



HHS Public Access

Author manuscript

IEEE EMBS Int Conf Biomed Health Inform. Author manuscript; available in PMC 2016 October 11.

Published in final edited form as:

IEEE EMBS Int Conf Biomed Health Inform. 2016 February ; 2016: 136–139. doi:10.1109/BHI.

2016.7455853

Assessing the Comorbidity Gap between Clinical Studies and Prevalence in Elderly Patient Populations

Zhe He,

School of Information and Institute for Successful Longevity at Florida State University, Tallahassee, FL 32308 USA (phone: 850-644-5775)

Neil Charness,

Department of Psychology and Institute for Successful Longevity at Florida State University, Tallahassee, FL 32308 USA

Jiang Bian, and

Department of Health Outcomes and Policy, University of Florida, Gainesville, FL 32610 USA

William R. Hogan

Department of Health Outcomes and Policy, University of Florida, Gainesville, FL 32610 USA

Zhe He: zhe.he@cci.fsu.edu; Neil Charness: charness@psy.fsu.edu; Jiang Bian: bianjiang@ufl.edu; William R. Hogan: hoganwr@ufl.edu

Abstract

Well-designed and well-conducted clinical studies represent gold standard approaches for generating medical evidence. However, elderly populations are systematically underrepresented in studies across major chronic medical conditions, which has hampered the generalizability (external validity) of studies to the real-world patient population. It is the norm that intervention studies often require a homogeneous cohort to test their hypotheses; therefore older adults with co-medications and comorbidities are often excluded. The purpose of this study is to assess the gap between clinical studies on comorbidities and prevalence in elderly populations derived from the National Health and Nutrition Examination Survey (NHANES) and the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) dataset. A comorbidity gap between them was observed and reported in this work.

I. Introduction

Clinical studies are conducted for testing the efficacy and safety of a certain treatment (e.g., a medication, a device, or a procedure) for one or more medical conditions. They have been generating gold standard evidence in modern medical research. Nevertheless, it is expensive to conduct a clinical trial, with estimates of \$600 million dollars per clinical trial [1]. However, many trials failed to balance internal validity and external validity for generalizing results to the real-world target population [2]. It is widely reported that the elderly population is systematically underrepresented in clinical studies across major conditions including cardiovascular diseases [3], cancer [4], dementia [5], and diabetes [6]. Intervention studies usually require a homogeneous cohort to facilitate causal analysis. Thus, studies are often designed to minimize confounding factors by excluding patients with co-medications and complex comorbidities. However, this practice may also limit the external validity (or a

priori generalizability) of a clinical study. Boyd *et al.* pointed out the lack of evidence-based foundation for assessing quality of care in elders with comorbidities in clinical practice guidelines focused on a single disease [7]. We therefore hypothesize that there is a gap between the high prevalence of comorbidities in the elderly population and the clinical research that focuses on these prevalent comorbidities. Thus, in this work, we first analyze the prevalence and types of comorbidities in the elderly population. We use the patient data in Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II), one of the largest clinical databases that are freely available. It comprised high-resolution clinical information of more than 32,000 ICU admissions. Realizing that patients in ICU may be sicker than the general population we also used nationally representative data from National Health and Nutrition Examination Survey (NHANES). Then, we identify clinical studies in ClinicalTrials.gov that investigate these prevalent comorbidities. We compare them to see how well clinical studies investigate these prevalent comorbidities in elderly population. This work builds a foundation for understanding the comorbidity gap within aging populations that may motivate future studies to address potential health disparities and provide better clinical guidelines.

II. Background

A. ClinicalTrials.gov

ClinicalTrials.gov is the official clinical study and results registry created and maintained by the U.S. National Library of Medicine [8]. Mandated by the Food and Drug Administration Amendments Act (FDAAA) of 1997 at FDAAA 801, all the trials of drugs, device, and biologic other than Phase I trials have to be registered in ClinicalTrials.gov. As of November 3, 2015, 201,914 studies with sites in 190 countries are registered. Study summaries in ClinicalTrials.gov include structured study descriptors such as study title, sponsor, study phase, intervention, condition, as well as unstructured eligibility criteria that specify the characteristics of subjects to be included or excluded in the study. It is a valuable public resource for analyzing existing clinical studies to inform future study design.

