MINI-SENTINEL SYSTEMATIC EVALUATION OF HEALTH OUTCOME OF INTEREST DEFINITIONS FOR STUDIES USING ADMINISTRATIVE DATA

DEPRESSION REPORT

Prepared by: Lisa Townsend, PhD,1 James T Walkup, PhD,1 Stephen Crystal, PhD,1 and Mark Olfson, MD, MPH2


December 8, 2010

Mini-Sentinel is a pilot project sponsored by the U.S. Food and Drug Administration (FDA) to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. Mini-Sentinel is one piece of the Sentinel Initiative, a multi-faceted effort by the FDA to develop a national electronic system that will complement existing methods of safety surveillance. Mini-Sentinel Collaborators include Data and Academic Partners that provide access to health care data and ongoing scientific, technical, methodological, and organizational expertise. The Mini-Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223200910006I.
Mini-Sentinel Systematic Evaluation Of Health Outcome Of Interest Definitions
For Studies Using Administrative Data

Depression Report

I. EXECUTIVE SUMMARY ........................................................................................................................................... 3
   A. OVERVIEW OF PROJECT .................................................................................................................................. 3
   B. SUMMARY OF FINDINGS ............................................................................................................................... 3
   C. RECOMMENDATION FOR ALGORITHMS AND SUGGESTION FOR FUTURE RESEARCH ........................................ 3

II. PROJECT OBJECTIVES ............................................................................................................................................. 4

III. BACKGROUND ......................................................................................................................................................... 4

IV. METHODS ................................................................................................................................................................. 4
   A. SEARCH STRATEGY .............................................................................................................................................. 4
   B. ABSTRACT REVIEW .............................................................................................................................................. 5
      1. Abstract Review Methods ............................................................................................................................... 5
      2. Abstract Exclusion Criteria ............................................................................................................................ 5
   C. FULL-TEXT REVIEW ........................................................................................................................................... 5
      1. Full-Text Review Methods ........................................................................................................................... 5
      2. Full-Text Exclusion Criteria ........................................................................................................................ 6
   D. MINI-SENTINEL INVESTIGATOR SURVEY ........................................................................................................ 6
   E. EVIDENCE TABLE CREATION .......................................................................................................................... 6
   F. CLINICIAN OR TOPIC-EXPERT CONSULTATION ............................................................................................ 6

V. RESULTS ........................................................................................................................................................................ 6
   A. SEARCH STRATEGY AND RESULTS .................................................................................................................. 6
   B. ABSTRACT REVIEWS ......................................................................................................................................... 12
   C. FULL-TEXT REVIEWS ....................................................................................................................................... 12
   D. MINI-SENTINEL INVESTIGATOR SURVEY ..................................................................................................... 13
   E. EVIDENCE INCLUDED IN TABLE ....................................................................................................................... 13
   F. SUMMARY AND DISCUSSION OF ALGORITHMS AND VALIDATION ............................................................. 13
   G. SUMMARY OF EXCLUDED POPULATIONS AND DIAGNOSES ..................................................................... 16
   H. EVIDENCE TABLE ............................................................................................................................................. 17
      I. TOPIC EXPERT CONSULTATION .................................................................................................................... 21

VI. SUMMARY AND CONCLUSIONS ............................................................................................................................ 22
   A. RECOMMENDATIONS FOR ALGORITHMS ....................................................................................................... 22
   B. SUGGESTIONS FOR FUTURE RESEARCH BASED ON EVIDENCE GAPS ....................................................... 23

VII. REFERENCES ............................................................................................................................................................. 24

VIII. APPENDICES ............................................................................................................................................................ 28
   A. APPENDIX A: ABSTRACTS OF STUDIES INCLUDED IN EVIDENCE TABLE .................................................. 28
   B. APPENDIX B: LIST OF CITATIONS SELECTED FOR FULL-TEXT REVIEW BUT NOT INCLUDED, BY REASONS FOR EXCLUSION ............................................................................................................................................... 33
      1. Studies Excluded Due to Poorly Defined Algorithms ................................................................................... 33
      2. Studies Excluded Due to Lack of Validation or Reporting of Validation Statistics ....................................... 43
   C. APPENDIX C: LIST AND DEFINITIONS OF ICD OR PROCEDURAL CODES INCLUDED IN ALGORITHMS ......... 53
   D. APPENDIX D: COMPLETE RESULTS FROM FRAYNE, ET AL. (12) .............................................................. 54
I. EXECUTIVE SUMMARY

A. OVERVIEW OF PROJECT

The Food and Drug Administration (FDA) Mini-Sentinel contract is a pilot program that aims to conduct active surveillance to detect and refine safety signals that emerge for marketed medical products. To perform this active surveillance, it is necessary to develop and understand the validity of algorithms for identifying health outcomes of interest in administrative data. Thus, the goal of this project was to identify algorithms used to detect selected health outcomes of interest using administrative data sources and describe the performance characteristics of these algorithms as reported by the studies in which they were used. This report summarizes the process and findings of the depression algorithm review.

B. SUMMARY OF FINDINGS

Incomplete recognition of depression in routine clinical practice constrains the performance of electronic health information to identify depressive disorders. No single algorithm demonstrated sufficiently strong operating characteristics to be recommended for general use. However, the results suggest that for some purposes including preliminary case finding, claims-based algorithms may have acceptable properties. Among adolescent outpatients with asthma, for example, 31.5% who were identified with a claims based algorithm met formal criteria for major depressive disorder or dysthymia (prevalence 7.5%). Sensitivities of approximately 50% were also achieved in two studies of claims-based algorithms in relation to independent assessments of depression.

Most adults who receive administrative codes for depression report depressive symptoms and a majority have notations of depression in their medical records. Because antidepressants are prescribed for a wide variety of psychiatric disorders and some general medical conditions, reliance on only claims for antidepressant medications tends to capture many patients without depression. The current value of administrative depression codes appears to be strongest for identifying selected cases with a reasonable probability of having clinically recognized depression and depressive symptoms.

C. RECOMMENDATION FOR ALGORITHMS AND SUGGESTION FOR FUTURE RESEARCH

No one algorithm can be confidently recommended for achieving high agreement with depression as measured by independent assessment. The highest agreement with clinically diagnosed depression was achieved by an algorithm that required over a 12 month period at least 2 first listed ICD-9 codes for 296.2 (major depressive episode, single episode), 296.3 (major depressive episode, recurrent episode), 300.4 (dysthymic disorder), or 311 (depression not elsewhere classified) as well as a filled prescription for an antidepressant medication (1). In a large population of primary care patients, the chance corrected agreement of this algorithm was moderate (Kappa: 0.464) (1, 2).

The validity of algorithms to identify independently assessed depression is ultimately dependent upon the quality of clinical depression case identification. For this reason, it is suggested that research should focus on practices that systematically screen for depression. Routine screening adults and youth for depression, though consistent with recommendations from the US Preventive Services Task Force (3, 4), is not currently widespread.
II. PROJECT OBJECTIVES

The primary objective of this project was to identify studies that have validated algorithms used to identify various health outcomes of interest (HOIs) using administrative data from the United States or Canada, and to summarize the results of those validation studies. If fewer than five validation studies were identified, a secondary objective was to identify non-validated algorithms that have been used to identify the HOIs using administrative data.

III. BACKGROUND

The Food and Drug Administration (FDA) Mini-Sentinel contract is a pilot program that aims to conduct active surveillance to detect and refine safety signals that emerge for marketed medical products. In order to perform this work, the program needed to identify algorithms used to detect various health outcomes of interest using administrative data sources and identify the performance characteristics of these algorithms as measured in the studies in which they were used. The data sources of interest were limited to those from the United States or Canada to increase their relevance to the Mini-Sentinel data sources, which are all from the United States. The Mini-Sentinel Protocol Core developed a preliminary list of approximately 140 potential health outcomes of interest, based on several criteria. These criteria included: 1) previous validation studies had been identified in a textbook chapter reviewing the validity of drug and diagnosis data used in pharmacoepidemiologic studies (5), 2) a list of designated medical events from a proposed FDA rule on the safety reporting requirements for human drug and biological products (6), 3) the Observational Medical Outcomes Partnership (OMOP) (5) had commissioned reports on algorithms used to identify the health outcome using administrative data (7).

From the original list of 140 HOIs, the Protocol Core worked with FDA to select 20 for which reviews of algorithms would be completed. HOIs for which OMOP had already commissioned reports were purposefully excluded in order to avoid duplication of effort.

Depression was one of the 20 HOIs selected for review. This report describes the review process and findings for the depression definition algorithms.

