# Defining Depression Cohorts Using the EHR: Multiple Phenotypes Based on ICD-9 Codes and Medication Orders

Wendy Marie Ingram*[1,2], Anna M. Baker[3], Jason P. Brown[4], Fernando S. Goes[5], Peter P. Zandi[1], Sharon Larson[2,6]

1.  Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
2.  Department of Psychiatry, Geisinger Health System, Danville, Pennsylvania, USA
3.  Department of Psychology, Bucknell University, Lewisburg, Pennsylvania, USA
4.  Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, Pennsylvania, USA
5.  Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
6.  College of Population Health, Thomas Jefferson University, Philadelphia, Pennsylvania, USA

* Corresponding author:
    Wendy Marie Ingram
    E-mail address: wingram5@jhu.edu
    Phone number: 602-403-5024
    ORCID: https://orcid.org/0000-0002-1643-3367

## Abstract

The importance of methodology when conducting high-quality health behavior research cannot be overstated. Electronic Health Records (EHR) allow researchers to conduct unprecedented large-scale studies. However, careful consideration must be given to how to define patient cohorts, specifically when utilizing EHR for examining patients with depression. Because depression has been linked to increased morbidity and mortality in many disease groups and leads to higher health care utilization, better methods for identifying patients with depression must be developed in order to rigorously study the impact of depression on these outcomes. Identifying patients using only ICD9 codes for depression may result in inclusion of clinically depressed patients in comparison groups. Thus, more nuanced electronic phenotypes may better delineate patients that are receiving treatment for depression. We demonstrate the utility of a new method involving multiple depression phenotypes on a 10.75-year cohort from an integrated health system (n=287,281). Here we recommend a novel and easily adaptable method of categorizing patients. In this method, four groups are identified using ICD-9 codes and medication orders from an EHR which have varying levels of depression likelihood and severity: Dep ICD9, Rx no ICD9, Rx non-dep, and No Dep. We then measure a variety of EHR-based features including utilization patterns, medication orders, comorbidities, mortality data and symptom assessment scores to establish convergent validity of these groups. This superior and simple method allows for large scale studies of depressed patients, while accounting for the limitations associated with using specific electronic phenotypes for analysis of data from the EHR.

**Keywords:** Depression; Electronic health records; Phenotypic algorithms; Clinical informatics

**Conflict of Interest:** The authors declare that they have no conflict of interest.

## 1 Introduction

Electronic Health Records (EHR) allow researchers to conduct unprecedented, large scale studies. As a result, researchers using EHR data have an increased need for algorithms that correctly identify patient populations. These algorithms have been successfully created and disseminated in many patient populations[1–3]. However, among studies that have developed algorithms in specific disease groups, there is often notable heterogeneity with poor validity, justification, and description for future use[4]. In addition, there are some phenotypes that are more difficult to identify based on their characterization in the EHR[5]. Thus, the creation of algorithms defining specific patient cohorts in the EHR should be driven by sound theoretical justification and assessed for validity prior to dissemination[6].

Depression is a particularly difficult phenotype to define and studies often use heterogeneous criteria when utilizing EHR variables to identify patients with depression[7–12]. Many studies use only ICD-9 codes related to depression to identify patients[12, 13], while some researchers either have access to validated depression instrument scores or can include self-report depression symptom questionnaires[7–9] in their study design such as the Hamilton Depression Rating Scale[14], and the widely used Patient Health Questionnaire[15]. Following the United States Preventative Services Task Force (USPSTF) recommendation that all adults over the age of 18 should be screened for depression annually, many healthcare systems are starting to implement depression screening[16]; however, implementation has been fairly recent, is not standardized, and how this information is recorded in the EHR is variable. It is important to note that researchers have identified inconsistencies when examining ICD-9 code identification and other indicators[10, 11] such as PHQ9 scores or antidepressant medication orders.

Depression accounts for $43 billion in medical costs annually and is one of the leading causes of disability[17]. Depression has been linked to worse disease outcomes and increased healthcare utilization in many different disease groups including asthma[18], cardiac disease[19], rheumatoid arthritis[20], chronic kidney disease[21], and diabetes[22] among others, necessitating careful consideration. Using rigid inclusion definitions to identify depressed patients may exclude those with clinical features that are receiving treatment but do not meet these definitions. Traditional use of only ICD-9 codes to identify depressed patients[12, 13] may be flawed for many reasons including patient or provider bias against recording depression codes, systematic administrative data sequestration protocols, and possible inclusion of subclinical or undiagnosed depression in comparison groups. In addition, there are a variety of ICD-9 codes[23] that could be used by researchers to define depression and choices made by researchers about which codes are chosen for a particular study should be data driven and relevant to the particular cohort sought. For depression alone there are 17 possible ICD-9 codes that could be flagged[23].

Examination of antidepressant medication orders may help clarify depressive phenotypes, however, using only antidepressant medication would also result in incomplete identification of patients[24]. Practice guidelines provide detailed recommendations for the assessment and treatment of depression including psychopharmacology and other psychosocial interventions[25]. However, in clinical settings, antidepressant medication is frequently given without formal assessment or diagnosis[26], and severity is often unknown[27]. Though there are some off-label uses of antidepressants (e.g., tricyclic antidepressants for pain and aminoketone antidepressants for smoking cessation)[28], they are arguably fairly easy to discern and remove from cohorts[28, 29]. Thus, researchers may benefit from utilizing antidepressant orders, in addition to ICD-9 codes, to better identify patient groups. Because depression has been linked to increased morbidity and mortality in many disease groups and leads to higher health care utilization[18–22], better methods for identifying patients with depression must be developed in order to rigorously study the impact of depression on these outcomes.

Standardized methods allow for comparisons to be made across studies and better synthesis of research in a particular area. Our approach aimed to develop a methodology to generate electronic phenotypes that would limit the exclusion of clinically relevant patients from a cohort with rigid inclusion criteria, and avoid diluting control comparison groups with those potentially undiagnosed with depression but receiving treatment for related symptoms. The current study demonstrates the utility of using multiple depression phenotypes on a 10.75-year cohort from an integrated health system. In our method, depressed patients are identified by ICD-9 diagnosis and a control group is identified by excluding patients with antidepressant medication orders. Those with one or more antidepressant medications are grouped into two categories based on recent antidepressant orders and off-label diagnosis codes during encounters in which an antidepressant was ordered. We aim to investigate the outcomes of those that fall into each group and observe patterns of healthcare utilization and other EHR based features. We hypothesize that using this multiple depression phenotypes method of defining cohorts will reveal clinically important differences in certain outcome

variables, supporting the need for careful and consistent definitions of depressed patients and comparison groups for large studies involving administrative data.