B. National Health and Nutrition Examination Survey

National Health and Nutrition Examination Survey (NHANES) is a continuous cross-sectional health survey conducted by the National Center for Health Statistics of CDC [9]. It employs a sophisticated multi-stage sampling process. The survey samples are first interviewed at home, followed by a physical and laboratory test in a mobile examination center. Its rigorous quality control standards ensure high-quality data collection and national representativeness. NHANES publishes its survey results including individual-level interview and test results every two years. It has been widely used in epidemiology and observational studies.

C. MIMIC II

The Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) dataset is a free, public resource of de-identified patient data collected between 2001 and 2013 from a variety of Intensive Care Units (medical, surgical, coronary care, and neonatal) in a single tertiary teaching hospital [10]. It contains clinical data collected from bedside workstations and

hospital archives. In this work, we used MIMIC-II Clinical Database version 2.6 (April 2011; 32,536 subjects).

III. Methods

A. Dataset Preparation

We first downloaded 196,393 XML-format clinical study summaries from ClinicalTrials.gov as of August 26 2015. We extracted structured fields of the study summaries including study phase, study design, intervention, and start date. We downloaded the MeSH-based medical conditions annotation file of from the AACT (Aggregate Analysis of ClinicalTrials.gov) database (version: March 27, 2015), developed by US Food and Drug Administration and Duke University [11]. In the annotation file, each NCTID (the unique identifier of a study in ClinicalTrials.gov) is associated with one or more medical conditions encoded by a MeSH term. Using the native term mapping of the UMLS, we unified different MeSH terms of the same medical condition into a UMLS Concept Unique Identifier (CUI). The CUIs of the condition and its descendants were used to identify clinical studies that investigate the respective condition. We included all the studies (both interventional and observational) with a start date between January 2003 and March 2015.

To ensure statistical power of the analysis, we combined five two-year survey cycle data of NHANES between 2003 and 2012, and then imported the data of both NHANES and MIMIC II into a MySQL database.

B. Data Analysis

Previously, Charlson *et al.* [12] identified 17 conditions to be associated with one-year mortality and assigned each a weight (Charlson weight). The comorbidity index (CCI), the sum of the weight of all the conditions of a patient for assessing the risk of mortality, has been widely adopted by researchers to measure the burden of a disease [13]. They suggested that CCI would be most useful for assessing the impact of comorbid conditions on mortality for patients in longitudinal studies. It is an important indicator to help investigators detect and avoid overly restrictive eligibility criteria. Deyo *et al.* [14] and Romano *et al.* [15] later adapted CCI in administrative data with dissimilar sets of ICD-9-CM codes. Ghali *et al.* [16] compared these two adaptations and concluded that their power of predicting one year mortality is virtually identical. Recently, Quan *et al.* [13] updated the weight (Quan weight) for these 17 conditions, out of which 5 (see Table I) were reassigned a weight of 0 because they were not associated with one year mortality in their analysis.

In this work, using the ICD-9-CM coding algorithm developed by Deyo *et al.* [14], we first identified patients in MIMIC-II with one or more of the 17 conditions. Then we assigned both Charlson weight [12] and Quan weight [13] to each patient. The conditions not included in CCI were assigned a weight of 0. We then added up the weights for distinct disease categories for each patient. Table 1 shows the number and percentage of patients in MIMIC-II with one of the 17 diagnostic categories, and the weights for each diagnostic category. Note that some patients may be admitted to ICU multiple times. If a patient was diagnosed in the same diagnostic category in different sequences or in multiple ICU

admissions, the case was counted only once in the CCI calculation. We stratified the CCI distributions by age groups. We then analyzed the prevalence of the comorbidities consisting of the 17 conditions in Charlson's CCI among elderly patients in MIMIC-II. We also analyzed the number of clinical studies between January 2003 and March 2015 that investigate these comorbidities and the percentage of them that consider elderly patients according to the structured "age" eligibility criterion.