IV. METHODS

A. SEARCH STRATEGY

The general search strategy was developed based on prior work by OMOP and its contractors, and modified slightly for these reports. Originally, OMOP contracted with two organizations to perform reviews of 10 HOIs. Because the search strategies used by each organization resulted in very different sets of articles, OMOP investigators reviewed the PubMed indexing of the articles deemed useful in final reports and developed a strategy that would identify the majority of these citations while maintaining efficiency in the number of abstracts that would need to be reviewed. Mini-Sentinel investigators made minor changes to this strategy that would result in the identification of more citations, and confirmed empirically that the majority of relevant articles from one set of OMOP reports (angioedema) (8, 9) would be identified using this approach. The base search strategy was then combined with PubMed terms representing the HOIs. Medical subject heading (MeSH) terms were generally preferred as HOI search terms due to their likely specificity. Text word searches were sometimes used, particularly when the MeSH search resulted in a small number of citations for review. The workgroup also searched the
database of the Iowa Drug Information Service (IDIS) using a similar search strategy to identify other relevant articles that were not found in the PubMed search. For a limited number of outcomes where very few citations were identified from PubMed and IDIS searches, Embase searches were conducted. Search results were restricted to articles published on or after January 1, 1990.

University of Iowa investigators compiled the search results from different databases and eliminated duplicate results using a citation manager program. The results were then output into two sets of files, one containing the abstracts for review and the other for documenting abstract review results.

The search strategy and results for depression are detailed in the Results section. The PubMed search was conducted on May 14, 2010, and the IDIS searches on June 11, 2010.

B. ABSTRACT REVIEW

1. Abstract Review Methods

Each abstract was reviewed independently by two investigators to determine whether the full-text article should be reviewed. Exclusion criteria were documented sequentially (i.e. if exclusion criterion 1 was met then the other criteria were not documented). If the reviewers disagreed on whether the full-text should be reviewed, then it was selected for review. Inter-rater agreement on whether to include or exclude an abstract was calculated using a Cohen’s kappa statistic. The goal was to review any administrative database study that used data from the United States or Canada and studied the HOI, as validation components of studies are not necessarily included in the abstract and other relevant citations might be identified from the references of such studies.

2. Abstract Exclusion Criteria
   1. Did not study the HOI.
   2. Not an administrative database study. Eligible sources included insurance claims databases as well as other secondary databases that identify health outcomes using billing codes.
   3. Data source not from the United States or Canada.

C. FULL-TEXT REVIEW

1. Full-Text Review Methods

Full-text articles were reviewed independently by two investigators, with a goal of identifying validation studies described in the article itself or from the reference section of the article. Citations from the article’s references were selected for full-text review if they were cited as a source for the HOI algorithm, or were otherwise deemed likely to be relevant. Full-text review exclusion criteria were applied sequentially, since if fewer than 5 validation studies were identified, up to 10 of the articles excluded based on the second criterion would need to be incorporated into the final report. If there was disagreement on whether a study should be included, the two reviewers attempted to reach consensus on inclusion by discussion. If the reviewers could not agree, a third investigator would be consulted to make the final decision.
2. **Full-Text Exclusion Criteria**
   1. Poorly described HOI identification algorithm that would be difficult to operationalize.
   2. No validation of outcome definition or reporting of validity statistics.

D. **MINI-SENTINEL INVESTIGATOR SURVEY**

Mini-Sentinel investigators were surveyed to request information on any published or unpublished studies that validated an algorithm to identify an HOI in administrative data. Studies that would not be excluded by one of the aforementioned criteria were included in the final report.

E. **EVIDENCE TABLE CREATION**

A single investigator abstracted each study for the final evidence table. The data included in the table were confirmed by a second investigator for accuracy.

F. **CLINICIAN OR TOPIC-EXPERT CONSULTATION**

A clinician or topic-expert was consulted to review the results of the evidence table and discuss how they compare and contrast to diagnostic methods currently used in clinical practice. This included whether certain diagnostic codes used in clinical practice were missing from the algorithms, and the appropriateness of the validation definitions compared to diagnostic criteria currently used in clinical practice. A summary of this consultation was included in the results.

V. **RESULTS**

A. **SEARCH STRATEGY AND RESULTS**

The following summarizes the search results obtained from PubMed and IDIS searches. The PubMed search identified 1664 citations (Table 1), and the two IDIS searches identified 138 unique citations (Table 2). The total number of unique citations from the combined searches was 1731. An additional PubMed search was conducted at a later date to amend the original search strategy with names of relevant databases that were not included in the original search. This search identified 30 citations (Table 3).
Table 1. PubMed Search Strategy and Results: Performed on 05/14/10

Results = 1164

<table>
<thead>
<tr>
<th>Search</th>
<th>Query</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>(((&quot;Depression&quot;[Mesh] OR &quot;Depressive Disorder&quot;[Mesh:NoExp]) OR &quot;Depressive Disorder, Major&quot;[Mesh]) OR &quot;Dysthmic Disorder&quot;[Mesh]) OR &quot;Seasonal Affective Disorder&quot;[Mesh]</td>
<td>115292</td>
</tr>
<tr>
<td>#2</td>
<td>Limits: Humans, English, Publication Date from 1990/01/01 to 2010/06/01</td>
<td>6118555</td>
</tr>
<tr>
<td>#</td>
<td>Search Query</td>
<td>Results</td>
</tr>
<tr>
<td>----</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>#6</td>
<td>Search #1 and #3 and #4 Limits: Humans, English, Publication Date from 1990/01/01 to 2010/06/01</td>
<td>2840</td>
</tr>
<tr>
<td>#7</td>
<td>Search #6 not #5 Limits: Humans, English, Publication Date from 1990/01/01 to 2010/06/01</td>
<td>1664</td>
</tr>
</tbody>
</table>
### Table 2. IDIS Search Strategy and Results: Performed on 06/11/10

**Results** = 138

<table>
<thead>
<tr>
<th>Search 1: 17 Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADVANCED SEARCH</td>
</tr>
<tr>
<td><strong>All Fields:</strong></td>
</tr>
</tbody>
</table>
| "Premier" OR "Solucient" OR "Cerner" OR "Ingenix" OR "LabRx" OR "IHCIS" OR "marketscan" OR "market scan" OR "Medstat" OR "Thomson" OR "pharmetrics" OR "healthcare" OR "united healthcare" OR "UnitedHealthcare" OR "UHC" OR "GPRD" OR "general practice research database" OR "Research Database" OR "Group Health" OR "HCUP" OR "(Healthcare Cost" AND "Utilization Project") OR "(Health Care Cost" AND "Utilization Project") OR "MEPS" OR "Medical Expenditure Panel Survey" OR "NAMCS" OR "National Hospital Ambulatory Medical Care Survey" OR "National Ambulatory Medical Care Survey" OR "NHIS" OR "National Health Interview Survey" OR "Kaiser" OR "HMO Research" OR "Health Maintenance Organization" OR "HMO" OR "Cleveland Clinic" OR "Lovelace" OR "Department of Defense" OR "Henry Ford" OR "(Denmark" AND "Epidemiology)" OR "i3 Drug Safety" OR "i3" OR "Aetna" OR "Humana" OR "Wellpoint" OR "IMS" OR "Intercontinental Marketing Services" OR "IMS Health" OR "Geisinger" OR "GE Healthcare" OR "MQIC" OR "PHARMO" OR "Institute for Drug Outcome Research" OR "Pilgrim" OR "Puget Sound" OR "Regenstrief" OR "Saskatchewan" OR "Tayside" OR "MEMO" OR "Medicines Monitoring Unit" OR "Veterans Affairs" OR "Partners Healthcare" OR "Mayo Clinic" OR "Rochester Epidemiology" OR "Indiana Health Information Exchange" OR "Indiana Health" OR "Intermountain" OR "THIN" OR "The health improvement network" OR "blue cross" OR "health partners" OR "health plan" OR "health services" OR "Nationwide Inpatient Sample" OR "National Inpatient Sample" OR "medicaid" OR "medicare" OR "MediPlus" OR "Outcome Assessment" OR "insurance database" OR "insurance databases" OR "Data Warehouse" OR "ICD-9" OR "International statistical classification" OR "International classification of diseases" OR "ICD-10" OR "Database Management Systems" OR "Medical Records Systems, Computerized" OR "CPT" OR "Current procedural terminology" OR "drug surveillance" OR "(claims" AND "administrative") OR "(data" AND "administrative") OR "Databases, Factual" OR "Databases" OR "Medical Record Linkage" OR "ICD-9-CM" OR "ICD-10-CM"

