

Determining Compound Comorbidities for Heart Failure from Hospital Discharge Data

Indra Neil Sarkar, PhD, MLIS^{1,2,3} and Elizabeth S. Chen, PhD^{1,2,4}

¹Center for Clinical and Translational Science; ²Department of Microbiology and Molecular Genetics; ³Department of Computer Science; and, ⁴Department of Medicine
University of Vermont, Burlington, VT

Abstract

The course of treatment and ultimate clinical outcome often depends on a holistic understanding of the patient status, which often requires cataloguing of concomitant conditions (“comorbidities”). A number of approaches have been developed to quantify the effect of comorbidities (e.g., the Charlson Comorbidity Index); however, reported metrics have been based on pair-wise analyses of co-occurring conditions. This study explored the potential to develop “compound co-morbidities” (CCMs) as a knowledge construct to represent multiple comorbidities, which accommodates for relative prevalence, statistical significance, and rate of increased cost. In the context of congestive heart failure, which is a leading cause for hospital admissions nationally (particularly for the elderly), CCMs were developed and analyzed based on hospital discharge data for an entire state population (Vermont). The results suggest that CCMs may be a valuable construct for characterizing complex co-morbidity relationships that may not be captured using conventional pair-wise approaches.

Introduction

Co-occurring conditions (“comorbidities”) in chronically ill individuals can have a significant impact on the development of clinical plans as well as directly impact the ultimate clinical outcome¹⁻⁵. A number of measures have been developed for assessing the importance of co-morbid conditions in light of available data⁵⁻⁷. Most of these measures originate from review of medical charts, and many of them have since been adapted to leverage available administrative data^{8,9}. Regardless of which measure is chosen, comorbidity measures have been shown to reliably predict potential health care costs and mortality¹⁰. To date, comorbidity measures have been built on the study of pair-wise relationships between potential comorbidities and then assembled into risk scores that can be used to predict outcomes.

It has been suggested that the aging population, as well as improved medical care, have resulted in a higher prevalence of multiple comorbidities¹. Furthermore, there has been minimal effort in the development of centralized resources that can better enable the study of concurrent illnesses¹. Amidst the continued popularity of well-proven and reliable comorbidity indices (e.g., the Charlson Comorbidity Index [CCI¹¹]), there is an apparent need for approaches that would enable the analysis of co-morbidity of concurrent conditions^{1,3,4,12}.

Congestive Heart Failure (HF) is reportedly the leading diagnosis of hospital discharges in elderly patients around the globe¹³⁻¹⁵. It is very common for HF patients to have multiple comorbidities, which can have a very direct effect on the course of treatment, clinical outcome, and potential costs^{14,16,17}. The top conditions commonly associated with HF include Chronic Obstructive Pulmonary Disease (COPD), Chronic or Acute Renal Failure, Cerebrovascular Accident, and Dementia. Within the United States, approximately 40% of HF patients have five or more comorbidities, accounting for 80% of HF related hospital stays¹⁸.

Chart review is arguably the most reliable means to gather information about patients; however, this can often be a laborious and time-consuming process when considering the practical issues of data abstraction. There has been previous work in developing automated techniques to extract co-morbidities in alignment with existing comorbidity indices (e.g., for the CCI¹⁹). Administrative data sets have been shown to be a valuable proxy for studying the effect of comorbidities and their relationship to costs and general clinical outcomes^{8,9,20}. Most approaches utilizing administrative data sets make use of encoded claims data. Although it is known that administrative data sets might very well underestimate the full clinical profile of a patient cohort²¹, it does offer the substrate to at least get aggregate perceptions that can be used to guide subsequent detailed inquiries.

In this paper, we propose a new knowledge structure, compound comorbidities (CCMs), for studying multiple, simultaneous comorbidities. CCMs were developed for HF patients from hospital discharge data for the state of Vermont across three years (2007, 2008, and 2009). The results suggest that CCMs may be used to better understand complex conditions from three perspectives: (1) Prevalence; (2) Importance; and, (3) Cost.

Methods

The goal of this study was to develop a knowledge construct (Compound Comorbidities [CCMs]) for analyzing multiple comorbidities simultaneously. All processing was done using the Ruby scripting language. Hospital discharge data for exploring CCM aspects were acquired from the Vermont Uniform Hospital Discharge Data Set (VUHDDS²²), which contains data collected under state statute (18 V.S.A. §§ 9410, 9456 and 9457) from all hospitals in the state of Vermont. Focusing on patients with a diagnosis of heart failure, CCMs were generated and analyzed relative to three metrics to quantify the prevalence of a CCM, its importance, and its relative cost.

The three most recent available years (2007, 2008, and 2009) of the VUHDDS public use data sets for both inpatient and outpatient visits were downloaded for local processing. For each patient record in VUHDDS, up to 20 ICD-9-CM encoded diagnoses are available. Using a Ruby script, each set of ICD-9-CM codes were grouped into clinical categories based on the 2012 version of the Clinical Classifications Software (CCS; which is part of the Healthcare Cost and Utilization Project sponsored by the Agency for Healthcare Research and Quality²³). From the full CCS-encoded dataset, HF discharges were identified as those containing the appropriate CCS diagnosis code (108: “Congestive Heart Failure; Nonhypertensive”).

