (18.14)	26.73 (19.53)	26.95 (19.51)	27.73 (19.45)	34.46 (20.70)	35.07 (20.54)	35.26 (19.97)	41.57 (19.91)	41.56 (19.46)	42.16 (19.03)	53.29 (16.56)	53.19 (16.32)	53.73 (16.20)	<0.001
Sex = F (%)	13,105 (43.3)	16,530 (42.8)	18,034.0 (42.2)	13,764 (44.5)	16,852 (43.5)	17,813 (44.0)	9,101 (42.0)	11,898 (42.2)	12,360 (42.8)	6,115 (39.4)	7,829 (39.1)	7,982 (39.7)	4,384 (37.6)	5,736 (37.8)	5,881 (38.1)	13,263 (34.3)	17,580 (34.0)	19,417 (34.2)	<0.001
Total Cost (mean (sd))	690.67 (1703.33)	702.49 (1806.55)	700.9 (1683.92)	1,080.70 (2484.10)	1,034.87 (2471.51)	1,077.93 (2655.20)	1,559.66 (3180.62)	1,590.85 (4657.31)	1,588.51 (5182.02)	2,125.37 (5252.64)	2,004.86 (3998.93)	2,135.32 (4787.20)	2,730.58 (5496.69)	2,647.52 (7163.53)	2,761.64 (6534.50)	8,881.47 (19629.12)	9,122.26 (22492.75)	9,034.93 (21146.63)	<0.001
Total Cost Winsorized (mean (sd))	666.46 (1445.08)	675.39 (1496.60)	680.6 (1462.31)	1,011.52 (1760.24)	971.83 (1718.43)	1,001.25 (1755.38)	1,430.23 (2074.23)	1,410.48 (2058.28)	1,393.31 (2030.49)	1,812.74 (2268.49)	1,757.97 (2221.25)	1,822.32 (2329.88)	2,226.14 (2447.35)	2,142.27 (2451.98)	2,244.19 (2535.17)	4,532.11 (3462.69)	4,439.91 (3469.54)	4,511.44 (3452.41)	<0.001
Top 10 Single Chronic Conditions																			
Allergy, ENT and other upper resp disorders = Yes (%)	3,584 (37.6)	4,745 (39.2)	3,952.0 (37.5)	8,282 (33.6)	9,849 (31.7)	9,183 (30.5)	6,947 (34.5)	8,687 (33.0)	8,357 (32.2)	4,829 (32.0)	5,912 (30.4)	5,755 (30.0)	3,348 (29.2)	4,184 (27.9)	4,128 (27.3)	10,846 (28.1)	14,227 (27.6)	15,457 (27.3)	<0.001
Asthma, COPD, other chronic lung disease = Yes (%)	1,905 (20.0)	2,540 (21.0)	2,237.0 (21.2)	6,160 (25.0)	7,623 (24.5)	7,157 (23.7)	5,462 (27.1)	7,100 (27.0)	6,957 (26.8)	3,947 (26.2)	4,725 (24.3)	4,744 (24.7)	2,757 (24.0)	3,355 (22.4)	3,258 (21.5)	11,223 (29.1)	14,242 (27.6)	15,071 (26.6)	<0.001
Obesity = Yes (%)	1,668 (17.5)	2,050 (16.9)	1,817.0 (17.2)	3,721 (15.1)	4,914 (15.8)	4,562 (15.1)	3,660 (18.2)	4,556 (17.3)	4,535 (17.5)	2,964 (19.6)	3,592 (18.5)	3,815 (19.9)	2,355 (20.5)	2,996 (20.0)	3,002 (19.8)	9,510 (24.7)	12,700 (24.6)	14,285 (25.3)	<0.001
Degenerative eye problem (gl$uc/eye) = Yes (%)	1,599 (16.8)	1,843 (15.2)	1,528.0 (14.5)	3,168 (12.8)	4,015 (12.9)	3,888 (12.9)	2,962 (14.7)	3,731 (14.2)	3,716 (14.3)	2,819 (18.7)	3,217 (16.6)	3,178 (16.6)	2,512 (21.9)	3,100 (20.7)	3,171 (20.9)	13,597 (35.2)	17,900 (34.7)	19,181 (33.9)	<0.001
Hyperlipidemia = Yes (%)	1,075 (11.3)	1,424 (11.8)	1,242.0 (11.8)	2,279 (9.2)	3,015 (9.7)	2,805 (9.3)	3,187 (15.8)	4,391 (16.7)	4,098 (15.8)	4,106 (27.2)	5,322 (27.4)	5,172 (26.9)	4,313 (37.6)	5,776 (38.6)	5,640 (37.2)	23,041 (59.7)	30,954 (60.0)	33,748 (59.7)	<0.001
Depression and depressive disorders = Yes (%)	874 (9.2)	1,032 (8.5)	842.0 (8.0)	1,836 (7.4)	2,397 (7.7)	2,440 (8.1)	2,068 (10.3)	2,684 (10.2)	2,696 (10.4)	1,957 (13.0)	2,552 (13.1)	2,520 (13.1)	1,879 (16.4)	2,332 (15.6)	2,263 (14.9)	9,689 (25.1)	12,888 (25.0)	14,194 (25.1)	<0.001
Hypertension = Yes (%)	599 (6.3)	883 (7.3)	866.0 (8.2)	1,782 (7.2)	2,281 (7.3)	2,272 (7.5)	2,876 (14.3)	3,647 (13.9)	3,450 (13.3)	3,943 (26.1)	4,971 (25.6)	4,511 (23.5)	4,344 (37.8)	5,595 (37.3)	5,512 (36.4)	25,667 (66.5)	33,962 (65.9)	36,516 (64.6)	<0.001
Esophageal disorder and GI ulcers = Yes (%)	762 (8.0)	1,005 (8.3)	975.0 (9.2)	1,723 (7.0)	2,229 (7.2)	1,947 (6.5)	2,000 (9.9)	2,541 (9.7)	2,138 (8.2)	2,119 (14.0)	2,778 (14.3)	2,126 (11.1)	2,025 (17.6)	2,532 (16.9)	2,365 (15.6)	12,247 (31.7)	16,229 (31.5)	17,492 (30.9)	<0.001
Malnutrition and F/E cond (not obesity/overweight) -includes disorders of metabolism = Yes (%)	828 (8.7)	1,079 (8.9)	907.0 (8.6)	1,675 (6.8)	2,102 (6.8)	2,053 (6.8)	2,030 (10.1)	2,743 (10.4)	2,712 (10.5)	2,147 (14.2)	2,953 (15.2)	2,958 (15.4)	2,047 (17.8)	2,780 (18.6)	2,911 (19.2)	10,069 (26.1)	13,677 (26.5)	15,208 (26.9)	<0.001
Diabetes mellitus = Yes (%)	308 (3.2)	476 (3.9)	415.0 (3.9)	860 (3.5)	1,225 (3.9)	1,100 (3.6)	1,259 (6.3)	1,693 (6.4)	1,622 (6.3)	1,923 (12.7)	2,341 (12.0)	2,100 (10.9)	2,224 (19.4)	2,895 (19.3)	2,822 (18.6)	15,873 (41.1)	21,065 (40.9)	22,381 (39.6)	<0.001

Table 2: Top clusters of two and three chronic conditions using overall list of 69 conditions.

Singlet Chronic Condition	Average Yearly Membership	Unadjusted %	Age Adjusted %
Hypertension	20,939	18%	28%
Hyperlipidemia	20,614	18%	26%
Allergy, ENT and other upper resp disorders	18,921	16%	16%
Asthma, COPD, other chronic lung disease	15,665	14%	14%
Degenerative eye problem (glauc/eye)	13,499	12%	16%

Dyad Chronic Conditions	Average Yearly Membership	Unadjusted %	Age Adjusted %
Hypertension & Hyperlipidemia	12,829	11%	18%
Hypertension & Diabetes mellitus	8,740	8%	12%
Hyperlipidemia & Diabetes mellitus	8,249	7%	11%
Hypertension & Degenerative eye problem (glauc/eye)	6,367	6%	10%
Hyperlipidemia & Degenerative eye problem (glauc/eye)	6,193	5%	9%

Triad Chronic Conditions	Average Yearly Membership	Unadjusted %	Age Adjusted %
Diabetes mellitus, Hypertension, & Hyperlipidemia	6,704	6%	9%
Hypertension, Degenerative eye problem (glauc/eye), & Hyperlipidemia	4,727	4%	7%
Osteoarthritis, Hypertension, & Hyperlipidemia	3,918	3%	6%
Esophageal disorder and GI ulcers, Hypertension, &Hyperlipidemia	3,688	3%	5%
Diabetes mellitus, Hypertension, & Degenerative eye problem (glauc/eye)	3,666	3%	6%

Table 3: Observed/Expected ratios of chronic conditions among common (A) and uncommon clusters (B)

A: **Top 10 Dyads by Observed / Expected rate with at least an average yearly membership of 30 members or more**

Dyad	Unadjusted Frequency	Adjusted Frequency	Adjustment Magnitude	Expected Frequency	Observed / Expected	Average Yearly Cost
Acute myocardial infarction & Pulmonary heart disease	<0.001	0.001	0.001	0.00003	31.25	$89,321
Thrombosis and Embolism & Non-thrombotic, non-athlerosclerotic vascular disease	<0.001	0.001	0.001	0.00004	23.81	$68,047
Pulmonary heart disease & Congestive heart failure	0.002	0.004	0.002	0.00020	20.00	$58,355
Acute myocardial infarction & Congestive heart failure	0.001	0.002	0.001	0.00010	20.00	$66,271
Acute myocardial infarction & Cardiomyopathy and Structural Heart Disease	0.001	0.002	0.001	0.00010	19.23	$66,547
Paralysis & Epilepsy	0.001	0.001	<0.001	0.00005	19.23	$49,312
Paralysis & Organic brain problem (dementia)	<0.001	0.001	0.001	0.00006	16.67	$66,829
Congenital Heart Disease & Heart valve disorder	0.006	0.01	0.004	0.00062	16.03	$11,581
Paralysis & Immunity disorder	<0.001	0.001	0.001	0.00006	15.63	$86,182
Pulmonary heart disease & Cardiomyopathy and Structural Heart Disease	0.002	0.003	0.001	0.00021	14.42	$58,302

B: **Top 10 Dyads by Observed / Expected rate with at least an average yearly membership of 1,000 members or more**

Dyad	Unadjusted Frequency	Adjusted Frequency	Adjustment Magnitude	Expected Frequency	Observed / Expected	Average Yearly Cost
Heart valve disorder & Coronary atherosclerosis	0.009	0.016	0.007	0.0025	6.31	$20,892
Conduction disorder or cardiac dysrhythmia & Coronary atherosclerosis	0.011	0.020	0.009	0.0033	6.15	$26,589
Peripheral atherosclerosis & Coronary atherosclerosis	0.011	0.020	0.009	0.0035	5.70	$20,361
Cerebrovascular Disease & Coronary atherosclerosis	0.009	0.017	0.008	0.0030	5.69	$23,613
Anxiety disorders & Depression and depressive disorders	0.029	0.035	0.006	0.0073	4.77	$9,950
Depression and depressive disorders & Bipolar disorder	0.014	0.017	0.003	0.0036	4.76	$10,975
Anxiety disorders & Bipolar disorder	0.010	0.012	0.002	0.0025	4.76	$11,384
Peripheral atherosclerosis & Other central and peripheral nervous system disorders	0.014	0.022	0.008	0.0050	4.38	$17,150
Other central and peripheral nervous system disorders & Back problem	0.016	0.022	0.006	0.0051	4.30	$13,320
Cerebrovascular Disease & Other central and peripheral nervous system disorders	0.011	0.018	0.007	0.0043	4.21	$23,530

Figure 1: Distribution of individual annual healthcare expenditures as a function of number of chronic conditions.

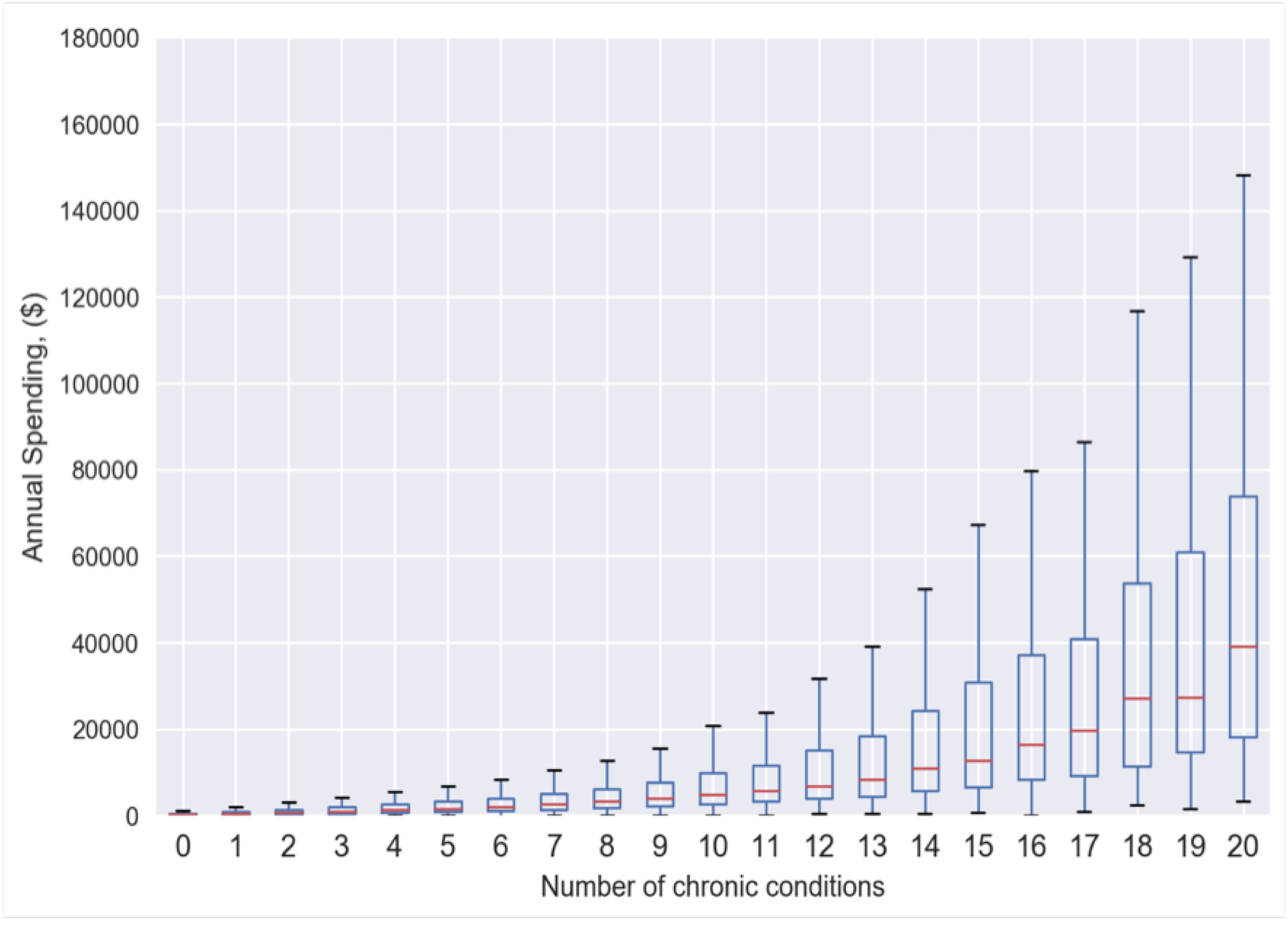
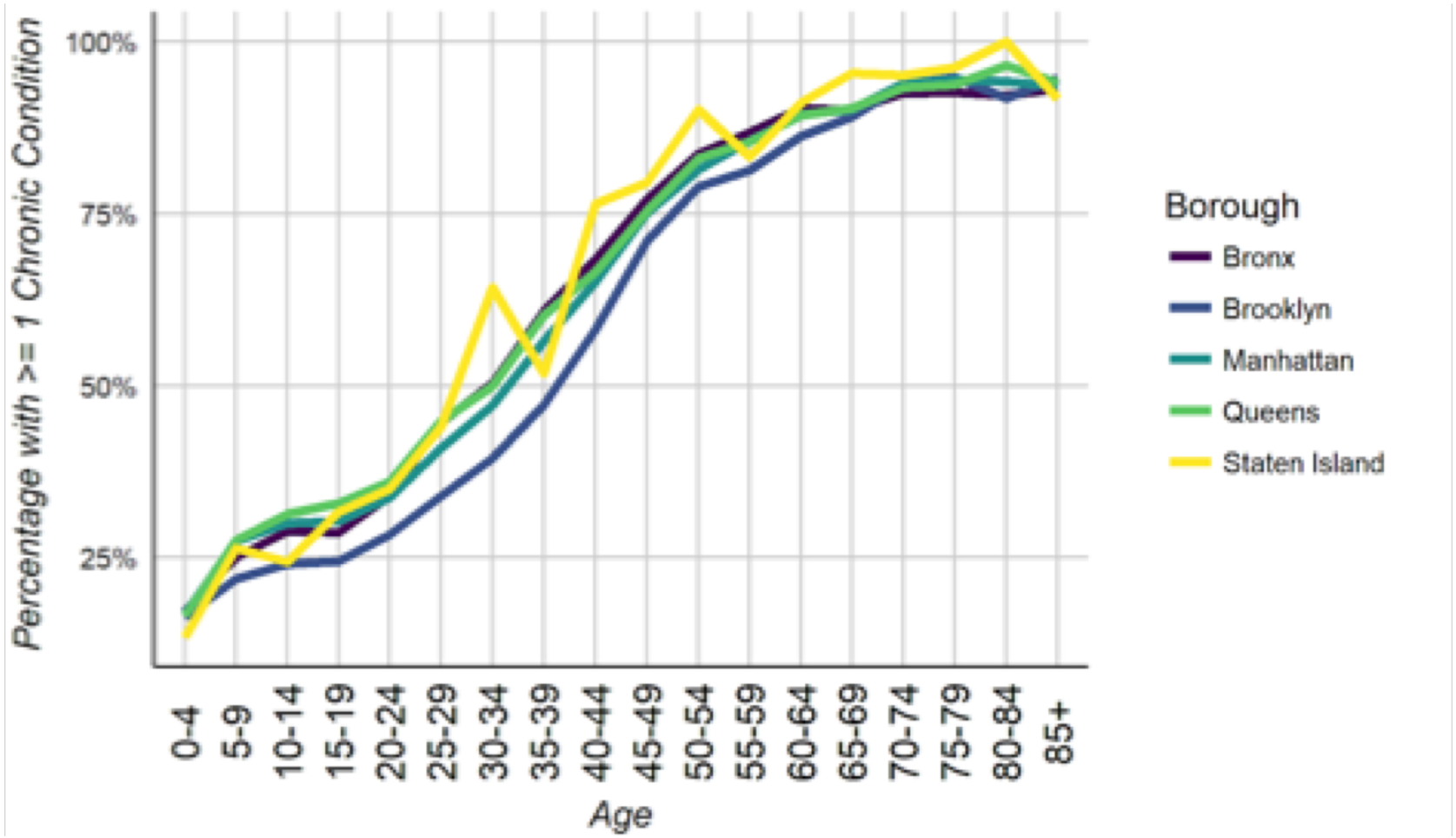


Figure 2: Frequency of multiple chronic conditions by age across selected boroughs of New York City. 50% prevalence of multiple chronic conditions seen at age 30-34 for all boroughs except for Brooklyn that reaches 50% at 35-39. Disparities between boroughs observed.