**AND Descriptor:**

"SIDE EF PSYCHIATRIC 88" not ("CASE REPORT ADULT 0" or "FDA APPOVAL PACKAGE 155" OR "FDA BLACK BOX WARNING 165" OR "PIVOTAL STUDY 162" OR "FDA ADVISORY COMMITTEE 164" or "CASE REPORT PEDIATRIC 1" or "CASE REPORT GERIATRIC 2" or "REVIEW ADULT 6" or "STUDY NON-CLINICAL 8" or "REVIEW PEDIATRIC 21" or "REVIEW GERIATRIC 23" or "STUDY RANDOMIZE ADULT 135" or "STUDY RANDOMIZE PEDIATRIC 136" or "STUDY RANDOMIZE GERIATRIC 137" or "CROSS-OVER 144" or "META-ANALYSIS 145" or "N-OF-ONE TRIAL 146" or "PRACTICE GUIDELINE 156" or "SYSTEMATIC REVIEW 161" or "ANNOTATED BIBLIOGRAPHY 167" or "PRIORITY CLIN PRACT GUIDE 168")

**AND NOT Author:**

"Editorial" OR "(Letter to Ed)"

**AND Abstract:**

Depression

Years: 1990-2010

Records = 17

| Search 2: 128 Results |

---

HOI Evidence Reviews - 9 - Depression Report
ADVANCED SEARCH

All Fields:

"Premier" OR "Solucient" OR "Cerner" OR "Ingenix" OR "LabRx" OR "IHCIS" OR "marketscan" OR "market scan" OR "Medstat" OR "Thomson" OR "pharmetrics" OR "healthcare" OR "united healthcare" OR "UnitedHealthcare" OR "UHC" OR "GPRD" OR "general practice research database" OR "Research Database" OR "Group Health" OR "HCUP" OR ("Healthcare Cost" AND "Utilization Project") OR ("Health Care Cost" AND "Utilization Project") OR "MEPS" OR "Medical Expenditure Panel Survey" OR "NAMCS" OR "National Hospital Ambulatory Medical Care Survey" OR "National Ambulatory Medical Care Survey" OR "NHIS" OR "National Health Interview Survey" OR "Kaiser" OR "HMO Research" OR "Health Maintenance Organization" OR "HMO" OR "Cleveland Clinic" OR "Lovelace" OR "Department of Defense" OR "Henry Ford" OR ("Denmark" AND "Epidemiology") OR "i3 Drug Safety" OR "i3" OR "Aetna" OR "Humana" OR "Wellpoint" OR "IMS" OR "Intercontinental Marketing Services" OR "IMS Health" OR "Geisinger" OR "GE Healthcare" OR "MQIC" OR "PHARMO" OR "Institute for Drug Outcome Research" OR "Pilgrim" OR "Puget Sound" OR "Regenstrief" OR "Saskatchewan" OR "Tayside" OR "MEMO" OR "Medicines Monitoring Unit" OR "Veterans Affairs" OR "Partners Healthcare" OR "Mayo Clinic" OR "Rochester Epidemiology" OR "Indiana Health Information Exchange" OR "Indiana Health" OR "Intermountain" OR "THIN" OR "The health improvement network" OR "blue cross" OR "health partners" OR "health plan" OR "health services" OR "Nationwide Inpatient Sample" OR "National Inpatient Sample" OR "medicaid" OR "medicare" OR "MediPlus" OR "Outcome Assessment" OR "insurance database" OR "insurance databases" OR "Data Warehouse" OR "ICD-9" OR "international statistical classification" OR "international classification of diseases" OR "ICD-10" OR "Database Management Systems" OR "Medical Records Systems, Computerized" OR "CPT" OR "Current procedural terminology" OR "drug surveillance" OR ("claims" AND "administrative") OR "data" OR "Databases, Factual" OR "Databases" OR "Medical Record Linkage" OR "ICD-9-CM" OR "ICD-10-CM"

AND Disease:

300.4 or 311. (note: DISORDER, DEPRESSIVE NEC 311. DEPRESSION, NEUROTIC 300.4)

and not Descriptor: "CASE REPORT ADULT 0" or “FDA APPROVAL PACKAGE 155” OR “FDA BLACK BOX WARNING 165” OR "PIVOTAL STUDY 162" OR "FDA ADVISORY COMMITTEE 164" or "CASE REPORT PEDIATRIC 1" or "CASE REPORT GERIATRIC 2" or "REVIEW ADULT 6" or "STUDY NON-CLINICAL 8" or "REVIEW PEDIATRIC 21" or "REVIEW GERIATRIC 23" or "STUDY RANDOMIZE ADULT 135" or "STUDY RANDOMIZE PEDIATRIC 136" or "STUDY RANDOMIZE GERIATRIC 137" or "CROSS-OVER 144" or "META-ANALYSIS 145" or "N-OF-ONE TRIAL 146" or "PRACTICE GUIDELINE 156" or "SYSTEMATIC REVIEW 161" or "ANNOTATED BIBLIOGRAPHY 167" or "PRIORITY CLIN PRACT GUIDE 168"

AND NOT Author:

"Editorial" OR "(Letter to Ed)"

Years: 1990-2010

Records = 128
Table 3. Search to Update the Original PubMed Search with Additional Database Names: Performed on 07/06/10

Results = 30

<table>
<thead>
<tr>
<th>Search</th>
<th>Query</th>
<th>Results</th>
</tr>
</thead>
</table>
B. ABSTRACT REVIEWS

Of the 1761 abstracts reviewed, 286 were selected for full-text review; 219 were excluded because they did not study depression, 994 were excluded because they were not administrative database studies, and 262 were excluded because the data source was not from the United States or Canada. Cohen’s kappa for agreement between reviewers on inclusion vs exclusion of abstracts was 0.77.

C. FULL-TEXT REVIEWS

Of the 286 full-text articles reviewed, 10 were included in the final evidence tables; 142 were excluded because the HOI identification algorithm was poorly defined, and 134 were excluded because they included no validation of the outcome definition or reporting of validity statistics. Reviewers identified 36 citations for review from full-text article references. Of these, 1 was included in the final report; 9 did not study depression, 12 were not database studies, 5 were excluded because the HOI algorithm was poorly defined, and 9 were excluded because they included no validation of the outcome definition or reporting of validity statistics. The final evidence table thus includes 11 articles. Cohen’s kappa for agreement between reviewers on inclusion vs exclusion of full-text articles reviewed was 0.95.
D. MINI-SENTINEL INVESTIGATOR SURVEY

Mini-Sentinel investigators provided 0 published and 0 unpublished reports of validation studies that had been completed by their teams.

E. EVIDENCE INCLUDED IN TABLE

Of the 11 studies included in the table, 10 were identified from the initial search strategy and 1 was identified through references of articles that underwent full-text review. The table is organized from top to bottom with listing first those studies that focused on positive predictive value, followed by those which include information about positive predictive value and sensitivity, and finally one study with only information regarding sensitivity of the algorithm.

F. SUMMARY AND DISCUSSION OF ALGORITHMS AND VALIDATION

Codes Used in Algorithms. All 11 publications listed in the evidence table included algorithms with various combinations of four different ICD-9 diagnostic codes depression, NOS (311), dysthymic disorder (300.4), and major depressive disorder, single episode (296.2) or recurrent (296.3) to define depression (See Appendix C). Two studies further permitted adjustment disorder with brief depressive reaction (309.0), adjustment reaction with prolonged depressive reaction (309.1), adjustment reaction with mixed emotional features (309.28), and depressive type psychosis (298.0) (10, 11) (See Appendix C). For some of the algorithms in one study (12), the following codes were also included: other specified alcohol-induced mental disorder (291.89), other specified drug-induced mental disorders (292.89), bipolar affective disorder, depressed (296.5), and bipolar affective disorder, mixed (296.6). Algorithms in three studies also required a claim for a filled antidepressant prescription (1, 13, 14).

The algorithms for individual studies varied with respect to specific selected codes, treatment setting (outpatient or inpatient), code listing (principal or first listed vs secondary), billing health care professional, number of required codes, and timing of the codes. The heterogeneity of the algorithms across the studies constrains strict comparisons. For example, the study which defined depression on the basis of at least 2 outpatient or 1 inpatient codes for depression (296.2, 296.3, 300.4, 311) in a 12 month period likely included a narrower patient population (13) than the study which required only a single depression claim (296.2, 296.3, 311) (15).