Using a Ruby script, all possible combinations of CCS codes were generated for each year (allowing for up to four possible CCS code combinations) for both the full dataset (ds) and the HF dataset (Dx). Combinations that occurred less than 5 times or for less than 1% of the size of the respective dataset (whichever was larger) were removed and not considered in the subsequent analyses. These combinations were referred to as Compound Comorbidities (CCMs) and also referenced as d^c , where c is the number of diseases in a particular combination. Note that the equations below use Einstein notation for simpler representation (i.e., superscript c does not mean that the affected value is “to the exponent c ,” it instead means the affected value “at the index c ”).

The prevalence ($PREV$) of each combination in the disease dataset (Dx) was then calculated, using a formula similar to TF*IDF (which is often used in information retrieval to determine the relative importance of a given term [TF] relative to the inverse of the universe of documents [IDF]²⁴):

$$PREV(d_{Dx}^c) = \sum d_{Dx}^c \times \log \left(\frac{R_{ds}}{\sum d_{ds}^c} \right) \quad (\text{Eq. 1})$$

Where, d_{Dx}^c corresponds to a CCM of size c for the disease of interest (Dx), R_{ds} corresponds to the number of records (R) in the full dataset (ds), and d_{ds}^c is the number of times the particular disease combination (d_{Dx}^c) occurs in the full dataset. Relative to the full set of possible CCMs ($\forall c$) for both the disease data set (Dx) and the full data set (ds), the Odds Ratio (OR) was then calculated for each CCM:

$$OR(d_{Dx}^c) = \frac{\sum d_{Dx}^x \times (\sum d_{ds}^{\forall c} - \sum d_{Dx}^c - \sum d_{ds}^c)}{(\sum d_{ds}^{\forall c} - \sum d_{ds}^c) \times (\sum d_{Dx}^{\forall c} - \sum d_{Dx}^c)} \quad (\text{Eq. 2})$$

The VUHDDS public data set includes costs associated with each discharge event. These costs were averaged for each CCM and then used to determine the relative costs (RC), which was calculated as the relative increase of a CCM of size c compared to CCMs of size $c-1$ for the disease relative to the rate of increase for the data set for the same sizes:

$$RC(d_{Dx}^c) = \frac{(COST_{d_{Dx}^{c-1}} / COST_{d_{Dx}^c})}{(COST_{d_{ds}^{c-1}} / COST_{d_{ds}^c})} \quad (\text{Eq. 3})$$

For each CCM, an aggregate score ($SCORE$) was calculated as a product of the $PREV$, OR , and RC values:

$$SCORE(d_{Dx}^c) = PREV(d_{Dx}^c) \times OR(d_{Dx}^c) \times RC(d_{Dx}^c) \quad (\text{Eq. 4})$$

Which resulted in a single value that accommodated the three dimensions that were the focus of the present study: Prevalence, Relevance, and Cost. The sum of all CCMs were then analyzed according to each dimension individually, as well as compared to the Charlson Comorbidity Index (CCI) for each CCM.

The CCI for each CCM was calculated based on the sum of Charlson Disease Weights (CDW) using the Deyo, *et al.*⁹ mapping of ICD-9-CM codes to CDW values and then mapping to respective CCS codes (see Table 1).

Table 1: Mapping of Charlson Diagnostic Categories associated with Charlson Disease Weights (CDW) to Clinical Classifications System (CCS) Codes based on the Deyo, *et al.*⁹ ICD-9-CM to CDW mapping.

	Charlson Diagnostic Category	CCS Code: Description
CDW = 1	Myocardial Infarction	100: Acute Myocardial Infarction 101: Coronary atherosclerosis and other heart disease
	Congestive Heart Failure	108: Congestive heart failure; nonhypertensive
	Peripheral Vascular Disease	114: Peripheral and visceral atherosclerosis 115: Aortic; peripheral; and visceral artery aneurysms 117: Other circulatory disease 248: Gangrene
	Cerebrovascular Disease	109: Acute cerebrovascular disease 110: Occlusion or stenosis of precerebral arteries 111: Other and ill-defined cerebrovascular disease 112: Transient cerebral ischemia 113: Late effects of cerebrovascular disease
	Dementia	653: Delirium, dementia, and amnesic and other cognitive disorders
	Chronic Pulmonary Disease	127: Chronic obstructive pulmonary disease and bronchiectasis 132: Lung disease due to external agents
	Rheumatologic Disease	202: Rheumatoid arthritis and related disease 210: Systemic lupus erythematosus and connective tissue disorders 211: Other connective tissue disease
	Peptic Ulcer Disease	139: Gastroduodenal ulcer (except hemorrhage)
	Mild Liver Disease	6: Hepatitis 150: Liver Disease; alcohol-related 151: Other liver diseases
	Diabetes	663: Screening and history of mental health and substance abuse 49: Diabetes mellitus without complications
CDW = 2	Diabetes with Chronic Complications Hemiplegia or Paraplegia Renal Disease	50: Diabetes mellitus with complications 82: Paralysis 156: Nephritis; nephrosis; renal sclerosis 157: Acute and unspecified renal failure 158: Chronic renal failure 161: Other diseases of kidney and ureters
	Any Malignancy, Including Leukemia and Lymphoma	11,12,13,14,15,16,17,18,19,20,21,22,24,25,26,27,28, 29,30,31,32,33,34,45,36,37,38,39,40,41: Cancer of (Respective): Head/Neck, Esophagus, Stomach, Colon, Rectum and Anus, Liver and Intrahepatic Bile Duct, Pancreas, other GI Organs and Peritoneum, Bronchus and Lung, Other Respiratory and Intrathoracic, Bone and Connective Tissue, Melanoma of Skin, Breast, Uterus, Cervix, Ovary, Other Female Genital Organs, Prostate, Testis, Other Male Genital Organs, Brain and Nervous System, Thyroid 40: Hodgkin's Disease 41: Cancer; other and Unspecified Primary
CDW = 3	Moderate or Severe Liver Disease	138: Esophageal Disorders 151: Other Liver Diseases 153: Gastrointestinal Hemorrhage
CDW = 6	Metastatic Solid Tumor	42: Secondary malignancies 43: Malignant neoplasm without specification of site
	AIDS	5: HIV infection