## 2 Methods

*2.1 Data acquisition, analysis, and graphing.* This is a retrospective observational study using de-identified Electronic Health Records (EHR) of a general patient population. The study included health care encounter data between January 1st, 2005 and September 30th, 2015 (10.75 years) for patients seen in the Geisinger Health System, an integrated health care system located primarily in central Pennsylvania. This is a stable patient population whose EHRs have been collected in a central data warehouse and are available for clinical and research purposes, described previously elsewhere[30–33]. Patients 18 years or older at the beginning of the study (January 1st, 2005), who had a Geisinger Primary Care Physician (PCP) at any point during the study period, and had at least one outpatient visit within the system were included in the cohort (n=287,281) (Table 1). Demographic information, medication order histories, and details of outpatient, Emergency Department (ED), and inpatient encounters were pulled from a central data warehouse and de-identified by an approved data broker in the Geisinger Phenomics and Clinical Data Analytics Core under the oversight of the Geisinger Internal Review Board as non-human subjects research. Analysis and graphing were conducted with R (2017, R Core Team, Vienna, Austria), Rstudio (Boston, MA), and GraphPad Prism 6 (La Jolla, CA).

| | Characteristics | Total | % | Dep ICD9 | % | Comparison | % | Rx no ICD9 | % | Rx non-dep | % | No Dep | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 287281 | % | 59097 | % | 228184 | % | 32126 | % | 31029 | % | 165029 | % |
| **Sex** | Male | 131159 | 45.7 | 18497 | 31.3 | 112662 | 49.4 | 11373 | 35.4 | 12790 | 41.2 | 88499 | 53.6 |
| | Female | 156088 | 54.3 | 40591 | 68.7 | 115497 | 50.6 | 20753 | 64.6 | 18235 | 58.8 | 76509 | 46.4 |
| **Race** | White | 275199 | 95.8 | 57753 | 97.7 | 217446 | 95.3 | 31459 | 97.9 | 30126 | 97.1 | 155861 | 94.4 |
| | Black | 7221 | 2.5 | 969 | 1.6 | 6353 | 2.8 | 428 | 1.3 | 602 | 1.9 | 5222 | 3.2 |
| | Asian | 2294 | 0.8 | 141 | 0.2 | 2153 | 0.9 | 95 | 0.3 | 121 | 0.4 | 1937 | 1.2 |
| | Native | 1393 | 0.5 | 141 | 0.2 | 1252 | 0.5 | 93 | 0.3 | 108 | 0.3 | 1051 | 0.6 |
| | Other | 1174 | 0.4 | 93 | 0.2 | 1081 | 0.5 | 51 | 0.2 | 72 | 0.2 | 958 | 0.6 |
| **Marital Status** | Married | 171941 | 59.9 | 30555 | 51.7 | 141386 | 62.0 | 19654 | 61.2 | 18081 | 58.3 | 103651 | 62.8 |
| | Single | 58545 | 20.4 | 11470 | 19.4 | 47075 | 20.6 | 5550 | 17.3 | 6234 | 20.1 | 35291 | 21.4 |
| | Widowed | 26607 | 9.3 | 7103 | 12.0 | 19504 | 8.5 | 3303 | 10.3 | 2959 | 9.5 | 13242 | 8.0 |
| | Divorced | 24691 | 8.6 | 8008 | 13.6 | 16683 | 7.3 | 3005 | 9.4 | 3079 | 9.9 | 10599 | 6.4 |
| | Other | 156 | 0.1 | 19 | 0.0 | 137 | 0.1 | 5 | 0.0 | 10 | 0.0 | 122 | 0.1 |
| **Age at Beginning of Study** | 18-30 | 60099 | 20.9 | 11463 | 19.4 | 48636 | 21.3 | 6640 | 20.7 | 6426 | 20.7 | 35570 | 21.6 |
| | 31-45 | 85041 | 29.6 | 18071 | 30.6 | 66970 | 29.3 | 9950 | 31.0 | 9756 | 31.4 | 47264 | 28.6 |
| | 46-65 | 98734 | 34.4 | 21012 | 35.6 | 77722 | 34.1 | 10424 | 32.4 | 10196 | 32.9 | 57102 | 34.6 |
| | 66+ | 34180 | 11.9 | 6569 | 11.1 | 27611 | 12.1 | 3970 | 12.4 | 3598 | 11.6 | 20043 | 12.1 |
| **Length of Observation** | 1 year or less | 37974 | 13.2 | 3473 | 5.9 | 34501 | 15.1 | 831 | 2.6 | 3291 | 10.6 | 30379 | 18.4 |
| | 1-4 years | 64132 | 22.3 | 10319 | 17.5 | 53813 | 23.6 | 5865 | 18.3 | 5886 | 19.0 | 42062 | 25.5 |
| | 4-7 years | 42814 | 14.9 | 8631 | 14.6 | 34183 | 15.0 | 5028 | 15.7 | 4529 | 14.6 | 24626 | 14.9 |
| | 7-10.75 years | 142361 | 49.6 | 36674 | 62.1 | 105687 | 46.3 | 20402 | 63.5 | 17323 | 55.8 | 67962 | 41.2 |

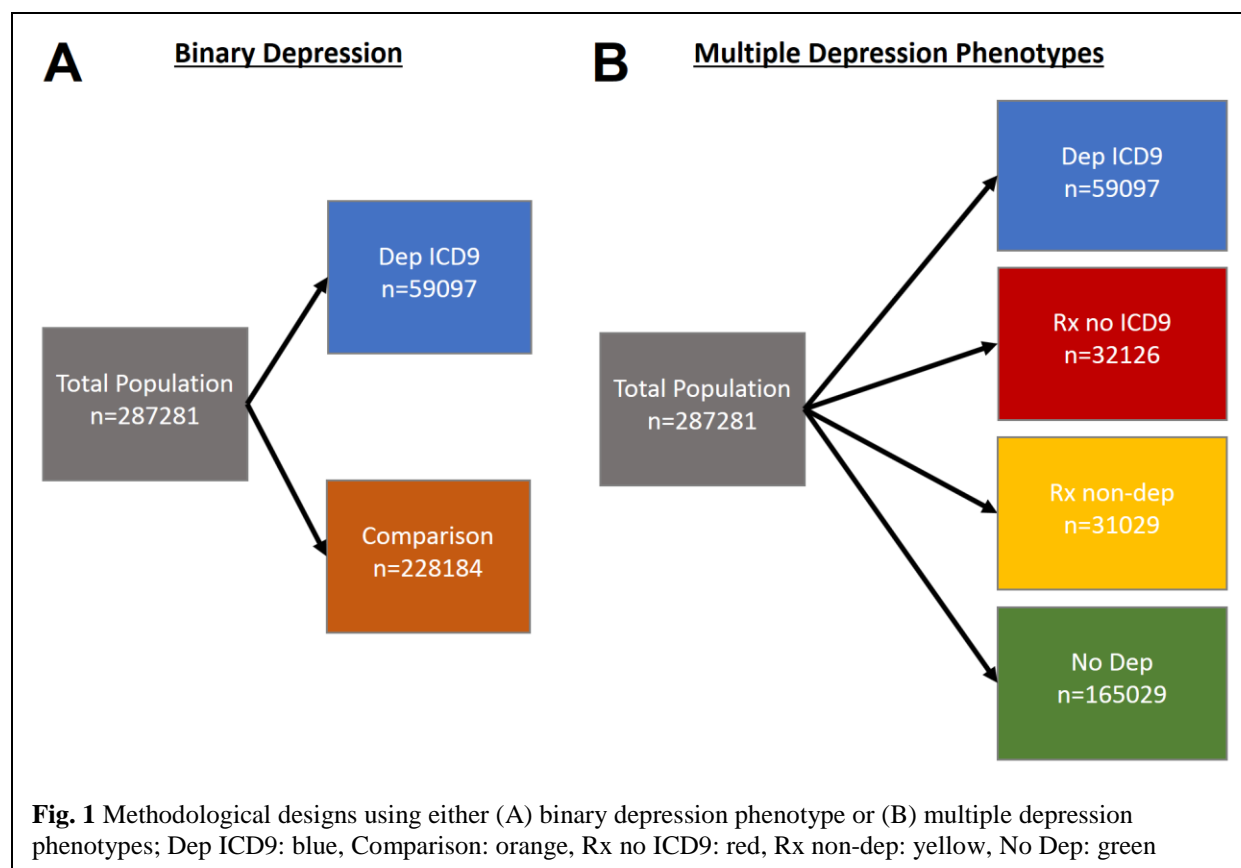**Table 1.** Summary statistics of total population and depression phenotypes

*2.2 Multiple depression phenotypes logic.* Using domain knowledge of depression clinical care, we designed a method of partitioning patients into four clinically relevant groups as they relate to depression. While a purely empirical methodology using feature selection and multivariate regression models may be worth developing, we argue that evaluating every possible feature and integrating them into a cohesive singular model is not necessary for many studies aiming to investigate depression using administrative clinical EHRs. In addition, EHRs can vary tremendously in the amount and type of data captured on each patient. Advanced modeling may be appropriate and useful in some situations while not in others. Those that have highly transient populations, data exclusively from primary, secondary,

or tertiary care, or EHRs that have only recently begun to capture patient data in a minable fashion would benefit from a simple but verified strategy for categorizing patients based on depression to allow for portability and generalizability.

