General Hospital Psychiatry 36 (2014) 563–569

Contents lists available at ScienceDirect

General Hospital Psychiatry

journal homepage: http://www.ghpjournal.com

Baseline screening tools as indicators for symptom outcomes and health services utilization in a collaborative care model for depression in primary care: a practice-based observational study

Nathan D. Shippee, Ph.D. [a,*], Brooke H. Rosen, B.S. [b], Kurt B. Angstman, M.D. [c], Manuel E. Fuentes, M.D. [d], Ramona S. DeJesus, M.D. [e], Steven M. Bruce, M.D. [c], Mark D. Williams, M.D. [f]

[a] Division of Health Policy and Management, University of Minnesota, 420 Delaware St. SE, D375 Mayo MMC 729, Minneapolis, MN 55455
[b] Mayo Medical School, Mayo Clinic
[c] Department of Family Medicine, Mayo Clinic
[d] Department of Psychiatry, Clinica Alemana/Universidad del Desarrollo, Santiago, Chile
[e] Department of Primary Care Internal Medicine, Mayo Clinic
[f] Department of Psychiatry and Psychology, Mayo Clinic

ARTICLE INFO

Article history:
Received 7 February 2014
Revised 6 June 2014
Accepted 27 June 2014

Keywords:
Collaborative care
Depression
Mental and behavioral health screening tools
Utilization
Anxiety

ABSTRACT

Objective: Within a practice-based collaborative care program for depression, we examined associations between positive baseline screens for comorbid mental and behavioral health problems, depression remission and utilization after 1 year.
Methods: This observational study of 1507 depressed adults examined baseline screens for hazardous drinking (Alcohol Use Disorders Identification Test score≥8), severe anxiety (Generalized Anxiety Disorder 7-item score ≥15) and bipolar disorder [Mood Disorders Questionnaire (MDQ) positive screen]; 6-month depression remission; primary care, psychiatric, emergency department (ED) and inpatient visits 1 year postbaseline; and multiple covariates. Analyses included logistic and zero-inflated negative binomial regression.
Results: At unadjusted baseline, 60.7% had no positive screens beyond depression, 31.5% had one (mostly severe anxiety), 6.6% had two and 1.2% had all three. In multivariate models, positive screens reduced odds of remission versus no positive screens [e.g., one screen odds ratio (OR)=0.608, p=.000; all three OR=0.152, p=.018]. Screening positive for severe anxiety predicted more postbaseline visits of all types; severe anxiety plus hazardous drinking predicted greater primary care, ED and inpatient; severe anxiety plus MDQ and the combination of all three positive screens both predicted greater psychiatric visits (all p<.05). Regression-adjusted utilization patterns varied across combinations of positive screens.
Conclusions: Positive screens predicted lower remission. Severe anxiety and its combinations with other positive screens were common and generally predicted greater utilization. Practices may benefit from assessing collaborative care patients presenting with these screening patterns to determine resource allocation.

1. Background

Depression and anxiety are common challenges in primary care, and evidence suggests that recognition and management of these conditions is inconsistent and that mental health resources are scarce [1–4]. These realities have led many practices to improve outcomes through evidence-based programs such as collaborative care for depression, psychotherapy for anxiety disorders and others [5–12]. The use of tracking registries in these programs, along with quality measurement generally, have driven increased use of mental and behavioral screening and tracking tools to assess baseline problems and outcomes [13–16]. As might be expected, positive screens for anxiety, alcohol misuse and other mental and behavioral problems are associated with poorer outcomes in practice-based implementation of collaborative care [17].

However, with the need for sustainability in practice-based settings [18], collaborative care programs are also increasingly being evaluated for cost containment via exploration of primary and secondary cost reduction [19,20]. Beyond an outcomes–utilization relationship found in prior studies [21,22], obtaining a clearer picture of the ties between mental/behavioral symptomatology and service use requires fully leveraging existing program data, including baseline screening tools. Gaining practical insight about patients' potential use of health care moving forward is also important from a care management perspective, since better awareness of patterns of presentation and utilization may help organize resources toward patients based on their expected combinations of visits for different services. Assessing utilization across service settings is key to this effort, since some utilization is needed and beneficial (e.g., services for

* Corresponding author. Tel.: +1-612-624-3579.
E-mail address: nshippee@umn.edu (N.D. Shippee).

http://dx.doi.org/10.1016/j.genhosppsych.2014.06.014
0163-8343/

management of symptoms and prevent worsening of existing conditions), whereas other kinds of use may be targets for reduction.