Results

Across the three years of data that were analyzed, an average of $35,869 \pm 8441$ CCMs were identified from the total possible CCMs for all possible conditions (for 285 possible CCS categories allowing for CCMs up to size 4 [i.e., ${}_{285}C_4 = 269,145,735$]). For CCMs specific to HF, an average of 3182 ± 428 CCMs were identified. Figure 1 shows the distribution of average numbers of CCMs for each CCM size (c) considered for each year. Based on these data, it can be observed that there is a significant increase (almost eight fold) in the number of CCMs from size 1 to size 2. For CCMs of size 3, there was also an approximate increase of 50% of possible CCMs; however, for CCMs of size 4, the number of CCMs encountered significantly dropped (by more than 50%). Overall, the analysis reveals that the number of CCMs of size 2 or more increases each successive year; those of size 1 (which are equivalent to classic co-morbidities) remained relatively constant year to year. These observations can be seen in Figure 1.

The average Prevalence (PREV), Odds Ratio (OR), and Relative Cost (RC) were compared to the calculated Charlson Comorbidity Index (CCI), as shown in Figure 1. Overall, it is apparent that CCMs of size 1 are of highest prevalence, with very similar prevalence values for CCMs of higher sizes. It is also observable that there is a consistent pattern of higher prevalence for CCIs of value 2 (this is especially the case for CCMs of size 1, and less so, but still identifiable pattern for CCMs of higher sizes). This supports the principle of CCI based risk scores, which identify and weigh co-morbidities in an additive function.