As MIMIC-II includes only ICU patients, who may be sicker than the general patient population, we further analyzed the comorbidity of a nationally representative population in NHANES. NHANES contains self-reported medical conditions instead of ICD-9-CM codes. Further, few conditions in CCI are reported in NHANES. We therefore chose patients with seven major medical conditions that are reported in NHANES: *diabetes*, *asthma*, *arthritis*, *congestive heart failure*, *myocardial infarction*, *stroke*, and *cancer*. After normalizing the two-year sample weights based on the analytical guideline of NHANES [17], the patients in NHANES can represent the national population in the midpoint of the combined survey cycle. To understand how existing clinical studies investigate these prevalent comorbidities, we performed a preliminary meta-analysis of these studies on their study characteristics.

IV. Results

We report on the distribution of CCI for MIMIC-II patients stratified by age groups as shown in Fig. 1. Note that MIMIC-II does not provide the age of a patient directly. We computed the age of the patients based on their date of birth and the date of admission. Using Quan weight, the mean CCI is 1.5 for all the patients ($n=31,090$), 2.3 for elderly patients (≥ 65 years old, $n=11,849$), and 1.1 for non-elderly patients (≤ 64 years old, $n=19,281$). Using Charlson weight, the mean CCI is 1.7 for all the patients ($n=31,090$), 2.5 for elderly patients (≥ 65 years old, $n=11,849$), and 1.2 for non-elderly patients (≤ 64 years old, $n=19,281$). Quan's CCI is consistently lower than Charlson's, but in both cases the CCI of elderly patients is higher than that of younger patients. As illustrated in Fig. 1, the numbers of elderly patients with CCI = 0 and 1 are noticeably different between Quan's and Charlson's because five conditions used in Charlson weight are not considered to be associated with one-year mortality by Quan.

When considering the original 17 conditions defined by Charlson, there are 10,001 elderly patients (84.4% of all the elderly patients) with at least one of these conditions. The top 5 most prevalent conditions are *Congestive heart failure* (44.3%), *Myocardial infarction* (28.7%), *Diabetes* (28.4%), *Chronic pulmonary disease* (25.9%), and *Moderate or severe liver disease* (25.9%). Table II lists top 8 prevalent comorbidities consisting of the 17 conditions in elderly patients in MIMIC-II. Note that some patients may be counted in different comorbidities because they have more than two of the 17 conditions. There are 751 combinations of comorbidities of elderly patients in MIMIC-II. The most prevalent comorbidity is *congestive heart failure + myocardial infarction*, which has been investigated only by 14 clinical studies so far.

Among 25,616 clinical studies on multiple comorbidities, 18,609 (72.6%) allow elderly patients according to the age criterion, but only a few of them investigate the top prevalent

comorbidities in elderly population in MIMIC-II. The reason may be that the conditions in CCI are correlated with high mortality within one year, thus, investigating these comorbidities may be of high risk. It is worth noting that even if the number of studies that investigate these prevalent comorbidities is low, the majority of them consider elderly patients according to their age criterion.

As NHANES is a national representative sample of patients, the comorbidities in this dataset should better represent the national population than MIMIC-II. Table III shows the percentage of patients in NHANES with each of seven conditions. We further stratified the analysis by age groups. As expected, except for asthma, the morbidity of all the conditions in the elderly population is significantly higher than the younger population.

About 35% of all the people have one or more conditions. Table IV reports the number of conditions for patients who have at least one of the seven conditions. About 78% of these patients who are < 65 years old have only one condition, whereas only 48% of these patients who are \geq 65 years old have only one condition. Compared with patients < 65 years old, the percentage of elderly patients with multiple comorbidities is significantly higher.

Table V reports the most prevalent comorbidities for the elderly patients derived from the NHANES data, including the number of patients, the percentage of patients out of the 28,785,361 elderly patients who has at least one of the seven conditions, and the percentage of clinical studies that consider the comorbidities. The top 6 prevalent comorbidities all include arthritis. However, the numbers of clinical studies that investigate these conditions are quite small.