Validation Criteria and Method. The studies used three general approaches to the validation of the algorithms: structured diagnostic interviews of depressive disorders (10, 16), self-report items or forms (11, 12, 15, 17, 18), and depression diagnoses in the medical record (1, 13, 14, 17, 19). Some of the self report validations were based on a single item for depression or depressed mood (15, 17) or whether patients had ever been told by a doctor that they had depression (12, 17). One study (17) used a 20-item self-report depression scale (CES-D) (20) and two studies (11, 18) used the PHQ-9, which is a brief nine item validated screen for major depression (21). One of the studies permitted direct comparison of four different validation criteria: a CES-D score of ≥6, self report depression item, being told by a physician of a depression diagnosis, and a medical record diagnosis of depression (17).

Validation Algorithm. Four of the 11 studies provided sufficient information to derive a chance corrected measure of agreement (kappa) between the algorithm and a criterion standard. Landis and Koch have suggested the following kappa assessment standards: 0.0-0.20 (slight), 0.21-0.40 (fair), 0.41-0.60 (moderate), and 0.61-0.80 (substantial) (2). In the three studies which used an independent
assessment of depression based on structured diagnostic interview or symptom checklist as the criterion standard, agreement was either in the slight (16, 18) or fair (10, 16) range. The only study which permitted calculation of kappa and used a physician diagnosis of depression in the medical record as the criterion standard achieved a moderate level of chance corrected agreement (1).

Beyond overall agreement as measured by kappa, one key criterion by which to evaluate the performance of the algorithms is to consider the proportion of people identified as depressed by a claims-based measure whom have significant depression confirmed by a second information source (Positive Predictive Value). Ten of the 11 studies provide positive predictive values that ranged from 48.6% (14) to 98.8% (13). Because positive predictive values vary as a function of prevalence, the interpretation of the positive predictive values from several of the studies (12-15, 17, 19) is undermined by the absence of prevalence data.

Four studies included information about PPV and prevalence. One study, which compared claims-based algorithms with physician diagnoses in the medical record as the criterion standard (1), did not include an independent assessment of depression. In another study, the PPV (66.4%) using the PHQ-9 as the criterion standard was only marginally higher than the prevalence (55.8%) (18). Two studies, one of emergently admitted older adult medical inpatients (16) and one of adolescents with asthma (10), included prevalence information, an independent criterion standard, and PPV values that were substantially greater than base prevalence. In the study of older adults, PPVs ranged from 52.6% to 62.5% (prevalence 30.8%) (16), while in the study of adolescents PPVs ranged from 31.5% (prevalence 7.7%) to 39.4% (prevalence 14.6%) (10).

Another key criterion by which to evaluate claims-based algorithms is to consider what proportion of people with depression identified by an independent source will have depression identified by the algorithm (sensitivity). Five of the studies included information on sensitivity (1, 10, 11, 16, 18). One of these studies, however, used physician notes in medical records as the criterion standard (1) and did not include an independent measure of depression. Of the remaining 4 studies, the sensitivities ranged from 15.8% to 51.1%. Therefore, the most sensitive algorithms detected approximately one-half of patients who were determined to be depressed by independent assessment (10, 18).

Operating characteristics varied as a function of the validation algorithm. These relationships were most readily apparent in studies that tested different algorithms against the same criterion standard within the same patient population. In a study of Veterans treated for diabetes mellitus, for example, broadening the algorithm from unipolar depression codes to include also bipolar disorder and substance-related mood disorder codes markedly increased the percentage of patients captured by the algorithm from 4.5% to 16.5%, but reduced the positive predictive value (0.90 to 0.82). In this study, the criterion standard was patient report of being told by a doctor that the patient had depression (12).

In one primary care study in which physician medical record diagnosis of depression was the criterion standard, narrowing an algorithm markedly altered the operating characteristics (1). In the first algorithm, patients were required to have at least 2 events either of which could be an outpatient encounter with a primary diagnosis of depression or a pharmacy claim for an antidepressant medication. Thus patients with antidepressant claims without an outpatient depression diagnosis met algorithm requirements. The second algorithm required one outpatient claim with a primary depression diagnosis and another event which could be either a second outpatient depression diagnosis or an antidepressant claim. Perhaps not surprisingly, the less stringent first algorithm had a lower specificity (65.4% vs 88.4%),
but a much higher sensitivity (95.0% vs 52.1%) than the more stringent second algorithm. The kappa was slightly higher with the first (0.464) than the second (0.425) algorithm (1).

**Selected Patients.** The studies also varied with respect to patient population. Beyond differences in patient age, described below, clinical differences were evident among the study populations: three studies were limited to patients treated for diabetes mellitus (11, 12, 18), one focused exclusively on patients with a history of asthma treatment (10), and one involved only disabled persons (15). The patient populations included a range of sources payment for medical services. While some of the studies were limited to patients with Medicaid coverage (15, 18), others included only patients with private insurance (10, 11, 13). Two studies were based on patients receiving care that was financed and provided by the Veterans Health Administration (12, 14). One study of Veterans care indicated that supplementing Veterans Health Administration data with Medicare data enhanced the rate of depression detection (12) (Appendix D).

The validity of claims based measures likely varies with characteristics of the patient population. Several studies have examined whether patient and service characteristics influence the clinical recognition of depression in general medical settings (23-30). A consistent finding to emerge from this literature is that clinical detection erodes with declining depression symptom severity (23-25). Some indices of socio-economic disadvantage also appear to decrease the likelihood of clinical detection of depression. Depressed Medicaid financed patients (26), adults with less formal education (23, 24), African Americans (26), and patients with fewer (24) or shorter (26) primary care visits have each been reported to have a comparatively high risk of missed clinical depression diagnoses. An absence of current or past psychotropic treatment has also been linked to a failure to diagnose depression in primary care (23, 25).

Acute co-occurring general medical disorders may further compromise clinical recognition of depression (24, 27-29). Within the context of an individual visit, somatic complaints related to general medical problems may compete with depressive symptoms for clinical attention (24, 28) or lead physicians to misattribute signs and symptoms of depression to other medical problems (29). Over time, however, co-occurring chronic medical conditions may increase clinical recognition of depression (30). This effect may be mediated by a tendency for patients with chronic medical conditions to make frequent medical visits to a consistent source of care whom they trust (30). For these and related reasons, the validity of claims based measures of depression likely depends upon the clinical and service characteristics of the patient population.

**Age of Study Population.** One study was limited to adolescents (10), one included children as well as adults (19), and one included only patients at least 65 years of age (16). The remaining 8 studies involved non-elderly as well as elderly adults. The average patient age in these studies was between 50 and 60 years.

None of the studies included validation information stratified by patient age. Because positive predictive value depends upon the prevalence of the underlying condition, it is likely that PPV will be lower in children and adolescents than in adults because of the lower treated prevalence of depression in children and adolescents than adults (31). Clinical recognition of depression may also vary with patient age (30).

**Patient Sex.** Except for two studies of Veterans (12, 14), all of the studies involved a predominance of female patients. None of the studies, however, provided validation data stratified by patient sex. After
controlling for measures of severity and impairment, patient sex has not been found to be related to the rate of treatment of major depression (32).

**Time Period of Data Collection.** The 11 reviewed studies were published between 1997 and 2010. The earliest data were based on care delivered in 1986 (19) and the most recent data were derived from 2006 (18), though two studies did not specify the dates of service delivery (11, 16). During this period, there has been a substantial increase the proportion of Americans treated for depression (33, 34).

**Principal vs Secondary Diagnosis.** Only one of the studies included algorithms which required that the depression code appear in the principal or primary position (1). All of the other studies did not indicate the position of the depression codes. Whether the validity of depression codes varies with code position has not been subject to systematic study.

**Hospitalization Diagnosis vs Outpatient Encounter.** The studies included in this report examined depression outcomes based on billing codes for inpatient treatment (12, 13, 19), outpatient visits (1, 12, 13, 14, 16, 17) or emergency encounters (12, 14, 17). None of the studies separately assessed or compared the validation of depression outcomes from different treatment settings, though one study compared validity measures from primary care visit codes with mental health service codes (12).

### G. SUMMARY OF EXCLUDED POPULATIONS AND DIAGNOSES

As indicated in section F above, five studies focused exclusively on populations with specific conditions including diabetes (11, 12, 18), asthma (10), and disability (15). Of the other six studies, two were limited to patients who had received inpatient treatment (16, 19) and one was limited to Veterans and focused on new onsets of depression (14).