Supplementary Table 1: Top clusters of chronic conditions by age and gender segments. (A) largest clusters by member count, (B) most costly clusters 30 people or greater, (C) most costly clusters 1000 people or greater.

A: Top 10 clusters by frequency

Gender	Age	Chronic Condition 1	Chronic Condition 2	Total Attributable Cost	Average Yearly Cost	Average Yearly Membership
F	50-65	Hypertension	Hyperlipidemia	$93,122,272	$7,172	4,329
F	50-65	Hypertension	Diabetes Mellitus	$72,878,330	$8,557	2,839
F	50-65	Hyperlipidemia	Diabetes Mellitus	$65,165,290	$8,143	2,668
M	50-65	Hypertension	Hyperlipidemia	$61,719,638	$7,948	2,589
F	50-65	Hypertension	Degenerative eye problem (glauc/eye)	$54,012,310	$8,240	2,185
F	50-65	Hypertension	Osteoarthritis	$66,447,600	$10,166	2,179
F	50-65	Hyperlipidemia	Degenerative eye problem (glauc/eye)	$49,533,370	$7,674	2,152
F	50-65	Hyperlipidemia	Osteoarthritis	$56,171,247	$9,295	2,014
F	50-65	Hypertension	Esophageal disorder and GI ulcers	$60,965,767	$10,297	1,974
F	50-65	Hyperlipidemia	Esophageal disorder and GI ulcers	$53,619,011	$9,194	1,944

B: Top 10 clusters by average yearly cost with 30 total member counts or more

Gender	Age	Chronic Condition 1	Chronic Condition 2	Total Attributable Cost	Average Yearly Cost	Average Yearly Membership
M	35-50	Anemia and other non-cancer heme disorders	Conduction disorder or cardiac dysrhythmia	$8,390,439	$90,220	31
F	50-65	Pulmonary heart disease	Anemia and other non-cancer heme disorders	$7,542,310	$83,803	30
M	35-50	Congestive heart failure	Malnutrition and F/E cond (not obesity/overweight)-includes disorders of metabolism	$7,068,459	$77,675	30
M	50-65	Conduction disorder or cardiac dysrhythmia	Immunity disorder	$9,800,142	$76,564	43
M	35-50	Other central and peripheral nervous system disorders	Immunity disorder	$6,917,900	$73,595	31
F	50-65	Congestive heart failure	Anemia and other non-cancer heme disorders	$14,346,180	$70,671	68
F	50-65	Congestive heart failure	Chronic skin ulcer	$7,852,354	$69,490	38
M	35-50	Anemia and other non-cancer heme disorders	Kidney and Vesicoureteral Disorders (excluding renal failure)	$6,174,594	$68,607	30
M	65+	Cardiomyopathy and Structural Heart Disease	Anemia and other non-cancer heme disorders	$9,505,337	$67,414	47
M	35-50	Congestive heart failure	Conduction disorder or cardiac dysrhythmia	$7,077,370	$67,404	35

C: Top 10 clusters by average yearly cost with 1,000 member counts or more

Gender	Age	Chronic Condition 1	Chronic Condition 2	Total Attributable Cost	Average Yearly Cost	Average Yearly Membership
F	50-65	Hypertension	Coronary atherosclerosis	$45,703,351	$14,486	1,052
F	50-65	Osteoarthritis	Other central and peripheral nervous system disorders	$43,931,227	$14,013	1,045
F	50-65	Hypertension	Other central and peripheral nervous system disorders	$61,441,249	$13,433	1,525
F	50-65	Hypertension	Asthma, COPD, other chronic lung disease	$57,028,007	$12,193	1,559
F	50-65	Hyperlipidemia	Other central and peripheral nervous system disorders	$50,790,724	$12,105	1,399
M	65+	Hypertension	Hyperlipidemia	$40,441,616	$11,933	1,130
F	50-65	Esophageal disorder and GI ulcers	Diabetes mellitus	$39,891,768	$11,827	1,124
F	50-65	Hyperlipidemia	Asthma, COPD, other chronic lung disease	$45,652,105	$11,757	1,294
F	50-65	Esophageal disorder and GI ulcers	Osteoarthritis	$42,200,192	$11,745	1,198
F	50-65	Diabetes mellitus	Osteoarthritis	$43,479,202	$11,591	1,250

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1	Multiple chronic conditions at a major urban health system: a descriptive analysis of frequencies, costs and patterns
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2	Lines 1-39
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4,5	Lines 96-145
Objectives	3	State specific objectives, including any prespecified hypotheses	5	Lines 162-168
Methods				
Study design	4	Present key elements of study design early in the paper	5	Lines 162-168
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6	Lines 202-208
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6	Lines 202-216
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	N/A	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6	Lines 219-222
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6	Lines 202-216

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Bias	9	Describe any efforts to address potential sources of bias	10	Line 375, 385
Study size	10	Explain how the study size was arrived at	6	Line 205

Continued on next page

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5,6	Line 173-191
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7	Line 225-231
		(*b*) Describe any methods used to examine subgroups and interactions	6	Line 187-188
		(*c*) Explain how missing data were addressed	6	Line 214
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	7	Line 225
		(*e*) Describe any sensitivity analyses	N/A	
Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7	Line 239
		(b) Give reasons for non-participation at each stage	7	Line 255
		(c) Consider use of a flow diagram	N/A	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7	Line 236-244
		(b) Indicate number of participants with missing data for each variable of interest	7	Line 255
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	N/A	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	N/A	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	N/A	
		Cross-sectional study—Report numbers of outcome events or summary measures	7,8	Table 1,2,3
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included		
		(*b*) Report category boundaries when continuous variables were categorized	6	Line 180
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A	

Continued on next page

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8	Line 290
Discussion				
Key results	18	Summarise key results with reference to study objectives	8	Line 303-341
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9	Line 357-386
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10	Line 388-397
Generalisability	21	Discuss the generalisability (external validity) of the study results	10	Line 373
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	1	This work was supported by Teva Pharmaceuticals for the Multiple Chronic Conditions Initiative with the Arnhold Institute for Global Health. Dr. Heller also reports support from the NIH Fogarty International Center (R21 TW010452-01).

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

Multiple chronic conditions at a major urban health system: a descriptive retrospective analysis of frequencies, costs and comorbidity patterns

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029340.R1
Article Type:	Original research
Date Submitted by the Author:	14-Aug-2019
Complete List of Authors:	Majumdar, Usnish; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health Hunt, Christophe; Mount Sinai Health Partners Doupe, Patrick; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health Baum, Aaron; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health; Icahn School of Medicine at Mount Sinai, Health System Design and Global Health Heller, David; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health; Icahn School of Medicine at Mount Sinai, Health System Design and Global Health Levine, Erica; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health Kumar, Rashi; Healthfirst Futterman, Robert; Healthfirst Hajat, Cother; UAE University, Kishore, Sandeep; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health; Icahn School of Medicine at Mount Sinai, Health System Design and Global Health
Primary Subject Heading:	Health informatics
Secondary Subject Heading:	Health services research
Keywords:	HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PUBLIC HEALTH

SCHOLARONE™
Manuscripts

Multiple chronic conditions at a major urban health system: a descriptive retrospective analysis of frequencies, costs and comorbidity patterns

Usnish Majumdar[1], Christophe Hunt[2], Patrick Doupe[1], Aaron Baum[1], David J. Heller[1], Erica Levine[1], Rashi Kumar[3], Robert Futterman[3], Cother Hajat[4], Sandeep P. Kishore[1]

1. Arnhold Institute for Global Health, Icahn School of Medicine at Mount Sinai, New York, New York, USA
2. Mount Sinai Health Partners, Mount Sinai Health System, New York, New York
3. Healthfirst, New York, New York, USA
4. Institute of Public Health, United Arab Emirates University, Al Ain, United Arab Emirates

Corresponding author:

Sandeep P. Kishore
1216 Fifth Avenue, Room 556
New York, New York 10129
Phone: 212-824-8957
Fax: 212 - 824 – 7995
Email Address: sunny.kishore@gmail.com

Word Count (max 4500): 3,728

Tables: 4

Figures: 2

Supplementary Material:
The code we used is available here: https://github.com/usnish/mcc_scripts

Funding:
This work was supported by Teva Pharmaceuticals for the Multiple Chronic Conditions Initiative with the Arnhold Institute for Global Health. Dr. Heller also reports support from the NIH Fogarty International Center (R21 TW010452-01).

Competing Interests Statement:
We declare no significant competing financial, professional, or personal interests that might have influenced the design, performance, interpretation, or presentation of the analyses described in this manuscript apart from the above. Teva Pharmaceuticals played no role in the conception, analysis, or writing of this manuscript, nor the decision to publish.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Abstract

Objective

To (1) examine the burden of multiple chronic conditions (MCC) in an urban health system, and (2) propose a methodology to identify sub-populations of interest based on diagnosis groups and costs.

Design: Retrospective cross-sectional study.

Setting: Mount Sinai Health System, set in all five boroughs of New York City, USA.

Participants: 192,085 adult (18+) plan members of capitated Medicaid contracts between the Healthfirst managed care organization and the Mount Sinai Health System in the years 2012-2014.

Methods

We classified adults as having 0,1, 2, 3, 4, or 5+ chronic conditions from a list of 69 chronic conditions. After summarizing the demographics, geography, and prevalence of MCC within this population, we then described groups of patients (clusters) using a novel methodology: we combinatorially defined 18,768 potential clusters of patients by a pair of chronic conditions, a sex, and an age group, and then ranked clusters by 1) frequency, 2) cost and 3) ratios of observed to expected frequencies of co-occurring chronic conditions. We then compiled pairs of conditions that occur more frequently together than otherwise expected.

Results

61.5% of the study population suffers from two or more chronic conditions. The most frequent dyad was hypertension and hyperlipidemia (19%**)** and the most frequent triad was diabetes, hypertension and hyperlipidemia (10%). Women aged 50-65 with hypertension and hyperlipidemia were the leading cost segment in the study population. Costs and prevalence of MCC increase with number of conditions and age. The disease dyads associated with the largest observed/expected ratios were pulmonary disease and myocardial infarction. Inter-borough range MCC prevalence was 16%.

Conclusions

In this low-income, urban population, MCC is more prevalent (61%) than nationally (42%), motivating further research and intervention in this population. By identifying potential target populations in an interpretable manner, this clustering methodology has utility for health services analysts.

Strengths and limitations of this study

Strengths of the study:

- Large, robust dataset of patients with high prevalence of chronic disease

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

- New descriptive/analytic approach identifies unanticipated overlap of conditions
- Methodology applicable to other similar settings, including urban health systems

Weaknesses of the study:

- Cross-sectional data precludes causal analysis
- Use of cost claims data rather than clinical diagnosis

Data sharing statement

Data is available upon request from the corresponding author of the manuscript.

Patient and public involvement section

The study was a retrospective review using administrative claims data. The patients and public were not involved in this study.

Contributorship Statement

SPK conceived of the study. C Hajat advised on technical analysis. UM, C Hunt, and PD completed analyses. AB, EL, DJH, RK, and RF provided technical input to the manuscript. UM wrote the manuscript. SPK, EL, DJH, C Hunt, and C Hajat edited drafts of the manuscript.

Introduction

The management of multiple chronic conditions (MCC, here defined as the association of two or more chronic health conditions) constitutes a formidable clinical and financial challenge. An increasingly large proportion of the United States population lives with MCC, including 42% of adults overall and 81% of those over the age of 65 years [1]. In the US, MCC patients account for more than 70% of all healthcare spending [2]. In patients over 65 years old, costs increase exponentially with each additional chronic condition, suggesting that there are additional costs associated with the complexity or inefficiency of care for MCC. [3–10].

Health systems have responded to these challenges with clinical and financial innovations. Clinical innovations include new models of care coordination, joint clinical guidelines for MCC patients and alternative delivery models which include bundling of services [10–14]. Financial innovations include value-based payments and bundled payment schemes. One growing form of value-based financial transformation involves capitation, where a fixed "budget" for each patient is agreed upon between the payer and the health system. Accordingly, the health system is incentivized to bring costs down while still maintaining a small margin of profit. In this context, a standard methodology to evaluate the potential interactions between conditions could be mutually beneficial. Importantly, risk adjustment generates appropriately large budgets for high-cost and complex patients, and by doing so accounts for changes in severity over time and incentivizes providing coverage to these high-cost individuals. Existing systems of risk adjustment employed by the Centers for Medicaid & Medicare Services (CMS) predict medical and pharmaceutical spending using

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

demographics and diagnosis codes, and are employed in a standardized fashion for Medicare Advantage patients. State-managed Medicaid plans can choose to employ any of many different risk adjustment models, some of which are based on the Medicare Advantage models [15].

Especially important in the setting of value-based payment schemes like capitation is the appropriate selection of sub-populations to receive clinical interventions. While increasingly popular nationally, measures targeting patients who are chronically hospitalized (sometimes known as “super-utilizers”) have demonstrated mixed cost savings, in part because of difficulties targeting patients who could benefit from interventions [14,16,17].

While there exist numerous sophisticated statistical methods for clustering populations of patients - such as random forests, single decision trees, k-means, and hierarchical cluster analysis - these methods suffer from limited interpretability, result instability, immense computing overhead and/or tendency for overfitting [18–20]. Rather than relying on complex statistical models that require significant computing overhead, we propose a simple descriptive method that can be applied to any population for whom medical claims are available. Because its requisites are computationally simple, this methodology can be easily scaled to larger populations.

Prior studies of spending and MCC have focused on synergy in spending between conditions, or on a specific slice of a population, or type of spending - for example. on inpatient or outpatient spending, or on those older than 65 [7–9,21–24]. Notably, literature on MCC patterns and trends among younger, lower socioeconomic status, and vulnerable populations remains scarce, despite their carrying a significant share of chronic disease burden and, accordingly, financial risk in value-based schemes [25]. Additionally, under global capitation both inpatient and outpatient costs must be considered together.
In order to develop a methodology that would yield interpretable insights for both clinical interventions and financial incentives, we sought to first iteratively but simply generate many different sub-populations within the study population and then sort them via either clinically meaningful or financially relevant mechanisms. Clinical interventions can be developed from epidemiological information about which conditions are observed more frequently together than expected [26]. We theorized that observed/expected (independent) ratios would reveal groups of patients distinct from those based purely on frequency or cost. Combinations of chronic conditions could have shared risk factors (e.g. hypertension and diabetes), shared etiology (e.g. hypertension and congestive heart failure) or could be independent altogether (e.g. hypertension and arthritis). By contrast, financial interventions can be developed from cost information about which conditions and combinations of conditions occur in the most costly groups of patients. In practical terms, targeting the highest cost combinations of conditions (and therefore clusters of patients) could lead to proactive interventions to reduce avoidable or excess utilization.