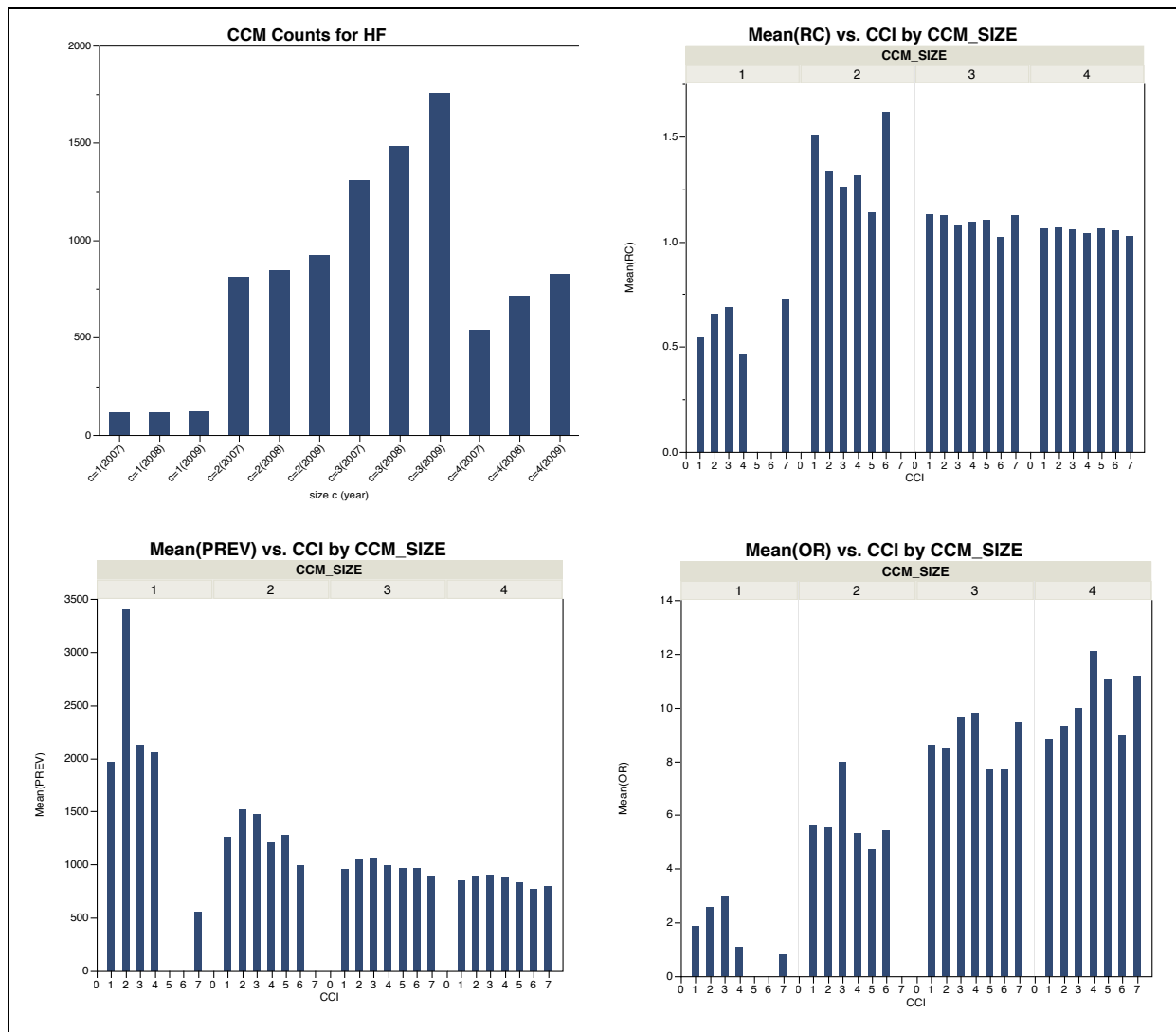


Figure 1: Histograms of average number of HF CCMs based on the analyzed hospital discharge data. Top-left is the average total number of HF CCMs for each size c and year of data analyzed. Top-right is the average Relative Cost (RC) relative to CCI values. Bottom-left is the average Prevalence (PREV) of HF CCMs relative to Charlson Comorbidity Index (CCI) values. Bottom-right is the average Odds Ratio (OR) of HF CCMs relative to CCI values.

Interestingly, for relative cost (RC), the rate of increase is noticeable from CCMs of size 1 to 2, but then stays even across all CCI bins for higher size CCMs. This seems to imply that the rate of cost increase for comorbidities to multiple comorbidities is significant, but that regardless of the number of concurrent comorbidities the rate of cost increases remains the same. However, it is important to note that RC does not reflect full cost; it only reflects the relative change in cost.

A markedly different pattern emerges when studying the correlation between the average OR values and CCIs (bottom left of Figure 1). In this analysis of the data, it can be seen that the general pattern is that higher OR values are associated with higher size CCMs. This suggests that CCMs of higher size have potentially higher significance in the context of the meaning of the correlation with the disease condition of interest (in this case, HF). The effect is further apparent when looking at all data for a given year (i.e., not separated by CCI values; Figure 2). The overall mean from year to year decreases for PREV, but increases for OR. This supports the notion that there are single comorbidities that do occur at a high frequency, but are of lower relative importance when compared to simultaneously co-occurring comorbidities of lesser frequencies. Similarly, amidst the lower frequency values for higher size CCMs, their relative importance increases. In both cases, these findings are supported across the three years analyzed. Tables 2 and 3 show the top five CCMs according to PREV or OR, respectively. For PREV, the top five consist of CCMs mostly of size 1 (with an occasional one of size 2). Conversely, for OR, the top five CCMs consist of size 3 or 4.

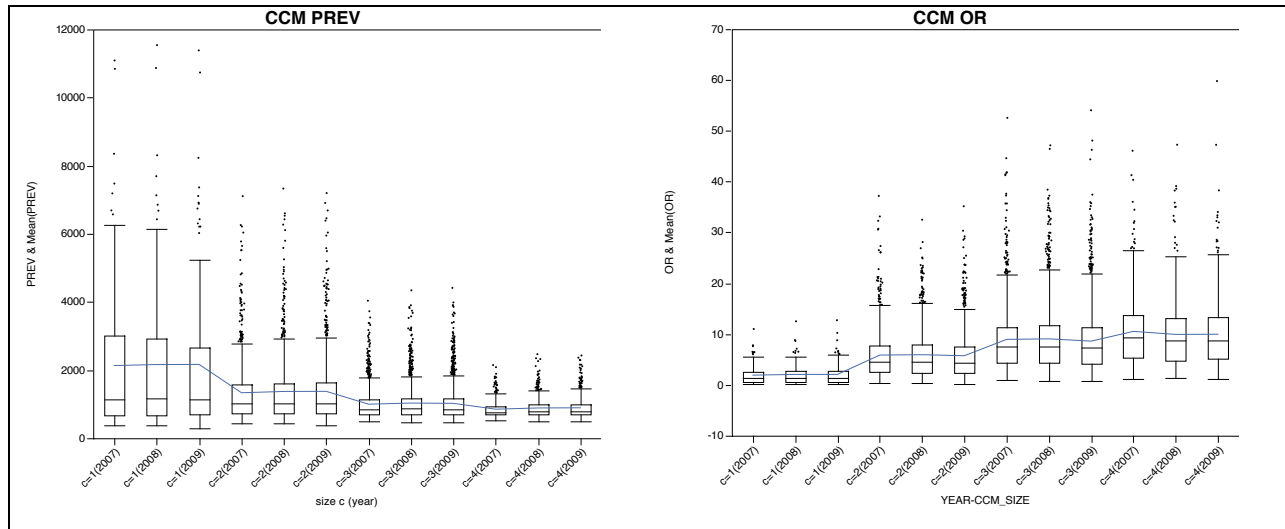


Figure 2: Boxplots of HF CCM prevalence (*PREV*; left) and odds ratio (*OR*; right) for Vermont hospital discharge data relative to each size *c* and year analyzed. Average value across data is shown as a line.