The present paper examines baseline self-assessments in a collaborative care program for depression in primary care and their association with clinical outcomes and health care utilization. We addressed the following specific aims:

Aim 1 To assess the effects of positive screens for anxiety, bipolar disorder and alcohol use cumulatively predict lower symptomatic remission of depression in collaborative care, or if the specific combinations of screenings have different effects on remission. Severity on baseline mental and behavioral screens is associated with persistent depression in collaborative care [17]. Here, we examined combinations of baseline screens as cohorts to assess how the number and kind of positive screens predict 6-month remission.

Aim 2 To examine how combinations of positive screens predict utilization in terms of visits for primary care, psychiatric specialty, emergency department (ED) and inpatient services over 1-year postscreening. Mental and behavioral health disorders are generally associated with greater utilization [23,24]. In this paper, we assessed whether combinations of positive screens for comorbid problems would differ in their effects on service use.

2. Methods

2.1. Data and sample

This study uses self-reported and electronic health record (EHR) data for patients enrolled in collaborative care for depression at a large, multisite primary care practice in the midwest. Descriptions of this program can be found elsewhere [17,19]. Basically, the collaborative care program was designed to target adult patients in a large primary care practice. Entry criteria for patients included a diagnosis of major depression or dysthymia, a patient health questionnaire, 9-item (PHQ-9) of 10 or more, age 18 or above and no previous diagnosis of bipolar disorder.

The program involved a care coordinator meeting with the patient to gather initial information using screening tools and a standard set of intake information that was presented the following week to the psychiatrist who would review all patients weekly (without physically seeing those patients). Care recommendations for each patient were then given to the primary care provider, who was responsible for all prescriptions. The care coordinator then would connect with the patient on a regular basis with motivational interviewing and behavioral activation techniques used to link health goals with patient-centered goals. Patients needing more direct mental health services were assisted in finding them (therapy or medication management).

For the present study, we assessed patients eligible for collaborative care between March 2008 and March 2011 (total=2525; see Appendix A). Patients were excluded if (a) we did not have research authorization to use their data and (b) the patient opted out of collaborative care (total remaining: 1709). Some patients were missing baseline screening data (n=202 missing), resulting in a cohort of 1507 patients; of these, 1090 also had a 6-month PHQ-9 (with missing questionnaire items due to incomplete follow-up information in the practice-based setting). We outline our use of these two groups below, as well as strategies to assess and adjust for potential bias due to missing data.

2.2. Measures

We employed binary (0–1) variables for each psychosocial screening tool to indicate a severe or threshold level for the respective problem based on existing scoring standards, specifically: (a) severe anxiety based on a Generalized Anxiety Disorder, 7-item (GAD-7) score of 15 or greater as per Spitzer et al. [25]; (b) positive screen for bipolar based on the Mood Disorders Questionnaire (MDQ [26,27]) based on answering "yes" to 7 or more of the 13 behavioral/symptomatic items in Question 1, *yes* to Question 2 indicating cooccurrence of items from Question 1 and *moderate problem* or *serious problem* to Question 3 (relating to life disruption caused by items from Question 1); and (c) hazardous drinking based on a score of 8 or greater on the full Alcohol Use Disorders Identification Test (AUDIT [28]). Importantly, it was not our goal to address the sensitivity or specificity of these measures but, rather, to study the relation of a positive baseline screen on one or more of the scales in collaborative care for depression with symptomatic remission and, especially, utilization. We further created cohorts based on all combinations of positive screens, that is, those who had none; those who had only a positive MDQ, only severe anxiety on the GAD-7 or only hazardous drinking on the AUDIT; those with any combination of only two positive screens; and those with all three positive screens. Finally, we created a simple count of the number of positive screens per patient (0=no screeners positive/severe/hazardous; 1=any single screen being positive, 2=any two positive; and 3=all three positive).

We used 6-month scores on the PHQ-9 scale for depression at 6 months to calculate 6-month remission of depression (as measured by PHQ-9 score of <5 at 6months postbaseline). We did not use baseline PHQ-9 (e.g., an indication of severe depression) alongside other screeners for multiple reasons, including truncated distribution of PHQ-9 scores (since all individuals had a PHQ-9 of 10 or more to be included in the study), the potential for perfect prediction in 6-month remission models (i.e., all patients with severe depression, or all with less-than-severe depression, might have the same outcome on our binary logistic model).

Utilization measures were obtained from EHR review for consenting patients meeting inclusion criteria, and consisted of visit counts for 1 year postbaseline for the following: primary care, psychiatric specialty, ED and inpatient visits (we omitted nonpsychiatric specialty visits due to lack of granularity and limited interpretability). Contacts with care coordinators in collaborative care were not counted as "visits" in these categories. We also used the prior-year count of each visit type as a covariate.