Accordingly, in this manuscript we (1) develop a descriptive methodology to identify and describe unique clusters of MCC patients, and (2) apply the methodology in an urban health system using administrative claims data derived from a population of managed

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Medicaid patients at the Mount Sinai health system under global capitation -- a low-income, urban population unlike those previously studied. We also describe the general cost and geographic characteristics of this population, with potential use in future clustering applications.

Methods

Clustering

Clusters refer to groups of patients who meet certain disease criteria, demographic criteria, or both. For example, a cluster of patients could be defined by a dyad of diseases (i.e. hypertension and hyperlipidemia), an age range (ages 35-50 years), and sex (males). That cluster would consist of male patients aged 35-50 years with both hypertension and hyperlipidemia. As described, these clusters are not mutually exclusive (i.e. one patient can belong to several clusters). We systematically investigated every possible cluster of patients defined by a combination of two chronic conditions (among 69), an age group (18-35, 35-50, 50-65, 65+), and sex, yielding 18,768 potential clusters. For each of these clusters of patients, we computed a number of cluster characteristics by which to rank them: total cost attributable to cluster, average cost per person in cluster, and observed:expected ratio of disease dyads in each cluster. The total cost attributable to the patients in each cluster was computed using claims provided by the payer. This calculation includes all costs for these patients, not just those attributable to the diseases defining the cluster. Clusters were also ranked by average cost per person per year of plan enrollment represented in the cluster. For each pair of diseases defining a cluster, an *observed:expected ratio* was computed by dividing the observed frequency of the pair of diseases in the study population by the expected frequency (multiplying together the individual frequencies of each disease in the pair). We chose a cutoff of 30 cluster members as the lower limit for understanding probable outcomes through a pilot program [27].

Chronic Conditions Lists

We completed a review of pre-existing approaches and opted to work with a defined list of 69 chronic condition categories from the Agency for Healthcare Research and Quality (AHRQ) [28–30]. This condition list was chosen because (1) it included the most expansive list developed by a consensus body of physicians, enabling us to detect uncommon combinations of conditions, and (2) it aligns with other federal multiple chronic condition projects.

Data Set and Inclusion Criteria

We used claims data from patients operating under a capitated contract between Mount Sinai Health System and Healthfirst, the largest managed care organization for federal Medicaid funds in New York State. These data include all medical claims from 2012 to 2014 including 6,676,867 claims for 213,091 plan members. This period spans from the first full year of claims following the start of the Mount Sinai-Healthfirst contract to the last year when claims were made with the International Classification of Diseases version 9 (ICD-9). Costs represent paid amounts, not charged amounts.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

We used the Agency for Healthcare Quality and Research (AHRQ) Healthcare Cost and Utilization Project (HCUP) mapping of 4,427 ICD-9 codes to 69 clinically-relevant chronic condition categories. We omitted 2015 data because ICD-10 codes were used inconsistently alongside ICD-9 codes, and the HCUP mapping of ICD-10 codes to chronic condition categories is incomplete. We performed a complete case analysis and excluded participants with missing age or gender. The study was approved through Institutional Review Board of the Icahn School of Medicine at Mount Sinai.

Variables

We studied age, gender, location, chronic condition codes, number of chronic conditions, and total cost of care during the member's plan enrollment. Multiple chronic conditions were studied as dyads and triads. The analysis of different combinations of cluster criteria was limited by processing power and computational cost.

Statistical Analyses

The observed frequency of each cluster was age-adjusted using the New York State age distribution. Clusters were segmented by gender. Estimates were calculated for clusters defined by chronic condition codes, gender, age, and total cost of care. Claims were aggregated by patient-year via SQL, and subsequent cleaning, analysis, and plotting was performed with R and Python (code available in Supplementary Information).

Results

Prevalence of MCC by selected characteristics

61.5% of the study population (61.6% in women, 61.4% in men) lives with two or more chronic conditions, as compared to 42% nationally. **Table 1** displays demographic data of the sample (n = 143,297 patients). Median age was 47 years (25th percentile = 30; 75th percentile = 61), and 54.6% (78,199) were female. We identified the most prevalent combinations of two and three chronic conditions. Each dyad or triad result represents the prevalence of patients with that combination of chronic conditions, including those that also have additional conditions (for example, a patient with hypertension, hyperlipidemia, and diabetes would be counted in a single triad, and also within both the hypertension-hyperlipidemia and hyperlipidemia-diabetes dyads).

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 1.

	0 Chronic Condition(s)			1 Chronic Condition(s)			2 Chronic Condition(s)			3 Chronic Condition(s)			4 Chronic Condition(s)			5+ Chronic Condition(s)		
	2012	2013	2014	2012	2013	2014	2012	2013	2014	2012	2013	2014	2012	2013	2014	2012	2013	2014
n	10,732	15,092	17,416	9,960	13,544	16,286	7,271	9,887	11,698	5,644	7,741	9,014	4,531	6,125	7,338	15,641	20,984	26,968
Age (mean (sd))	30.88 (11.34)	31.25 (11.69)	31.76 (12.11)	34.74 (13.30)	35.39 (13.87)	35.95 (14.06)	38.92 (14.40)	38.73 (14.56)	38.97 (14.80)	43.08 (15.06)	43.23 (15.19)	42.76 (15.09)	46.91 (15.35)	46.52 (15.02)	46.49 (14.99)	54.45 (14.65)	54.03 (14.42)	54.36 (14.56)
Sex = F (%)	3,678 (34.3)	5,030 (33.3)	5,906 (33.9)	3,487 (35.0)	4,720 (34.8)	5,926 (36.4)	2,506 (34.5)	3,282 (33.2)	4,181 (35.7)	1,872 (33.2)	2,583 (33.4)	3,168 (35.1)	1,581 (34.9)	2,135 (34.9)	2,643 (36.0)	5,393 (34.5)	7,103 (33.8)	9,335 (34.6)
Total Cost (mean (sd))	847.26 (2002.96)	898.89 (2074.48)	860.56 (2006.20)	1,327.86 (2893.45)	1,268.02 (2740.95)	1,266.16 (2904.01)	1,758.23 (3416.84)	1,800.90 (4907.68)	1,777.87 (5717.80)	2,221.77 (5122.45)	2,001.08 (3586.80)	2,093.73 (4167.59)	2,634.04 (5017.66)	2,588.73 (6606.52)	2,606.87 (5742.42)	8,968.28 (19991.33)	8,673.94 (20181.73)	8,415.85 (18111.71)
Total Cost Winsorized (mean (sd))	842.51 (1949.60)	894.49 (2022.10)	858.48 (1973.22)	1,280.61 (2364.27)	1,227.41 (2312.96)	1,216.23 (2288.50)	1,661.55 (2623.90)	1,659.58 (2615.23)	1,622.92 (2551.18)	1,987.38 (2825.16)	1,861.01 (2651.65)	1,927.38 (2738.23)	2,306.30 (2865.16)	2,232.10 (2882.43)	2,304.93 (2888.36)	5,061.43 (4455.64)	4,866.61 (4416.25)	4,945.13 (4352.52)
Top 10 Single Chronic Conditions (%)																		
Allergy, ENT and other upper resp disorders	0 (0)	0 (0)	0 (0)	911 (19.8)	1,109 (17.8)	1,338 (18.5)	1,084 (20.8)	1,430 (19.8)	1,574 (19.0)	1,111 (22.9)	1,360 (20.4)	1,604 (21.1)	1,029 (24.5)	1,276 (22.7)	1,536 (22.6)	4,419 (28.8)	5,614 (27.2)	7,478 (28.3)
Asthma, COPD, other chronic lung disease	0 (0)	0 (0)	0 (0)	654 (14.2)	876 (14.1)	960 (13.3)	758 (14.6)	1,098 (15.2)	1,229 (14.9)	770 (15.9)	965 (14.5)	1,129 (14.8)	729 (17.4)	922 (16.4)	1,098 (16.1)	4,344 (28.3)	5,526 (26.8)	7,079 (26.8)
Obesity	0 (0)	0 (0)	0 (0)	199 (4.3)	314 (5.0)	334 (4.6)	438 (8.4)	633 (8.8)	706 (8.5)	698 (14.4)	965 (14.5)	1,040 (13.7)	856 (20.4)	1,242 (22.0)	1,411 (20.7)	6,482 (42.2)	8,974 (43.5)	11,143 (42.1)
Degenerative eye problem (glauc/eye)	0 (0)	0 (0)	0 (0)	287 (6.2)	418 (6.7)	451 (6.2)	517 (9.9)	679 (9.4)	756 (9.1)	683 (14.1)	909 (13.7)	921 (12.1)	715 (17.1)	955 (17.0)	1,139 (16.7)	5,128 (33.4)	6,810 (33.0)	8,907 (33.7)
Hyperlipidemia	0 (0)	0 (0)	0 (0)	597 (13.0)	846 (13.6)	999 (13.8)	598 (11.5)	827 (11.5)	983 (11.9)	754 (15.5)	926 (13.9)	1,125 (14.8)	839 (20.0)	1,069 (19.0)	1,339 (19.7)	5,531 (36.0)	7,431 (36.0)	9,571 (36.2)
Depression and depressive disorders	0 (0)	0 (0)	0 (0)	476 (10.4)	604 (9.7)	700 (9.7)	1,106 (21.3)	1,448 (20.1)	1,637 (19.8)	1,613 (33.2)	2,159 (32.5)	2,276 (29.9)	1,834 (43.8)	2,442 (43.4)	2,790 (41.0)	10,781 (70.2)	14,320 (69.5)	18,018 (68.1)
Hypertension	0 (0)	0 (0)	0 (0)	438 (9.5)	633 (10.2)	700 (9.7)	1,079 (20.7)	1,604 (22.2)	1,764 (21.3)	1,579 (32.5)	2,246 (33.8)	2,542 (33.4)	1,753 (41.8)	2,492 (44.2)	2,899 (42.6)	9,735 (63.4)	13,393 (65.0)	16,891 (63.8)
Esophageal disorder and GI ulcers	0 (0)	0 (0)	0 (0)	284 (6.2)	444 (7.1)	577 (8.0)	569 (10.9)	913 (12.7)	1,171 (14.2)	717 (14.8)	1,169 (17.6)	1,449 (19.0)	736 (17.6)	1,103 (19.6)	1,536 (22.6)	4,069 (26.5)	5,971 (29.0)	8,502 (32.1)
Malnutrition and F/E cond (not obesity/overweight)-includes disorders of metabolism	0 (0)	0 (0)	0 (0)	564 (12.3)	793 (12.7)	909 (12.6)	799 (15.4)	1,106 (15.3)	1,317 (15.9)	831 (17.1)	1,091 (16.4)	1,384 (18.2)	765 (18.2)	1,025 (18.2)	1,274 (18.7)	3,767 (24.5)	5,199 (25.2)	7,459 (28.2)
Diabetes mellitus	0 (0)	0 (0)	0 (0)	183 (4.0)	197 (3.2)	250 (3.5)	309 (5.9)	405 (5.6)	427 (5.2)	437 (9.0)	592 (8.9)	651 (8.6)	524 (12.5)	724 (12.9)	825 (12.1)	4,965 (32.3)	6,592 (32.0)	8,847 (33.4)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

These overlapping clusters of patients, ranked by age-adjusted frequency, are reported in **Table 2.** Of these, 16,044 clusters contained at least one patient - with the largest cluster containing an average of 4,329 patients per year. The most common dyad was hypertension and hyperlipidemia (19% age adjusted), and the most common triad was hypertension, hyperlipidemia and diabetes **(**10% age adjusted**).**

Table 2.

Singlet Chronic Condition	Average Yearly Membership	Unadjusted %	Age Adjusted %
Hypertension	20,724	29%	28%
Hyperlipidemia	19,932	28%	26%
Diabetes mellitus	11,801	16%	16%
Degenerative eye problem (glauc/eye)	11,153	16%	15%
Allergy, ENT, and other upper resp disorders	10,938	15%	12%
Dyad Chronic Conditions	**Average Yearly Membership**	**Unadjusted %**	**Age Adjusted %**
Hypertension & Hyperlipidemia	12,808	18%	18%
Hypertension & Diabetes mellitus	8,707	12%	12%
Hyperlipidemia & Diabetes mellitus	8,203	11%	11%
Hypertension & Degenerative eye problem (glauc/eye)	6,332	9%	10%
Hyperlipidemia & Degenerative eye problem (glauc/eye)	6,116	9%	9%
Triad Chronic Conditions	**Average Yearly Membership**	**Unadjusted %**	**Age Adjusted %**
Diabetes mellitus, Hypertension, & Hyperlipidemia	6,778	9%	9%
Hypertension, Degenerative eye problem (glauc/eye), & Hyperlipidemia	4,792	7%	7%
Osteoarthritis, Hypertension, & Hyperlipidemia	4,087	6%	6%
Esophageal disorder and GI ulcers, Hypertension, &Hyperlipidemia	3,828	5%	5%
Diabetes mellitus, Hypertension, & Degenerative eye problem (glauc/eye)	3,727	5%	5%

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Healthcare expenditures

Figure 1 shows healthcare expenditure among patients with different numbers of chronic conditions. Patients with missing demographic data have been excluded (12.4% of all patients). Costs increase by over 40% with each additional condition, as does the patient-to-patient variance in yearly cost.

Clusters by Age, Sex, Costs

Supplementary Table 1 indicates the top clusters and characteristics by chronic conditions by age and gender using the classification outlined in the Methods section. The lists are presented by top 10 highest frequency (3A), top 10 dyads with the highest costs and at least an average of 30 members per year (3B) and by top 10 dyads with the highest cost and at least an average of 1,000 members per year.

Adjusting the minimum threshold number of patients constituting a cluster alters the kinds of diseases represented. For example, if the minimum size of a cluster is 30 members, the highest cost segment becomes males age 35-50 with “Anemia and other non-cancer hematological disorder” & “conduction disorder or cardiac dysrhythmia”. However, if this threshold is raised to 1,000 members, the highest cost segment becomes females age 50-65 with “Hypertension & Coronary atherosclerosis”. In general, smaller clusters (>30-1000 members) tended to be higher in average individual cost, but lower in total cost, than the larger clusters (>1,000 members).

Table 3 shows all clusters segmented by age (5 categories) and gender (male/female). This table indicates dyads of chronic conditions organized by observed/expected ratios. This data reveal a different relationship of chronic conditions to one another than the frequency and cost tables. By selecting clusters of patients with at least 30 included, we demonstrate relationships between unexpected diseases in small yet high-cost groups of patients. For example, paralysis and immunity disorders occur at 16.6 times the expected rate, accounting for an average yearly cost of $81,414. By selecting clusters of patients with at least 1,000, we demonstrate relationships that are more commonly observed (and more frequently expected), such as between peripheral atherosclerosis and coronary atherosclerosis, or between anxiety disorders and bipolar disorder.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 3.

Top 10 Dyads by Observed / Expected rate with at least an average yearly membership of 30 members or more

Dyad	Unadjusted Frequency	Adjusted Frequency	Adjustment Magnitude	Expected Frequency	Observed / Expected	Average Yearly Cost	Total Cost	Total Members
Acute myocardial infarction & Pulmonary heart disease	0.001	0.001	0	2.80E-05	35.7	$89,321	$11,790,348	132
Thrombosis and Embolism & Non-thrombotic, non-athlerosclerotic vascular disease	0.001	0.001	0	4.20E-05	23.8	$68,541	$9,184,538	134
Pulmonary heart disease & Congestive heart failure	0.003	0.004	0.001	0.000175	22.9	$56,526	$38,098,314	674
Paralysis & Epilepsy	0.001	0.001	0	4.80E-05	20.8	$52,895	$9,732,621	184
Acute myocardial infarction & 'Cardiomyopathy and Structual Heart Disease	0.001	0.002	0.001	0.0001	20.0	$66,547	$21,095,470	317
Acute myocardial infarction & Congestive heart failure	0.002	0.002	0	0.0001	20.0	$66,271	$24,453,854	369
Congenital Heart Disease & Heart valve disorder	0.01	0.01	0	0.000546	18.3	$11,172	$22,979,895	2,057
Pulmonary heart disease & 'Cardiomyopathy and Structual Heart Disease	0.003	0.003	0	0.000175	17.1	$55,752	$34,510,492	619
Paralysis & Organic brain problem (dementia)	0.001	0.001	0	6.00E-05	16.7	$60,838	$6,935,557	114
Paralysis & Immunity disorder	0	0.001	0.001	6.00E-05	16.7	$81,415	$8,711,389	107

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Top 10 Dyads by Observed / Expected rate with at least an average yearly membership of 1,000 members or more

Dyad	Unadjusted Frequency	Adjusted Frequency	Adjustment Magnitude	Expected Frequency	Observed / Expected	Average Yearly Cost	Total Cost	Total Members
Heart valve disorder & Coronary atherosclerosis	0.014	0.016	0.002	0.002535	6.31	$20,896	$64,547,753	3,089
Conduction disorder or cardiac dysrhythmia & 'Coronary atherosclerosis	0.017	0.02	0.003	0.003185	6.28	$26,595	$97,260,685	3,657
Cerebrovascular Disease & Coronary atherosclerosis	0.014	0.017	0.003	0.00286	5.94	$23,622	$72,803,180	3,082
Peripheral atherosclerosis & 'Coronary atherosclerosis	0.017	0.02	0.003	0.00351	5.70	$20,381	$75,512,538	3,705
Anxiety disorders & 'Depression and depressive disorders	0.042	0.033	-0.009	0.006365	5.18	$10,143	$92,526,384	9,122
Depression and depressive disorders & Bipolar disorder	0.021	0.016	-0.005	0.003135	5.10	$11,218	$50,471,365	4,499
Anxiety disorders & Bipolar disorder	0.015	0.011	-0.004	0.002211	4.98	$11,539	$36,800,083	3,189
Cerebrovascular Disease & Other central and peripheral nervous system disorders	0.017	0.018	0.001	0.004004	4.50	$23,374	$86,954,477	3,720
Peripheral atherosclerosis & Other central and peripheral nervous system disorders	0.022	0.022	0	0.004914	4.48	$17,088	$81,155,040	4,749
Other central and peripheral nervous system disorders & Back problem	0.025	0.022	-0.003	0.005005	4.40	$13,315	$72,770,548	5,465

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Age, Spatial distribution and rising risk for patients with multiple chronic conditions

Figure 2 shows frequency of multiple chronic conditions as a function of age across the 5 counties in New York City. A 50% prevalence of MCC is seen at age 30-34 in the Bronx, a lower-income borough of the city, whereas in Brooklyn at in the same 30-34 age-group, the prevalence is only 34%**.**

Discussion

In this paper, we argue that this simple descriptive clustering methodology has utility for resource planning, care coordination, and care delivery. This methodology would be especially useful in the context of public and private benefits schemes focused on low-income populations.