First, for group "Dep ICD9", we define those that most clearly meet the criteria of depression if specific ICD-9 codes for depression are in their EHR in the following ways: 1) once or more in an ED or Inpatient discharge diagnosis; or on their problem list record; or 2) two or more times within a 2-year period as an outpatient discharge diagnosis (Fig 1B). Only one discharge diagnosis is required in ED and inpatient settings and on the problem list because the clinical threshold of assigning these diagnoses is higher than in an outpatient setting as reported by system physicians. In an outpatient setting, spurious diagnoses of depression are more likely and thus we required two relatively close diagnoses within a two-year period to avoid categorizing those with transient depressive symptoms or misdiagnosed adjustment disorder as depressed. We chose to include the following 17 ICD-9 codes for this study: 296.20, 296.21, 296.22, 296.23, 296.24, 296.25, 296.26, 296.30, 296.31, 296.32, 296.33, 296.34, 296.35, 296.36, 296.82, 301.12, 311. We did not include any 300, 309, or 648.4 codes for anxiety, adjustment disorder, or mental disorders complicating pregnancy, respectively, although some of these codes specify depressed mood. These codes may be selected or excluded per the specific purpose and interest of each study. For example, those interested in studying exclusively severe depression likely meeting DSM criteria for Major Depressive Disorder would want to exclude 301.12 and 311. Second, for group "Rx no ICD9", we define those that are most likely to be recently receiving treatment for depression or related symptoms as those who have received two or more antidepressant medication orders within the study period, here defined as January 1st, 2005 through September 30th, 2015. However, this excludes those that are in group "Rx non-dep", defined as having one or fewer medication orders for antidepressants during the study period or if they have had two or more, then the discharge diagnosis codes during such encounters include codes for common uses not necessarily related to depression. Here we define those ICD-9 codes as 305.1 or 356, tobacco use disorder[34] or hereditary and idiopathic neuropathy[35], respectively. While tobacco use disorder is associated with depression and other serious mental illnesses, we believe that association alone is not a strong enough reason to include these patients as likely receiving antidepressants for treatment of depression symptoms. However, because of the possibility of comorbidity and other social factors, it is reasonable to exclude these individuals from the clearly non-depressed group. Finally, group "No Dep" is defined as those least likely to have depression or depression symptoms and do not meet the inclusion criteria of group "Dep ICD9" and have never received an antidepressant medication order. This multiple depression phenotype method contrasts with traditional methods comparing those in the "Dep ICD9" group to a simple "Comparison" group diluting the clearly non-depressed patients with those who have received antidepressant medications (Fig 1A).

*2.3 Utilization patterns.* To visualize the general trends of each group's health care utilization, the age at each visit was calculated for each patient, excluding those of age 91 and older (for de-identification purposes, those that reached age 90 during the study period had their date of birth masked and ages could not be calculated). We then calculated the number of visits in the outpatient, ED, and inpatient setting at each age and standardized this by the total number of patients from that group, resulting in the visit frequency (Fig 2). Patients that were seen multiple times at a given age will thus contribute more than once to the numerator of this visit frequency. We used a dynamic open cohort study design, meaning that patients may have entered or left the health system at any time during the study period. This allows us to take a more naturalistic view of a patient population from the health care system service area perspective, however it is then impossible to know when exactly patients are entering or leaving the system, preventing us from standardizing these visits by the known number of patients at each age contributing to the total number of patients at risk of being seen in each setting.

*2.4 Convergent validation.* We sought to evaluate the validity of our multiple depression phenotypes method by analyzing several metrics available in our data set that we expect to be significantly different between groups. Pharmacologically, there are certain prescriptions that would be expected to be prescribed at higher rates to those with depression than those without. Depressed patients tend to be prescribed antipsychotics to augment their antidepressant medication. Anxiety is also a well-known highly comorbid condition in depression and thus many depressed patients would be expected to be prescribed antianxiety agents. We used medication order history to determine the percent of patients in each group that had ever received an antipsychotic or an antianxiety agent (Fig 3A). With severe mental illness, it is common that patients are first diagnosed with and treated for depression before receiving a diagnosis for disorders such as bipolar disorder and schizophrenia. We used the discharge diagnosis codes from outpatient, ED, and inpatient encounter records to calculate the percent of patients in each group that had ever received an ICD-9 code for bipolar disorder or schizophrenia (Fig 3B). Depressed patients are known to have higher rates of suicidality and higher mortality than non-depressed patients. We used the discharge diagnosis codes from outpatient, ED, and