Among 25,616 clinical studies on multiple comorbidities, only 48 intervention studies and 33 observational studies investigate these 14 prevalent comorbidities of elderly in MIMIC-II and NHANES. Among 48 intervention studies, 18 (37.5%) are drug trials, 14 (29.2%) are behavioral trials. Among 43 studies with specified primary purpose, 27 (62.8%) are treatment studies, followed by health services research (14%) and supportive care research (14%). On average intervention studies have 4.9 inclusion criteria (1–12, SD: 3.1) and 6.5 exclusion criteria (0–27, SD: 6.2). Even though most of these studies do consider elderly patients according to their age criterion, they may use other restrictive (e.g., ‘cognitive impairment’) or vague (e.g., ‘patients not expected to survive their hospitalization’) exclusion criteria to exclude elderly patients with complications. The scale of such phenomena is worthy of further investigation.

V. Discussion and Conclusion

This paper presents a preliminary analysis of the comorbidity gap in clinical studies of the elderly population based on both an ICU sample and a nationally representative population. We observed that very few clinical studies investigate prevalent comorbidities in the elderly population. As current clinical practice guidelines are inadequate to address the needs of older patients with complex comorbid conditions [18], future trials should include elderly individuals with representative comorbidities to provide evidence for clinical guidelines [7]. To assist such effort, we will develop informatics methods to understand the factors

associated with the comorbidity gap and provide suggestions to policy makers and clinical investigators for optimizing clinical study design. In future work, we also plan to develop data-driven methods to identify overly restrictive eligibility criteria that can be relaxed to improve the representativeness of elderly patients in clinical studies.

This study has limitations. We used the MeSH-based condition annotation of AACT [11], which may have erroneous or missing condition information. Future research is warranted to evaluate its accuracy of capturing the comorbidity information of clinical studies in ClinicalTrials.gov.

Acknowledgments

We would like to thank Zhiwei Chen for processing the MIMIC-II data. This work was partially supported by an AWS in Education Research Grant Award (PI: He). This work was partially supported by NIH/NCATS Clinical and Translational Science Award UL1TR001427 (PIs: Nelson & Shenkman). The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health.

References

1. Transforming Clinical Research in the United States. The National Academies Collection: Reports funded by National Institutes of Health. Washington (DC): 2010. Challenges and Opportunities: Workshop Summary.
2. Kukull WA, Ganguli M. Generalizability: The trees, the forest, and the low-hanging fruit. *Neurology*. 2012; 78(23):1886–91. [PubMed: 22665145]
3. Sardar MR, Badri M, Prince CT, et al. Underrepresentation of women, elderly patients, and racial minorities in the randomized trials used for cardiovascular guidelines. *JAMA Intern Med*. 2014; 174(11):1868–70. [PubMed: 25264856]
4. Lewis JH, Kilgore ML, Goldman DP, et al. Participation of patients 65 years of age or older in cancer clinical trials. *J Clin Oncol*. 2003; 21(7):1383–9. [PubMed: 12663731]
5. Schoenmaker N, Van Gool WA. The age gap between patients in clinical studies and in the general population: a pitfall for dementia research. *Lancet Neurol*. 2004; 3(10):627–30. [PubMed: 15380160]
6. Cigolle CT, Blaum CS, Halter JB. Diabetes and cardiovascular disease prevention in older adults. *Clin Geriatr Med*. 2009; 25(4):607–41. vii–viii. [PubMed: 19944264]
7. Boyd CM, Darer J, Boulton C, et al. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. *JAMA*. 2005; 294(6):716–24. [PubMed: 16091574]
8. [ClinicalTrials.gov](http://www.clinicaltrials.gov/). Nov 1. 2015 Available from: <http://www.clinicaltrials.gov/>
9. CDC. National Health and Nutrition Examination Survey Data. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; Nov 1. 2015 Available from: <http://www.cdc.gov/nchs/nhanes.htm>
10. Saeed M, Villarreal M, Reisner AT, et al. Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database. *Crit Care Med*. 2011; 39(5):952–60. [PubMed: 21283005]
11. Tasneem A, Aberle L, Ananth H, et al. The database for aggregate analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty. *PLoS One*. 2012; 7(3):e33677. [PubMed: 22438982]
12. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chron Dis*. 1987; 27:387–404.
13. Quan H, Li B, Couris CM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. 2011; 173(6):676–82. [PubMed: 21330339]

14. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol.* 1992; 45:613–9. [PubMed: 1607900]
15. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol.* 1993; 46(10):1075–9. [PubMed: 8410092]
16. Ghali WA, Hall RE, Rosen AK, et al. Searching for an Improved Clinical Comorbidity Index for Use with ICD-9-CM Administrative Data. *J Clin Epidemiol.* 1996; 49(3):273–8. [PubMed: 8676173]
17. CDC. National Health and Nutrition Examination Survey: Analytic Guidelines, 1999–2010. Available from: http://www.cdc.gov/nchs/data/series/sr_02/sr02_161.pdf
18. Mutasingwa DR, Ge H, Upshur RE. How applicable are clinical practice guidelines to elderly patients with comorbidities? *Can Fam Physician.* 2011; 57(7):e253–62. [PubMed: 21753084]

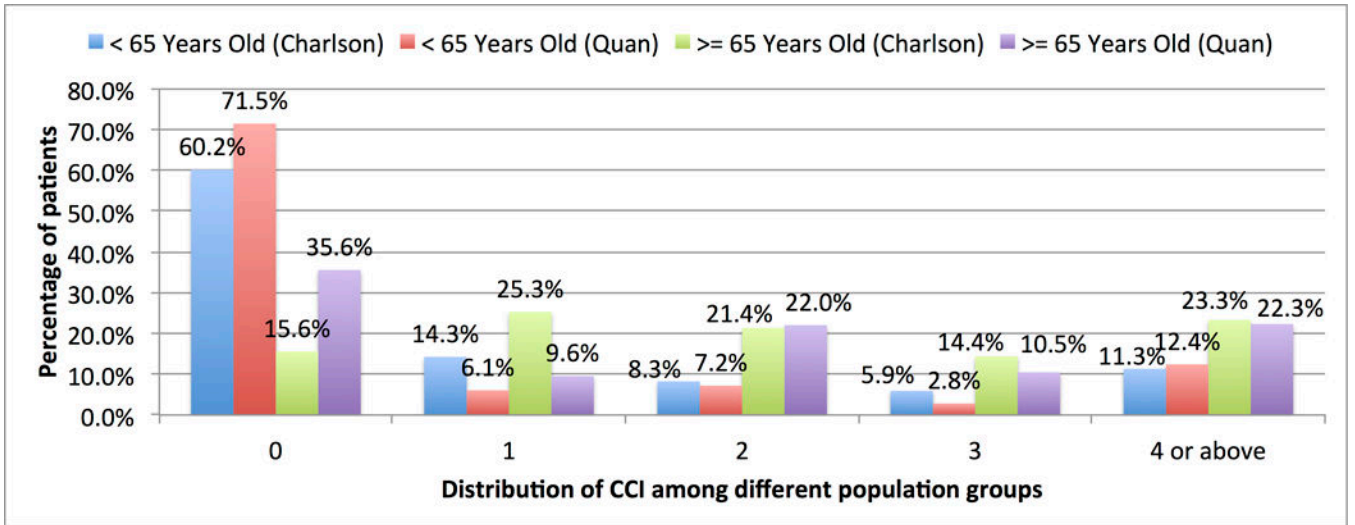


Figure 1.
Distribution of CCI of patients in MIMIC-II.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table I

Prevalence of conditions in patients of MIMIC-II.

Diagnostic category (Charlson weight, Quan weight)	# patients in MIMIC-II (n=31,090)
<i>Myocardial infarction</i> (1, 0)	4,399 (14.1%)
Congestive heart failure (1, 2)	6,168 (19.8%)
<i>Peripheral vascular disease</i> ^a (1, 0)	1,410 (4.5%)
<i>Cerebrovascular disease</i> (1, 0)	2,590 (8.3%)
Dementia (1, 2)	112 (0.4%)
Chronic pulmonary disease (1, 1)	4,186 (13.5%)
Rheumatologic disease (1, 1)	450 (1.4%)
<i>Peptic ulcer disease</i> (1, 0)	471 (1.5%)
Mild liver disease (1, 2)	911 (2.9%)
<i>Diabetes without chronic complications</i> (1, 0)	4,888 (15.7%)
Hemiplegia or paraplegia (2, 2)	312 (1.0%)
Renal disease (2, 1)	328 (1.1%)
Diabetes with chronic complications (2, 1)	1,159 (3.5%)
Any malignancy, including leukemia and lymphoma (2, 2)	2,522 (8.1%)
Moderate or severe liver disease (3, 4)	3,172 (10.2%)
Metastasis disease (6, 6)	1,392 (4.5%)
HIV (6, 4)	0 (0%)

^aThe conditions with names in italic were reassigned the weight 0 by Quan et al. [13].