We find that 61.5% of our population lives with two or more chronic conditions as compared to 42% nationally, motivating efforts to build MCC interventions and tools in the Medicaid population [2]. Using an established list of conditions, we found that total costs increase with each condition added, consistent with findings from other research groups [31–38]. We also found that the most frequent dyad of co-occurring chronic conditions was hypertension and hyperlipidemia (19% age-adjusted) and the most frequent triad was diabetes, hypertension and hyperlipidemia (10% age-adjusted), each in turn more frequent in our study population than nationally (13.6%, as estimated from NHANES in 2010, and 6.3%, from NHANES in 2012) [39,40]. This is a striking finding, considering that the NHANES cohort includes a larger proportion of older adults than our study. As NHANES includes fixed sample-size targets and weighting to generate a national sample of households that is representative of the US adult population, the median age at the time of these studies was 37.2, significantly older than the median of 26 in our dataset. This age discrepancy could be due to two reasons: (1) As adults who are dual-eligible for Medicaid and Medicare are often re-directed to managed Medicare contracts, our study population under-represents adults over 65. (2) Studies of chronic conditions in adults using NHANES tend to utilize a minimum age of 20, as people aged 19 or younger are categorized as ‘youth’; compared to the age cutoff of 17 or younger in our study population [5,38].

Women aged 50-65 with hypertension and hyperlipidemia were the leading cost segment in the health system for dyads. Overall, women age 50-65 and hypertension, osteoarthritis, hyperlipidemia were the leading triad in terms of prevalence and cost. The most significant observed/expected ratio dyads were pulmonary disease and myocardial infarction. We provided various approaches to grouping these chronic conditions in service of broader research objectives to identify conditions that drive multiplicative, rather than additive, health or cost burdens.

The O/E approach provides a clinically oriented view of examining which conditions occur disproportionately together. For example, we find that in our study population, anemia, pulmonary heart disease, congestive heart failure and conduction disorders occur together more frequently than expected. We also observe that patients’ costs balloon when they have these conditions. This would suggest an area where healthcare systems need to focus – screening, dedicated counseling, resources and research dollars. For instance, by targeting patients with conditions like anemia and pulmonary heart disease that do not appear to be physiologically related, care managers can minimize fractures in care. If taken together with our finding that MCC burden differs by locale (**Figure 2)** health systems should elect to co-locate specialty clinics, share clinical teams, and develop joint management protocols for these conditions. While these kinds of innovations have been prototyped around episodic procedural care, such as knee and hip replacements, they have yet to be adopted in managing MCC [13,41,42]. Meanwhile, patients with multiple chronic conditions are already requesting these changes [43]. Importantly, this approach yields specific chronic disease targets beyond the most frequent conditions.

Conditions like anemia and pulmonary heart disease are not currently considered among the interaction terms included in existing CMS models (which focus instead on predicting indicators of severe disease like sepsis, pulmonary embolism, or seizure disorders), but may be more locally appropriate measures of disease severity or spending in this population. Further validation would be required of these novel disease interactions in a larger or different sample population.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

At the same time, the sorting of clusters by highest cost and frequency provides a simple view of groups where minor interventions could result in larger-scale cost-savings, particularly for health systems facing value-based financing schemes. Addressing the top clusters of patients with bundled financial incentives could supplement the clinical innovations described above. Indeed, recent analyses of the Medicare Shared Savings plan have found that a significant proportion of savings were derived from incremental cost interventions that applied to large swathes of the insured population [44].

Notably, these results differ from a separate analysis by our research team using a list of 12 chronic conditions in the Behavioral Risk Factor Surveillance Survey conducted by the Centers for Disease Control. In this work, we found that from 2011-2016, 50.6% of adults in New York State had two or more chronic conditions. The most prevalent dyads we identified were hypertension and high cholesterol (17% and most prevalent triad was hypertension, high cholesterol and arthritis (4.5%). Prevalence of MCC in NYC neighborhoods ranged from 33.5 to 60.6% [45].

Our findings apply not only to the reform of existing programs for low-income and vulnerable populations, but also the design of novel ones, in the Mount Sinai system and beyond. For example, Mount Sinai offers Healthfirst (and other) patients who require inpatient-level care an alternative: a Hospitalization-at-Home (HaH) program in lieu of inpatient admission [46,47]. Evaluation to date demonstrates that this HaH program delivers superior patient outcomes (including shorter length of stay) and greater patient satisfaction than in-hospital care, though costs have not yet been compared [46]. The HaH program focused on only nine diagnoses at its founding in 2014, but has since expanded in size and breadth of care across multiple New York hospitals, treating myriad other conditions across eight domains of care, such as post-surgical care, palliative care, and sub-acute rehabilitation, among others [47]. Rapid and timely data on the prevalence and overlap of these (largely chronic) diseases and their risk factors will be instrumental to the program's ongoing cost-effective scale-up. Such data could prove even more valuable in low- and middle- income countries, where the burden of chronic disease is rapidly expanding, but models for the integrated care of more than one chronic condition are few and small in scope [48].

The limitations of our proposed approach include the following: (1) the use of health insurance claims itself limits the epidemiologic utility of the analyses. Claims are effectively billing receipts and therefore have limited reliability in reporting disease states [49]. Additionally, we did not control for variations in coding by center or physician. We plan to integrate these claims data with EMR data going forward in order to retrieve higher quality epidemiological insights. (2) Our analysis is limited by the study period. Data from 2012-2014 is likely not recent enough to enact present-day interventions in a health system -- this is largely because the mapping of ICD-10 codes to chronic condition categories has not been finalized, with some remaining discontinuities between ICD-9 and ICD-10-based classifications, limiting our ability to use data from 2015 onwards. We plan to include more recent data once the mapping is completed, as well as prototype this methodology using the CMS Chronic Condition Warehouse algorithm, which functions with ICD-10 codes but includes fewer conditions (27 rather than 69). [50] Additionally, we did not examine epidemiologic trends through time, as a period as short as 3 years is not long enough to elucidate relationships between diseases that share etiology (i.e. hypertension, stroke). (3) The generalizability of our analysis is limited by the geospatial distribution of patients in the study population -- because provider attribution is accomplished regionally, our data set includes the subset of New York City patients who live near Mount Sinai practices. As a result, in the current data set, the majority of patients are located in just 10 of 176 ZIP codes. Future analyses using a data set such as an all-payer claims database would allow researchers to define clusters by region and ZIP code. Accordingly, this study population of managed Medicaid patients is not necessarily representative of the Medicaid or U.S. population at-large, or the fee-for-service Medicaid population served by Mount Sinai. (4) We did not include pharmacy claims in our analysis, which will result in an

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

underestimation of spending. This underestimation is most significant regarding conditions that require expensive medications (i.e. high-cost injectables for HIV and hepatitis C). However, we also note that risk adjustment methodologies employed by Medicaid Advantage and State Medicaid programs tend to predict spending on pharmaceuticals separate from other costs. [15] (5) Lastly, a significant portion (12.4%) of our study population was excluded on account of missing demographic data, introducing some bias into which clusters of patients were highlighted. Any more pragmatic application of this methodology would also require an approach to patients with missing data.

Taken together, these analyses have implications for health systems, financiers, and researchers working to address MCC, and provide a common methodology for targeting populations for financial and clinical intervention. Most notably, this tool yields a simple, transparent methodology for selecting coherent, clearly-defined populations of patients for intervention, and can be applied to any commercial claims dataset. With application in the right contexts, this methodology could help improve the selection strategy of super-utilizer clinics and other clinical innovations, yielding further advancements in our health systems' management of chronic conditions. Payors may increasingly rely on interaction of diseases to help identify appropriate levels of reimbursement based on predicted risk of hospitalization or mortality for patients. Ultimately, however, more research is needed to evaluate this methodology's utility in business scenarios, and applicability to different sizes and kinds of patient populations.

Figure Legends / Captions

Table 1. Demographics of Medicaid patients at Mount Sinai Health System belonging to Healthfirst capitated contracts.
Table 2. Top clusters of two and three chronic conditions using overall list of 69 conditions.
Figure 1. Distribution of individual annual healthcare expenditures as a function of number of chronic conditions.
Supplementary Table 1. Top clusters of chronic conditions by age and gender segments. (A) largest clusters by member count, (B) most costly clusters 30 people or greater, (C) most costly clusters 1000 people or greater.
Table 3. Observed/Expected ratios of chronic conditions among common (A) and uncommon clusters (B)
Figure 2. Frequency of multiple chronic conditions by age across selected boroughs of New York City. 50% prevalence of multiple chronic conditions seen at age 30-34 for all boroughs except for Brooklyn that reaches 50% at age 35-39. Disparities between boroughs observed.

References

1 Buttorff C, Ruder T, Bauman M. Multiple Chronic Conditions in the United States. RAND Corporation 2017. doi:10.7249/tl221

2 Gerteis J, Izrael D, Deitz D, *et al.* Multiple Chronic Conditions Chartbook. Rockville, MD: Agency for Healthcare Research and Quality; 2014. https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/prevention-chronic-care/decision/mcc/mccchartbook.pdf

3 Hoffman C. Persons with chronic conditions. Their prevalence and costs. *JAMA* 1996;**276**:1473–9.

4 Schneider KM, O'Donnell BE, Dean D. Prevalence of multiple chronic conditions in the United States' Medicare population. *Health Qual Life Outcomes* 2009;**7**:82.

5 Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med* 2002;**162**:2269–76.

6 Lehnert T, Heider D, Leicht H, *et al.* Review: health care utilization and costs of elderly

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

persons with multiple chronic conditions. *Med Care Res Rev* 2011;**68**:387–420.

7 Cortaredona S, Ventelou B. The extra cost of comorbidity: multiple illnesses and the economic burden of non-communicable diseases. *BMC Med* 2017;**15**:216.

8 Brilleman SL, Purdy S, Salisbury C, *et al.* Implications of comorbidity for primary care costs in the UK: a retrospective observational study. *Br J Gen Pract* 2013;**63**:e274–82.

9 He Z, Bian J, Carretta HJ, *et al.* Prevalence of Multiple Chronic Conditions Among Older Adults in Florida and the United States: Comparative Analysis of the OneFlorida Data Trust and National Inpatient Sample. *J Med Internet Res* 2018;**20**:e137.

10 Hajat C, Stein E. The global burden of multiple chronic conditions: A narrative review. *Prev Med Rep* 2018;**12**:284–93.

11 Hajat C, Kishore SP. The case for a global focus on multiple chronic conditions. *BMJ Glob Health* 2018;**3**:e000874.

12 Hajat C, Stein E, Yach D. Multiple chronic conditions: the global state. https:// lnkd. in/ etb9eAJ

13 Keswani A, Koenig KM, Bozic KJ. Value-based Healthcare: Part 1-Designing and Implementing Integrated Practice Units for the Management of Musculoskeletal Disease. *Clin Orthop Relat Res* 2016;**474**:2100–3.

14 Berkowitz SA, Parashuram S, Rowan K, *et al.* Association of a Care Coordination Model With Health Care Costs and Utilization: The Johns Hopkins Community Health Partnership (J-CHiP). *JAMA Netw Open* 2018;**1**:e184273–e184273.

15 March 31, 2016, HHS-Operated Risk Adjustment Methodology Meeting Discussion Paper. Centers for Medicare & Medicaid Services, Center for Consumer Information & Insurance Oversight https://www.cms.gov/CCIIO/Resources/Forms-Reports-and-Other-Resources/Downloads/RA-March-31-White-Paper-032416.pdf

16 Kanzaria HK, Hoffman JR. Hot-Spotters Aren't 'The Problem'...But They Are Emblematic of the Failure of U.S. Healthcare. *J Gen Intern Med* 2017;**32**:6–8.

17 Lee NS, Whitman N, Vakharia N, *et al.* High-Cost Patients: Hot-Spotters Don't Explain the Half of It. *J Gen Intern Med* 2017;**32**:28–34.

18 Breiman L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Stat Sci* 2001;**16**:199–231.

19 Nicholson K, Bauer M, Terry A, *et al.* The Multimorbidity Cluster Analysis Tool: Identifying Combinations and Permutations of Multiple Chronic Diseases Using a Record-Level Computational Analysis. *J Innov Health Inform* 2017;**24**:962.

20 Ng SK, Tawiah R, Sawyer M, *et al.* Patterns of multimorbid health conditions: a systematic review of analytical methods and comparison analysis. *Int J Epidemiol* 2018;**47**:1687–704.

21 Crystal S, Johnson RW, Harman J, *et al.* Out-of-pocket health care costs among older Americans. *J Gerontol B Psychol Sci Soc Sci* 2000;**55**:S51–62.

22 Fishman P, Von Korff M, Lozano P, *et al.* Chronic care costs in managed care. *Health Aff* 1997;**16**:239–47.

23 Hoffman C. Persons with chronic conditions. Their prevalence and costs. *JAMA* 1996;**276**:1473–9.

24 Moxey ED, O'Connor JP, Novielli KD, *et al.* Prescription drug use in the elderly: a descriptive analysis. *Health Care Financ Rev* 2003;**24**:127–41.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

25 Shaw KM, Theis KA, Self-Brown S, *et al.* Chronic Disease Disparities by County Economic Status and Metropolitan Classification, Behavioral Risk Factor Surveillance System, 2013. *Prev Chronic Dis* 2016;**13**:E119.

26 Schäfer I, Kaduszkiewicz H, Wagner H-O, *et al.* Reducing complexity: a visualisation of multimorbidity by combining disease clusters and triads. *BMC Public Health* 2014;**14**:1285.

27 Hertzog MA. Considerations in determining sample size for pilot studies. *Res Nurs Health* 2008;**31**:180–91.

28 Harrison C, Britt H, Miller G, *et al.* Examining different measures of multimorbidity, using a large prospective cross-sectional study in Australian general practice. *BMJ Open* 2014;**4**:e004694.

29 Fortin M, Stewart M, Poitras M-E, *et al.* A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Ann Fam Med* 2012;**10**:142–51.

30 Le Reste JY, Nabbe P, Manceau B, *et al.* The European General Practice Research Network presents a comprehensive definition of multimorbidity in family medicine and long term care, following a systematic review of relevant literature. *J Am Med Dir Assoc* 2013;**14**:319–25.

31 Crystal S, Johnson RW, Harman J, *et al.* Out-of-pocket health care costs among older Americans. *J Gerontol B Psychol Sci Soc Sci* 2000;**55**:S51–62.

32 Fishman P, Von Korff M, Lozano P, *et al.* Chronic care costs in managed care. *Health Aff* 1997;**16**:239–47.

33 Hoffman C. Persons with chronic conditions. Their prevalence and costs. *JAMA* 1996;**276**:1473–9.

34 Moxey ED, O'Connor JP, Novielli KD, *et al.* Prescription drug use in the elderly: a descriptive analysis. *Health Care Financ Rev* 2003;**24**:127–41.

35 Rice DP, LaPlante MP. Medical expenditures for disability and disabling comorbidity. *Am J Public Health* 1992;**82**:739–41.

36 Schneider KM, O'Donnell BE, Dean D. Prevalence of multiple chronic conditions in the United States' Medicare population. *Health Qual Life Outcomes* 2009;**7**:82.

37 Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med* 2002;**162**:2269–76.

38 Lehnert T, Heider D, Leicht H, *et al.* Review: health care utilization and costs of elderly persons with multiple chronic conditions. *Med Care Res Rev* 2011;**68**:387–420.