Demographic covariates included an indicator for female (1=female, 0=male) and age in years at baseline.

2.3. Analyses

Beyond descriptive analyses, we conducted bivariate analyses using cross-tabulation and Kruskal–Wallis [29] and chi-square tests of association as appropriate to confirm associations for screening-based cohorts and depression outcomes and, more centrally, the association of those cohorts with utilization. Second, we conducted two sets of logistic regression for 6-month remission of depression, one using simple counts of positive screens and another using screening-based cohort indicators entered separately. Third, we conducted a series of regression models for count outcomes, specifically zero-inflated negative binomial (ZINB) models, to predict each utilization measure separately, with screening-based cohorts entered separately alongside covariates. Zero-inflated models allow for having many zeroes (common in utilization counts) by estimating two equations: first, a binary "zero-inflation" equation predicting a zero count (i.e., a logit model with the outcome of having a zero count versus ≥1 visit; independent variables entered here, based on performance in preliminary logistic models and sensitivity analyses, were prior-year visits of the respective type and age); and second, a "count" equation predicting number of visits (here, with independent variables being screening cohort indicators, age, sex and prior-year visits of the respective type). Comparisons of model fit using log-likelihood and Bayesian and Akaike Information criteria across zero-inflated and

nonzero-inflated Poisson and negative binomial models, as well as statistical tests for comparing count models in Stata, generally indicated consistent support for zero-inflated (accounting for many "zero" counts of visits) negative binomial (allowing for overdispersion) models across utilization types, although in our own sensitivity analyses, the significance and size of coefficients differed little across models. Finally, for interpretability, we examined regression-adjusted estimates for number of each kind of visits, based on the ZINB models above, according to screening-based cohorts.

2.4. Weighting

To assess and adjust for potential bias due to missing data, we conducted logistic regression models that predicted being present/nonmissing for (a) baseline screenings, predicted by age, sex and all available prior-year utilization counts (age was the only significant predictor) and (b) 6-month remission, based on demographics, prior-year utilization and screening-based cohorts [notably, "MDQ positive plus hazardous drinking" was the only significant screening combination: odds ratio (OR)=.123, p=.013].

Based on these regression models, we created inverse probability weights to adjust for bias due to missing screening data (screening weight) and missing remission data (remission weight); the screening weight was applied in the creation of the remission weight (since screening-based cohorts were included as predictors).

Finally, we applied these weights to the main regression models (remission weight applied to the logistic regression predicting remission and the screening weight applied to the ZINB utilization models). ORs, interval-rate ratios, and p values were generally (but not always) smaller at the second or third decimal place in weighted versus unweighted models. Given some differences in point estimates and the fact that weighting did not result in counterintuitive or dubious findings, we present findings from weighted regression models here.

3. Results

Descriptive statistics (Table 1) revealed that most individuals in this collaborative care program had zero (n=915) or only one (n=474) additional positive baseline screen (364 of whom screened only for severe anxiety on the GAD-7). On average, this group was 40years old and mainly female (though age and sex had significant associations with screening combinations; p=0.000 for both). Unadjusted rates for 6-month remission of depression (n=1,090, overall: 53%) generally decreased as the number of positive screens increased (chi-square p=0.000). The main exception being the combination of positive MDQ plus hazardous drinking (only two cases had 6-month PHQ-9 data; neither remitted). Mean counts of visits varied less systematically by screening combinations (though Kruskal–Wallis tests for all but inpatient visits were p<.05). Generally, additional positive screens were associated with higher mean visits, but there were exceptions. For instance, compared to no positive screens, hazardous drinking had slightly lower means for primary care and psychiatric visits but higher mean ED visits, whereas all three positive screens had a lower mean primary care visit count but higher mean visits of other types.

Logistic regression models for 6-month remission of depression (Table 2) confirmed that positive screenings (compared to the referent of zero positive screens) generally decreased the odds of remission. Model 1, using simple counts of baseline screens entered separately, showed significant, incremental decreases in odds of remission for additional screens versus the referent of no positive screens (all p<.05); having all three positive screens was associated with 84.8% lower odds of remission (p=.018). Model 2, using discrete screening-related groups rather than a count, echoed this pattern, with ORs generally similar for the number of positive screens even with different combinations (e.g., ORs for combinations of only one positive screen ranged between .565 and .616). Two cases with MDQ positive plus hazardous drinking were dropped because both predicted nonremission (perfect prediction). Hazardous drinking on the AUDIT trended as expected but only approached marginal significance (OR=0.565, p=0.059); positive MDQ alone, age and sex were not significantly associated with remission.