**Fig. 1** Methodological designs using either (A) binary depression phenotype or (B) multiple depression phenotypes; Dep ICD9: blue, Comparison: orange, Rx no ICD9: red, Rx non-dep: yellow, No Dep: green
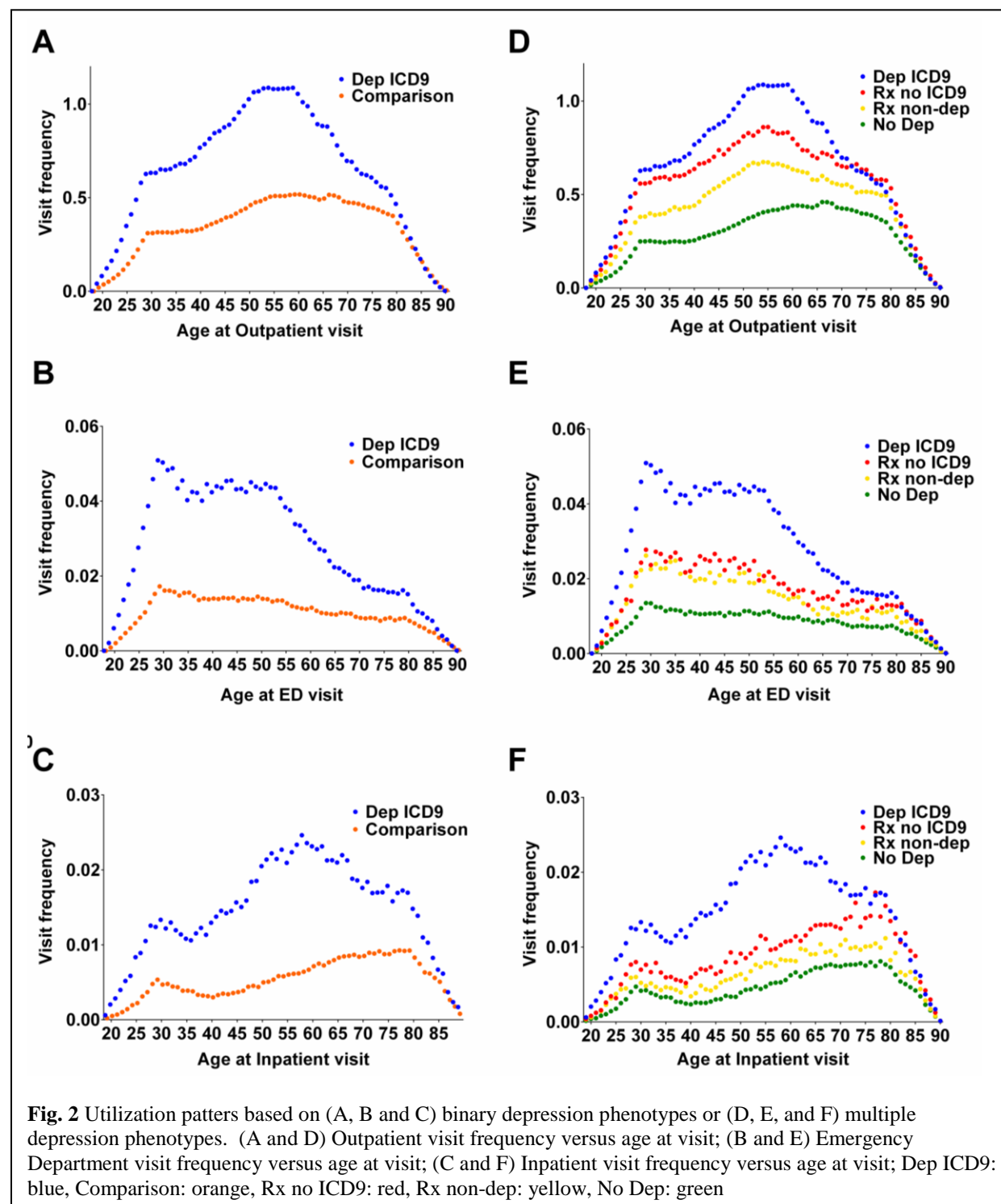
inpatient encounter records and determined the percent of patients in each group that had ICD-9 codes for suicides and suicide attempts (Fig 3C). We used the date of death to determine the percent of patients from each group that had a record of death in their EHR (Fig 3D). Substance abuse has also been associated with depression. We used the discharge diagnosis codes from outpatient, ED, and inpatient encounter records and determined the percent of patients in each group that had ICD-9 codes for alcohol and drug abuse or dependence (Fig 3E). Other studies have shown that depressed patients tend to have more serious comorbidities than non-depressed patients[36]. The Charlson Comorbidity Index (CCI) contains 19 categories of serious comorbidities and predicts the 10-year mortality of patients[37]. Using the discharge diagnoses from outpatient, ED, and inpatient encounter records during the study period, we calculated the CCI score for each patient and then the mean CCI score for each phenotype group (Fig 3F). In 2012, Geisinger Health System began implementing universal screening for depression with the Patient Health Questionnaire 2 (PHQ2). Upon affirming either of the two screener questions, patients are asked an additional 7 questions (PHQ9). The PHQ9 is a validated instrument for assessing current depression symptom severity and has a tiered rating scale based on total score (0: no depression, 1-9: mild, 10-14: moderate, 15-19: moderately severe, 20+: severe). For all those that had one or more PHQ2 or PHQ9 data in their EHR, we identified their maximum score and determined the percent of patients that received PHQ2 or PHQ9s in each group for each score (Fig 4).

## 3 Results
Here we present and validate a method for defining depression cohorts based on ICD-9 codes and medication orders which involves the identification of multiple depression phenotypes. Through convergent validity of EHR based outcomes and features, we show that this method clearly demonstrates superiority to a simple ICD-9 code based definition of depression and a comparison group that does not take antidepressant medication orders into account.

*3.1 Multiple depression phenotypes: definition and summary statistics.* There are many possible ways to define depression cohorts. A common method is to use a strict ICD-9 code based algorithm to define those with depression and use the remaining individuals that do not meet that criteria as a comparison control group. Applying this standard methodology (Fig 1A) to the 287,281-person patient population described here results in 20.6% (n=59,097) of the population meeting the definition of depression before or during the study period. A major critique of then comparing
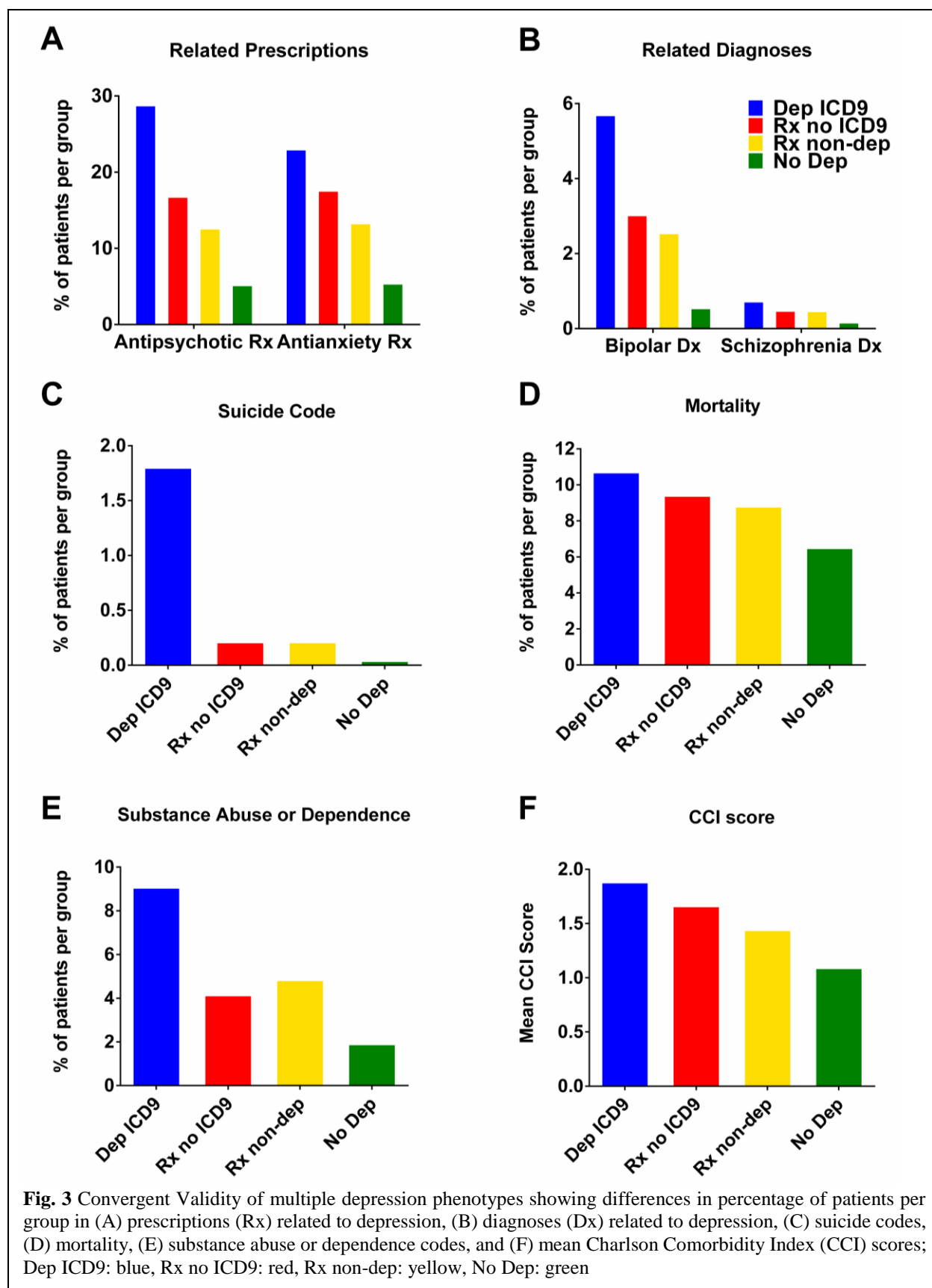
this group to a simple comparison group that do not meet that definition is that an additional 22% (n=63,155) of the total patient population have received at least one antidepressant order. This results in a "control" population that potentially has been diluted by patients with subclinical or undiagnosed depression which is likely to affect conclusions drawn from subsequent analyses. We then developed a method in which we identify multiple depression phenotypes that results in a total of four ordinal groups, varying in likelihood and severity of depression, named: "Dep ICD9", "Rx no ICD9", "Rx non-dep", and "No Dep" (Fig 1B). Of the total patient population, each group accounts for 20.6%, 11.2%, 10.8%, and 57.4%, respectively.