39 Song Y, Liu X, Zhu X, *et al.* Increasing trend of diabetes combined with hypertension or hypercholesterolemia: NHANES data analysis 1999-2012. *Sci Rep* 2016;**6**:36093.

40 Egan BM, Li J, Qanungo S, *et al.* Blood pressure and cholesterol control in hypertensive hypercholesterolemic patients: national health and nutrition examination surveys 1988-2010. *Circulation* 2013;**128**:29–41.

41 Bleich SN, Sherrod C, Chiang A, *et al.* Systematic Review of Programs Treating High-Need and High-Cost People With Multiple Chronic Diseases or Disabilities in the United States, 2008-2014. *Prev Chronic Dis* 2015;**12**:E197.

42 Bandara S, Lynch G, Cooke C, *et al.* Using Care Bundles to Improve Surgical Outcomes and Reduce Variation in Care for Fragility Hip Fracture Patients. *Geriatr Orthop Surg Rehabil* 2017;**8**:104–8.

43 Bayliss EA, Edwards AE, Steiner JF, *et al.* Processes of care desired by elderly patients with multimorbidities. *Fam Pract* 2008;**25**:287–93.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

44 McWilliams JM, Chernew ME, Landon BE. Medicare ACO Program Savings Not Tied To Preventable Hospitalizations Or Concentrated Among High-Risk Patients. *Health Aff* 2017;**36**:2085–93.

45 Newman D, Levine E, Kishore SP. Prevalence of multiple chronic conditions in New York State, 2011-2016. *PLoS One* 2019;**14**:e0211965.

46 Federman AD, Soones T, DeCherrie LV, *et al.* Association of a Bundled Hospital-at-Home and 30-Day Postacute Transitional Care Program With Clinical Outcomes and Patient Experiences. *JAMA Intern Med* 2018;**178**:1033–40.

47 DeCherrie LV, Wajnberg A, Soones T, *et al.* Hospital at Home-Plus: A Platform of Facility-Based Care. *J Am Geriatr Soc* 2019;**67**:596–602.

48 Joshi R, Alim M, Kengne AP, *et al.* Task shifting for non-communicable disease management in low and middle income countries--a systematic review. *PLoS One* 2014;**9**:e103754.

49 Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997;**127**:666–74.

50 HCUP CCS. Healthcare Cost and Utilization Project (HCUP). 2017.www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp (accessed 27 Oct 2017).

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

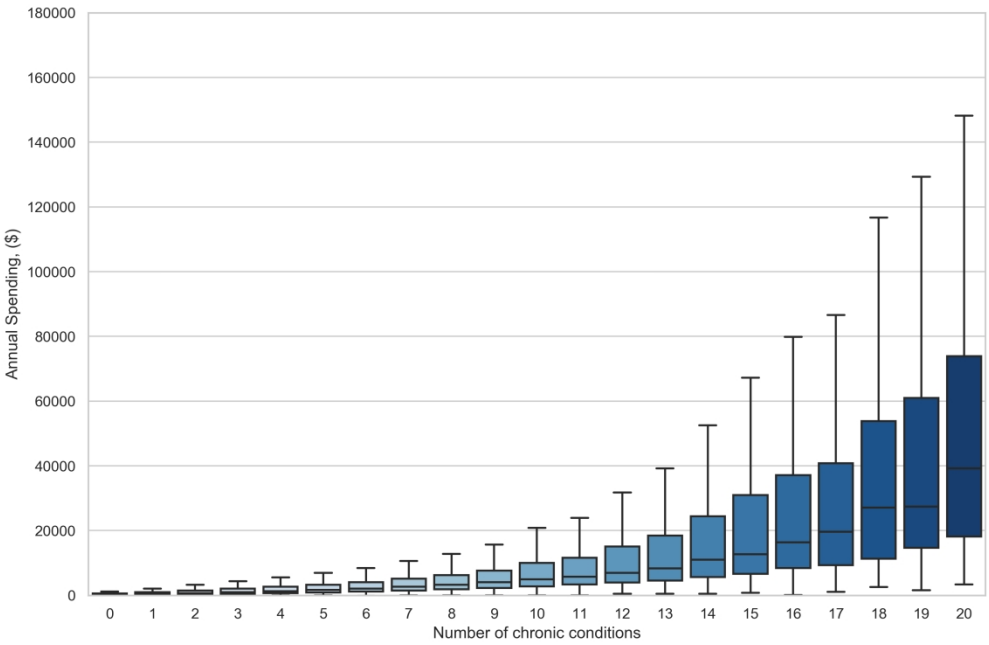


Figure 1 -- reuploaded as high-resolution Tiff

263x171mm (300 x 300 DPI)

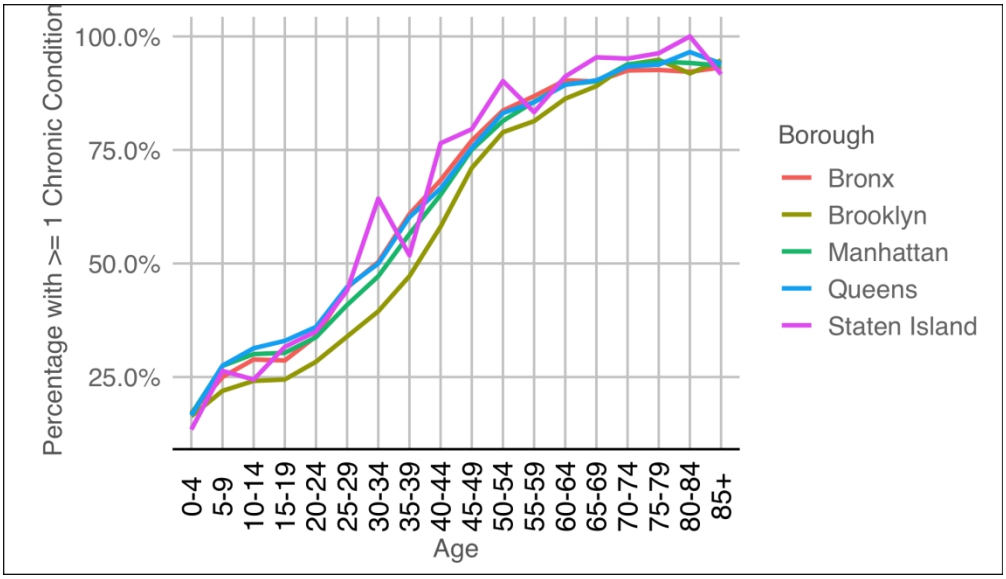


Figure 2 -- reuploaded as high-resolution Tiff

686x392mm (72 x 72 DPI)

Supplementary Table 1

Top 10 clusters by frequency

Gender	Age	Condition 1	Condition 2	Total Attributable Cost	Average Yearly Cost	Average Yearly Membership
F	50-65	Hypertension	Hyperlipidemia	$93,122,272	$7,172	4,329
F	50-65	Hypertension	Diabetes Mellitus	$72,878,330	$8,557	2,839
F	50-65	Hyperlipidemia	Diabetes Mellitus	$65,165,290	$8,143	2,668
M	50-65	Hypertension	Hyperlipidemia	$61,719,638	$7,948	2,589
F	50-65	Hypertension	Degenerative eye problem (glauc/eye)	$54,012,310	$8,240	2,185
F	50-65	Hypertension	Osteoarthritis	$66,447,600	$10,166	2,179
F	50-65	Hyperlipidemia	Degenerative eye problem (glauc/eye)	$49,533,370	$7,674	2,152
F	50-65	Hyperlipidemia	Osteoarthritis	$56,171,247	$9,295	2,014
F	50-65	Hypertension	Esophageal disorder and GI ulcers	$60,965,767	$10,297	1,974
F	50-65	Hyperlipidemia	Esophageal disorder and GI ulcers	$53,619,011	$9,194	1,944

Top 10 clusters by average yearly cost with 30 total member counts or more

Gender	Age	Condition 1	Condition 2	Total Attributable Cost	Average Yearly Cost	Average Yearly Membership
M	35-50	Anemia and other non-cancer heme disorders	Conduction disorder or cardiac dysrhythmia	$8,390,439	$90,220	31
F	50-65	Pulmonary heart disease	Anemia and other non-cancer heme disorders	$7,542,310	$83,803	30
M	35-50	Congestive heart failure	Malnutrition and F/E cond (not obesity/overweight) -includes disorders of metabolism	$7,068,459	$77,675	30
M	50-65	Conduction disorder or cardiac dysrhythmia	Immunity disorder	$9,800,142	$76,564	43
M	35-50	Other central and peripheral nervous system disorders	Immunity disorder	$6,917,900	$73,595	31

For peer review only

F	50-65	Congestive heart failure	Anemia and other non-cancer heme disorders	$14,346,180	$70,671	68
F	50-65	Congestive heart failure	Chronic skin ulcer	$7,852,354	$69,490	38
M	35-50	Anemia and other non-cancer heme disorders	Kidney and Vesicoureteral Disorders (excluding renal failure)	$6,174,594	$68,607	30
M	65+	Cardiomyopathy and Structural Heart Disease	Anemia and other non-cancer heme disorders	$9,505,337	$67,414	47
M	35-50	Congestive heart failure	Conduction disorder or cardiac dysrhythmia	$7,077,370	$67,404	35

Top 10 clusters by average yearly cost with 1,000 member counts or more

Gender	Age	Condition 1	Condition 2	Total Attributable Cost	Average Yearly Cost	Average Yearly Membership
F	50-65	Hypertension	Coronary atherosclerosis	$45,703,351	$14,486	1,052
F	50-65	Osteoarthritis	Other central and peripheral nervous system disorders	$43,931,227	$14,013	1,045
F	50-65	Hypertension	Other central and peripheral nervous system disorders	$61,441,249	$13,433	1,525
F	50-65	Hypertension	Asthma, COPD, other chronic lung disease	$57,028,007	$12,193	1,559
F	50-65	Hyperlipidemia	Other central and peripheral nervous system disorders	$50,790,724	$12,105	1,399
M	65+	Hypertension	Hyperlipidemia	$40,441,616	$11,933	1,130
F	50-65	Esophageal disorder and GI ulcers	Diabetes mellitus	$39,891,768	$11,827	1,124
F	50-65	Hyperlipidemia	Asthma, COPD, other chronic lung disease	$45,652,105	$11,757	1,294

For peer review only

F	50-65	Esophageal disorder and GI ulcers	Osteoarthritis	$42,200,192	$11,745	1,198
F	50-65	Diabetes mellitus	Osteoarthritis	$43,479,202	$11,591	1,250

For peer review only

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1	Multiple chronic conditions at a major urban health system: a descriptive analysis of frequencies, costs and patterns
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2	Lines 1-39
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4,5	Lines 96-145
Objectives	3	State specific objectives, including any prespecified hypotheses	5	Lines 162-168
Methods				
Study design	4	Present key elements of study design early in the paper	5	Lines 162-168
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6	Lines 202-208
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6	Lines 202-216
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	N/A	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6	Lines 219-222
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6	Lines 202-216

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Bias	9	Describe any efforts to address potential sources of bias	10	Line 375, 385
Study size	10	Explain how the study size was arrived at	6	Line 205

Continued on next page

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5,6	Line 173-191
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7	Line 225-231
		(*b*) Describe any methods used to examine subgroups and interactions	6	Line 187-188
		(*c*) Explain how missing data were addressed	6	Line 214
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	7	Line 225
		(*e*) Describe any sensitivity analyses	N/A	
Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7	Line 239
		(b) Give reasons for non-participation at each stage	7	Line 255
		(c) Consider use of a flow diagram	N/A	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7	Line 236-244
		(b) Indicate number of participants with missing data for each variable of interest	7	Line 255
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	N/A	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	N/A	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	N/A	
		Cross-sectional study—Report numbers of outcome events or summary measures	7,8	Table 1,2,3
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included		
		(*b*) Report category boundaries when continuous variables were categorized	6	Line 180
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A	

Continued on next page

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8	Line 290
Discussion				
Key results	18	Summarise key results with reference to study objectives	8	Line 303-341
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9	Line 357-386
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10	Line 388-397
Generalisability	21	Discuss the generalisability (external validity) of the study results	10	Line 373
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	1	This work was supported by Teva Pharmaceuticals for the Multiple Chronic Conditions Initiative with the Arnhold Institute for Global Health. Dr. Heller also reports support from the NIH Fogarty International Center (R21 TW010452-01).

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

Multiple chronic conditions at a major urban health system: a retrospective cross-sectional analysis of frequencies, costs and comorbidity patterns

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029340.R2
Article Type:	Original research
Date Submitted by the Author:	31-Aug-2019
Complete List of Authors:	Majumdar, Usnish; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health Hunt, Christophe; Mount Sinai Health Partners Doupe, Patrick; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health Baum, Aaron; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health; Icahn School of Medicine at Mount Sinai, Department of Health System Design and Global Health Heller, David; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health; Icahn School of Medicine at Mount Sinai, Department of Health System Design and Global Health Levine, Erica; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health Kumar, Rashi; Healthfirst Futterman, Robert; Healthfirst Hajat, Cother; UAE University, Kishore, Sandeep; Icahn School of Medicine at Mount Sinai, Arnhold Institute for Global Health; Brigham and Women's Hospital / Harvard Medical School, Department of Medicine
Primary Subject Heading:	Health informatics
Secondary Subject Heading:	Health services research
Keywords:	HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PUBLIC HEALTH

SCHOLARONE™
Manuscripts

Multiple chronic conditions at a major urban health system: a retrospective cross-sectional analysis of frequencies, costs and comorbidity patterns

Usnish Majumdar[1], Christophe Hunt[2], Patrick Doupe[1], Aaron Baum[1], David J. Heller[1], Erica Levine[1], Rashi Kumar[3], Robert Futterman[3], Cother Hajat[4], Sandeep P. Kishore[1,5]

1. Arnhold Institute for Global Health, Icahn School of Medicine at Mount Sinai, New York, New York, USA
2. Mount Sinai Health Partners, Mount Sinai Health System, New York, New York
3. Healthfirst, New York, New York, USA
4. Institute of Public Health, United Arab Emirates University, Al Ain, United Arab Emirates
5. Brigham & Women's Hospital/ Harvard Medical School, Boston, MA, USA

Corresponding author:

Sandeep P. Kishore
1216 Fifth Avenue, Room 556
New York, New York 10129
Phone: 212-824-8957
Fax: 212 - 824 – 7995
Email Address: sunny.kishore@gmail.com

Word Count (max 4500): 3,810

Tables: 4

Figures: 2

Funding:
This work was supported by Teva Pharmaceuticals for the Multiple Chronic Conditions Initiative with the Arnhold Institute for Global Health. Dr. Heller also reports support from the NIH Fogarty International Center (R21 TW010452-01).

Competing Interests Statement:
We declare no significant competing financial, professional, or personal interests that might have influenced the design, performance, interpretation, or presentation of the analyses described in this manuscript apart from the above. Teva Pharmaceuticals played no role in the conception, analysis, or writing of this manuscript, nor the decision to publish.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Abstract

Objective
To (1) examine the burden of multiple chronic conditions (MCC) in an urban health system, and (2) propose a methodology to identify sub-populations of interest based on diagnosis groups and costs.

Design: Retrospective cross-sectional study.

Setting: Mount Sinai Health System, set in all five boroughs of New York City, USA.

Participants: 192,085 adult (18+) plan members of capitated Medicaid contracts between the Healthfirst managed care organization and the Mount Sinai Health System in the years 2012-2014.

Methods
We classified adults as having 0,1, 2, 3, 4, or 5+ chronic conditions from a list of 69 chronic conditions. After summarizing the demographics, geography, and prevalence of MCC within this population, we then described groups of patients (segments) using a novel methodology: we combinatorially defined 18,768 potential segments of patients by a pair of chronic conditions, a sex, and an age group, and then ranked segments by 1) frequency, 2) cost and 3) ratios of observed to expected frequencies of co-occurring chronic conditions. We then compiled pairs of conditions that occur more frequently together than otherwise expected.

Results
61.5% of the study population suffers from two or more chronic conditions. The most frequent dyad was hypertension and hyperlipidemia (19%**)** and the most frequent triad was diabetes, hypertension and hyperlipidemia (10%). Women aged 50-65 with hypertension and hyperlipidemia were the leading cost segment in the study population. Costs and prevalence of MCC increase with number of conditions and age. The disease dyads associated with the largest observed/expected ratios were pulmonary disease and myocardial infarction. Inter-borough range MCC prevalence was 16%.

Conclusions
In this low-income, urban population, MCC is more prevalent (61%) than nationally (42%), motivating further research and intervention in this population. By identifying potential target populations in an interpretable manner, this segmenting methodology has utility for health services analysts.

Strengths and limitations of this study

Strengths of the study:

- Large, robust dataset of patients with high prevalence of chronic disease
- New descriptive/analytic approach identifies unanticipated overlap of conditions

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

- Methodology applicable to other similar settings, including urban health systems

Weaknesses of the study:

- Cross-sectional data precludes causal analysis
- Use of cost claims data rather than clinical diagnosis

Data sharing statement

Data is available upon request from the corresponding author of the manuscript.