Table 2: Top 5 CCMs Per Year that Occur with HF According to Prevalence (PREV)

	PREV	OR	RC	SCORE	SIZE	CCI	CCM [CCS Code: Description]
2007	11090	3.6	0.70	28128	1	2	101: Coronary atherosclerosis and other heart disease
	10844	4.0	0.68	29475	1	1	106: Cardiac dysrhythmias
	8351	3.1	0.58	15261	1	2	127: Chronic obstructive pulmonary disease and bronchiectasis
	7476	1.3	0.60	5873	1	1	98: Essential hypertension
	7188	2.0	0.54	7543	1	2	49: Diabetes mellitus without complication
2008	11540	4.0	0.70	33493	1	2	101: Coronary atherosclerosis and other heart disease
	10868	4.2	0.68	31865	1	1	106: Cardiac dysrhythmias
	8306	3.3	0.58	16211	1	2	127: Chronic obstructive pulmonary disease and bronchiectasis
	7693	1.5	0.60	6875	1	1	98: Essential hypertension
	7330	8.7	0.54	76437	2	2	101: Coronary atherosclerosis and other heart disease 106: Cardiac dysrhythmias
2009	11385	4.0	0.70	32446	1	2	101: Coronary atherosclerosis and other heart disease
	10737	4.1	0.68	29810	1	1	106: Cardiac dysrhythmias
	8231	3.3	0.58	16469	1	2	127: Chronic obstructive pulmonary disease and bronchiectasis
	7363	2.1	0.60	8457	1	2	49: Diabetes mellitus without complication
	7197	8.5	0.54	74358	2	2	101: Coronary atherosclerosis and other heart disease 106: Cardiac dysrhythmias

Table 3: Top 5 CCMs Per Year that Occur with HF According to Odds Ratio (OR)

	PREV	OR	RC	SCORE	SIZE	CCI	CCM [CCS Code: Description]
2007	730	52.5	1.18	45273	3	2	96: Heart valve disorders 101: Coronary atherosclerosis and other heart disease 131: Respiratory failure; insufficiency; arrest (adult)
	692	46.1	1.04	33021	4	2	96: Heart valve disorders 99: Hypertension with complications and secondary hypertension 101: Coronary atherosclerosis and other heart disease 106: Cardiac dysrhythmias
	776	44.6	1.18	40949	3	3	103: Pulmonary heart disease 106: Cardiac dysrhythmias 158: Chronic renal failure
	809	41.9	1.07	36253	3	3	106: Cardiac dysrhythmias 131: Respiratory failure; insufficiency; arrest (adult) 158: Chronic renal failure
	711	41.7	0.95	28295	3	1	96: Heart valve disorders 106: Cardiac dysrhythmias 131: Respiratory failure; insufficiency; arrest (adult)
2008	713	47.3	1.02	34406	4	4	96: Heart valve disorders 99: Hypertension with complications and secondary hypertension 127: Chronic obstructive pulmonary disease and bronchiectasis 158: Chronic renal failure
	769	47.1	1.10	40003	3	2	96: Heart valve disorders 99: Hypertension with complications and secondary hypertension 127: Chronic obstructive pulmonary disease and bronchiectasis
	982	46.4	1.12	51152	3	4	96: Heart valve disorders 127: Chronic obstructive pulmonary disease and bronchiectasis 158: Chronic renal failure
	840	39.1	1.01	33339	4	4	96: Heart valve disorders 101: Coronary atherosclerosis and other heart disease 106: Cardiac dysrhythmias 158: Chronic renal failure
	981	38.6	1.10	41497	4	4	101: Coronary atherosclerosis and other heart disease 105: Conduction disorders 106: Cardiac dysrhythmias 158: Chronic renal failure
2009	815	59.8	0.98	47922	4	4	99: Hypertension with complications and secondary hypertension 101: Coronary atherosclerosis and other heart disease 103: Pulmonary heart disease 158: Chronic renal failure
	889	54.0	1.01	48590	3	2	99: Hypertension with complications and secondary hypertension 101: Coronary atherosclerosis and other heart disease 103: Pulmonary heart disease
	980	48.1	1.05	49445	3	4	101: Coronary atherosclerosis and other heart disease 103: Pulmonary heart disease 158: Chronic renal failure
	732	47.2	1.02	35141	4	3	99: Hypertension with complications and secondary hypertension 103: Pulmonary heart disease 106: Cardiac dysrhythmias 158: Chronic renal failure
	877	46.3	0.96	39058	3	1	99: Hypertension with complications and secondary hypertension 103: Pulmonary heart disease 106: Cardiac dysrhythmias