Only one of the studies was based on a national population (12). The other studies were derived from a Canadian province (19), 2 acute care hospitals (16), 3 Veterans Health Administration medical centers (14), large managed care organizations (1, 18), a private health plan or health maintenance organization (10, 11, 13) or a group of 9 primary care clinics (17).
### HOI Evidence Reviews - 17 - Depression Report

#### H. EVIDENCE TABLE

<table>
<thead>
<tr>
<th>Citation</th>
<th>Study Population and Time Period</th>
<th>Description of Outcome Studied</th>
<th>Algorithm</th>
<th>Validation/Adjudication Procedure and Operational Definition</th>
<th>Validation Statistics</th>
</tr>
</thead>
</table>
| Frayne et al., 2010 | National sample (n=133,068) of Veterans who were treated for diabetes mellitus and responded to a health survey, 98.1% male, 76.6% white, mean age-66.3 years, 1998-1999. | Patient self-report of depression: “Has a doctor ever told you that you have depression?” | Diagnosis of ICD-9: 296.2x, 296.3x, 311 (Algorithm H). | Proportion of patients meeting algorithm criteria told that they have depression (PPV) and proportion not meeting algorithm criteria not told that they have depression (NPV). | PPV=84%  
NPV=72%  
(Algorithm H: see Appendix D for complete results) |
| Kramer et al., 2003 | Veterans (n=109) from 3 VA medical centers with ≥1 outpatient claim for a depressive disorder (296.2-296.26, 296.3-296.36, or 311) after 180 days without such a claim or an AD fill, 95.2% male, mean age 55.5 years, 1999-2001. | Presence of depression diagnosis in medical record during 180 day period after the new depressive disorder claim | ≥1 outpatient claims for depression (ICD-9: 296.2, 296.3, or 311) in any service setting after 189 days without such a claim or an antidepressant fill. | Estimate of PPV for new onset depression by determining proportion of cases with depression diagnosis in medical record during 180 days prior to new index claim for depression. | PPV= 48.6% (53 of 109)  
(New onset depression) |
| Rawson et al., 1997 | Randomly selected inpatients with a first listed discharge diagnosis of depression (311) in administrative files, Saskatchewan, 30.1% male, male median age: 43 years, female median age: 51 years, 1986 | A diagnosis of depression (311) or a depression-related disorder (ICD-9: 296.1, 296.4, 296.6, 300.4, 309, 311) in the medical record discharge note. | First listed inpatient discharge diagnosis of depression (311) in administrative files. | Proportion with first listed discharge diagnosis of depression with ICD-9 311 (PPV₁) or depression-related diagnosis (ICD-9: 296.1, 296.4, 296.6, 300.4, 309, 311) in medical record discharge note. | PPV₁= 58.3% (91 of 156)  
PPV₂= 93.6% (146 of 156) |
| Smith et al., 2009 | Adult (19-64 years) work disabled Medicaid beneficiaries with depression claims | Self-rated depressed mood item. | ≥1 Medicaid claims for ICD-9: 296.2, 296.3, or 311 during 12 months prior to | Proportion of beneficiaries with depression claims who report depressed mood (PPV₁). Also the corresponding proportion | PPV₁= 87.9% (175 of 199)  
PPV₂= 91.1% (153 |
who responded to an employment and disability survey, 65.3% female, 87.4% white, 19.1% married, mean age: 43.4 years, 2003 or 2005

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>Methods</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solberg et al., 2006</td>
<td>Adults (&gt;19 years) from a large private health plan (N=135,842). 5 random samples (N=20) meeting different algorithms for depression, 2000.</td>
<td>Medical record diagnosis of depression.</td>
<td>D1. Prevalent depression: ≥2 outpatient or ≥1 inpatient ICD-9 296.2, 296.3, 300.4, 311 in 12 months; D2. Antidepressant (AD) treatment (A and B and C): A. 6-months no AD prior to new AD fill; B. ≥1 depression code 3 months before or after AD; C. ≥1 more depression code or AD fill in 1.5 years before or after new AD. Proportion of selected plan members with D1 claims-based algorithm found to have depression diagnosis in medical record (PPV1) and proportion with D2 algorithm found in medical record to have started a new antidepressant treatment episode for depression (PPV2). PPV1= 98.8% (79 of 80); PPV2= 65.0% (13 of 20) and 90.0% (18 of 20)</td>
</tr>
<tr>
<td>Solberg et al., 2003</td>
<td>Adult (≥18 years) outpatients (n=274) from 9 staff model primary care clinics in a metropolitan area with depression code, no antidepressant prescription 6 months, no diagnosis bipolar, schizophrenia, or alcoholism past year, 74.1% female, 91.6% white, 60.6% married, mean age: 52.4 years, 1998-1999.</td>
<td>1. Depressive symptoms (CES-D ≥6) (PPV3) 2. Self-reported current depression (PPV3) 3. Reported told at index visit has depression (PPV3) 4. Chart audit ICD-9 311 (only code available for depression) code, no other 311 codes in previous 6 months, no AP fills in previous 6 months. Proportion of patients meeting administrative code definition of depression who met each of the four outcomes (CES-D score, self-reported current depression, told by health care professional at visit had depression, and chart audit with depression diagnosis or treatment at index visit. PPV1=71.5% (196 of 274); PPV2=71.5% (196 of 274); PPV3=54.6% (149 of 274); PPV4=94.9% (260 of 274)</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Population</td>
<td>Methodology</td>
<td>Relevant Code</td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>McCusker et al., 2008</td>
<td>Emergently admitted medical inpatients ages ≥65 years (n=185) from 2 university-affiliated acute care hospitals in Montreal, over sampled for depression, excluding patients with cognitive impairment, 64.3% female, mean age 80.0 years, 67.6% disabled, 48.1% lives alone</td>
<td>DIS assessed major depressive disorder of &gt; 6 months or &lt; 6 months duration.</td>
<td>During 12 months after index inpatient admission, 3 Algorithms: 1. Outpatient claim for physicians services for ICD-9: 311, 300.4. 2. Antidepressant prescription. 3. Psychiatrist visit</td>
</tr>
<tr>
<td>Kahn et al., 2008</td>
<td>Adult (≥18 years) Medicaid behavioral health managed care organization enrollees (n=249) with a diagnosed mental disorder and diabetes, 63.1% female, mean age 52.2 years, 2006.</td>
<td>PHQ-9 assessed depression by mail survey</td>
<td>Depression (ICD-9 code 311) in the encounter data.</td>
</tr>
<tr>
<td>Katon et al., 2006</td>
<td>Adolescent (11-17 years) primary care outpatients (n=769) with history of asthma treatment, excluded patients treated for bipolar</td>
<td>C-DISC assessed DSM-IV major depression or dysthymia; C-DISC assessed DSM-IV panic, ≥1 ICD-9: 296.2,296.3, 298.0, 300.4, 309.0, 309.1, 309.28, or 311 “depressive disorder</td>
<td>Proportion of patients with utilization claim for depression who met C-DISC criteria for a depressive disorder (PPV1); Proportion of patients with utilization claim for depression who</td>
</tr>
<tr>
<td>Study</td>
<td>Population</td>
<td>Methodology</td>
<td>Diagnosis</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Spetell et al. 2003</td>
<td>Primary care physician panel members (≥12 years) from a large MCO (n=892,786) selected for meeting algorithm 1 or 2 and members matched by age, gender, and number of comorbid conditions not meeting algorithms, 59.6% female, median age algorithm positive: 44 years, median age algorithm negative: 41 years, 1997.</td>
<td>Algorithm 1. In 12 months, ≥2: A. First listed ICD-9 296.2, 296.3, 300.4, 311 B. AD fill Algorithm 2. In 12 months, A. above and ≥1 of A. or B. Bipolar disorder, depressive psychosis or lithium fills excluded. A or B above during 12 months before study period also excluded.</td>
<td>Physician diagnosis of depression in medical record during the 12 month study period.</td>
</tr>
<tr>
<td>Katon et al., 2004</td>
<td>Adult HMO patients (n=4385) with treatment of diabetes mellitus and major depression as assessed by PHQ-9, 60% female, mean</td>
<td>In 12 months before assessment: 1. ICD-9 code for ≥1 depression: 296.2,296.3, 298.0, 300.4,</td>
<td>PHQ-9 positive screen for current major depression.</td>
</tr>
</tbody>
</table>
I. TOPIC EXPERT CONSULTATION

The goal of identifying depression from claims data is constrained by incomplete clinical detection. In the community, just over one-half of adults with major depression receive treatment for their symptoms during the course of 1 year (32, 35). In clinical samples as well, primary care physicians recognize as depressed only about one half of all their depressed patients who present for treatment (36). The detection rate may be even lower in some clinical subgroups including patients with medical morbidity (30%) (29) and Veterans (40%) (37). Deficiencies in the clinical diagnosis of depression impose a ceiling on the performance of claims based efforts to detect depression. One countervailing consideration is that clinically detected or treated depression may be more severe than undetected or untreated depression (32, 38). Screening initiatives (39) and other efforts to improve the detection and management of depression in primary care practice (40) may have an incidental salutary effect on the validity of electronic health record based detection of depression.