Patient and public involvement section

The study was a retrospective review using administrative claims data. The patients and public were not involved in this study.

Contributorship Statement

SPK conceived of the study. C Hajat advised on technical analysis. UM, C Hunt, and PD completed analyses. AB, EL, DJH, RK, and RF provided technical input to the manuscript. UM wrote the manuscript. SPK, EL, DJH, C Hunt, and C Hajat edited drafts of the manuscript.

Introduction

The management of multiple chronic conditions (MCC, here defined as the association of two or more chronic health conditions) constitutes a formidable clinical and financial challenge. An increasingly large proportion of the United States population lives with MCC, including 42% of adults overall and 81% of those over the age of 65 years [1]. In the US, MCC patients account for more than 70% of all healthcare spending [2]. In patients over 65 years old, costs increase exponentially with each additional chronic condition, suggesting that there are additional costs associated with the complexity or inefficiency of care for MCC. [3–10].

Health systems have responded to these challenges with clinical and financial innovations. Clinical innovations include new models of care coordination, joint clinical guidelines for MCC patients and alternative delivery models which include bundling of services [10–14]. Financial innovations include value-based payments and bundled payment schemes. One growing form of value-based financial transformation involves capitation, where a fixed "budget" for each patient is agreed upon between the payer and the health system. Accordingly, the health system is incentivized to bring costs down while still maintaining a small margin of profit. In this context, a standard methodology to evaluate the potential interactions between conditions could be mutually beneficial. Importantly, risk adjustment generates appropriately large budgets for high-cost and complex patients, and by doing so accounts for changes in severity over time and incentivizes providing coverage to these high-cost individuals. Existing systems of risk adjustment employed by the Centers for Medicaid & Medicare Services (CMS) predict medical and pharmaceutical spending using demographics and diagnosis codes, and are employed in a standardized fashion for

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Medicare Advantage patients. State-managed Medicaid plans can choose to employ any of many different risk adjustment models, some of which are based on the Medicare Advantage models [15].

Especially important in the setting of value-based payment schemes like capitation is the appropriate selection of sub-populations to receive clinical interventions. While increasingly popular nationally, measures targeting patients who are chronically hospitalized (sometimes known as “super-utilizers”) have demonstrated mixed cost savings, in part because of difficulties targeting patients who could benefit from interventions [14,16,17].

While there exist numerous sophisticated statistical methods for segmenting populations of patients - such as random forests, single decision trees, k-means, and hierarchical segment analysis - these methods suffer from limited interpretability, result instability, immense computing overhead and/or tendency for overfitting [18–20]. Rather than relying on complex statistical models that require significant computing overhead, we propose a simple descriptive method that can be applied to any population for whom medical claims are available. Because its requisites are computationally simple, this methodology can be easily scaled to larger populations.

Prior studies of spending and MCC have focused on synergy in spending between conditions, or on a specific slice of a population, or type of spending - for example. on inpatient or outpatient spending, or on those older than 65 [7–9,21–24]. Notably, literature on MCC patterns and trends among younger, lower socioeconomic status, and vulnerable populations remains scarce, despite their carrying a significant share of chronic disease burden and, accordingly, financial risk in value-based schemes [25]. Additionally, under global capitation both inpatient and outpatient costs must be considered together.
In order to develop a methodology that would yield interpretable insights for both clinical interventions and financial incentives, we sought to first iteratively but simply generate many different sub-populations within the study population and then sort them via either clinically meaningful or financially relevant mechanisms. Clinical interventions can be developed from epidemiological information about which conditions are observed more frequently together than expected [26]. We theorized that observed/expected (independent) ratios would reveal groups of patients distinct from those based purely on frequency or cost. Combinations of chronic conditions could have shared risk factors (e.g. hypertension and diabetes), shared etiology (e.g. hypertension and congestive heart failure) or could be independent altogether (e.g. hypertension and arthritis). By contrast, financial interventions can be developed from cost information about which conditions and combinations of conditions occur in the most costly groups of patients. In practical terms, targeting the highest cost combinations of conditions (and therefore segments of patients) could lead to proactive interventions to reduce avoidable or excess utilization.

Accordingly, in this manuscript we (1) develop a descriptive methodology to identify and describe unique segments of MCC patients, and (2) apply the methodology in an urban health system using administrative claims data derived from a population of managed Medicaid patients at the Mount Sinai health system under global capitation -- a low-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

income, urban population unlike those previously studied. We also describe the general cost and geographic characteristics of this population, with potential use in future segmenting applications.

Methods

Segmenting

Segments refer to groups of patients who meet certain disease criteria, demographic criteria, or both. For example, a segment of patients could be defined by a dyad of diseases (i.e. hypertension and hyperlipidemia), an age range (ages 35-50 years), and sex (males). That segment would consist of male patients aged 35-50 years with both hypertension and hyperlipidemia. As described, these segments are not mutually exclusive (i.e. one patient can belong to several segments). We systematically investigated every possible segment of patients defined by a combination of two chronic conditions (among 69), an age group (18-35, 35-50, 50-65, 65+), and sex, yielding 18,768 potential segments. For each of these segments of patients, we computed a number of segment characteristics by which to rank them: total cost attributable to segment, average cost per person in segment, and observed:expected ratio of disease dyads in each segment. The total cost attributable to the patients in each segment was computed using claims provided by the payer. This calculation includes all costs for these patients, not just those attributable to the diseases defining the segment. Segments were also ranked by average cost per person per year of plan enrollment represented in the segment. For each pair of diseases defining a segment, an *observed:expected ratio* was computed by dividing the observed frequency of the pair of diseases in the study population by the expected frequency (multiplying together the individual frequencies of each disease in the pair). We chose a cutoff of 30 segment members as the lower limit for understanding probable outcomes through a pilot program [27].

Chronic Conditions Lists

We completed a review of pre-existing approaches and opted to work with a defined list of 69 chronic condition categories from the Agency for Healthcare Research and Quality (AHRQ) [28–30]. This condition list was chosen because (1) it included the most expansive list developed by a consensus body of physicians, enabling us to detect uncommon combinations of conditions, and (2) it aligns with other federal multiple chronic condition projects.

Data Set and Inclusion Criteria

We used claims data from patients operating under a capitated contract between Mount Sinai Health System and Healthfirst, the largest managed care organization for federal Medicaid funds in New York State. These data include all medical claims from 2012 to 2014 including 6,676,867 claims for 213,091 plan members. This period spans from the first full year of claims following the start of the Mount Sinai-Healthfirst contract to the last year when claims were made with the International Classification of Diseases version 9 (ICD-9). Costs represent paid amounts, not charged amounts.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

We used the Agency for Healthcare Quality and Research (AHRQ) Healthcare Cost and Utilization Project (HCUP) mapping of 4,427 ICD-9 codes to 69 clinically-relevant chronic condition categories. We omitted 2015 data because ICD-10 codes were used inconsistently alongside ICD-9 codes, and the HCUP mapping of ICD-10 codes to chronic condition categories is incomplete. We performed a complete case analysis and excluded participants with missing age or gender. The study was approved through Institutional Review Board of the Icahn School of Medicine at Mount Sinai.

Variables

We studied age, gender, location, chronic condition codes, number of chronic conditions, and total cost of care during the member's plan enrollment. Multiple chronic conditions were studied as dyads and triads. The analysis of different combinations of segment criteria was limited by processing power and computational cost.

Statistical Analyses

The observed frequency of each segment was age-adjusted using the New York State age distribution. Segments were segmented by gender. Estimates were calculated for segments defined by chronic condition codes, gender, age, and total cost of care. Claims were aggregated by patient-year via SQL, and subsequent cleaning, analysis, and plotting was performed with R and Python (code available at https://github.com/usnish/mcc_scripts).

Results

Prevalence of MCC by selected characteristics

61.5% of the study population (61.6% in women, 61.4% in men) lives with two or more chronic conditions, as compared to 42% nationally. **Table 1** displays demographic data of the sample (n = 143,297 patients). Median age was 47 years (25th percentile = 30; 75th percentile = 61), and 54.6% (78,199) were female. We identified the most prevalent combinations of two and three chronic conditions. Each dyad or triad result represents the prevalence of patients with that combination of chronic conditions, including those that also have additional conditions (for example, a patient with hypertension, hyperlipidemia, and diabetes would be counted in a single triad, and also within both the hypertension-hyperlipidemia and hyperlipidemia-diabetes dyads).

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 1. Demographics of Medicaid patients at Mount Sinai Health System belonging to Healthfirst capitated contracts.

	0 Chronic Condition(s)			1 Chronic Condition(s)			2 Chronic Condition(s)			3 Chronic Condition(s)			4 Chronic Condition(s)			5+ Chronic Condition(s)		
	2012	2013	2014	2012	2013	2014	2012	2013	2014	2012	2013	2014	2012	2013	2014	2012	2013	2014
n	10,732	15,092	17,416	9,960	13,544	16,286	7,271	9,887	11,698	5,644	7,741	9,014	4,531	6,125	7,338	15,641	20,984	26,968
Age (mean (sd))	30.88 (11.34)	31.25 (11.69)	31.76 (12.11)	34.74 (13.30)	35.39 (13.87)	35.95 (14.06)	38.92 (14.40)	38.73 (14.56)	38.97 (14.80)	43.08 (15.06)	43.23 (15.19)	42.76 (15.09)	46.91 (15.35)	46.52 (15.02)	46.49 (14.99)	54.45 (14.65)	54.03 (14.42)	54.36 (14.56)
Sex = F (%)	3,678 (34.3)	5,030 (33.3)	5,906 (33.9)	3,487 (35.0)	4,720 (34.8)	5,926 (36.4)	2,506 (34.5)	3,282 (33.2)	4,181 (35.7)	1,872 (33.2)	2,583 (33.4)	3,168 (35.1)	1,581 (34.9)	2,135 (34.9)	2,643 (36.0)	5,393 (34.5)	7,103 (33.8)	9,335 (34.6)
Total Cost (mean (sd))	847.26 (2002.96)	898.89 (2074.48)	860.56 (2006.20)	1,327.86 (2893.45)	1,268.02 (2740.95)	1,266.16 (2904.01)	1,758.23 (3416.84)	1,800.90 (4907.68)	1,777.87 (5717.80)	2,221.77 (5122.45)	2,001.08 (3586.80)	2,093.73 (4167.59)	2,634.04 (5017.66)	2,588.73 (6606.52)	2,606.87 (5742.42)	8,968.28 (19991.33)	8,673.94 (20181.73)	8,415.85 (18111.71)
Total Cost Winsorized (mean (sd))	842.51 (1949.60)	894.49 (2022.10)	858.48 (1973.22)	1,280.61 (2364.27)	1,227.41 (2312.96)	1,216.23 (2288.50)	1,661.55 (2623.90)	1,659.58 (2615.23)	1,622.92 (2551.18)	1,987.38 (2825.16)	1,861.01 (2651.65)	1,927.38 (2738.23)	2,306.30 (2865.16)	2,232.10 (2882.43)	2,304.93 (2888.36)	5,061.43 (4455.64)	4,866.61 (4416.25)	4,945.13 (4352.52)
Top 10 Single Chronic Conditions (%)																		
Allergy, ENT and other upper resp disorders	0 (0)	0 (0)	0 (0)	911 (19.8)	1,109 (17.8)	1,338 (18.5)	1,084 (20.8)	1,430 (19.8)	1,574 (19.0)	1,111 (22.9)	1,360 (20.4)	1,604 (21.1)	1,029 (24.5)	1,276 (22.7)	1,536 (22.6)	4,419 (28.8)	5,614 (27.2)	7,478 (28.3)
Asthma, COPD, other chronic lung disease	0 (0)	0 (0)	0 (0)	654 (14.2)	876 (14.1)	960 (13.3)	758 (14.6)	1,098 (15.2)	1,229 (14.9)	770 (15.9)	965 (14.5)	1,129 (14.8)	729 (17.4)	922 (16.4)	1,098 (16.1)	4,344 (28.3)	5,526 (26.8)	7,079 (26.8)
Obesity	0 (0)	0 (0)	0 (0)	199 (4.3)	314 (5.0)	334 (4.6)	438 (8.4)	633 (8.8)	706 (8.5)	698 (14.4)	965 (14.5)	1,040 (13.7)	856 (20.4)	1,242 (22.0)	1,411 (20.7)	6,482 (42.2)	8,974 (43.5)	11,143 (42.1)
Degenerative eye problem (glauc/eye)	0 (0)	0 (0)	0 (0)	287 (6.2)	418 (6.7)	451 (6.2)	517 (9.9)	679 (9.4)	756 (9.1)	683 (14.1)	909 (13.7)	921 (12.1)	715 (17.1)	955 (17.0)	1,139 (16.7)	5,128 (33.4)	6,810 (33.0)	8,907 (33.7)
Hyperlipidemia	0 (0)	0 (0)	0 (0)	597 (13.0)	846 (13.6)	999 (13.8)	598 (11.5)	827 (11.5)	983 (11.9)	754 (15.5)	926 (13.9)	1,125 (14.8)	839 (20.0)	1,069 (19.0)	1,339 (19.7)	5,531 (36.0)	7,431 (36.0)	9,571 (36.2)
Depression and depressive disorders	0 (0)	0 (0)	0 (0)	476 (10.4)	604 (9.7)	700 (9.7)	1,106 (21.3)	1,448 (20.1)	1,637 (19.8)	1,613 (33.2)	2,159 (32.5)	2,276 (29.9)	1,834 (43.8)	2,442 (43.4)	2,790 (41.0)	10,781 (70.2)	14,320 (69.5)	18,018 (68.1)
Hypertension	0 (0)	0 (0)	0 (0)	438 (9.5)	633 (10.2)	700 (9.7)	1,079 (20.7)	1,604 (22.2)	1,764 (21.3)	1,579 (32.5)	2,246 (33.8)	2,542 (33.4)	1,753 (41.8)	2,492 (44.2)	2,899 (42.6)	9,735 (63.4)	13,393 (65.0)	16,891 (63.8)
Esophageal disorder and GI ulcers	0 (0)	0 (0)	0 (0)	284 (6.2)	444 (7.1)	577 (8.0)	569 (10.9)	913 (12.7)	1,171 (14.2)	717 (14.8)	1,169 (17.6)	1,449 (19.0)	736 (17.6)	1,103 (19.6)	1,536 (22.6)	4,069 (26.5)	5,971 (29.0)	8,502 (32.1)
Malnutrition and F/E cond (not obesity/overweight)-includes disorders of metabolism	0 (0)	0 (0)	0 (0)	564 (12.3)	793 (12.7)	909 (12.6)	799 (15.4)	1,106 (15.3)	1,317 (15.9)	831 (17.1)	1,091 (16.4)	1,384 (18.2)	765 (18.2)	1,025 (18.2)	1,274 (18.7)	3,767 (24.5)	5,199 (25.2)	7,459 (28.2)
Diabetes mellitus	0 (0)	0 (0)	0 (0)	183 (4.0)	197 (3.2)	250 (3.5)	309 (5.9)	405 (5.6)	427 (5.2)	437 (9.0)	592 (8.9)	651 (8.6)	524 (12.5)	724 (12.9)	825 (12.1)	4,965 (32.3)	6,592 (32.0)	8,847 (33.4)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

These overlapping segments of patients, ranked by age-adjusted frequency, are reported in **Table 2.** Of these, 16,044 segments contained at least one patient - with the largest segment containing an average of 4,329 patients per year. The most common dyad was hypertension and hyperlipidemia (19% age adjusted), and the most common triad was hypertension, hyperlipidemia and diabetes **(**10% age adjusted**).**

Table 2. Top segments of two and three chronic conditions, ranked by age-adjusted frequency using overall list of 69 conditions.

Singlet Chronic Condition	**Average Yearly Membership**	**Unadjusted %**	**Age Adjusted %**
Hypertension	20,724	29%	28%
Hyperlipidemia	19,932	28%	26%
Diabetes mellitus	11,801	16%	16%
Degenerative eye problem (glauc/eye)	11,153	16%	15%
Allergy, ENT, and other upper resp disorders	10,938	15%	12%
Dyad Chronic Conditions	**Average Yearly Membership**	**Unadjusted %**	**Age Adjusted %**
Hypertension & Hyperlipidemia	12,808	18%	18%
Hypertension & Diabetes mellitus	8,707	12%	12%
Hyperlipidemia & Diabetes mellitus	8,203	11%	11%
Hypertension & Degenerative eye problem (glauc/eye)	6,332	9%	10%
Hyperlipidemia & Degenerative eye problem (glauc/eye)	6,116	9%	9%
Triad Chronic Conditions	**Average Yearly Membership**	**Unadjusted %**	**Age Adjusted %**
Diabetes mellitus, Hypertension, & Hyperlipidemia	6,778	9%	9%
Hypertension, Degenerative eye problem (glauc/eye), & Hyperlipidemia	4,792	7%	7%
Osteoarthritis, Hypertension, & Hyperlipidemia	4,087	6%	6%
Esophageal disorder and GI ulcers, Hypertension, &Hyperlipidemia	3,828	5%	5%
Diabetes mellitus, Hypertension, & Degenerative eye problem (glauc/eye)	3,727	5%	5%

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Healthcare expenditures

Figure 1 shows healthcare expenditure among patients with different numbers of chronic conditions. Patients with missing demographic data have been excluded (12.4% of all patients). Costs increase by over 40% with each additional condition, as does the patient-to-patient variance in yearly cost.