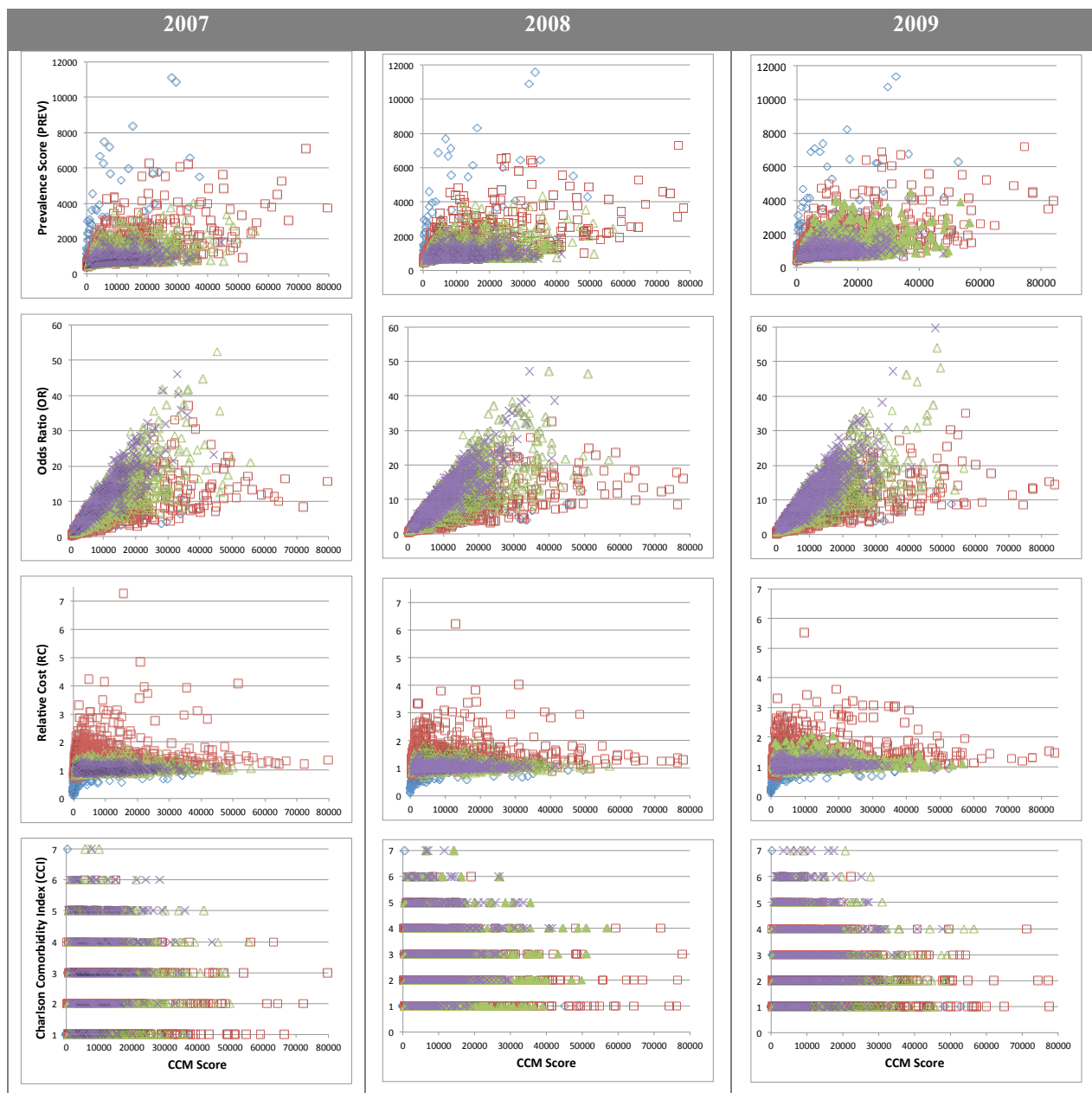


Figure 3: The relative distribution of CCM Scores relative to prevalence (*PREV*), odds ratio (*OR*), relative cost (*RC*), and Charlson Comorbidity Index (*CCI*). Diamonds (blue), squares (red), triangles (green), and crosses (purple) respectively represent CCMs of sizes 1,2,3, and 4.

Comparing the CCM Scores relative to the three individual axes (*PREV*, *OR*, and *RC*) as well as to the *CCI* reveals a similar pattern across all three years analyzed (Figure 3). Because the CCM Score is a product of the three axes of importance for this study, it provides a means to identify those CCMs that might warrant further investigation. For example, studying the outliers in a given chart in Figure 3, such as a CCM that does not cluster near the mean of other CCMs of the same size, might enable the identification of CCMs associated with unusual prevalence, importance, cost, or burden of disease (as could be determined based on *PREV*, *OR*, *RC*, or *CCI* scores, respectively).

Discussion

The study of comorbidities is an essential aspect of modern medicine. The comorbidity profile of a patient can have a direct impact on the ability to develop testable clinical hypotheses, suggest prognosis of clinical interventions, and provide valuable insights about clinical outcomes¹. Within the United States, there has emerged an acute need to understand comorbidity conditions, especially in light of escalating healthcare costs and the significant societal impacts of the chronically ill^{5,12,20}. It has been suggested that comorbidities can be a major contributor to costly complications and thus be a potentially significant factor in the overall cost of care^{5,12,20,25}. It has been further suggested that there is a need to develop approaches for accommodating multiple simultaneous comorbidities^{3,4,25,26}; however, to date there has been no direct attempt at developing such constructs. This study thus reflects perhaps one of the first attempts at developing a framework for formally linking co-occurring comorbidities that allows for subsequent aggregate analyses. The three axes that the construct (which we have termed “Compound Comorbidities” or CCMs) accommodate are: (1) Prevalence [*PREV*: a vector space normalized approach to quantify how often a CCM occurs in a dataset]; (2) Odds Ratio [*OR*: the importance of a CCM relative to a chosen condition]; and, (3) Relative Cost [*RC*: the relative increase in cost of a CCM compared to a specified baseline increase in cost].