**Fig. 2** Utilization patters based on (A, B and C) binary depression phenotypes or (D, E, and F) multiple depression phenotypes. (A and D) Outpatient visit frequency versus age at visit; (B and E) Emergency Department visit frequency versus age at visit; (C and F) Inpatient visit frequency versus age at visit; Dep ICD9: blue, Comparison: orange, Rx no ICD9: red, Rx non-dep: yellow, No Dep: green

6

Summary statistics on patient sex, race, marital status, age at beginning of the study and length of known observation are shown for the total patient population and all possible groups described above in Fig. 1 (Table 1). The general patient population is 54.3% female, 95.8% white, 59.9% married, median age 45, and median known length of observation is just less than 7 years. There are expected as well as potentially confounding differences between the various groups. It is well known that depression occurs more commonly in females than males, which is reflected by the ratio in each group: Dep ICD9 (2.2:1), Rx no ICD9 (1.8:1), Rx non-dep (1.4:1), and No Dep (0.9:1). The Comparison group has a ratio of 1:1, which confirms the masking of higher ratios seen when Rx no ICD9 and Rx non-dep phenotypes are extracted. The Dep ICD9 group is the only group that has a large difference in marital status of married (51.7%) when compared with the No Dep (62.8%) and lesser depression phenotypes Rx no ICD9 (61.2%) and Rx non-dep (58.3%). Age at the beginning of the study does not differ dramatically between phenotype groups, however length of observation does. Because length of observation differs dramatically in the ratio of patients in each group between those seen 1 year or less to those seen 7 or more years, this measure should be calculated and accounted for when performing analyses using this method. The ratios for the multiple depression phenotypes are as follows: Dep ICD9 (10.5:1), Rx no ICD9 (24.4:1), Rx non-dep (5.3:1), and No Dep (2.2:1). The ratio for the comparison group is 3.1:1, which completely masks the very large difference we see in the Rx no ICD9 group, which by definition of having two or more antidepressant orders during the study period increases the likelihood that they have been seen more times over a longer period. Similarly, part of the definition of Dep ICD9 requires two or more outpatient ICD-9 codes during a 2-year period and thus many who meet that definition will inherently have been seen more than once, increasing the likelihood that they were seen for a longer period of time.

*3.2 Healthcare system utilization patterns differ based on depression phenotypes.* Next, we calculated the age of each patient at each outpatient, ED, and inpatient visit and calculated the visit frequency standardized by the total number of patients in each group. The dynamic open cohort design of this particular study allows for a system service area perspective of a patient population, however based on the assumptions and study objectives, others may impose alternative inclusion criteria in order to be able to calculate person years and conduct more traditional epidemiological studies. Here we make the imperfect assumption that all patients contribute to the entire 10.75-year period to demonstrate the general trends of healthcare system utilization patterns in these three settings and compare the simple binary method of depression cohort definition (Fig 1A) with the multiple depression phenotype method (Fig 1B). Outpatient visit utilization (Fig 2A and D) demonstrates the most striking differences between the groups Dep ICD9, Rx no ICD9, Rx non-dep, and No Dep. A clear but minor delineation between outpatient visit frequency exists between ages 25 through 75 for Dep ICD9 and Rx no ICD9, but that difference diminishes at age 75 and above. This distinction of populations and their utilization patters would be completely obscured by including them in the simple Comparison group. In addition, the clear distinction between the Rx non-dep and No Dep group through the majority of ages suggests that this is also a sufficiently different, higher utilizing population than the No Dep group. For ED visit frequency (Fig 2 B and E) and inpatient visit frequency (Fig 2C and F), we see greater differences between the Dep ICD9 and Rx no ICD9/Rx non-dep groupings. For ED visit frequency, Rx no ICD9 and Rx non-dep are indistinguishable from one another, falling at approximately half the visit frequency of Dep ICD9 until around age 65. However, Rx no ICD9 and Rx non-dep are almost twice as high as the No Dep group between ages 18 through 55. The lowest over all visit frequency for all three types of utilization, as expected, is inpatient utilization. While the Dep ICD9 group clearly has much higher visit frequency than any of the other groups until after the age of 75, the Rx no ICD9 group, similar to the outpatient setting, gradually increases and reaches the same frequency as the Dep ICD9 group in these later years. The Rx non-dep group has a minor but clear intermediate visit frequency between Rx no ICD9 and No Dep, both of which are lost when they are included in one Comparison group. The peak at age 30 in all groups is due to women during pregnancy (gendered data not shown). These results demonstrate the utility of using multiple depression phenotypes to define patient cohorts and highlight a few ages and settings in which this methodology would improve study designs by providing a cleaner control group.

*3.3 Convergent validity of medication orders, comorbidities, mortality, and symptom severity.* Finally, we analyzed several EHR-based features that would be expected to differ between groups if they were truly creating ordinal categorizations of multiple depression phenotypes. The percent of patients in each group that ever received an antipsychotic (28.8, 16.8, 12.6, 5.2) or antianxiety agent (23.0, 17.6, 13.3, 5.4) decreases from Dep ICD9 to No Dep as expected (Fig 3A). The percentage of patients from each group that received at least one bipolar ICD-9 diagnosis code decreased substantially between the Dep ICD9 group (5.7) and the Rx no ICD9 and Rx non-dep groups (3.0 and 2.6, respectively) (Fig 3B). As expected, all of these were much higher than the No Dep group (.6). Schizophrenia diagnosis codes were much lower percentages but were indeed higher in the Dep ICD9, Rx no ICD9, and Rx non-dep groups (0.7, 0.5, 0.5, respectively), compared with the No Dep group (0.2). Almost all suicide ICD-9 codes were
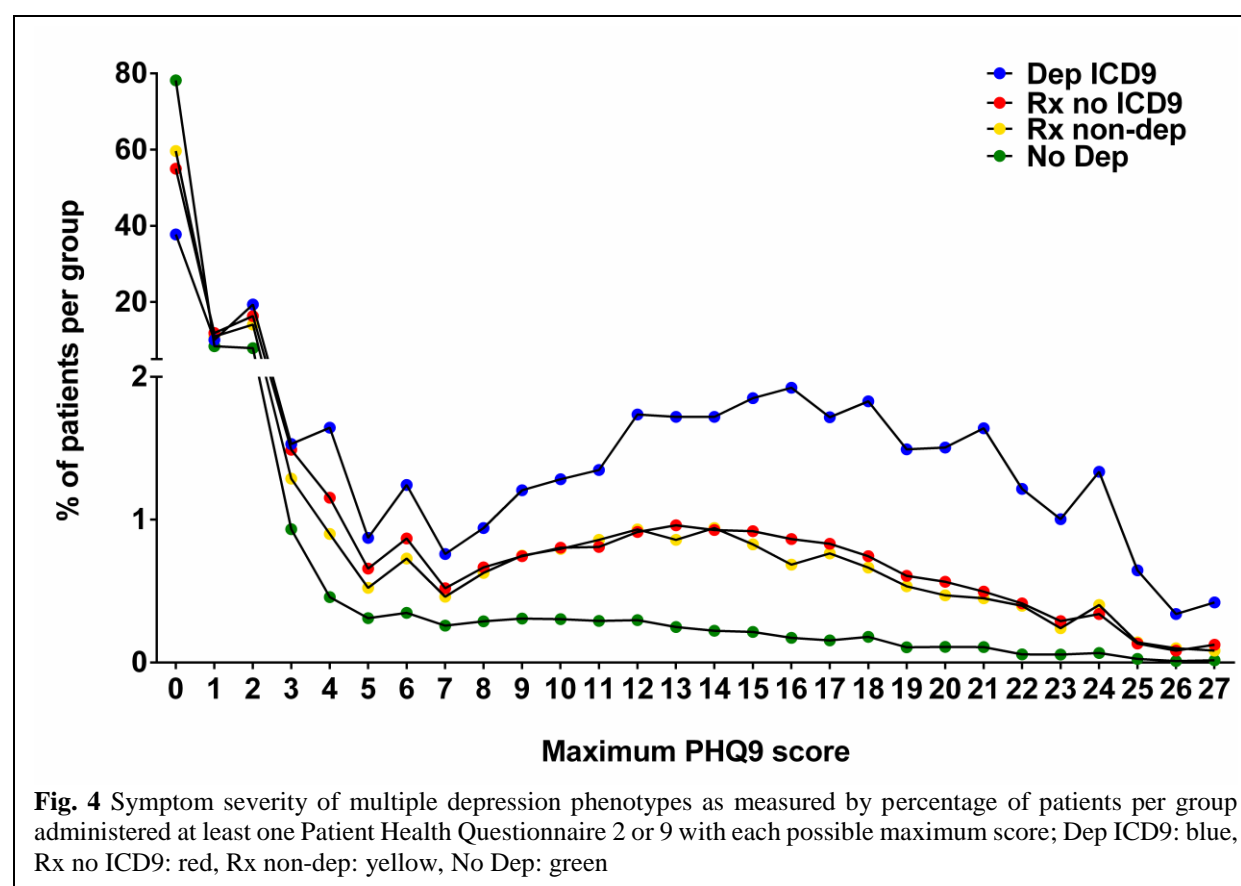
7