In ZINB regression results for visit counts 1 year after baseline (Table 3), a positive screening for severe anxiety alone was significantly associated with greater visits for all service settings (e.g., compared to no positive screenings, it was associated with an ED visit rate 62% higher). Severe anxiety plus MDQ positive was associated with visit rates about 50% higher for psychiatric (p<.05) and ED visits (approached marginal significance; p<.1). Severe anxiety plus hazardous drinking on the AUDIT predicted higher visit rates for all but psychiatric visits (all p<.05), while the combination of all three positive screens was associated with an increase of 173% in the ED visit rate. In zero-inflation equations, more prior-year visits in the

Table 1
Descriptive statistics for depressed patients entering collaborative care, by screening-based groups

Positive screens:	Total	None	One positive screening (n=474)			Two positive screenings (n=100)			All three	p value*
			MDQ positive	Severe anxiety (GAD-7)	Hazardous drinking (AUDIT)	MDQ positive, hazardous drinking	MDQ positive, severe anxiety	Severe anxiety, hazardous drinking		
Total *N*	1507	915	30	364	80	11	46	43	18	
Age (mean)	40	42	38	38	33	25	37	38	31	0.000
Female (%)	71%	74%	37%	75%	58%	55%	67%	49%	44%	0.000
6-month depression outcome										
N	1090	685	20	262	50	2	31	29	11	
Remission (%)	53%	59%	45%	48%	44%	0%	26%	28%	18%	
Visits, 1-year following screening (n=1507)										
Primary care (mean #)	2.6	2.5	3.3	3	2	2.9	2.8	3.4	1.7	0.004
Psychiatric	0.9	0.8	0.9	1.1	0.7	1.3	1.3	1.6	2.2	0.010
ED	0.5	0.3	0.9	0.7	0.5	1.1	0.7	0.7	0.7	0.000
Inpatient	0.2	0.2	0.3	0.3	0.2	0.2	0.2	0.4	0.4	0.076
Any psychiatric visit (%)	36%	34%	40%	37%	33%	45%	50%	40%	61%	0.111
Any inpatient visit	15%	12%	20%	18%	15%	9%	22%	23%	22%	0.082
Any emergency room visit	28%	22%	37%	36%	31%	64%	43%	40%	33%	0.000

* Tests for association with screening-based groups: Kruskal–Wallis for continuous variables and chi-square for binary variables.

Table 2
Logistic regression for 6-month remission of depression

	Model 1			Model 2		
	Count of positive screens			Discrete screening cohorts		
	OR	*p* value	95% Confidence interval (CI)	OR	*p* value	95% CI
No positive screens	(reference)			(reference)		
Any one positive screen	0.608	0.000	0.465–0.795			
MDQ positive				0.574	0.235	0.230–1.434
Severe anxiety (GAD-7)				0.619	0.001	0.464–0.827
Hazardous drinking (AUDIT)				0.565	0.059	0.313–1.022
Any two positive screens	0.214	0.000	0.115–0.397			
MDQ positive, Hazardous drinking				[a]		
MDQ positive, severe anxiety				0.244	0.001	0.106–0.561
Severe anxiety, hazardous drinking				0.262	0.002	0.114–0.604
All three positive screens	0.152	0.018	0.032–0.725	0.151	0.018	0.032–0.720
Age	1.006	0.128	0.998–1.014	1.006	0.169	0.998–1.014
Female	1.118	0.428	0.849–1.473	1.110	0.464	0.839–1.469
Intercept	1.046	0.842	0.671–1.631	1.078	0.742	0.689–1.686
N	1090			1088		

Notes: Regression models used inverse probability weighting to adjust for potential bias due to missing 6-month data.
[a] Two cases dropped due to perfect prediction.

respective categories were associated with lower odds of membership in the "zero visits" group for following-year visits in all models; age was only significant for inpatient zero counts ($p<.001$).

For interpretability, Fig. 1 presents adjusted visit counts based on ZINB regression estimates in Table 3. We include estimates only for screening-based groups that had at least one significant association with visit counts, and will highlight only certain combinations for illustration. First, the combination of all three positive screens and its adjusted visit count to primary care (1.9) was the only instance in which a combination of any positive screens had a lower adjusted visit count compared to no positive screens. In contrast, having all positive screens also meant an adjusted 2.3 psychiatry visits, compared to less than one for no positive screens. Severe anxiety alone was in some cases associated with higher adjusted visit counts than its combinations with other positive screens (e.g., with MDQ positive for primary care).