The calculation of ORs in the present study used the complete set of possible CCMs across all diseases from a given year’s data. This method for calculating the OR reduces the confidence in calculating reliable p-values using a Fisher’s exact test and thus reduces the potential interpretation of the true importance of CCMs (which will undoubtedly vary between disease conditions). Some of this can be accounted for by adjusting the importance of a given CCM’s OR relative to its *PREV* compared to the *PREV* of other disease CCMs. Nonetheless, to better assess the importance of a given CCM, future work will require the development of simulations to more accurately calculate a given CCMs OR.

In contrast to the Charlson Comorbidity Index (CCI), which produces an index score that is used to approximate the disease comorbidity burden, CCMs provide a holistic view of a co-morbidity profile that may offer subtle insights that are particular to a disease population. This enables one to assess the relative impact of a given comorbidity that may not directly impact the comorbidity index. For example, if considering the HF CCM that consists of “Conduction Disorders” and “Diabetes Mellitus,” the addition of “Disorders of lipid metabolism” results in an RC of 1.01 (i.e., average increase of 1% in costs). However, if instead the added co-morbidity is “Hypertension with complications and secondary hypertension,” the RC increases to 1.51 (i.e., the relative costs increases 51% on average). In both cases, the CCI for the given CCM is 2.

It is important to note that the aim of this study is not to necessarily displace existing and well-proven comorbidity indices (e.g., CCI). Instead, it is to provide a complementary construct that may offer a holistic view of the constellation of comorbidities associated with an individual. Moreover, there may be benefit in incorporating existing indices into the aggregate score for CCMs. Future work may thus include an additional fourth axis of consideration that incorporates comorbidity indices like CCI. Because indices like CCI have target conditions that are weighted according to importance, additional work will require the harmonization of existing comorbidity indices that accommodate the largest number of target conditions possible, while still remaining computationally tractable.

A major infrastructural step that was developed as part of this study was the encoding of ICD-9-CM hospital discharge data into Clinical Classification Software (CCS) codes. CCS codes were also then mapped to the CCI scoring scheme (as developed by Deyo, *et al.*⁹). This second step allowed for the results to be organized by CCI values, which enabled study of the distribution of the CCM axes relative to CCI values. This encoding step could easily be generalized to other ICD-9-CM encoded datasets (e.g., insurance claim data sets or hospital billing data). CCS also has a mapping of codes to ICD-10, which suggests that the approach used here can be used prospectively. A mapping model will be an essential step for subsequent analyses, since it provides a common mapping of concepts to a single level hierarchy. However, there may be limitations or practical considerations of using CCS or any chosen mapping hierarchy. Nonetheless, for the purposes of studying aggregate relationships, such as reflected in CCMs, CCS is sufficient.

An interesting artifact of CCMs that might warrant further investigation is the total possible number of CCMs actually exists for a particular condition. The potential universe of CCMs is 269,145,735 (${}_{285}C_4$) for CCMs up to size 4; however, the average total number of CCMs found in this data analyzed for this study was 35,869±8441. One area for future work would be to assess the total number of CCMs in different types of data sets (e.g., discharge data, insurance claims data, billing data, and electronic health record data). When considering a particular condition, such

as congestive heart failure (HF), the average number of possible CCMs dropped to 3182±428. Thus, it may be interesting to study the relative number of CCMs for individual conditions – for example, this might help identify those conditions that are “CCM-rich” (have more possible CCMs) versus those that are “CCM-poor” (have fewer possible CCMs).

This study focused specifically on identifying CCMs associated with HF, which is an increasingly occurring condition in the aging population of the United States and around the world. This particular study was focused on the initial development of CCMs as a construct for understanding complex comorbidity relationships associated with HF. However, there may be merit in augmenting the approach to allow for filtering of particular condition types. For example, it is generally accepted that HF patients will have other cardiac comorbidities (which the results show here; as can be seen in Table 3). An interesting future study might be to filter CCMs to exclude cardiac conditions, or up-weight non-cardiac conditions in the importance score (which right now is solely based on a statistical odds ratio test). Within the context of HF, it may have significant clinical importance, since it has been suggested that almost 40% HF patients often have more than 5 non-cardiac comorbidities¹⁸. CCM based analysis may offer a systematic means to identify such co-morbidities and offer an empirical ranking of relative importance.

A final aspect of CCMs as proposed in this study is their inherent inclusion of relative increase in cost. It is certainly true that clinical metrics should primarily consider clinical outcomes and clinical significance; however, and especially in the current economic climate and attention to rapidly escalating health care costs, cost of care may be an additional important feature to explicitly incorporate. The intent here is not to identify those CCMs that are of high cost and effectively lead to the oft-feared “death panels²⁷” motivated by cost savings. Instead, the identification of CCMs that impact cost significantly may help identify system wide challenges that could be addressed, such as better advance care planning²⁷. This has been suggested for existing comorbidity indices (e.g., for CCI²⁰); however, the approach provided here provides a more inherent inclusion of cost relative to the full constellation of possible comorbidities. The future of health care and its reform will depend on an understanding of the balance between health care costs and achievable clinical outcomes. CCMs may offer one such tool in quantifying and accommodating clinical importance with relative costs, thereby facilitating the identification of areas that may be of high interest for improvement.