**Fig. 3** Convergent Validity of multiple depression phenotypes showing differences in percentage of patients per group in (A) prescriptions (Rx) related to depression, (B) diagnoses (Dx) related to depression, (C) suicide codes, (D) mortality, (E) substance abuse or dependence codes, and (F) mean Charlson Comorbidity Index (CCI) scores; Dep ICD9: blue, Rx no ICD9: red, Rx non-dep: yellow, No Dep: green

found in the records of patients in the Dep ICD9 group (n=1063; 1.8%), while the Rx no ICD9 and Rx non-dep groups each had less than a quarter of a percent of patients with suicide codes (Fig 3C). This is consistent with the literature associating suicidality with severe depression. Related to both suicidality and associated worse outcomes in those with depression, we also see differences between each group and the percentage of patients with a date of death. Percent mortality was 10.7, 9.4, 8.8, and 6.5, respective to groups Dep ICD9, Rx no ICD9, Rx non-dep, and No Dep (Fig 3D). Substance abuse and dependence, here defined as any drug or alcohol excluding tobacco, is also highly associated with depression and is highest in Dep ICD9 with 9.1% of patients having at least one code during the study period (Fig 3E). Rx no ICD9 has slightly fewer patients with such codes (4.1%) when compared to Rx non-dep (4.8%), but this is not terribly surprising as one of the main criteria for group Rx non-dep was that they were prescribed antidepressants during visits with an ICD-9 code of 503.1, tobacco use disorder and it is common for patients to have more than one type of substance use disorder. The No Dep group, in contrast had only 1.9% of patients with substance use disorder codes. Also, consistent with other literature, we find that mean Charlson Comorbidity Index scores are higher depending on depression severity group (Fig 3F). Dep ICD9, Rx no ICD9, Rx non-dep, and No Dep groups had mean scores and standard error of 1.88±0.01, 1.66±0.01, 1.44±0.01, and 1.09±0.005, respectively.

A subset of the study population has been administered the validated depression tool the Patient Health Questionnaire (PHQ), either the 2-question or 9-question version. Some patients had been assessed more than once and so we identified the maximum score in their EHR and calculated the percent of patients having received at least one PHQ2/9 for each group with a maximum score of each possible score (Fig 4). Interestingly, patients in Dep ICD9 were the least likely to have been screened (32.1%; n=18961), followed by those in Rx no ICD9 (32.3%; n=10367), Rx non-dep (38.4%; n=11922), and finally No Dep (44.2%; n=72909). This may be due to the fact that patients with overt symptoms of depression are not being routinely screened with tools designed to detect depression but are not diagnostic and instead are given more thorough depression inventories and diagnostic assessments. Despite the higher proportion of patients in the No Dep group receiving the PHQ9, most them scored a maximum of 0, indicating no depression (78.2%), and only 2.6% scored 10 or higher, indicating moderate, moderately severe, or severe depression. Those in the Rx non-dep and Rx no ICD9 groups were remarkably similar at all possible scores, only deviating from each other by 0.01% to 4.65%, with 59.6% and 55.0% having a score of zero and 10.1% and 10.8% having scores of 10 or more. This is in great contrast to the Dep ICD9 group which had 37.7% of patients with a maximum score of 0 and nearly a quarter (24.7%) of patients with a maximum score of 10 or above.

## 4 Discussion

*4.1 Outcomes.* When we examine healthcare system utilization outcomes including ED, inpatient, and outpatient visits, our method of defining multiple depression phenotypes shows clear differences in those who are classified with depression (Dep ICD9) compared to those who do not have depression (No Dep). We also see a separation between Rx no ICD9 (those who likely have depression based on multiple antidepressant medication orders for depression associated reasons) and Rx non-dep (those who have a single antidepressant medication order or have clear discharge diagnosis indicating a likely off-label use), particularly when looking at outpatient visit frequency. Thus, the patients we have identified as likely depressed based on medication use do utilize outpatient services differently than those without depression (No Dep) and those with antidepressant medications likely given for off-label reasons (Rx non-dep). Furthermore, because there are differences between Dep ICD9 and Rx no ICD9 groups based on utilization and other outcome measures, it is possible that Dep ICD9 is identifying a more severe depressed patient cohort group compared to Rx no ICD9. Physicians may be more likely to give ICD-9 diagnosis codes to patients that have more severe symptoms compared to patients given antidepressants but not a diagnosis code. Interestingly, there is a pattern that emerged when examining these outcomes at different ages. Later in life, the two groups that identify those with depression (Dep ICD9 and Rx no ICD9) follow a similar pattern in utilization. This indicates that for older patients, multiple medication orders and ICD-9 diagnosis identify patients that are likely depressed and engaging in similar frequency of inpatient visits compared to those without steady medication use or non-depressed individuals.

Particularly striking is the difference between the standard Comparison group (Fig 1A) and No Dep (Fig 1B: the control group in the multiple depression phenotypes method) when comparing outcomes. The outcomes in the control group change when you separate those who have medication orders into another group. This is important to consider when defining patient groups and why the multiple depression phenotype method we are proposing is a better method to use if controls without depression are the targeted comparison group. These data show that the Comparison group had higher utilization compared to No Dep, likely due to those with depression being included in the control group. In addition, the outcomes between Dep ICD9 and the Comparison group in older cohorts is practically identical, particularly in inpatient visits by age 70. When you compare this to the multiple depression phenotype method it is clear that Rx no ICD9 (patients on antidepressants) may be contributing to the apparently similar utilization rates in these two groups. By not removing those with antidepressant medication orders from the control group, the results may be obscuring the truth. We argue these data show that it is important to use the multiple depression phenotypes method when researchers are aiming to examine the impact depression has on EHR outcomes. By failing to remove those with antidepressant medication orders from a control group, researchers may come to the wrong conclusions, and at a minimum, the impact depression has may be masked by imprecise phenotyping.