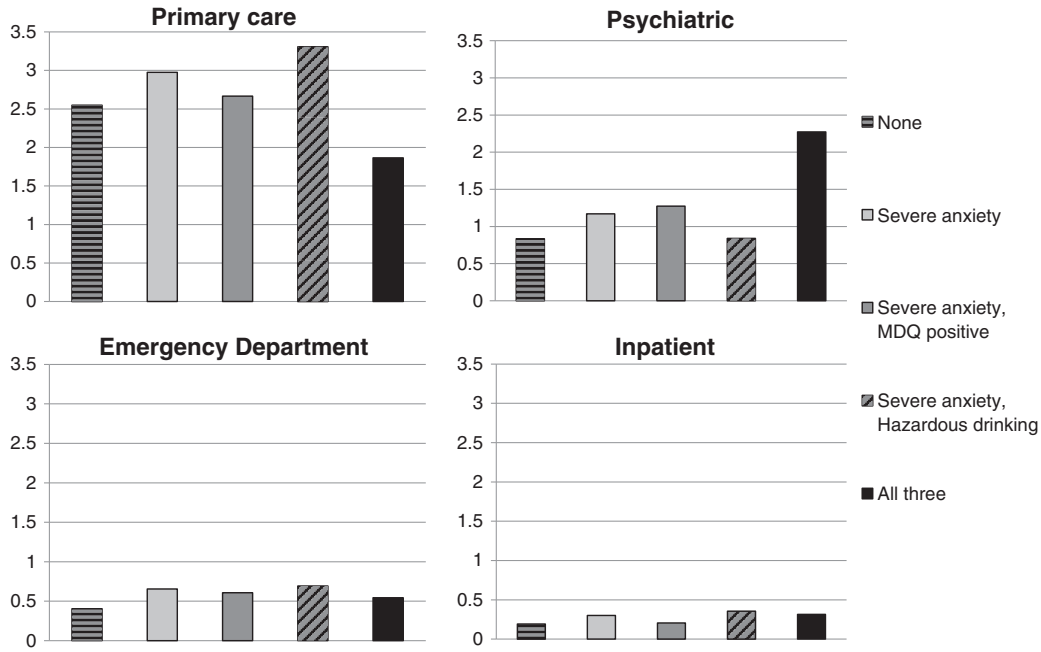
4. Conclusions

The increased interest and prevalence of collaborative care programs has brought with it a need for measurement tools to assist in clarifying patient needs and planning for appropriate allocation of scarce mental health resources. In this analysis of patients in collaborative care for depression, positive screening tools for comorbid psychiatric problems at baseline (positive MDQ, severe anxiety on the GAD-7 and hazardous drinking on the AUDIT) were generally associated with lower remission of depression and higher visits in primary care, psychiatric, emergency and inpatient services over 1 year. Whether these combinations of positive screens truly indicate combinations of bipolar, anxiety and alcohol misuse disorders, they did signal worse symptomatic outcomes and differences in utilization over time.

Table 3
ZINB models for visits after 1 year among patients in collaborative care for depression

	Primary care visits			Psychiatric visits			ED visits			Inpatient visits		
	IRR	95% CI		IRR	95% CI		IRR	95% CI		IRR	95% CI	
Visit count equation												
Positive screening cohorts												
None	(Reference)											
MDQ positive	1.21	.87	1.68	1.02	0.47	2.24	1.35	0.82	2.22	1.25	0.62	2.54
Severe anxiety (GAD-7)	1.17*	1.03	1.32	1.40*	1.05	1.86	1.62***	1.31	2.02	1.57*	1.10	2.25
Hazardous drinking (AUDIT)	0.83	0.64	1.07	0.87	0.53	1.41	1.14	0.78	1.65	1.10	0.60	2.01
MDQ positive, hazardous drinking	1.34	.65	2.73	2.04	0.74	5.61	1.40	0.61	3.18	1.70	0.20	14.04
MDQ positive, severe anxiety	1.04	.81	1.35	1.52*	1.01	2.30	1.50†	0.93	2.43	1.07	0.65	1.77
Severe anxiety, hazardous drinking	1.30*	1.02	1.66	1.01	0.48	2.11	1.72*	1.10	2.69	1.86*	1.07	3.24
All three	0.73	0.50	1.08	2.73**	1.51	4.92	1.36	0.44	4.15	1.65	0.85	3.19
Age	1.00	0.99	1.00	1.01†	1.00	1.02	0.99*	0.98	0.999	1.00	0.99	1.01
Female	1.09	0.98	1.22	0.93	0.72	1.20	1.26*	1.003	1.59	0.77	0.55	1.07
Prior year visits, same category	1.10***	1.08	1.12	1.34***	1.22	1.47	1.34***	1.22	1.47	1.34***	1.23	1.47
Constant	1.97***	1.65	2.36	0.46**	0.29	0.73	0.51**	0.32	0.83	0.45*	0.21	0.98
Zero-inflation equation												
	Coef.	95% CI		Coef.	95% CI			95% CI		Coef.	95% CI	
Prior year visits, same category	−4.66***	−5.74	−3.58	−1.03*	−2.03	−0.02	−1.17***	−1.76	−0.59	−1.21***	−1.84	−0.58
Age	0.02	−0.01	0.05	0.11	−0.02	0.24	0.01	−0.01	0.03	−0.02**	−0.04	−0.01
Constant	1.57*	0.22	2.91	−9.00†	−19.51	1.51	−0.84	−1.91	0.23	1.44**	0.61	2.27