Conclusion

The study of comorbidities have to date been done based on pairwise analyses. This study developed a knowledge structure, termed “compound comorbidities” (CCMs), which reflect constellations of simultaneously occurring comorbidities that can be organized according to prevalence, importance, and cost. Based on the preliminary results, which focused on developing CCMs for congestive heart failure, there appears to be positive indication that CCMs may very well reflect a mechanism for simultaneous analysis of multiple comorbidities.

Acknowledgements

The Vermont Association of Hospitals and Health Systems-Network Services Organization (VAHHS-NSO) and the Vermont Department of Banking, Insurance, Securities and Health Care Administration (BISHCA) supplied hospital discharge data used in this study. All analyses, interpretations or conclusions based on these data are solely that of the authors. VAHHS-NSO and BISHCA disclaim responsibility for any such analyses, interpretations or conclusions. In addition, as the data have been edited and processed by VAHHS-NSO, BISHCA assumes no responsibility for errors in the data due to coding or processing by hospitals, VAHHS-NSO or any other organization, including the authors.

References

1. Parekh AK, Barton MB. The challenge of multiple comorbidity for the US health care system. *JAMA : the journal of the American Medical Association*. 2010 Apr 7;303(13):1303-4.
2. Fortin M, Stewart M, Poitras ME, Almirall J, Maddocks H. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Annals of family medicine*. 2012 Mar;10(2):142-51.

3. Tinetti ME, McAvay GJ, Chang SS, et al. Contribution of multiple chronic conditions to universal health outcomes. *Journal of the American Geriatrics Society*. 2011 Sep;59(9):1686-91.
4. Vogeli C, Shields AE, Lee TA, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. *Journal of general internal medicine*. 2007 Dec;22 Suppl 3:391-5.
5. Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. *Annals of family medicine*. 2009 Jul-Aug;7(4):357-63.
6. Quail JM, Lix LM, Osman BA, Teare GF. Comparing comorbidity measures for predicting mortality and hospitalization in three population-based cohorts. *BMC health services research*. 2011;11:146.
7. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. a critical review of available methods. *Journal of clinical epidemiology*. 2003 Mar;56(3):221-9.
8. Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. *Journal of clinical epidemiology*. 2000 Dec;53(12):1258-67.
9. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of clinical epidemiology*. 1992 Jun;45(6):613-9.
10. Perkins AJ, Kroenke K, Unutzer J, et al. Common comorbidity scales were similar in their ability to predict health care costs and mortality. *Journal of clinical epidemiology*. 2004 Oct;57(10):1040-8.
11. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*. 1987;40(5):373-83.
12. Fishman P, Von Korff M, Lozano P, Hecht J. Chronic care costs in managed care. *Health Aff (Millwood)*. 1997 May-Jun;16(3):239-47.
13. O'Connell JB. The economic burden of heart failure. *Clinical cardiology*. 2000 Mar;23(3 Suppl):III6-10.
14. Rich MW. Heart failure in the oldest patients: the impact of comorbid conditions. *The American journal of geriatric cardiology*. 2005 May-Jun;14(3):134-41.
15. Hunt SA, Abraham WT, Chin MH, et al. 2009 Focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the International Society for Heart and Lung Transplantation. *Journal of the American College of Cardiology*. 2009 Apr 14;53(15):e1-e90.
16. Metra M, Zaca V, Parati G, et al. Cardiovascular and noncardiovascular comorbidities in patients with chronic heart failure. *J Cardiovasc Med (Hagerstown)*. 2011 Feb;12(2):76-84.
17. Braunstein JB, Anderson GF, Gerstenblith G, et al. Noncardiac comorbidity increases preventable hospitalizations and mortality among Medicare beneficiaries with chronic heart failure. *Journal of the American College of Cardiology*. 2003 Oct 1;42(7):1226-33.
18. Lang CC, Mancini DM. Non-cardiac comorbidities in chronic heart failure. *Heart*. 2007 Jun;93(6):665-71.
19. Chuang JH, Friedman C, Hripcsak G. A comparison of the Charlson comorbidities derived from medical language processing and administrative data. *Proceedings / AMIA Annual Symposium AMIA Symposium*. 2002:160-4.
20. Charlson ME, Charlson RE, Peterson JC, Marinopoulos SS, Briggs WM, Hollenberg JP. The Charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. *Journal of clinical epidemiology*. 2008 Dec;61(12):1234-40.
21. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived from ICD-9-CCM administrative data. *Medical care*. 2002 Aug;40(8):675-85.
22. <http://www.bishca.state.vt.us/health-care/vermont-uniform-hospital-discharge-data-set-vuhdds>.
23. <http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>.
24. Salton G, McGill MJ. *Introduction to Modern Information Retrieval*. New York: McGraw-Hill; 1983.
25. Glynn LG, Valderas JM, Healy P, et al. The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. *Family practice*. 2011 Oct;28(5):516-23.
26. Nagl A, Witte J, Hodek JM, Greiner W. Relationship between multimorbidity and direct healthcare costs in an advanced elderly population : Results of the PRISCUS trial. *Zeitschrift fur Gerontologie und Geriatrie*. 2012 Feb;45(2):146-54.
27. Billings JA. The Need for Safeguards in Advance Care Planning. *Journal of general internal medicine*. 2012 Jan 12.