**Fig. 4** Symptom severity of multiple depression phenotypes as measured by percentage of patients per group administered at least one Patient Health Questionnaire 2 or 9 with each possible maximum score; Dep ICD9: blue, Rx no ICD9: red, Rx non-dep: yellow, No Dep: green

*4.2 Convergent validity.* When examining this methodology, it is important to determine convergent validity to examine whether groups are identified appropriately. We have used multiple measures to do this and as reported earlier, outcomes are clearly different between those with clear depression and the controls. Furthermore, as discussed above, pulling out those with antidepressant medication orders (Rx no ICD9 and Rx non-dep) lowers control group utilization, supporting our methods. In addition, psychotropic medication use is higher in those with clear depression versus controls and removing those with medication orders reduces control group psychotropic medication use further. While there are differences in psychotropic medication orders between Rx no ICD9 and Rx non-dep, these differences are not as large. There does not seem to be a large difference in Schizophrenia diagnosis, but Dep ICD9 and No Dep groups differ dramatically for bipolar disorder diagnosis. This confirms that Dep ICD9 is likely identifying a more severely depressed cohort; those with bipolar disorder often are identified as depressed first and present to healthcare settings during depressive episodes[38].

10

When looking at suicide codes in the EHR, there are clear differences between Dep ICD9 and the other groups. Patients with only medication orders or in the No Dep group have very few suicide attempts. These results may also confirm that those in Dep ICD9 are likely experiencing more severe depression. There are differences in mortality, particularly between the Dep ICD9 and No Dep groups and it is important to keep in mind the sample size; changes in percentage scores by group indicate large numbers of patients, which is striking when considering how depression may impact mortality. Additionally, there tend to be more comorbidities in those with depression. Finally, when examining PHQ9 maximum score we see clear differences between Dep ICD9, those with antidepressants only, and No Dep groups, confirming that the multiple depression phenotype method is identifying distinctly different groups and that it may be better to remove those with medication orders from the Comparison group. These results also confirm that those with ICD-9 diagnoses are likely experiencing more severe depression compared to those with only medication orders. These results combined show good convergent validity, confirming that we have likely identified patient cohorts with varying levels of depression severity correctly using this method.

*4.3 Limitations and future directions.* While the size of the population in this study and access to variables allows us to perform in depth analysis using a variety of methods on a large clinical sample, there are several limitations to keep in mind. The sample was mostly made up of white patients from rural communities from central Pennsylvania, limiting the generalizability of our findings. In addition, there are other healthcare systems in the service area and these data are likely not capturing all utilization by patients, even though patients were limited to include those with a Geisinger Primary Care Physician. Claims data may provide more detailed information about healthcare utilization. As noted above, defining patient cohorts is difficult and based on indicators in the EHR, there is no way for us to know how diagnoses were reached, severity of patients, and whether those with antidepressant medication orders are truly meeting criteria for a depressive disorder. While our methods do indicate that our control group likely contains a cohort relatively free from those with severe depression, it is possible that this group still contains patients with depression that has not been identified by the healthcare system. Presentation bias due to extreme symptoms and/or social stigma may prevent those with severe depression from seeking medical care. Universal depression screening that has been implemented more recently (and used in this study when possible for convergent validity) will likely better identify patients with depression and in future research, identifying depressed patient cohorts should take this data into consideration if it is available for use. Using our methodology in future studies will help determine its usefulness for the field, can add further nuance and clarity in defining this patient cohort, and can help us understand the impact depression has on the healthcare system and on patient outcomes to better address these issues.

*4.5 Conclusions.* These methodology and results have implications for defining depressed cohorts and assessing outcomes in the EHR. The multiple depression phenotypes methodology we are recommending is useful for researchers who want to examine depressed patient outcomes compared to clean control populations. For researches who want to look at nuanced differences and take depression severity into account, it appears that the multiple depression phenotypes method is identifying different groups between those with just medication orders and those with an ICD-9 diagnosis, with one big difference being that those with an ICD-9 diagnosis are likely more severe and have worse outcomes. There seem to be some differences between those with two antidepressant medication orders versus those with only one antidepressant medication order or off label use, though this is not seen in all outcomes. Thus, depending on the research question, it may be prudent to separate Rx no ICD9 and Rx non-dep (or exclude Rx non-dep) while in other cases it may be fine combining those groups. Researchers can use this multiple depression phenotype method in a few different ways: comparing Dep ICD9 and No Dep for those with clear depression versus clean controls, or a more nuanced method comparing Dep ICD9, Rx no ICD9, Rx non-dep, and No Dep. We argue that a version of the multiple depression phenotype method is superior to the method often used identifying those with ICD-9 codes and a default control Comparison group. The multiple depression phenotype method allows researchers to use the combination of groups that makes the most sense for their targeted outcomes. Lastly, this method is easy for researchers to use, clearly defined, shows good convergent validity, and would work with many EHR systems, so long as they contain linked ICD codes and medication orders for individual patients. Preliminary analyses show worse outcomes in patients with depression, underscoring the need for continued research on patients with depression using clear, validated methodology.

**References**
1.    Morley KI, Wallace J, Denaxas SC, Hunter RJ, Patel RS, Perel P, Shah AD, Timmis AD, Schilling RJ, Hemingway H (2014) Defining Disease Phenotypes Using National Linked Electronic Health Records: A Case Study of Atrial Fibrillation. PLoS One 9: . doi: 10.1371/journal.pone.0110900

2.  Affolter K, Gligorich K, Samadder NJ, Samowitz WS, Curtin K (2017) Feasibility of Large-Scale Identification of Sessile Serrated Polyp Patients Using Electronic Records: A Utah Study. Dig Dis Sci 6:1455–1463 . doi: 10.1007/s10620-017-4543-9

3.  Newton-Dame R, McVeigh KH, Schreibstein L, Perlman S, Lurie-Moroni E, Jacobson L, Greene C, Snell E, Thorpe LE (2016) Design of the New York City Macroscope: Innovations in Population Health Surveillance Using Electronic Health Records. eGEMs 4: . doi: 10.13063/2327-9214.1265

4.  Sallakh MA Al, Vasileiou E, Rodgers SE, Lyons RA, Sheikh A, Davies GA (2017) Defining asthma and assessing asthma outcomes using electronic health record data: a systematic scoping review. Eur Respir J 49:1700204 . doi: 10.1183/13993003.00204-2017

5.  McDonald HI, Shaw C, Thomas SL, Mansfield KE, Tomlinson LA, Nitsch D (2016) Methodological challenges when carrying out research on CKD and AKI using routine electronic health records. Kidney Int 90:943–949 . doi: 10.1016/j.kint.2016.04.010

6.  Cole AM, Stephens KA, Keppel GA, Estiri H, Baldwin L-M (2016) Extracting Electronic Health Record Data in a Practice-Based Research Network: Processes to Support Translational Research across Diverse Practice Organizations. eGEMs 4: . doi: 10.13063/2327-9214.1206

7.  Pathak J, Simon G, Li D, Biernacka JM, Jenkins GJ, Chute CG, Hall-Flavin DK, Weinshilboum RM (2014) Detecting Associations between Major Depressive Disorder Treatment and Essential Hypertension using Electronic Health Records. AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci 2014:91–96