Notes: Regression models used inverse probability weighting to adjust for potential bias due to missing baseline screening data. IRR=incident-rate ratio; Coef=Coefficient.
† p<.10.
* p<.05.
** p<.01.
*** p<.001.



Notes: Based on zero-inflated negative binomial models, controlling for age, sex, and prior year visits, and weighted to adjust for potential bias due to missing data. Only screening combinations with significant effects (P<.05) for at least one visit type are shown.

Fig. 1. Adjusted visit counts by patients in collaborative care for depression, 1-year postbaseline (n=1507). Notes: Based on ZINB models, controlling for age, sex and prior year visits, and weighted to adjust for potential bias due to missing data. Only screening combinations with significant effects (p<.05) for at least one visit type are shown.

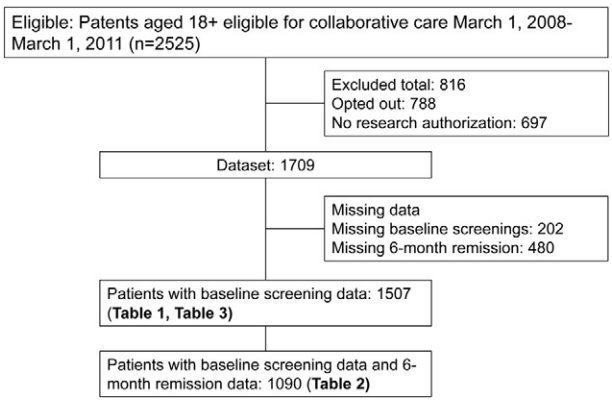
Individuals in collaborative care for depression who had no other positive screeners had higher rates of remission than the sample overall, whereas those with one or more positive screens (and especially combinations including severe anxiety) had lower odds of 6-month remission of depression, as expected based on prior literature [17]. However, these associations were generally consistent for the number, rather than kind, of positive screens, with unadjusted remission rates and ORs from differently configured multivariate models indicating similar patterns. Thus, rather than specific type of response, parsimonious counts of these baseline screens — representing a volume of symptom burden — may be worth attending to in understanding likelihood of remission as a plan of care is created for patients entering care coordination.

Regarding visits for different services over 1 year postbaseline, overall, in multivariate models, just under half of the baseline screenings' associations with different visits (18 of 32 total associations tested) were not significant. Of those screening combinations that did have significant effects, the common factor was a screening for severe anxiety on the GAD-7, which alone was a robust predictor of greater visits for all service settings compared to no positive screens. While this might suggest limited statistical power due to small sizes for other groups, severe anxiety plus hazardous drinking on the AUDIT — a combination with only 43 individuals — also significantly predicted most visit types, and even the 18 individuals with all three positive screens had greater psychiatric visits. Beyond statistical power, differing associations with utilization (and limited significant effects in multivariate models) may be due to multiple factors. First, nonsignificant results may also be the results of mixed causal pathways. For example, hazardous drinking on the AUDIT alone, though not significant, often trended toward lower utilization than no positive screens, and existing evidence in the literature supports this negative association [30]. If screenings for hazardous drinking (or potentially the positive MDQ indicating a substance use disorder; see below) represent patient-driven reasons for decreased use of some (potentially necessary) visits, but more acute health problems and need for other visits, then these screens may be encompassing multiple, sometimes contradicting, effects on utilization. A second issue may be that screens had different recall periods (e.g., 2weeks for GAD-7 vs. lifetime for MDQ), with longer recalls lowering accuracy [31,32] and, thereby, potentially masking associations between real symptoms and utilization. A related third issue is these tools' varying sensitivity and specificity in assessing specific disorders (and thus indications of specific morbidities with different implications for service use). For instance, the MDQ has a high false-positive rate, and may also indicate borderline personality disorder, posttraumatic stress disorder and other anxiety conditions, substance use, unipolar depression and other disorders [33–35]; thus, in addition to limited use in predicting bipolar disorders themselves, a positive MDQ may represent a heterogeneous group of comorbid behavioral issues with varying effects on utilization. Other issues not able to be assessed here which may confound, moderate or otherwise mask the true effects of screeners here include comorbid medical conditions, differing coverage or access to care (which may prevent individuals from using care or change the way they use care despite having positive screens) and sizes of groups. In particular, comorbidity of physical and mental health is likely relevant in fully understanding findings regarding utilization, but we are not able to address it with the data here. Still, even with these issues in evidence, current findings indicate that the presence of a positive screen for severe anxiety appears in a sizable proportion of collaborative care patients in this practice setting and is consistently associated with greater utilization after controlling for other screens and prior-year utilization. Of course, some utilization is beneficial; higher visits for primary care might be crucial in preventing acute care and maintaining health, whereas psychiatry visits may indicate additional services for therapy for diagnosed problems (potentially facilitated through collaborative care). Therefore, severe anxiety and its combinations' association with higher ED and inpatient visits may be the greatest care management targets for collaborative care or similar programs.