Beyond problems with clinical detection, psychosocial considerations may further limit the accuracy of claims for depression. Concern over protecting patient confidentiality may lead some physicians to deliberately substitute alternative non-mental disorder diagnoses on claims and encounter forms for their patients receiving treatment for depression (41). Under reporting of depression may occur in a conscious effort by the clinician to reduce social stigma that might otherwise have adverse occupational or legal consequences for the patient (42, 43).

Constraints on clinical diagnosis may help to explain the range of sensitivities in the reported studies. When independent patient assessments of depression that capture clinically undetected depression serve as the criterion standard, sensitivities are low, ranging from 12.5% (16) to 51.1% (18). When medical record diagnoses are treated as the criterion standard, which excludes clinically undetected or unreported depression, the sensitivity of claims based algorithms reaches as high as 95.0% (1). These differences are also reflected in the kappa values.

Studies with independent assessments of depression (10-12, 15-18), rather than medical record diagnoses (1, 13, 14, 19) provide more credible evidence of algorithm validity. In this regard, the most rigorous studies involve using a structured diagnostic assessment (10, 16) followed by those that use a depression disorder screening instrument (18) as the criterion standard. Unfortunately, low chance corrected agreement of these studies precludes the algorithms in these studies from being recommended for general case identification, though the operating characteristics may be adequate for specific applications.
One strategy for broadening claims-based algorithms and capturing more patients with depression is to include pharmacy records indicating an antidepressant prescription fill (1, 11, 13, 14). Although this strategy tends to capture more depressed patients thereby increasing sensitivity, it comes at the expense of a larger number of false positive cases thereby lowering specificity and positive predictive value (1, 11, 13). These trade-offs arise from the large proportion of antidepressant prescriptions that are for psychiatric and general medical conditions other than depression. In one national study, only 27% of individuals treated with antidepressants reported receiving them for depression (44).

Positive predictive value depends upon the prevalence of the underlying condition. Many of the studies were performed in highly enriched samples with base prevalence of depression that greatly exceeds primary care populations. Without careful attention to the prevalence of depression in the base population, the positive predictive values may appear deceptively high. While this may be obvious in samples that are limited only to patients with depression (13, 17), it also distorts estimates in samples that are enriched by oversampling of depressed patients. In one study, for example, the base prevalence of depression in a sample selected for treatment of a mental disorder (28.5%) (18) was associated with a PPV of 66.4%. In another study which involved matching non-depressed patients to depressed patients, the base prevalence of depression from the resulting study sample was 26.0% and the PPV ranged from 49.1% to 60.6% depending upon the algorithm (1). Substantially lower positive predictive values would be expected when using databases populated by general primary care patients. The prevalence of major depression is approximately 5%–10% in primary care patients and 10%–14% in medical inpatients (45).

VI. SUMMARY AND CONCLUSIONS

A. RECOMMENDATIONS FOR ALGORITHMS

No single algorithm consistently demonstrated sufficiently strong operating characteristics in relation to independent assessments of depression to be recommended for general use. The proportion of people identified by claims-based measures that met independent criteria for depression varied with the algorithm and patient population. In a study of adolescent outpatients with asthma, 31.5% of adolescents identified with one algorithm met formal criteria for major depressive disorder or dysthymia (prevalence 7.5%) (10). In a study of elderly adult inpatients over-sampled for clinical depression diagnoses (prevalence 30.8%), 62.5% of patients identified with another algorithm met criteria for major depression (16). For some purposes, such as first stage screening for participation in clinical depression research, these positive predictive values may be sufficient. For other purposes, such as monitoring clinical relapse in individual patients, these positive predictive values may not be acceptable.

The algorithms also varied in the percentage of people with independently confirmed depressive disorders that they successfully identified as depressed. The algorithms with the highest sensitivity detected approximately one-half of patients who were determined to be depressed by independent assessment (10, 18). Here as well, these performance characteristics may be adequate for some clinical and research purposes, but not for others.

Electronic health records provide sufficient information to identify with a moderate level of agreement adult primary care patients who have been diagnosed with depression. In this regard, the algorithm with the strongest psychometric properties required over the course of 12 months at least 2 first listed codes for depression (296.2, 296.3, 300.4, 311) as well as an antidepressant prescription claim (kappa: 0.464).
This algorithm is recommended for the identification of clinically diagnosed depression on the basis of claims data. However, its operating characteristics for identifying independently assessed depression are not known.

B. SUGGESTIONS FOR FUTURE RESEARCH BASED ON EVIDENCE GAPS

Depression is an established adverse effect of several medications (46, 47). A well validated algorithm to identify depression from electronic health data would be valuable for post-marketing evaluation of drug safety.

Research priorities include:

1. Replication of an evaluation of the most promising depression algorithm (1) for identifying clinically diagnosed depression in a fee for service or other primary care setting.

2. Assessment of this depression algorithm in relation to an independent patient depression assessment standard. In view of the low agreement between depression codes and independently assessed depression in previous studies, future research should be conducted in general medical or specialty mental health settings that routinely screen for depression. In this context, it should be noted that even within specialty mental health settings correlations are modest between clinical diagnoses and structured diagnostic interviews (48). Some health systems have integrated measures of depression severity, such as the PHQ-9, within electronic health records (49). If implemented on a large scale, this would provide opportunities for validating algorithms based on traditional service and prescription claims and provide an alternative method of case ascertainment.

3. Because of increasing concern over depression-related adverse events in youth (50, 51) and the paucity of information that is currently available for this age group (10), priority should be given to developing administrative database algorithms to identify depression in children and adolescents.
VII. REFERENCES


35. Hasin DS, Goodwin RD, Stinson FS, Grant BF. Epidemiology of major depressive disorder: Results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch Gen Psychiatry.* 2005; 62: 1097-1106.


46. Patten SB, Barbui C. Drug-induced depression: a systematic review to inform clinical practice. 

47. Hull PR, D'Arcy C. Isotretinoin use and subsequent depression and suicide: Presenting the 


49. Klein EW, Hunt JS, LeBlanc BH. Depression screening interfaced with an electronic health record: 
   A feasibility study in a Primary care clinic using optical mark reader technology. *Prim Care 

50. Hammad TA, Langhren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant 
   drugs. *Arch Gen Psychiatry*. 2006; 63: 332-339.

51. Stone M, Laughren T, Jones ML, et al. Risk of suicidality in clinical trials of antidepressants in 
   339: b2880.
VIII. APPENDICES

A. APPENDIX A: ABSTRACTS OF STUDIES INCLUDED IN EVIDENCE TABLE


ABSTRACT: The authors estimated the validity of algorithms for identification of mental health conditions (MHCs) in administrative data for the 133,068 diabetic patients who used Veterans Health Administration (VHA) nationally in 1998 and responded to the 1999 Large Health Survey of Veteran Enrollees. They compared various algorithms for identification of MHCs from *International Classification of Diseases, 9th Revision* (ICD-9) codes with self-reported depression, posttraumatic stress disorder, or schizophrenia from the survey. (Positive predictive value (PPV) and negative predictive value (NPV) for identification of MHC varied by algorithm (0.65-0.86, 0.68-0.77, respectively). PPV was optimized by requiring ≥2 instances of MHC ICD-9 codes or by only accepting codes from mental health visits. NPV was optimized by supplementing VHA data with Medicare data. Findings inform efforts to identify MHC in quality improvement programs that assess health care disparities. When using administrative data in mental health studies, researchers should consider the nature of their research question in choosing algorithms for MHC identification.