8.  Huang SH, LePendu P, Iyer S V, Tai-Seale M, Carrell D, Shah NH (2014) Toward personalizing treatment for depression: predicting diagnosis and severity. J Am Med Inform Assoc 21:1–7 . doi: 10.1136/amiajnl-2014-002733

9.  Anderson HD, Pace WD, Brandt E, Nielsen RD, Allen RR, Libby AM, West DR, Valuck RJ (2015) Monitoring suicidal patients in primary care using electronic health records. J Am Board Fam Med 28:65–71 . doi: 10.3122/jabfm.2015.01.140181

10. Kobus AM, Harman JS, Do HD, Garvin RD (2013) Challenges to depression care documentation in an EHR. Fam Med 45:268–271

11. Madden JM, Lakoma MD, Rusinak D, Lu CY, Soumerai SB (2016) Missing clinical and behavioral health data in a large electronic health record (EHR) system. J Am Med Informatics Assoc 23:1143–1149 . doi: 10.1093/jamia/ocw021

12. Mayer MÁ, Gutiérrez Sacristán A, Leis Machin A, De La Peña S, Sanz F, Furlong LI (2017) Using Electronic Health Records to Assess Depression and Cancer Comorbidities. doi: http://dx.doi.org/10.3233/978-1-61499-753-5-236

13. Gill JM, Chen YX, Lieberman MI (2008) Management of Depression in Ambulatory Care for Patients with Medical Co-Morbidities: A Study from a National Electronic Health Record (EHR) Network. Int J Psychiatry Med 38:203–215 . doi: 10.2190/PM.38.2.g

14. Hamilton M (1967) Development of a Rating Scale for Primary Depressive Illness. Br J Soc Clin Psychol 6:278–296 . doi: 10.1111/j.2044-8260.1967.tb00530.x

15. Kroenke K, Spitzer RL, Williams JBW (2001) The PHQ-9. J Gen Intern Med 16:606–613 . doi: 10.1046/j.1525-1497.2001.016009606.x

16. (2009) Screening for depression in adults: U.S. preventive services task force recommendation statement. Ann Intern Med 151:784–792 . doi: 10.7326/0003-4819-151-11-200912010-00006

17. Maurer DM (2012) Screening for Depression. Am Fam Physician 85:139–144

18. Thomas M, Price D (2008) Impact of comorbidities on asthma. Expert Rev Clin Immunol 4:731–742 . doi: 10.1586/1744666X.4.6.731

19. Cm C, Jc H (2011) Depression and cardiac disease: a review. Cardiol Rev 19:130–142 . doi: 10.1097/CRD.0b013e31820e8106

20. Ang DC, Choi H, Kroenke K, Wolfe F (2005) Comorbid depression is an independent risk factor for mortality in patients with rheumatoid arthritis. J Rheumatol 32:1013–1019

21. Young BA, Von Korff M, Heckbert SR, Ludman EJ, Rutter C, Lin EHB, Ciechanowski PS, Oliver M, Williams L, Himmelfarb J, Katon WJ (2010) Association of major depression and mortality in Stage 5 diabetic chronic kidney disease. Gen Hosp Psychiatry 32:119–124 . doi: 10.1016/j.genhosppsych.2009.11.018

22. Pouwer F, Nefs G, Nouwen A (2013) Adverse Effects of Depression on Glycemic Control and Health Outcomes in People with Diabetes. Endocrinol Metab Clin North Am 42:529–544 . doi: 10.1016/j.ecl.2013.05.002

23. Slee VN (1978) The International Classification of Diseases: Ninth Revision (ICD-9). Ann Intern Med 88:424

. doi: 10.7326/0003-4819-88-3-424

24.  Gill JM, Klinkman MS, Chen YX (2010) Antidepressant medication use for primary care patients with and without medical comorbidities: a national electronic health record (EHR) network study. J Am Board Fam Med JABFM 23:499–508 . doi: 10.3122/jabfm.2010.04.090299

25.  Depression. | National Guideline Clearinghouse

26.  Farmer MM, Rubenstein L V., Sherbourne CD, Huynh A, Chu K, Lam CA, Fickel JJ, Lee ML, Metzger ME, Verchinina L, Post EP, Chaney EF (2016) Depression Quality of Care: Measuring Quality over Time Using VA Electronic Medical Record Data. J Gen Intern Med 31:36–45 . doi: 10.1007/s11606-015-3563-4

27.  Arroll B, Chin W, Martis W, Goodyear-Smith F, Mount V, Kingsford D, Humm S, Blashki G, MacGillivray S (2016) Antidepressants for treatment of depression in primary care: a systematic review and meta-analysis. J Prim Health Care 8:325–334 . doi: 10.1071/HC16008

28.  Wong J, Motulsky A, Abrahamowicz M, Eguale T, Buckeridge DL, Tamblyn R (2017) Off-label indications for antidepressants in primary care: descriptive study of prescriptions from an indication based electronic prescribing system. BMJ 356:j603 . doi: 10.1136/bmj.j603

29.  Wong J, Motulsky A, Eguale T, Buckeridge DL, Abrahamowicz M, Tamblyn R (2016) Treatment Indications for Antidepressants Prescribed in Primary Care in Quebec, Canada, 2006-2015. JAMA 315:2230–2232 . doi: 10.1001/jama.2016.3445

30.  Carey DJ, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL, Mirshahi U, Murray MF, Smelser DT, Gerhard GS, Ledbetter DH (2016) The Geisinger MyCode community health initiative: An electronic health record-linked biobank for precision medicine research. Genet Med 18:906–913 . doi: 10.1038/gim.2015.187

31.  Borthwick K, Smelser D, Bock J, Elmore J, Ryer E, Ye Z, Pacheco J, Carrell D, Michalkiewicz M, Thompson W, Pathak J, Bielinski S, Denny J, Linneman J, Peissig P, Kho A, Gottesman O, Parmar H, Kullo I, McCarty C, Böttinger E, Larson E, Jarvik G, Harley J, Bajwa T, Franklin D, Carey D, Kuivaniemi H, Tromp G (2015) Ephenotyping for Abdominal Aortic Aneurysm in the Electronic Medical Records and Genomics (emerge) Network: Algorithm Development and Konstanz Information Miner Workflow. Int J Biomed Data Min 4:1–8 . doi: 10.4172/2090-4924.1000113

32.  Maeng DD, Han JJ, Fitzpatrick MH, Boscarino JA (2017) Patterns of health care utilization and cost before and after opioid overdose: findings from 10-year longitudinal health plan claims data. Subst Abuse Rehabil 8:57–67 . doi: 10.2147/SAR.S135884

33.  Boscarino JA, Kirchner HL, Pitcavage JM, Nadipelli VR, Ronquest NA, Fitzpatrick MH, Han JJ (2016) Factors associated with opioid overdose: a 10-year retrospective study of patients in a large integrated health care system. Subst Abuse Rehabil 7:131–141 . doi: 10.2147/SAR.S108302

34.  Reid RD, Pritchard G, Walker K, Aitken D, Mullen K-A, Pipe AL (2016) Managing smoking cessation. CMAJ 188:E484–E492 . doi: 10.1503/cmaj.151510

35.  Chong MS, Bajwa ZH (2003) Diagnosis and treatment of neuropathic pain. J Pain Symptom Manage 25:S4–S11

36.  Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B (2007) Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet 370:851–858 . doi: 10.1016/S0140-6736(07)61415-9

37.  Deyo RA, Cherkin DC, Ciol MA (1992) Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 45:613–619 . doi: 10.1016/0895-4356(92)90133-8

38.  Kessing LV, Willer I, Andersen PK, Bukh JD (2017) Rate and predictors of conversion from unipolar to bipolar disorder: A systematic review and meta-analysis. Bipolar Disord 19:324–335 . doi: 10.1111/bdi.12513