Beyond an agnostic approach to the validity of positive screens as diagnostic tools, this study's limitations include somewhat limited sample size, which created limited cell size for certain combinations (such as MDQ positive plus hazardous drinking). In a related sense, the data are constrained by certain patterns of missing among screening tools and 6-month remission data, as well as bias for which we could not adjust. Specifically, we cannot test whether those eligible for collaborative care during the study period who opted out or did not provide authorization for the use of their data for research may be systematically different than those who we included in our sample. In addition, the lack of data on health conditions, and lack of availability or variability in other sociodemographic controls beyond age and sex, limits our ability to further explain and control for the factors guiding utilization patterns and hinders external validity. In particular, this study was conducted in the midwest US, with a largely White population; hence, findings may not be applicable to minority groups. Finally, utilization came from a single health system, and so utilization outside that system cannot be measured, despite the fact that patients may have had visits elsewhere. There are no published data on this issue of potential leakage regarding the primary care practice under study here. Individuals who are the local patients of the practice have two facilities to choose from in receiving care, but the two are very different (the practice under study here is part of an academic medical center and is much larger). While this provides for the possibility of leakage, the likelihood of this being significant is relatively low.

Despite limitations, this paper demonstrates that baseline screening tools, and their combinations, are associated with variations in rates of remission of depression and visits for health care services. As patients enter collaborative care programs for depression, psychiatrists in the role of supervising care coordinators may consider attending to the psychiatric symptom burden in deciding who needs more attention, and paying special attention to anxiety, alone and combined with other positive screens, may be wise in planning resource allocation to improve outcomes and sustainability. In sum, this study supports the utility of leveraging the data gathered from baseline screening tools to inform and anticipate outcomes and needs for services, which will help optimize the sustainability and effectiveness of collaborative care and other evidence-based programs.