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table II

Prevalent Comorbidities of Elderly Patients in MIMIC-II

Comorbid conditions	# of patients	Percentage of patients ^a	# of clinical studies (% considering elders)
Congestive heart failure + Myocardial infarction	1,521	15.2%	14 (100%)
Chronic pulmonary disease + Congestive heart failure	1,289	12.9%	22 (100%)
Congestive heart failure + Diabetes without chronic complications	1,244	12.4%	13 (100%)
Diabetes + Myocardial Infarction	781	7.8%	8 (100%)
Congestive heart failure + Moderate or severe liver disease	722	7.2%	1 (100%)
Chronic pulmonary disease + Myocardial infarction	619	6.2%	2 (100%)
Chronic pulmonary disease + Diabetes	617	6.2%	16 (100%)
Cerebrovascular disease + Congestive heart failure	441	4.4%	4 (100%)

^aThe denominator is 10,001 elderly patients with at least one of the 17 conditions.

Table III

Condition prevalence in NHANES

Medical Condition	Age		
	All (n=296,702,030)	>= 65 (n=37,139,322)	< 65 (n=259,562,708)
Diabetes	6.2% (n=18,272,012)	19.4% (n=7,198,001)	4.3% (n=11,074,011)
Asthma	14.2% (n=42,035,521)	11.3% (n=4,195,707)	14.6% (n=37,839,814)
Arthritis	17.6% (n=52,355,612)	54.1% (n=20,089,725)	12.4% (n=32,265,887)
Congestive heart failure	1.8% (n=5,259,757)	9.1% (n=3,369,678)	0.7% (n=1,890,079)
Myocardial infarction	2.5% (n=7,387,759)	11.2% (n=4,157,719)	1.2% (n=3,230,040)
Stroke	2.1% (n=6,174,893)	9.4% (n=3,502,919)	1.0% (n=2,671,974)
Cancer	6.6% (n=19,536,290)	25.5% (n=9,474,605)	3.9% (n=1,006,1685)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table IV

Number of Conditions for Patients in NHANES

# of conditions	Age		
	All (n=104,845,221)	≥ 65 (n=28,785,361)	< 65 (n=76,059,860)
7	0.03% (n=36,302)	0.09% (n=27,029)	0.01% (n=9,273)
6	0.1% (n=145,548)	0.4% (n=121,006)	0.03% (n=24,541)
5	0.6% (n=584,790)	1.1% (n=304,425)	0.4% (n=280,365)
4	2.2% (n=2,281,577)	4.8% (n=1,378,646)	1.2% (n=902,931)
3	6.7% (n=7,058,963)	13.2% (n=3,798,965)	4.3% (n=3,259,997)
2	20.9% (n=21,929,254)	32.9% (n=9,484,221)	16.4% (n=12,445,033)
1	69.4% (n=72,808,788)	47.5% (n=13,671,069)	77.8% (n=59,137,719)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table V

Prevalent Comorbidities of Elderly Patients in NHANES

Comorbid conditions	# of patients	Percentage of patients ^a	# of clinical studies (% considering elderly)
Arthritis + cancer	5,506,633	19.1%	5 (80%)
Arthritis + diabetes	4,241,994	14.7%	7 (85.7%)
Arthritis + asthma	2,615,833	9.1%	3 (100%)
Arthritis + myocardial infarction	2,407,460	8.4%	2 (0%)
Arthritis + stroke	2,199,728	7.6%	5 (80%)
Arthritis + congestive heart failure	2,162,934	7.5%	2 (50%)
Cancer + diabetes	1,822,775	6.3%	4 (75%)

^aThe denominator is 28,785,361 elderly patients with at least one of the seven conditions.