OBJECTIVES: To determine the prevalence of self-reported depression symptoms among diabetic individuals enrolled in Gold Choice, a Medicaid managed care organization specifically for people with mental health and/or substance abuse diagnoses; and to assess the sensitivity and specificity of individuals' self-report with encounter data. METHODS: The 9-item depression scale of the Patient Health Questionnaire (PHQ-9) was mailed to 454 Gold Choice members in Western New York diagnosed with diabetes; and 249 completed PHQ-9 forms were returned (55% response rate). The PHQ-9 forms were compared to primary care encounter data to determine whether the respondents had been diagnosed with depression. Descriptive and inferential statistical analysis was undertaken. RESULTS: The majority (56%) of individuals in the sample screened positive for depression (PHQ-9 > or = 10), and half (49%) did not have evidence of a diagnosis in their encounter data. The percentage of those diagnosed with depression rose with increasing PHQ-9 severity levels, with 63% of individuals with the most severe depression (PHQ-9 > or = 20) having a diagnosis. This trend was statistically significant, confirmed by independent sample t-tests and chi-square tests. The sensitivity of the PHQ-9 was moderate (66%), as was the specificity (52%). CONCLUSIONS: The results of this study suggest that depressive disorders may be under-recognized and under-treated amongst individuals with diabetes in the primary care setting. Half (51%) of those with PHQ-9 scores > or = 10 had depression diagnoses, suggesting poor compliance rates and/or a need for therapy reassessment.


OBJECTIVES: Youth with asthma have a high rate of anxiety and depressive disorders, and these comorbid disorders are associated with increased asthma symptom burden and functional impairment. This study examined the rates and predictors of recognition of anxiety and depressive disorders among youth (ages 11 to 17) with asthma who are seen in primary care settings as well as...
the quality of mental health care provided to those with comorbid anxiety and depression over a 12-month period. METHODS: This study used automated utilization and pharmacy data from a health maintenance organization to describe the rate of recognition of Diagnostic and Statistical Manual of Mental Disorders, edition IV, anxiety and depressive disorders and the quality of mental health care provided for the 17% of youth with asthma and comorbid anxiety and/or depression during the 12-month period prior to diagnosis. Psychiatric diagnoses were based on a telephone version of the Computerized Diagnostic Interview Schedule for Children (Version 4.0). RESULTS: Approximately 35% of youth with 1 or more anxiety and depressive disorders and 43% of those with major depression were recognized by the medical system during a 12-month period. Greater functional impairment (odds ratio [OR] 3.32, 95% confidence interval [CI] 1.25-8.79), higher severity on parent-rated anxiety and depressive symptoms (OR 2.49, 95% CI 1.04-6.00), and a greater number of primary care visits (OR 1.26, 95% CI 1.10-1.44) were associated with significantly higher recognition rates while having Medicaid or Washington state medical insurance was associated with lower rates of recognition (OR 0.27, 95% CI 0.08-0.92). Only approximately 1 in 5 youths with comorbid major depression received an adequate dosage and duration of antidepressant medication, and only 1 in 6 received a minimally adequate number of psychotherapy sessions (> or =4 visits). CONCLUSION: Rates of recognition of comorbid anxiety and depressive disorders are low in youth with asthma and few youth with asthma and comorbid anxiety and depression receive guideline-level mental health treatment.


OBJECTIVES: Major depression occurs in approximately 11% to 15% of patients with diabetes and is associated with poor glycemic control and adverse medical outcomes. This study examined the rates and predictors of recognition of depression among primary care patients with diabetes and comorbid major depression and the quality of depression care provided during a 12-month period. METHODS: This study used automated utilization, pharmacy, and laboratory data from a health maintenance organization to describe the rate of recognition of depression and quality of care provided for patients with major depression and diabetes in the 12-month period before diagnosis. Major depression was diagnosed based on the Patient Health Questionnaire (PHQ-9) that was included in a mail survey sent to 9,063 patients on the Group Health diabetes registry from 9 primary care clinics. RESULTS: Approximately 51% of patients with major depression and diabetes were recognized as depressed by the health care system. Women were more likely to be recognized (odds ratio [OR] 1.58, 95% confidence interval [CI] 1.26-1.97), as were those with dysthymia (OR 3.44, 95% CI 2.08-5.72), panic attacks (OR 1.55, 95% CI 1.19-2.19), patients with more than 7 primary care visits (OR 1.42, 95% CI 1.06-1.91) and patients reporting poor health (OR 1.62, 95% CI 1.04-2.53). Of the 51% of patients with major depression who were recognized, 43% received 1 or more antidepressant prescriptions but only 6.7% received 4 or more psychotherapy sessions during a 12-month period. DISCUSSION: There were large gaps in both recognition and quality of depression care provided to patients with major depression and diabetes within a health maintenance organization system.


Understanding the quality of routine care for adolescent depression constitutes the initial step in designing and implementing improvement strategies. This study assessed depression detection and type and duration of services for adolescents in mental health care settings. Medical record
diagnosis and standardized research interview results were compared for youth seeking mental health treatment. The majority of depressed adolescents received care consistent with guidelines and evidence. However, only 51% received appropriate medication; fewer than half received at least 8 sessions of outpatient care or follow-up after hospitalization. Males received significantly fewer components of quality care compared with females. Depression diagnoses in routine care may be facilitated by using structured interviews or questionnaires. Quality monitoring and improvement initiatives may also increase rates of care components that are consistent with guidelines and evidence. Methods tested in this study may facilitate the evaluation of quality improvement initiatives for adolescent depression or other mental health disorders.


OBJECTIVE: This study aimed to examine the recognition of depression in older medical inpatients by nonpsychiatric physicians over a 2-year period. METHODS: A cohort of medical inpatients aged 65 and above was recruited at two university-affiliated hospitals, with oversampling of depressed patients. Participants were assessed with research diagnoses of major or minor depression (DSM-IV) at admission and at 3, 6 and 12 months. Indicators of recognition during the 12 months before and the 12 months after admission, derived from administrative databases, included the following: depression diagnosis, antidepressant prescription and psychiatric referral. Multiple logistic regression analyses were used to identify factors associated with recognition. RESULTS: Among 185 patients with at least one research diagnosis of depression during the study, recognition rates ranged up to 56% during the 12 months before admission among patients with major depression lasting at least 6 months and up to 61% during the 12 months after admission among patients with persistent major depression. In both study periods, a greater number of physician visits and prescription of a psychotropic medication (non-antidepressant) were independently associated with recognition. CONCLUSIONS: A longitudinal approach to measuring recognition of late-life depression in ambulatory care settings indicates that persistent major depression is more likely to be recognized than previously reported.


Administrative data have long been used in psychiatric epidemiology and outcomes evaluation. This article examines the reliability of the recording of schizophrenia and depressive disorder in three Saskatchewan administrative health care utilization datafiles. Due to their comprehensive nature, these datafiles have been used in a wide range of epidemiologic studies. Close agreement was found between hospital computer data and patients' charts for personal and demographic factors (> or = 94.7%). Diagnostic concordance between computerized hospital data and medical charts was very good for schizophrenia (94%) but poor for depressive disorder (58%). Appropriate physician services were identified for 60% and 72% of hospital discharges for schizophrenia and depressive disorder, respectively, and exact diagnostic agreement between hospital and physician datafiles was 62% for schizophrenia and 66% for depressive disorder. Appropriate provincial mental health branch services were found for 83% and 38% of hospital discharges for schizophrenia and depressive disorder, respectively; exact diagnostic concordance between these datafiles was 75% for schizophrenia and 0% for depressive disorder. A significant number of patients with major or neurotic depression appeared to be wrongly coded as having depressive disorder in the hospital file. The differences in diagnostic agreement may also be partly a function of how the two conditions are differentially treated in the health system. These findings suggest that more specific and severe
psychiatric diagnoses are likely to be recorded accurately and consistently in the Saskatchewan datafiles. However, disorders with multiple manifestations or those for which there are several possible codes should be examined with caution and ways sought to validate them. Attention should also be paid to which service sectors are involved in the treatment of specific disorders.


Abstract: This cross-sectional study of adult survey respondents with disability and depression (n = 199) enrolled in Massachusetts' Medicaid program examined the association of adequately or inadequately prescribed antidepressant treatment and self-reported work status using conditional logistic regression, controlling for age, gender, race, marital status, education, receipt of SSI/SSDI, self-reported disabling condition, and health status. Confounding by severity was addressed by two methods: restriction of our sample and subsequent stratification by propensity score. Individuals receiving adequate antidepressant treatment had an increased odds of working compared to individuals receiving inadequate treatment, both in analyses in which restriction was used to limit confounding (OR = 3.45, 95% CI = 1.15-10.32, P < .03), and in analyses which combined restriction with adjustment by propensity score stratification (OR = 3.04, 95% CI = 1.01-9.62, P < .05). Among this sample of Medicaid enrollees with disability and depression, those receiving adequate antidepressant treatment were significantly more likely to report working.