Appendix A. Flow chart: Eligibility, exclusion and missing data



References

[1] Craven MA, Bland R. Depression in primary care: current and future challenges. Can J Psychiatry 2013;58(8):442–8.
[2] Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007;146(5):317–25.
[3] Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. Health Aff (Millwood) 2009;28(3):w490–501.
[4] Williams Jr J, Kerber C, Mulrow C, Medina A, Aguilar C. Depressive disorders in primary care. J Gen Intern Med 1995;10(1):7–12.
[5] Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. Arch Intern Med 2006;166(21):2314–21.
[6] Katon WJ, Lin EHB, Von Korff M, et al. Collaborative care for patients with depression and chronic illnesses. N Engl J Med 2010;363(27):2611–20.
[7] Archer J, Bower P, Gilbody S, et al. Collaborative care for depression and anxiety problems. Cochrane Database Syst Rev 2012;10[Art. No: CD006525].
[8] Roy-Byrne P, Craske MG, Sullivan G, et al. Delivery of evidence-based treatment for multiple anxiety disorders in primary care: a randomized controlled trial. JAMA 2010;303(19):1921–8.
[9] Sullivan G, Craske MG, Sherbourne C, et al. Design of the coordinated anxiety learning and management (calm) study: innovations in collaborative care for anxiety disorders. Gen Hosp Psychiatry 2007;29(5):379–87.
[10] AIMS Center. Aims center history. [cited 2014 28 May]; Available from: http://aims.uw.edu/who-we-are/aims-center-history; 2014.
[11] Institute for Clinical Systems Improvement. The diamond program: success in primary care depression treatment and extension to other health care challenges [white paper]. Bloomington, MN: Institute for Clinical Systems Improvement; 2013.
[12] SAMHSA-HRSA Center for Integrated Health Solutions. Behavioral health in primary care: integrating behavioral health into primary care. [cited 2014 28 May]; Available from: http://www.integration.samhsa.gov/integrated-care-models/behavioral-health-in-primary-care; 2014.
[13] Das AK, Olfson M, Gameroff MJ, et al. Screening for bipolar disorder in a primary care practice. JAMA 2005;293(8):956–63.
[14] O'Connor EA, Whitlock EP, Beil TL, Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. Ann Intern Med 2009;151(11):793–803.
[15] Pignone MP, Gaynes BN, Rushton JL, et al. Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;136(10):765–76.
[16] Snowden AM, Vetta E, LaFrance A, Mlodzik R, Xiong M. Mn community measurement 2012 health care quality report. [cited 2013 November 7]; Available from: http://mncm.org/wp-content/uploads/2013/04/2012_Final_HealthCareQualityReport_2.18.13.pdf; 2012.
[17] Angstman KB, Shippee ND, MacLaughlin KL, et al. Patient self-assessment factors predictive of persistent depressive symptoms 6 months after enrollment in collaborative care management. Depress Anxiety 2013;30(2):143–8.
[18] Katon W, Unützer J, Wells K, Jones L. Collaborative depression care: history, evolution and ways to enhance dissemination and sustainability. Gen Hosp Psychiatry 2010;32(5):456–64.
[19] Shippee ND, Shah ND, Angstman KB, et al. Impact of collaborative care for depression on clinical, functional, and work outcomes: a practice-based evaluation. J Ambul Care Manage 2013;36(1):13–23.
[20] Unutzer J, Katon WJ, Fan M, et al. Long-term cost effects of collaborative care for late-life depression. Am J Manag Care 2008;14(2):95–100.
[21] Angstman KB, Oberhelman S, Rohrer JE, et al. Depression remission decreases outpatient utilization at 6 and 12 months after enrollment into collaborative care management. Popul Health Manag 2013. http://dx.doi.org/10.1089/pop.2013.0004 [Epub ahead of print].
[22] Meunier MR, Angstman KB, MacLaughlin KL, et al. Impact of symptom remission on outpatient visits in depressed primary care patients treated with collaborative care management and usual care. Popul Health Manag 2013. http://dx.doi.org/10.1089/pop.2013.0057 [Epub ahead of print].
[23] Cohen B, Gima K, Bertenthal D, et al. Mental health diagnoses and utilization of va non-mental health medical services among returning Iraq and Afghanistan veterans. J Gen Intern Med 2010;25(1):18–24.
[24] Birnbaum HG, Kessler RC, Kelley D, et al. Employer burden of mild, moderate, and severe major depressive disorder: mental health services utilization and costs, and work performance. Depress Anxiety 2010;27(1):78–89.
[25] Spitzer RL, Kroenke K, Williams JW, Löwe B. A brief measure for assessing generalized anxiety disorder: the gad-7. Arch Intern Med 2006;166(10):1092–7.
[26] Hirschfeld RM. The mood disorder questionnaire: a simple, patient-rated screening instrument for bipolar disorder. Prim Care Companion J Clin Psychiatry 2002;4(1):9–11.
[27] Hirschfeld RM, Williams JB, Spitzer RL, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the mood disorder questionnaire. Am J Psychiatry 2000;157(11):1873–5.
[28] Saunders JB, Aasland OG, Babor TF, De La Fuente JR, Grant M. Development of the alcohol use disorders identification test (audit): who collaborative project on early detection of persons with harmful alcohol consumption-ii. Addiction 1993;88(6):791–804.
[29] Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952;47(260):583–621.
[30] Zarkin GA, Bray JW, Babor TF, Higgins-Biddle JC. Alcohol drinking patterns and health care utilization in a managed care organization. Health Serv Res 2004;39(3):553–70.
[31] Broderick JE, Schwartz JE, Vikingstad G, et al. The accuracy of pain and fatigue items across different reporting periods. Pain 2008;139(1):146–57.

[32] Wells JE, Horwood LJ. How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. Psychol Med 2004;34(06):1001–11.
[33] Zimmerman M, Galione JN, Chelminski I, Young D, Dalrymple K. Psychiatric diagnoses in patients who screen positive on the mood disorder questionnaire: implications for using the scale as a case-finding instrument for bipolar disorder. Psychiatry Res 2011;185(3):444–9.
[34] Zimmerman MGJ, Ruggero CJ, Chelminski I, Young D, Dalrymple K, McGlinchey JB. Screening for bipolar disorder and finding borderline personality disorder. J Clin Psychiatry 2010;71(9):1212–7.
[35] Parker G, Graham R, Rees A-M, Futeran S, Friend PA. A diagnostic profile of those who return a false positive assignment on bipolar screening measures. J Affect Disord 2012;141(1):34–9.