Segments by Age, Sex, Costs

Supplementary Table 1 indicates the top segments and characteristics by chronic conditions by age and gender using the classification outlined in the Methods section. The lists are presented by top 10 highest frequency (3A), top 10 dyads with the highest costs and at least an average of 30 members per year (3B) and by top 10 dyads with the highest cost and at least an average of 1,000 members per year.

Adjusting the minimum threshold number of patients constituting a segment alters the kinds of diseases represented. For example, if the minimum size of a segment is 30 members, the highest cost segment becomes males age 35-50 with “Anemia and other non-cancer hematological disorder” & “conduction disorder or cardiac dysrhythmia”. However, if this threshold is raised to 1,000 members, the highest cost segment becomes females age 50-65 with “Hypertension & Coronary atherosclerosis”. In general, smaller segments (>30-1000 members) tended to be higher in average individual cost, but lower in total cost, than the larger segments (>1,000 members).

Table 3 shows the top 10 segments including age (4 categories) and gender (male/female). This table indicates dyads of chronic conditions organized by observed/expected ratios. This data reveal a different relationship of chronic conditions to one another than the frequency and cost tables. By selecting segments of patients with at least 30 included, we demonstrate relationships between unexpected diseases in small yet high-cost groups of patients. For example, paralysis and immunity disorders occur at 16.6 times the expected rate, accounting for an average yearly cost of $81,414. By selecting segments of patients with at least 1,000, we demonstrate relationships that are more commonly observed (and more frequently expected), such as between peripheral atherosclerosis and coronary atherosclerosis, or between anxiety disorders and bipolar disorder.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 3. Observed/Expected ratios of chronic conditions among common (A) and uncommon segments (B)

Top 10 Dyads by Observed / Expected rate with at least an average yearly membership of 30 members or more

Dyad	Unadjusted Frequency	Adjusted Frequency	Adjustment Magnitude	Expected Frequency	Observed / Expected	Average Yearly Cost	Total Cost	Total Members
Acute myocardial infarction & Pulmonary heart disease	0.001	0.001	0	2.80E-05	35.7	$89,321	$11,790,348	132
Thrombosis and Embolism & Non-thrombotic, non-athlerosclerotic vascular disease	0.001	0.001	0	4.20E-05	23.8	$68,541	$9,184,538	134
Pulmonary heart disease & Congestive heart failure	0.003	0.004	0.001	0.000175	22.9	$56,526	$38,098,314	674
Paralysis & Epilepsy	0.001	0.001	0	4.80E-05	20.8	$52,895	$9,732,621	184
Acute myocardial infarction & 'Cardiomyopathy and Structual Heart Disease	0.001	0.002	0.001	0.0001	20.0	$66,547	$21,095,470	317
Acute myocardial infarction & Congestive heart failure	0.002	0.002	0	0.0001	20.0	$66,271	$24,453,854	369
Congenital Heart Disease & Heart valve disorder	0.01	0.01	0	0.000546	18.3	$11,172	$22,979,895	2,057
Pulmonary heart disease & 'Cardiomyopathy and Structual Heart Disease	0.003	0.003	0	0.000175	17.1	$55,752	$34,510,492	619
Paralysis & Organic brain problem (dementia)	0.001	0.001	0	6.00E-05	16.7	$60,838	$6,935,557	114
Paralysis & Immunity disorder	0	0.001	0.001	6.00E-05	16.7	$81,415	$8,711,389	107

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Top 10 Dyads by Observed / Expected rate with at least an average yearly membership of 1,000 members or more

Dyad	Unadjusted Frequency	Adjusted Frequency	Adjustment Magnitude	Expected Frequency	Observed / Expected	Average Yearly Cost	Total Cost	Total Members
Heart valve disorder & Coronary atherosclerosis	0.014	0.016	0.002	0.002535	6.31	$20,896	$64,547,753	3,089
Conduction disorder or cardiac dysrhythmia & 'Coronary atherosclerosis	0.017	0.02	0.003	0.003185	6.28	$26,595	$97,260,685	3,657
Cerebrovascular Disease & Coronary atherosclerosis	0.014	0.017	0.003	0.00286	5.94	$23,622	$72,803,180	3,082
Peripheral atherosclerosis & 'Coronary atherosclerosis	0.017	0.02	0.003	0.00351	5.70	$20,381	$75,512,538	3,705
Anxiety disorders & 'Depression and depressive disorders	0.042	0.033	-0.009	0.006365	5.18	$10,143	$92,526,384	9,122
Depression and depressive disorders & Bipolar disorder	0.021	0.016	-0.005	0.003135	5.10	$11,218	$50,471,365	4,499
Anxiety disorders & Bipolar disorder	0.015	0.011	-0.004	0.002211	4.98	$11,539	$36,800,083	3,189
Cerebrovascular Disease & Other central and peripheral nervous system disorders	0.017	0.018	0.001	0.004004	4.50	$23,374	$86,954,477	3,720
Peripheral atherosclerosis & Other central and peripheral nervous system disorders	0.022	0.022	0	0.004914	4.48	$17,088	$81,155,040	4,749
Other central and peripheral nervous system disorders & Back problem	0.025	0.022	-0.003	0.005005	4.40	$13,315	$72,770,548	5,465

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Age, Spatial distribution and rising risk for patients with multiple chronic conditions

Figure 2 shows frequency of multiple chronic conditions as a function of age across the 5 counties in New York City. A 50% prevalence of MCC is seen at age 30-34 in the Bronx, a lower-income borough of the city, whereas in Brooklyn at in the same 30-34 age-group, the prevalence is only 34%**.**

Discussion

In this paper, we argue that this simple descriptive segmenting methodology has utility for resource planning, care coordination, and care delivery. This methodology would be especially useful in the context of public and private benefits schemes focused on low-income populations.

We find that 61.5% of our population lives with two or more chronic conditions as compared to 42% nationally, motivating efforts to build MCC interventions and tools in the Medicaid population [2]. Using an established list of conditions, we found that total costs increase with each condition added, consistent with findings from other research groups [31–38]. We also found that the most frequent dyad of co-occurring chronic conditions was hypertension and hyperlipidemia (19% age-adjusted) and the most frequent triad was diabetes, hypertension and hyperlipidemia (10% age-adjusted), each in turn more frequent in our study population than nationally (13.6%, as estimated from NHANES in 2010, and 6.3%, from NHANES in 2012) [39,40]. This is a striking finding, considering that the NHANES cohort includes a larger proportion of older adults than our study. As NHANES includes fixed sample-size targets and weighting to generate a national sample of households that is representative of the US adult population, the median age at the time of these studies was 37.2, significantly older than the median of 26 in our dataset. This age discrepancy could be due to two reasons: (1) As adults who are dual-eligible for Medicaid and Medicare are often re-directed to managed Medicare contracts, our study population under-represents adults over 65. (2) Studies of chronic conditions in adults using NHANES tend to utilize a minimum age of 20, as people aged 19 or younger are categorized as ‘youth’; compared to the age cutoff of 17 or younger in our study population [5,38].

Women aged 50-65 with hypertension and hyperlipidemia were the leading cost segment in the health system for dyads. Overall, women age 50-65 and hypertension, osteoarthritis, hyperlipidemia were the leading triad in terms of prevalence and cost. The most significant observed/expected ratio dyads were pulmonary disease and myocardial infarction. We provided various approaches to grouping these chronic conditions in service of broader research objectives to identify conditions that drive multiplicative, rather than additive, health or cost burdens.

The O/E approach provides a clinically oriented view of examining which conditions occur disproportionately together. For example, we find that in our study population, anemia, pulmonary heart disease, congestive heart failure and conduction disorders occur together more frequently than expected. We also observe that patients’ costs balloon when they have these conditions. This would suggest an area where healthcare systems need to focus – screening, dedicated counseling, resources and research dollars. For instance, by targeting patients with conditions like anemia and pulmonary heart disease that do not appear to be physiologically related, care managers can minimize fractures in care. If taken together with our finding that MCC burden differs by locale (**Figure 2)** health systems should elect to co-locate specialty clinics, share clinical teams, and develop joint management protocols for these conditions. While these kinds of innovations have been prototyped around episodic procedural care, such as knee and hip replacements, they have yet to be adopted in managing MCC [13,41,42]. Meanwhile, patients with multiple chronic conditions are already requesting these changes [43]. Importantly, this approach yields specific chronic disease targets beyond the most frequent conditions.

Conditions like anemia and pulmonary heart disease are not currently considered among the interaction terms included in existing CMS models (which focus instead on predicting indicators of severe disease like sepsis, pulmonary embolism, or seizure disorders), but may be more locally appropriate measures of disease severity or spending in this population. Further validation would be required of these novel disease interactions in a larger or different sample population.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

At the same time, the sorting of segments by highest cost and frequency provides a simple view of groups where minor interventions could result in larger-scale cost-savings, particularly for health systems facing value-based financing schemes. Addressing the top segments of patients with bundled financial incentives could supplement the clinical innovations described above. Indeed, recent analyses of the Medicare Shared Savings plan have found that a significant proportion of savings were derived from incremental cost interventions that applied to large swathes of the insured population [44].

It is clear that the threshold itself - small, medium or large - for the volume of patients to analyze can be modified with effect on the resultant segments. While senior executives and health services analysts in population health may be interested in overall patterns, costs and adjusted risk of co-morbidity, specialty service lines may be focused more on tailored, smaller patient segments with unique disease patterns requiring integrated care. For example, the development of a value-based healthcare program in the US Navy involved the creation of integrated practice units to treat low back pain and osteoarthritis [45]. Our analysis across multiple thresholds animates how the thresholds can affect the resultant patterns produced.

Notably, these results differ from a separate analysis by our research team using a list of 12 chronic conditions in the Behavioral Risk Factor Surveillance Survey conducted by the Centers for Disease Control. In this work, we found that from 2011-2016, 50.6% of adults in New York State had two or more chronic conditions. The most prevalent dyads we identified were hypertension and high cholesterol (17% and most prevalent triad was hypertension, high cholesterol and arthritis (4.5%). Prevalence of MCC in NYC neighborhoods ranged from 33.5 to 60.6% [46].

Our findings apply not only to the reform of existing programs for low-income and vulnerable populations, but also the design of novel ones, in the Mount Sinai system and beyond. For example, Mount Sinai offers Healthfirst (and other) patients who require inpatient-level care an alternative: a Hospitalization-at-Home (HaH) program in lieu of inpatient admission [47,48]. Evaluation to date demonstrates that this HaH program delivers superior patient outcomes (including shorter length of stay) and greater patient satisfaction than in-hospital care, though costs have not yet been compared [47]. The HaH program focused on only nine diagnoses at its founding in 2014, but has since expanded in size and breadth of care across multiple New York hospitals, treating myriad other conditions across eight domains of care, such as post-surgical care, palliative care, and sub-acute rehabilitation, among others [48]. Rapid and timely data on the prevalence and overlap of these (largely chronic) diseases and their risk factors will be instrumental to the program's ongoing cost-effective scale-up. Such data could prove even more valuable in low- and middle- income countries, where the burden of chronic disease is rapidly expanding, but models for the integrated care of more than one chronic condition are few and small in scope [49].

The limitations of our proposed approach include the following: (1) the use of health insurance claims itself limits the epidemiologic utility of the analyses. Claims are effectively billing receipts and therefore have limited reliability in reporting disease states [50]. Additionally, we did not control for variations in coding by center or physician. We plan to integrate these claims data with EMR data going forward in order to retrieve higher quality epidemiological insights. (2) Our analysis is limited by the study period. Data from 2012-2014 is likely not recent enough to enact present-day interventions in a health system -- this is largely because the mapping of ICD-10 codes to chronic condition categories has not been finalized, with some remaining discontinuities between ICD-9 and ICD-10-based classifications, limiting our ability to use data from 2015 onwards. We plan to include more recent data once the mapping is completed, as well as prototype this methodology using the CMS Chronic Condition Warehouse algorithm, which functions with ICD-10 codes but includes fewer conditions (27 rather than 69) [51]. Additionally, we did not examine epidemiologic trends through time, as a period as short as 3 years is not long

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

enough to elucidate relationships between diseases that share etiology (i.e. hypertension, stroke). (3) The generalizability of our analysis is limited by the geospatial distribution of patients in the study population -- because provider attribution is accomplished regionally, our data set includes the subset of New York City patients who live near Mount Sinai practices. As a result, in the current data set, the majority of patients are located in just 10 of 176 ZIP codes. Future analyses using a data set such as an all-payer claims database would allow researchers to define segments by region and ZIP code. Accordingly, this study population of managed Medicaid patients is not necessarily representative of the Medicaid or U.S. population at-large, or the fee-for-service Medicaid population served by Mount Sinai. (4) We did not include pharmacy claims in our analysis, which will result in an underestimation of spending. This underestimation is most significant regarding conditions that require expensive medications (i.e. high-cost injectables for HIV and hepatitis C). However, we also note that risk adjustment methodologies employed by Medicaid Advantage and State Medicaid programs tend to predict spending on pharmaceuticals separate from other costs. [15] (5) Lastly, a significant portion (12.4%) of our study population was excluded on account of missing demographic data, introducing some bias into which segments of patients were highlighted. Any more pragmatic application of this methodology would also require an approach to patients with missing data.

Taken together, these analyses have implications for health systems, financiers, and researchers working to address MCC, and provide a common methodology for targeting populations for financial and clinical intervention. Most notably, this tool yields a simple, transparent methodology for selecting coherent, clearly-defined populations of patients for intervention, and can be applied to any commercial claims dataset. With application in the right contexts, this methodology could help improve the selection strategy of super-utilizer clinics and other clinical innovations, yielding further advancements in our health systems' management of chronic conditions. Payors may increasingly rely on interaction of diseases to help identify appropriate levels of reimbursement based on predicted risk of hospitalization or mortality for patients. Ultimately, however, more research is needed to evaluate this methodology's utility in business scenarios, and applicability to different sizes and kinds of patient populations.

Figure Legends / Captions

Table 1. Demographics of Medicaid patients at Mount Sinai Health System belonging to Healthfirst capitated contracts.
Table 2. Top segments of two and three chronic conditions, ranked by age-adjusted frequency using overall list of 69 conditions.
Figure 1. Distribution of individual annual healthcare expenditures as a function of number of chronic conditions.
Supplementary Table 1. Top segments of chronic conditions by age and gender segments. (A) largest segments by member count, (B) most costly segments 30 people or greater, (C) most costly segments 1000 people or greater.
Table 3. Observed/Expected ratios of chronic conditions among common (A) and uncommon segments (B)
Figure 2. Frequency of multiple chronic conditions by age across selected boroughs of New York City. 50% prevalence of multiple chronic conditions seen at age 30-34 for all boroughs except for Brooklyn that reaches 50% at age 35-39. Disparities between boroughs observed.

References

1 Buttorff C, Ruder T, Bauman M. Multiple Chronic Conditions in the United States. RAND Corporation 2017. doi:10.7249/tl221

2 Gerteis J, Izrael D, Deitz D, *et al.* Multiple Chronic Conditions Chartbook. Rockville, MD: Agency for Healthcare Research and Quality; 2014. https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/prevention-chronic-care/decision/mcc/mccchartbook.pdf

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

3 Hoffman C. Persons with chronic conditions. Their prevalence and costs. *JAMA* 1996;**276**:1473–9.

4 Schneider KM, O'Donnell BE, Dean D. Prevalence of multiple chronic conditions in the United States' Medicare population. *Health Qual Life Outcomes* 2009;**7**:82.

5 Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med* 2002;**162**:2269–76.

6 Lehnert T, Heider D, Leicht H, *et al.* Review: health care utilization and costs of elderly persons with multiple chronic conditions. *Med Care Res Rev* 2011;**68**:387–420.

7 Cortaredona S, Ventelou B. The extra cost of comorbidity: multiple illnesses and the economic burden of non-communicable diseases. *BMC Med* 2017;**15**:216.

8 Brilleman SL, Purdy S, Salisbury C, *et al.* Implications of comorbidity for primary care costs in the UK: a retrospective observational study. *Br J Gen Pract* 2013;**63**:e274–82.

9 He Z, Bian J, Carretta HJ, *et al.* Prevalence of Multiple Chronic Conditions Among Older Adults in Florida and the United States: Comparative Analysis of the OneFlorida Data Trust and National Inpatient Sample. *J Med Internet Res* 2018;**20**:e137.

10 Hajat C, Stein E. The global burden of multiple chronic conditions: A narrative review. *Prev Med Rep* 2018;**12**:284–93.

11 Hajat C, Kishore SP. The case for a global focus on multiple chronic conditions. *BMJ Glob Health* 2018;**3**:e000874.

12 Hajat C, Stein E, Yach D. Multiple chronic conditions: the global state. https:// lnkd. in/ etb9eAJ

13 Keswani A, Koenig KM, Bozic KJ. Value-based Healthcare: Part 1-Designing and Implementing Integrated Practice Units for the Management of Musculoskeletal Disease. *Clin Orthop Relat Res* 2016;**474**:2100–3.