The objective of this study was to demonstrate a method to accurately identify patients with specific conditions from claims data for care improvement or performance measurement. In an iterative process of trial case definitions followed by review of repeated random samples of 10 to 20 cases for diabetes, heart disease, or newly treated depression, a final identification algorithm was created from claims files of health plan members. A final sample was used to calculate the positive predictive value (PPV). Each condition had unacceptably low PPVs (0.20, 0.60, and 0.65) when cases were identified on the basis of only 1 International Classification of Diseases, ninth revision, code per year. Requiring 2 outpatient codes or 1 inpatient code within 12 months (plus consideration of medication data for diabetes and extra criteria for depression) resulted in PPVs of 0.97, 0.95, and 0.95. This approach is feasible and necessary for those wanting to use administrative data for case identification for performance measurement or quality improvement.


OBJECTIVES: To understand the process, outcomes, and patient satisfaction of usual primary care for patients given a diagnostic code for depression. STUDY DESIGN: Health plan data were used to identify patients with a diagnostic code for depression (and no such diagnosis in the preceding 6 months). Patients were surveyed by mail soon after the coded visit and again 3 months later about the care they had received; their charts were also audited. METHODS: The 274 patients in 9 primary care clinics who responded to both surveys reported on their personal characteristics, depression symptoms and history, the care received in that initial visit, and the follow-up care during the next 3 months. They also reported on their satisfaction with various aspects of that care. RESULTS: These patients were likely to be given antidepressant medications as their main or only treatment. Referral
for mental health therapies was not used often, even though referral is readily available in this setting; other types of self-management recommendations and support were even less frequent. Patient outcomes and levels of satisfaction during a 3-month follow-up period were unimpressive.

CONCLUSIONS: To successfully maintain a key role in the care of this important problem for their patients, primary care physicians may need to incorporate a more comprehensive and systematic approach to management that involves other team members and is more satisfying to patients.


BACKGROUND: Multiple factors limit identification of patients with depression from administrative data. However, administrative data drives many quality measurement systems, including the Health Plan Employer Data and Information Set (HEDIS). METHODS: We investigated two algorithms for identification of physician-recognized depression. The study sample was drawn from primary care physician member panels of a large managed care organization. All members were continuously enrolled between January 1 and December 31, 1997. Algorithm 1 required at least two criteria in any combination: (1) an outpatient diagnosis of depression or (2) a pharmacy claim for an antidepressant Algorithm 2 included the same criteria as algorithm 1, but required a diagnosis of depression for all patients. With algorithm 1, we identified the medical records of a stratified, random subset of patients with and without depression (n = 465). We also identified patients of primary care physicians with a minimum of 10 depressed members by algorithm 1 (n = 32,819) and algorithm 2 (n = 6,837). RESULTS: The sensitivity, specificity, and positive predictive values were: Algorithm 1: 95 percent, 65 percent, 49 percent; Algorithm 2: 52 percent, 88 percent, 60 percent. Compared to algorithm 1, profiles from algorithm 2 revealed higher rates of follow-up visits (43 percent, 55 percent) and appropriate antidepressant dosage acutely (82 percent, 90 percent) and chronically (83 percent, 91 percent) (p < 0.05 for all). CONCLUSIONS: Both algorithms had high false positive rates. Denominator construction (algorithm 1 versus 2) contributed significantly to variability in measured quality. Our findings raise concern about interpreting depression quality reports based upon administrative data.
B. APPENDIX B: LIST OF CITATIONS SELECTED FOR FULL-TEXT REVIEW BUT NOT INCLUDED, BY REASONS FOR EXCLUSION

1. Studies Excluded Due to Poorly Defined Algorithms


Simon GE, VonKorff M, Barlow W. Health care costs of primary care patients with recognized depression. *Arch Gen Psychiatry*. 1995; 52: 850-856.


2. **Studies Excluded Due to Lack of Validation or Reporting of Validation Statistics**


## C. APPENDIX C: LIST AND DEFINITIONS OF ICD OR PROCEDURAL CODES INCLUDED IN ALGORITHMS

<table>
<thead>
<tr>
<th>Type of Code</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-9</td>
<td>296.2</td>
<td>Major depressive disorder, single episode</td>
</tr>
<tr>
<td>ICD-9</td>
<td>296.3</td>
<td>Major depressive disorder, recurrent episode</td>
</tr>
<tr>
<td>ICD-9</td>
<td>298.0</td>
<td>Depressive type psychosis</td>
</tr>
<tr>
<td>ICD-9</td>
<td>300.4</td>
<td>Neurotic Depression (Dysthymic Disorder)</td>
</tr>
<tr>
<td>ICD-9</td>
<td>309.0</td>
<td>Adjustment reaction with brief depressive reaction</td>
</tr>
<tr>
<td>ICD-9</td>
<td>309.1</td>
<td>Adjustment reaction with prolonged depressive reaction</td>
</tr>
<tr>
<td>ICD-9</td>
<td>309.28</td>
<td>Adjustment reaction with mixed emotional features</td>
</tr>
<tr>
<td>ICD-9</td>
<td>311</td>
<td>Depressive disorder, not elsewhere classified</td>
</tr>
<tr>
<td>Medication</td>
<td></td>
<td>Antidepressant prescription claim</td>
</tr>
</tbody>
</table>
### APPENDIX D: COMPLETE RESULTS FROM FRAYNE, ET AL. (12)

<table>
<thead>
<tr>
<th>Algorithms A through H (ICD-9 Codes)</th>
<th>Percentage</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. 291.89, 292.84, 296.20–296.25, 296.30–296.35, 296.50–296.55, 296.60–296.65, 296.89, 300.4x, 309.0x, 309.28, 311.xx. One or more code, during year of patient survey and preceding year in VHA data.</td>
<td>16.4</td>
<td>0.82</td>
<td>0.74</td>
</tr>
<tr>
<td>B. 291.89, 292.84, 296.20–296.25, 296.30–296.35, 296.50–296.55, 296.60–296.65, 296.89, 300.4x, 309.0x, 309.28, 311.xx. One or more code, during year of patient survey in VHA data.</td>
<td>12.3</td>
<td>0.85</td>
<td>0.71</td>
</tr>
<tr>
<td>C. 291.89, 292.84, 296.20–296.25, 296.30–296.35, 296.50–296.55, 296.60–296.65, 296.89, 300.4x, 309.0x, 309.28, 311.xx. Two or more codes, during year of patient survey and preceding year in VHA data.</td>
<td>11.3</td>
<td>0.88</td>
<td>0.71</td>
</tr>
<tr>
<td>D. 291.89, 292.84, 296.20–296.25, 296.30–296.35, 296.50–296.55, 296.60–296.65, 296.89, 300.4x, 309.0x, 309.28, 311.xx. One or more code from a primary care visit only, during year of patient survey and preceding year in VHA data.</td>
<td>7.0</td>
<td>0.82</td>
<td>0.68</td>
</tr>
<tr>
<td>E. 291.89, 292.84, 296.20–296.25, 296.30–296.35, 296.50–296.55, 296.60–296.65, 296.89, 300.4x, 309.0x, 309.28, 311.xx. One or more code from a mental health visit only, during year of patient survey and preceding year in VHA data.</td>
<td>11.4</td>
<td>0.87</td>
<td>0.71</td>
</tr>
<tr>
<td>F. 291.89, 292.84, 296.20–296.25, 296.30–296.35, 296.50–296.55, 296.60–296.65, 296.89, 300.4x, 309.0x, 309.28, 311.xx. One or more code, during year of patient survey and preceding year in VHA or Medicare data</td>
<td>18.5</td>
<td>0.80</td>
<td>0.75</td>
</tr>
<tr>
<td>G. 296.20–296.25, 296.30–296.35, 296.50–296.55, 300.4x, 309.0x, 309.28, 311.xx. One or more code, during year of patient survey and preceding year in VHA data.</td>
<td>4.5</td>
<td>0.90</td>
<td>0.67</td>
</tr>
<tr>
<td>H. 296.2x, 296.3x, 311. One or more code, during year of patient survey and preceding year in VHA data.</td>
<td>14.0</td>
<td>0.84</td>
<td>0.72</td>
</tr>
</tbody>
</table>

After: Frayne SM, Sharkansky EJ, Wang D, Berlowitz DR, Rosen CS: Using administrative data to identify mental illness: What approach is best? Am J Med Quality 2010;25(1) 42–50. Criterion standard is a survey item in which patients are asked: “Has a doctor ever told you that you have [depression]?”