14 Berkowitz SA, Parashuram S, Rowan K, *et al.* Association of a Care Coordination Model With Health Care Costs and Utilization: The Johns Hopkins Community Health Partnership (J-CHiP). *JAMA Netw Open* 2018;**1**:e184273–e184273.

15 March 31, 2016, HHS-Operated Risk Adjustment Methodology Meeting Discussion Paper. Centers for Medicare & Medicaid Services, Center for Consumer Information & Insurance Oversight https://www.cms.gov/CCIIO/Resources/Forms-Reports-and-Other-Resources/Downloads/RA-March-31-White-Paper-032416.pdf

16 Kanzaria HK, Hoffman JR. Hot-Spotters Aren't 'The Problem'...But They Are Emblematic of the Failure of U.S. Healthcare. *J Gen Intern Med* 2017;**32**:6–8.

17 Lee NS, Whitman N, Vakharia N, *et al.* High-Cost Patients: Hot-Spotters Don't Explain the Half of It. *J Gen Intern Med* 2017;**32**:28–34.

18 Breiman L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Stat Sci* 2001;**16**:199–231.

19 Nicholson K, Bauer M, Terry A, *et al.* The Multimorbidity Cluster Analysis Tool: Identifying Combinations and Permutations of Multiple Chronic Diseases Using a Record-Level Computational Analysis. *J Innov Health Inform* 2017;**24**:962.

20 Ng SK, Tawiah R, Sawyer M, *et al.* Patterns of multimorbid health conditions: a systematic review of analytical methods and comparison analysis. *Int J Epidemiol* 2018;**47**:1687–704.

21 Crystal S, Johnson RW, Harman J, *et al.* Out-of-pocket health care costs among older Americans. *J Gerontol B Psychol Sci Soc Sci* 2000;**55**:S51–62.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

22 Fishman P, Von Korff M, Lozano P, *et al.* Chronic care costs in managed care. *Health Aff* 1997;**16**:239–47.

23 Hoffman C. Persons with chronic conditions. Their prevalence and costs. *JAMA* 1996;**276**:1473–9.

24 Moxey ED, O'Connor JP, Novielli KD, *et al.* Prescription drug use in the elderly: a descriptive analysis. *Health Care Financ Rev* 2003;**24**:127–41.

25 Shaw KM, Theis KA, Self-Brown S, *et al.* Chronic Disease Disparities by County Economic Status and Metropolitan Classification, Behavioral Risk Factor Surveillance System, 2013. *Prev Chronic Dis* 2016;**13**:E119.

26 Schäfer I, Kaduszkiewicz H, Wagner H-O, *et al.* Reducing complexity: a visualisation of multimorbidity by combining disease clusters and triads. *BMC Public Health* 2014;**14**:1285.

27 Hertzog MA. Considerations in determining sample size for pilot studies. *Res Nurs Health* 2008;**31**:180–91.

28 Harrison C, Britt H, Miller G, *et al.* Examining different measures of multimorbidity, using a large prospective cross-sectional study in Australian general practice. *BMJ Open* 2014;**4**:e004694.

29 Fortin M, Stewart M, Poitras M-E, *et al.* A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Ann Fam Med* 2012;**10**:142–51.

30 Le Reste JY, Nabbe P, Manceau B, *et al.* The European General Practice Research Network presents a comprehensive definition of multimorbidity in family medicine and long term care, following a systematic review of relevant literature. *J Am Med Dir Assoc* 2013;**14**:319–25.

31 Crystal S, Johnson RW, Harman J, *et al.* Out-of-pocket health care costs among older Americans. *J Gerontol B Psychol Sci Soc Sci* 2000;**55**:S51–62.

32 Fishman P, Von Korff M, Lozano P, *et al.* Chronic care costs in managed care. *Health Aff* 1997;**16**:239–47.

33 Hoffman C. Persons with chronic conditions. Their prevalence and costs. *JAMA* 1996;**276**:1473–9.

34 Moxey ED, O'Connor JP, Novielli KD, *et al.* Prescription drug use in the elderly: a descriptive analysis. *Health Care Financ Rev* 2003;**24**:127–41.

35 Rice DP, LaPlante MP. Medical expenditures for disability and disabling comorbidity. *Am J Public Health* 1992;**82**:739–41.

36 Schneider KM, O'Donnell BE, Dean D. Prevalence of multiple chronic conditions in the United States' Medicare population. *Health Qual Life Outcomes* 2009;**7**:82.

37 Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med* 2002;**162**:2269–76.

38 Lehnert T, Heider D, Leicht H, *et al.* Review: health care utilization and costs of elderly persons with multiple chronic conditions. *Med Care Res Rev* 2011;**68**:387–420.

39 Song Y, Liu X, Zhu X, *et al.* Increasing trend of diabetes combined with hypertension or hypercholesterolemia: NHANES data analysis 1999-2012. *Sci Rep* 2016;**6**:36093.

40 Egan BM, Li J, Qanungo S, *et al.* Blood pressure and cholesterol control in hypertensive hypercholesterolemic patients: national health and nutrition examination surveys 1988-2010. *Circulation* 2013;**128**:29–41.

41 Bleich SN, Sherrod C, Chiang A, *et al.* Systematic Review of Programs Treating High-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Need and High-Cost People With Multiple Chronic Diseases or Disabilities in the United States, 2008-2014. *Prev Chronic Dis* 2015;**12**:E197.

42 Bandara S, Lynch G, Cooke C, *et al.* Using Care Bundles to Improve Surgical Outcomes and Reduce Variation in Care for Fragility Hip Fracture Patients. *Geriatr Orthop Surg Rehabil* 2017;**8**:104–8.

43 Bayliss EA, Edwards AE, Steiner JF, *et al.* Processes of care desired by elderly patients with multimorbidities. *Fam Pract* 2008;**25**:287–93.

44 McWilliams JM, Chernew ME, Landon BE. Medicare ACO Program Savings Not Tied To Preventable Hospitalizations Or Concentrated Among High-Risk Patients. *Health Aff* 2017;**36**:2085–93.

45 Hernandez A, Kaplan RS, Witkowski ML, et al. Navy Medicine Introduces Value-Based Health Care. Health Aff 2019;38:1393–400.

46 Newman D, Levine E, Kishore SP. Prevalence of multiple chronic conditions in New York State, 2011-2016. *PLoS One* 2019;**14**:e0211965.

47 Federman AD, Soones T, DeCherrie LV, *et al.* Association of a Bundled Hospital-at-Home and 30-Day Postacute Transitional Care Program With Clinical Outcomes and Patient Experiences. *JAMA Intern Med* 2018;**178**:1033–40.

48 DeCherrie LV, Wajnberg A, Soones T, *et al.* Hospital at Home-Plus: A Platform of Facility-Based Care. *J Am Geriatr Soc* 2019;**67**:596–602.

49 Joshi R, Alim M, Kengne AP, *et al.* Task shifting for non-communicable disease management in low and middle income countries--a systematic review. *PLoS One* 2014;**9**:e103754.

50 Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997;**127**:666–74.

51 HCUP CCS. Healthcare Cost and Utilization Project (HCUP). 2017.www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp (accessed 27 Oct 2017).

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

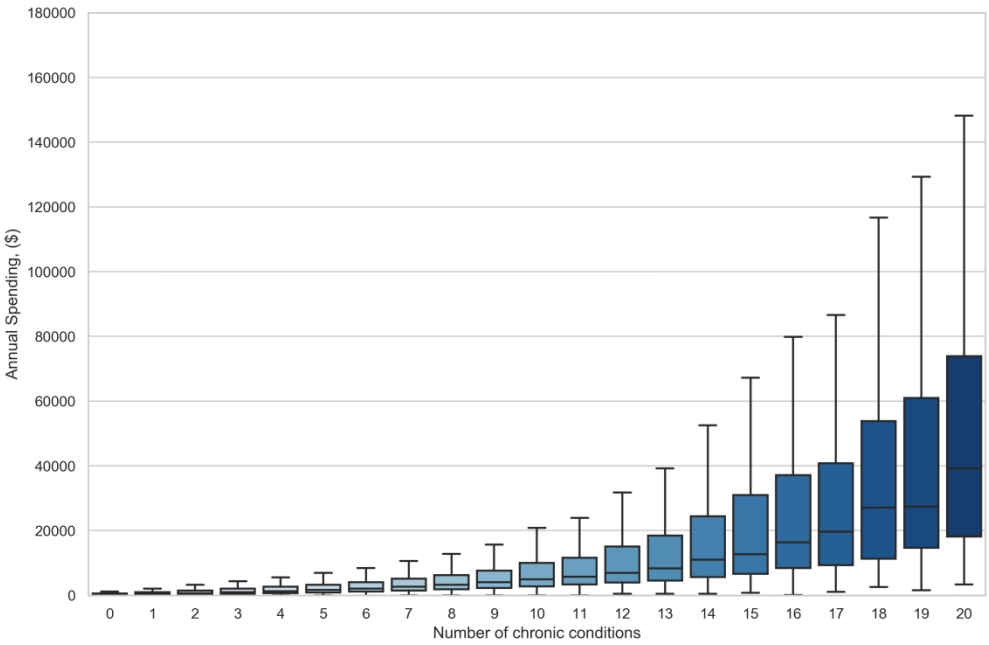


Figure 1 -- reuploaded as high-resolution Tiff

263x171mm (300 x 300 DPI)

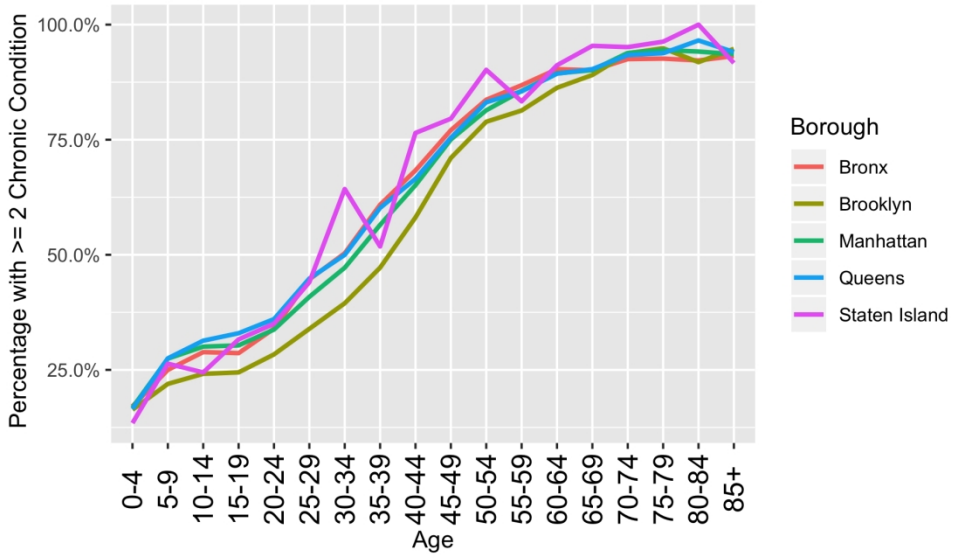


Figure 2 -- reuploaded as high-resolution Tiff; edited typographical error in y-axis title

686x392mm (72 x 72 DPI)

Supplementary Table 1

Top 10 clusters by frequency

Gender	Age	Condition 1	Condition 2	Total Attributable Cost	Average Yearly Cost	Average Yearly Membership
F	50-65	Hypertension	Hyperlipidemia	$93,122,272	$7,172	4,329
F	50-65	Hypertension	Diabetes Mellitus	$72,878,330	$8,557	2,839
F	50-65	Hyperlipidemia	Diabetes Mellitus	$65,165,290	$8,143	2,668
M	50-65	Hypertension	Hyperlipidemia	$61,719,638	$7,948	2,589
F	50-65	Hypertension	Degenerative eye problem (glauc/eye)	$54,012,310	$8,240	2,185
F	50-65	Hypertension	Osteoarthritis	$66,447,600	$10,166	2,179
F	50-65	Hyperlipidemia	Degenerative eye problem (glauc/eye)	$49,533,370	$7,674	2,152
F	50-65	Hyperlipidemia	Osteoarthritis	$56,171,247	$9,295	2,014
F	50-65	Hypertension	Esophageal disorder and GI ulcers	$60,965,767	$10,297	1,974
F	50-65	Hyperlipidemia	Esophageal disorder and GI ulcers	$53,619,011	$9,194	1,944

Top 10 clusters by average yearly cost with 30 total member counts or more

Gender	Age	Condition 1	Condition 2	Total Attributable Cost	Average Yearly Cost	Average Yearly Membership
M	35-50	Anemia and other non-cancer heme disorders	Conduction disorder or cardiac dysrhythmia	$8,390,439	$90,220	31
F	50-65	Pulmonary heart disease	Anemia and other non-cancer heme disorders	$7,542,310	$83,803	30
M	35-50	Congestive heart failure	Malnutrition and F/E cond (not obesity/overweight) -includes disorders of metabolism	$7,068,459	$77,675	30
M	50-65	Conduction disorder or cardiac dysrhythmia	Immunity disorder	$9,800,142	$76,564	43
M	35-50	Other central and peripheral nervous system disorders	Immunity disorder	$6,917,900	$73,595	31

For peer review only

F	50-65	Congestive heart failure	Anemia and other non-cancer heme disorders	$14,346,180	$70,671	68
F	50-65	Congestive heart failure	Chronic skin ulcer	$7,852,354	$69,490	38
M	35-50	Anemia and other non-cancer heme disorders	Kidney and Vesicoureteral Disorders (excluding renal failure)	$6,174,594	$68,607	30
M	65+	Cardiomyopathy and Structural Heart Disease	Anemia and other non-cancer heme disorders	$9,505,337	$67,414	47
M	35-50	Congestive heart failure	Conduction disorder or cardiac dysrhythmia	$7,077,370	$67,404	35

Top 10 clusters by average yearly cost with 1,000 member counts or more

Gender	Age	Condition 1	Condition 2	Total Attributable Cost	Average Yearly Cost	Average Yearly Membership
F	50-65	Hypertension	Coronary atherosclerosis	$45,703,351	$14,486	1,052
F	50-65	Osteoarthritis	Other central and peripheral nervous system disorders	$43,931,227	$14,013	1,045
F	50-65	Hypertension	Other central and peripheral nervous system disorders	$61,441,249	$13,433	1,525
F	50-65	Hypertension	Asthma, COPD, other chronic lung disease	$57,028,007	$12,193	1,559
F	50-65	Hyperlipidemia	Other central and peripheral nervous system disorders	$50,790,724	$12,105	1,399
M	65+	Hypertension	Hyperlipidemia	$40,441,616	$11,933	1,130
F	50-65	Esophageal disorder and GI ulcers	Diabetes mellitus	$39,891,768	$11,827	1,124
F	50-65	Hyperlipidemia	Asthma, COPD, other chronic lung disease	$45,652,105	$11,757	1,294

For peer review only

F	50-65	Esophageal disorder and GI ulcers	Osteoarthritis	$42,200,192	$11,745	1,198
F	50-65	Diabetes mellitus	Osteoarthritis	$43,479,202	$11,591	1,250

For peer review only

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1	Multiple chronic conditions at a major urban health system: a descriptive analysis of frequencies, costs and patterns
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2	Lines 1-39
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4,5	Lines 96-145
Objectives	3	State specific objectives, including any prespecified hypotheses	5	Lines 162-168
Methods				
Study design	4	Present key elements of study design early in the paper	5	Lines 162-168
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6	Lines 202-208
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6	Lines 202-216
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	N/A	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6	Lines 219-222
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6	Lines 202-216

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Bias	9	Describe any efforts to address potential sources of bias	10	Line 375, 385
Study size	10	Explain how the study size was arrived at	6	Line 205

Continued on next page

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5,6	Line 173-191
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7	Line 225-231
		(*b*) Describe any methods used to examine subgroups and interactions	6	Line 187-188
		(*c*) Explain how missing data were addressed	6	Line 214
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	7	Line 225
		(*e*) Describe any sensitivity analyses	N/A	
Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7	Line 239
		(b) Give reasons for non-participation at each stage	7	Line 255
		(c) Consider use of a flow diagram	N/A	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7	Line 236-244
		(b) Indicate number of participants with missing data for each variable of interest	7	Line 255
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	N/A	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	N/A	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	N/A	
		Cross-sectional study—Report numbers of outcome events or summary measures	7,8	Table 1,2,3
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included		
		(*b*) Report category boundaries when continuous variables were categorized	6	Line 180
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A	

Continued on next page

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8	Line 290
Discussion				
Key results	18	Summarise key results with reference to study objectives	8	Line 303-341
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9	Line 357-386
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10	Line 388-397
Generalisability	21	Discuss the generalisability (external validity) of the study results	10	Line 373
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	1	This work was supported by Teva Pharmaceuticals for the Multiple Chronic Conditions Initiative with the Arnhold Institute for Global Health. Dr. Heller also reports support from the NIH Fogarty International Center (R21 TW010452-01).

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml