Original Article

Author for correspondence:
Weng-Yee Chin, E-mail: chinwy@hku.hk

Tree analysis modeling of the associations between PHQ-9 depressive symptoms and doctor diagnosis of depression in primary care

Weng-Yee Chin[1], Eric Yuk Fai Wan[2], Christopher Dowrick[3], Bruce Arroll[4] and Cindy Lo Kuen Lam[5]

[1]Department of Family Medicine & Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong; [2]Nuffield Department of Population Health, University of Oxford, Oxford, UK; [3]Institute of Psychology Health and Society, University of Liverpool, Liverpool, UK; [4]Department of General Practice and Primary Health Care, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand and [5]Department of Family Medicine & Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

Abstract

Background. The aim of this study was to explore the relationship between patient self-reported Patient Health Questionnaire-9 (PHQ-9) symptoms and doctor diagnosis of depression using a tree analysis approach.

Methods. This was a secondary analysis on a dataset obtained from 10 179 adult primary care patients and 59 primary care physicians (PCPs) across Hong Kong. Patients completed a waiting room survey collecting data on socio-demographics and the PHQ-9. Blinded doctors documented whether they thought the patient had depression. Data were analyzed using multiple logistic regression and conditional inference decision tree modeling.

Results. PCPs diagnosed 594 patients with depression. Logistic regression identified gender, age, employment status, past history of depression, family history of mental illness and recent doctor visit as factors associated with a depression diagnosis. Tree analyses revealed different pathways of association between PHQ-9 symptoms and depression diagnosis for patients with and without past depression. The PHQ-9 symptom model revealed low mood, sense of worthlessness, fatigue, sleep disturbance and functional impairment as early classifiers. The PHQ-9 total score model revealed cut-off scores of >12 and >15 were most frequently associated with depression diagnoses in patients with and without past depression.

Conclusions. A past history of depression is the most significant factor associated with the diagnosis of depression. PCPs appear to utilize a hypothetical-deductive problem-solving approach incorporating pre-test probability, with different associated factors for patients with and without past depression. Diagnostic thresholds may be too low for patients with past depression and too high for those without, potentially leading to over and under diagnosis of depression.

Introduction

Depression is common in primary care settings with prevalence estimates ranging between 5–20% (Mitchell *et al.* 2009). In recent years, primary care physicians (PCPs) have come under increasing scrutiny for both underrecognition and overdetection (Höfler & Wittchen, 2000; Mitchell *et al.* 2009) of depression due to poor concordance with various gold standard diagnostic criteria (Ani *et al.* 2008; Cepoiu *et al.* 2008; McGrady *et al.* 2010). This highlights the challenges associated with making mental health diagnoses in primary care settings (Carey *et al.* 2014). Whilst underrecognition raises concern due to the high societal costs associated with disability, poor quality of life, morbidity and potentially mortality (Üstün *et al.* 2004; Cornelius *et al.* 2014), over-detection can also be problematic due to unnecessary treatments and the potential for adverse drug effects (Mitchell *et al.* 2011). Despite this, most PCPs do not use screening instruments or symptom inventories to identify depression, preferring to make decisions based on their own clinical assessment and judgment (Bermejo *et al.* 2005). One of the key challenges in deciding whether depression is present or not is the clinician's ability to differentiate whether the patient has a physical problem, a psychological problem, or both, in a setting where patients are more likely to present with somatic symptoms even when the problem is psychological (Tylee & Gandhi, 2005; Bekhuis *et al.* 2015). A large multi-center study found that 45–95% of primary care patients with depression presented with physical symptoms and 11% did not report any psychological symptoms (Simon *et al.* 1999). This has led some to speculate whether the criteria used for diagnosing depression in primary care should be different from those used in specialist settings (Armstrong & Earnshaw, 2004; Gask *et al.* 2008; Malhi *et al.* 2014; Goldberg *et al.* 2017).

Although many studies have explored the factors associated with depression recognition, most have used logistic regression analyses (Barbui & Tansella, 2006; Akhtar-Danesh & Landeen, 2007; Cepoiu *et al.* 2008; Henriques *et al.* 2009; Chin *et al.* 2014). Regression modeling, however, is limited because the results can be difficult to interpret when a large list of variables are assessed together. It also does not provide information about how factors interplay (Lemon *et al.* 2003). Decision tree (also known as classification tree) analysis is a non-parametric statistical procedure that identifies mutually exclusive subgroups within a study population that share common characteristics associated with an outcome. The analysis examines all possible independent or splitting variables and selects the one that results in binary groups that are most different in respect to the dependent variable. The tree continues to grow by assessing each of the remaining independent variables to determine which variable results in the best split. The process continues until a terminal node is reached. The end product is a visual multi-level output resembling the branches of a tree with the probabilities of having the dependent measure for each node (Lemon *et al.* 2003). Tree analyses have long been used for financial forecasting, but it can have useful applications to exploring large healthcare datasets (Podgorelec *et al.* 2002; Dowrick *et al.* 2011; Kuhn *et al.* 2014).

The aim of this study was to analyze the statistical associations between patient self-reported depressive symptoms and PCP diagnosis of depression to develop a model that might simulate the decision-making process involved in diagnosing depression. Insights gained from the model can help further our understanding of how doctors make a diagnosis of depression in primary care and identify areas for quality improvement. The findings can also further our understanding of what PCPs are diagnosing as depression.

Methods

Study design

This was a secondary analysis of the baseline data obtained from a cohort study examining the epidemiology of depressive disorders in Hong Kong's primary care setting. Baseline subject recruitment occurred between October 2010 and January 2012. The study protocol and findings from the primary analyses have been previously published (Chin *et al.* 2012; Chin *et al.* 2014; Chin *et al.* 2015*a*, *b*; 2016*a*, *b*).

Setting and subjects

PCPs working in clinics across Hong Kong were invited to collaborate as part of a practice-based research network. Doctors were identified using the mailing list of the Hong Kong College of Family Physicians and comprised PCPs working in various private practice settings (solo, group, out-patient departments of private hospitals), government-funded General Outpatient Clinics of the Hong Kong Hospital Authority, and non-profit, non-governmental organizations, representative of how primary care is delivered in Hong Kong. Characteristics of the 59 doctors who participated in the study and their clinics have been previously described (Chin *et al.* 2014). All consecutive, eligible patients presenting on one randomly allocated day each month over a 12-month recruitment period were approached to complete a self-administered survey. Patients were excluded if they were <18 years, did not understand English, Cantonese or Mandarin, had

cognitive or communication difficulties, had previously been recruited to the study, or did not consult the study doctor.

Study instruments

The Patient Health Questionnaire-9 (PHQ-9) is a ten-item questionnaire commonly used to screen, monitor, diagnose, and measure the severity of depressive symptoms in primary care and as an outcome instrument in treatment studies in specialist care settings. (Spitzer *et al.* 1999; Guo *et al.* 2017). It was used to collect the patients' self-reported symptoms of depression. The PHQ-9 scores each of the nine diagnostic criteria of the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) for depression from 0 (not at all) to 3 (nearly every day) over the previous 2 weeks, with a tenth item on functional impairment which does not contribute to the total score (Kroenke *et al.* 2001). The Chinese version of the PHQ-9 used for this study has been validated in Hong Kong (Yu *et al.* 2012).

Questions on patient socio-demographics, recent health service use and co-morbidities were adapted from previously performed health services studies (Lam *et al.* 2011). Information collected included: age (continuous); gender (Male; Female); marital status (Married; other including Single/Separated/Divorced/Widowed); employment status (Yes; No); household income (⩽HK\$ 30 000; >HK\$ 30 000); presence of chronic diseases (None; One; Two or more); self-reported past history of doctor-diagnosed depression or mental illness (Yes; No); family history of mental illness (Yes; No); and number of doctor visits in the past month (continuous).

PCP diagnosis of depression was collected using a case record form completed by the study doctor during the index consultation. Doctors were instructed to document whether they thought the patient had a clinically significant depressive disorder (yes/no). The PCP's diagnosis of depression was made spontaneously without the aid of any screening or diagnostic tools. Doctors were blinded to their patient's PHQ-9 responses. Only the doctor's diagnosis was recorded, and the contents of the medical consultation were not collected, so it is not known whether or not the patient's self-reported PHQ-9 symptoms were elicited during the clinical encounter.

Data analysis

Patient characteristics and PHQ-9 total scores stratified by depression diagnosis status were analyzed using descriptive statistics. Differences between groups were evaluated by independent *t* test for continuous variables and χ^2 for categorical variables. The association between PHQ-9 item responses and diagnosis of depression were examined using a logistic regression model. A second model was performed using PHQ-9 total scores in lieu of the individual item scores. Analyses were controlled for socio-demographic characteristics. Only complete data were used. Odds ratios (OR) for each factor were reported with a 95% confidence interval (CI). The goodness of fit was examined by the Hosmer and Lemeshow test (Paul *et al.* 2013) and presence of multicollinearity was checked using variance inflation factors.

For a more in-depth exploration of the correlations, a conditional inference decision tree model was developed through binary recursive partitioning of potential predictors that maximized between-group differences allowing for interactions, using the 'ctree' function in the *party* package in R (Hothorn *et al.* 2010). At each node of the tree, the recursive partitioning algorithm

was reapplied to select the classifier and the split in this classifier that allowed the maximal difference in the rate of diagnosis of depression between the two subgroups. The algorithm was reported recursively until the tree was grown to an optimal number of terminal leaves. Conditional inference tree is an extension of a classification and regression tree analysis where selection procedures are based on the associations between classifiers and the rate of diagnosis of depression using a significance test using a quadratic form of statistic rather than impurity functions such as Gini Index and entropy. This method ensures that the right sized tree is developed requiring no form of pruning or cross-validation (Hothorn *et al.* 2006). The model also allows for all data to be used with no need to deal with missing data. Unlike multivariable logistic regression, this method can maximize between-group differences, allowing for higher-order interactions between classifiers and a graphical display of the results. Two tree models were generated: the first using PHQ-9 symptoms, and the second using PHQ-9 total scores.

All significance tests were two-tailed and those with *p* values <0.05 were considered statistically significant. Statistical analyses were performed in Stata Version 13.0 and R Version 3.3.1.

Ethics

The Institutional Review Board of the University of Hong Kong/ Hospital Authority Hong Kong West Cluster (reference number UW 09–305), and all other relevant regional ethics committees approved this study.

Results

Overall, 10 179 patients consented to participate with a response rate of 81% of all eligible subjects approached. Of these, 9263 subjects who had complete data for both the doctor's diagnosis and all PHQ-9 items were included for further analysis. Amongst these, 594 patients (6.4%) received a diagnosis of depression. Among those with a PHQ-9 total score >9, 23.2% received a PCP diagnosis of depression. Amongst those with a PCP diagnosis of depression, 41.2% had PHQ-9 scores >9. Table 1 displays the subject characteristics categorized by depression diagnosis status. A comparison of the subjects who were included and excluded from the analyses found no statistical differences in characteristics between the two groups (Appendix 1).

Table 2 shows the results of the logistic regression analysis of the patient characteristics associated with a PCP diagnosis of depression. History of mental illness had the largest effect size with an OR of 10.97. Significant PHQ-9 items included depressed mood (OR 1.33), sleep disturbance (OR 1.17), appetite change (OR 1.15), worthlessness (OR 1.26) and functional impairment (OR 1.35). The total PHQ-9 score was significant but had a small effect size (OR 1.11). Other significant factors included female gender (OR 1.47), increasing age (OR 1.02), employment (OR 0.7), family history of mental illness (OR 1.69), doctor visit in the past month (OR 1.1). Of note, anhedonia and suicidal ideation were not significantly associated with a depression diagnosis. Variance inflation factors ranged from 1.03 to 2.18 indicating no multicollinearity. Hosmer and Lemeshow's test demonstrated the model fitted the data well with a *p* value of 1.00.

Figure 1 shows the PHQ-9 symptom tree model. The root of the tree was formed by past history of depression (yes/no), with the frequency of depressed mood forming the parent branch. In patients with past depression, subsequent branching occurred

with the presence of fatigue and feelings of worthlessness. In patients with no past depression, subsequent branching occurred with a frequency of sleep disturbance and severity of functional impairment. There were a total of 17 terminal nodes/groups with diagnosis rates ranging from 0.5% to 55%. For example, in patients with past depression and frequent to very frequent depressed mood and fatigue present, 55% were diagnosed as having depression. In patients with no past depression, frequent depressed mood and moderate to severe functional impairment, 27.3% were diagnosed as having depression.

Figure 2 shows the PHQ-9 total score tree model. Once again, past history of the depression formed the root of the tree with the PHQ-9 total score forming the parent branch. In patients with a past history of depression, branch splitting occurred at a PHQ-9 total score of >12. In patients with no past history of depression, a four-way split was observed at PHQ-9 total scores of ⩽5, 6–9, 10–15 and >15. There were 12 terminal nodes with depression diagnosis rates ranging from 0.4 to 70%. For example, in patients with past depression, 61.2% of subjects with PHQ-9 total score >12 were diagnosed with depression. In patients with no past depression, 26.8% of patients with PHQ-9 total scores >15 were diagnosed with depression.

Factors identified in the logistic regression were slightly different to the tree classifiers. This was expected as interaction effects between factors are fully taken into account in the decision tree analysis, whereas no interaction effect is assumed in the logistic regression.

Discussion

This is the first study to use an inference tree analysis to model the factors associated with depression diagnosis in primary care. Our models demonstrated the key criteria associated with PCP diagnosis of depression (including past history of depression, depressive symptoms and symptom severity), in the form of a tree diagram, illustrating how the variables interact. Our findings provide new insights into the complex processes involved in diagnosing depression in primary care. Although the study did not provide any evidence regarding the accuracy of PCP diagnosis, it revealed information on real-world clinician behaviors related to diagnosing depression.

From the waiting room questionnaire, only 23.2% of subjects with PHQ-9 total score >9 received a diagnosis of depression, while in those who received a diagnosis of depression, only 41.2% had PHQ-9 scores >9. This suggests that PCP diagnosis of depression in our setting has both low sensitivity and poor predictive value (if using the PHQ-9 scores with a cut-off of >9 as the diagnostic gold standard), which may have repercussions on the adequacy of depression care.

It was observed from the logistic regression that a past history of depression or other mental illness was strongly associated with a diagnosis of depression, whilst all other significant factors including depressive symptoms and depression severity had much smaller effect sizes with OR <2.0. The tree models further demonstrated this with a history of depression emerging as the parent branch.

The complexities associated with depression diagnoses in primary care were demonstrated in the PHQ-9 total score tree analysis where it was observed that many patients who had received a PCP diagnosis of depression had PHQ-9 total scores <9 with some diagnoses made at quite low PHQ-9 scores ⩽5. It is possible

Table 1. Subject characteristics of depression diagnosis status

Subject characteristics	Received a diagnosis of depression by the study doctor[a]		
	No (*N* = 8669)	Yes (*N* = 594)	*p* value
Gender (*n*, %)			**<0.001***
Male	3651 (96.1%)	150 (3.9%)	
Female	4849 (92.0%)	423 (8.0%)	
Age (Mean ± s.d.)	48.78 ± 18.04	53.28 ± 17.94	**<0.001***
Marital status (*n*, %)			**0.223**
Married	5214 (94.0%)	334 (6.0%)	
Single/Separated/Divorced/Widowed	3252 (93.3%)	232 (6.7%)	
Employment (*n*, %)			**<0.001***
No	3014 (90.7%)	309 (9.3%)	
Yes	5367 (95.7%)	243 (4.3%)	
Disease co-morbidity (*n*, %)			**<0.001***
None	4189 (95.4%)	204 (4.6%)	
One	2135 (93.7%)	144 (6.3%)	
Two or more	2102 (90.5%)	221 (9.5%)	
Family history mental illness (*n*, %)			**<0.001***
No	7610 (94.3%)	457 (5.7%)	
Yes	696 (87.1%)	103 (12.9%)	
Seen a doctor in past 4 weeks (Mean ± s.d.)	0.93 ± 1.38	1.66 ± 2.22	**<0.001***
Self-reported depression/other mental illness history (*n*, %)			**<0.001***
No	7871 (96.8%)	264 (3.2%)	
Yes	500 (62.4%)	301 (37.6%)	
Household income (*n*, %)			**<0.001***
≤HK\$ 30 000	2846 (95.4%)	138 (4.6%)	
>HK\$ 30 000	4559 (92.9%)	350 (7.1%)	
PHQ9 item[b] (Mean ± s.d.)			
Anhedonia	0.43 ± 0.71	0.98 ± 1.11	**<0.001***
Depressed mood	0.41 ± 0.67	1.17 ± 1.05	**<0.001***
Sleep disturbance	0.82 ± 1.01	1.60 ± 1.20	**<0.001***
Fatigue	1.03 ± 0.98	1.58 ± 1.14	**<0.001***
Appetite change	0.44 ± 0.74	0.84 ± 1.05	**<0.001***
Worthlessness	0.27 ± 0.59	0.82 ± 1.00	**<0.001***
Poor concentration	0.34 ± 0.69	0.87 ± 1.04	**<0.001***
Agitated or lethargic	0.24 ± 0.60	0.70 ± 1.02	**<0.001***
Suicidal ideation	0.04 ± 0.25	0.24 ± 0.63	**<0.001***
Functional impairment[c]	0.17 ± 0.80	0.84 ± 0.99	**<0.001***
Doctor's characteristics			
Age (Mean ± s.d.)	43.14 ± 11.31	46.01 ± 10.59	**<0.001***
Gender (*n*, %)			**0.007***
Male	6195 (93.2%)	455 (6.8%)	
Female	2474 (94.7%)	139 (5.3%)	
System setting (*n*, %)			0.187

(Continued)

Table 1. (Continued.)

	Received a diagnosis of depression by the study doctor[a]		
Subject characteristics	No (N = 8669)	Yes (N = 594)	p value
Private	6416 (93.8%)	425 (6.2%)	
Public	2253 (93.0%)	169 (7.0%)	

Bold indicates that the p-value is significant (i.e. p < 0.05).
[a]Diagnosis made at the time of the index visit and reported by the PCP on case record forms.
[b]'Over the last 2 weeks, how often have you been bothered by'; Response option = 0 to 3 (0 = Not present; 1 = Infrequent; 2 = Frequent; 3 = Very frequent).
[c]Response option ranges from 0 to 4 (0 = Not applicable; 1 = No difficulty; 2 = Some difficulty; 3 = Difficult; 4 = Very difficult).
*Significant differences (p < 0.05) between groups by independent t test or by χ^2 as appropriate.
Note: Missing value categories are omitted.

Table 2. Factors associated with doctor diagnosis of depression by non-linear mixed effect model

	Using PHQ-9 symptom items[a]			Incorporating PHQ-9 total score[a]		
	Odds ratio	(95% CI)	p value	Odds ratio	(95% CI)	p value
PHQ9 item[b] (Mean ± S.D.)						
Anhedonia	1.10	(0.95–1.27)	0.199	–		
Depressed mood	**1.33**	**(1.13–1.57)**	**0.001***	–		
Sleep disturbance	**1.17**	**(1.05–1.32)**	**0.005***	–		
Fatigue	0.92	(0.81–1.05)	0.236	–		
Appetite change	**1.15**	**(1.01–1.32)**	**0.034***	–		
Worthlessness	**1.26**	**(1.07–1.48)**	**0.006***	–		
Poor concentration	1.11	(0.97–1.28)	0.132	–		
Agitated or lethargic	1.03	(0.88–1.20)	0.708	–		
Suicidal ideation	0.89	(0.69–1.15)	0.373	–		
Functional impairment	**1.35**	**(1.15–1.59)**	**<0.001***	**1.38**	**(1.17–1.62)**	**<0.001***
Total PHQ-9 Score	–			**1.11**	**(1.08–1.15)**	**<0.001***
Gender (Male)						
Female	**1.47**	**(1.14–1.90)**	**0.003***	**1.45**	**(1.13–1.87)**	**0.004***
Age	**1.02**	**(1.01–1.03)**	**<0.001***	**1.02**	**(1.01–1.03)**	**<0.001***
Marital status (Married)						
Single/Separated/Divorced/Widowed	0.92	(0.73–1.18)	0.521	0.94	(0.74–1.20)	0.634
Employment (No)						
Yes	**0.70**	**(0.53–0.93)**	**0.012***	**0.68**	**(0.52–0.90)**	**0.007***
Disease co-morbidity (None)						
One	0.98	(0.72–1.33)	0.888	0.96	(0.71–1.31)	0.806
Two or more	1.01	(0.73–1.39)	0.962	0.97	(0.71–1.34)	0.858
Family history mental illness (No)						
Yes	**1.69**	**(1.25–2.29)**	**0.001***	**1.71**	**(1.26–2.31)**	**0.001***
Seen a doctor in past 4 weeks (Mean ± S.D.)	**1.10**	**(1.04–1.16)**	**0.001***	**1.10**	**(1.04–1.16)**	**0.001***
Self-reported depression/other mental illness history (No)						
Yes	**10.97**	**(8.65–13.90)**	**<0.001***	**11.19**	**(8.85–14.15)**	**<0.001***
Household income (⩽HK$ 30 000)						
>HK$ 30 000	0.96	(0.74–1.24)	0.751	0.96	(0.74–1.24)	0.750

Bold indicates that the p-value is significant (i.e. p < 0.05).
Hosmer and Lemeshow Tests suggested adequate model fit using PHQ-9 symptom items: χ^2 = 147.8, p value = 1.00; Incorporating PHQ-9 total score: χ^2 = 176.4, p value = 0.98.
[a]Diagnosis made at the time of the index visit and reported by the PCP on case record forms.
[b]'Over the last 2 weeks, how often have you been bothered by'.
*Statistically significant at p value <0.05 by linear mixed effects model.
Note: Doctor Intraclass Correlation Coefficient (ICC) in the model using PHQ-9 symptom items = 18.7%; in the model incorporating PHQ-9 total score = 18.6%.
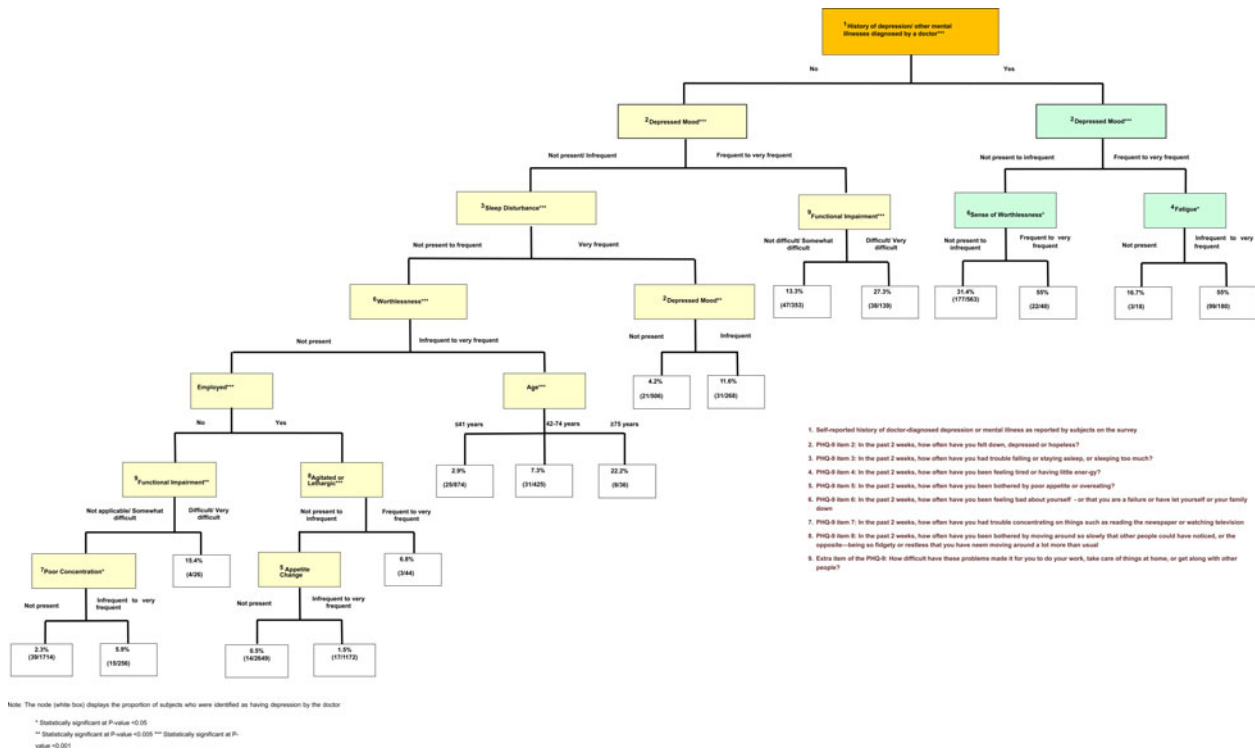
Fig. 1. Decision tree for depression diagnosis by doctors modeling PHQ-9 symptoms.

that these are cases of diagnostic inaccuracy, or alternately, they may be cases of remitted depression.

Our interpretation of the tree analysis models is that PCPs appear to use a hypothetical-deductive problem-solving approach incorporating pre-test probability with different diagnostic pathways for patients with and without past depression. The tree analyses found that in patients with past depression, fewer PHQ-9 items were associated with PCP diagnosis with diagnoses occurring at lower PHQ-9 symptom severities. This seems to indicate

that PCPs have a higher pre-test probability and lower thresholds for diagnosing depression in these patients. This may mean that PCPs are either more willing or confident to diagnose patients with known past depression or that patients with a past history of depression are much more willing to disclose their symptoms to the PCP, making it more likely for a positive diagnosis to occur. One potential risk is that PCPs may prematurely converge on a diagnosis of depression without performing sufficient further assessment and possible misdiagnoses. Future studies to assess
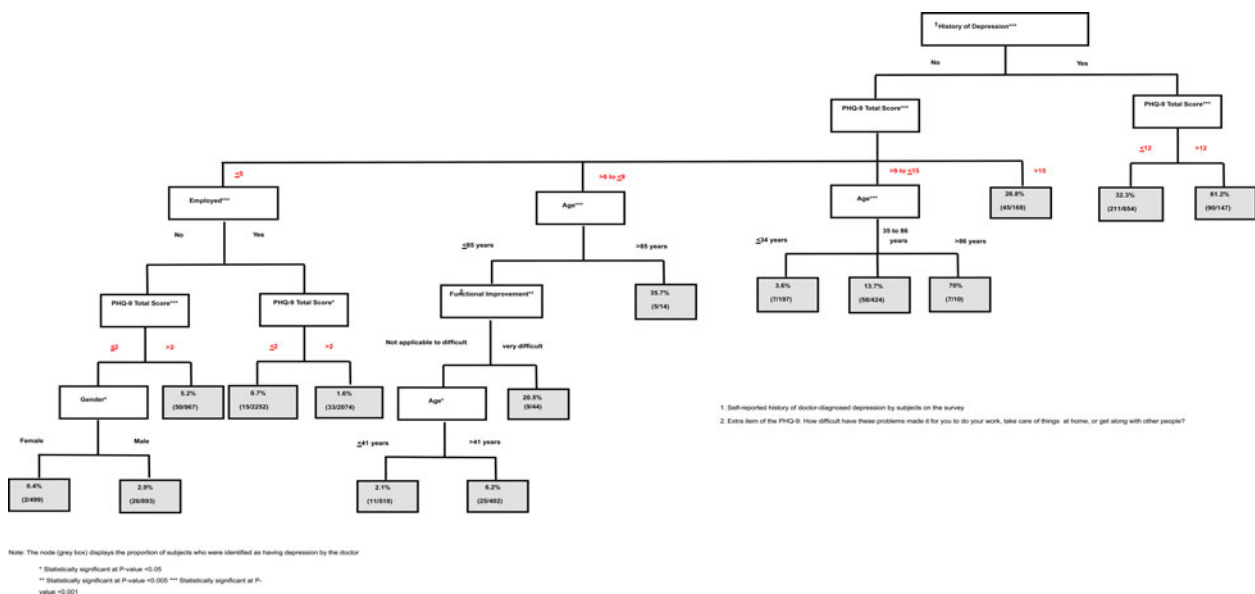

Fig. 2. Decision tree for depression diagnosis by doctors modeling PHQ-9 total score.

diagnostic accuracy using decision tree analyses could help to clarify if this is happening.

In contrast, for patients without a history of depression, the tree models had more branches suggesting that PCPs may take more factors into consideration before making a judgment on the diagnosis. This illustrates how diagnosing new cases of depression is a more complex clinical process. It was observed that in patients with no previous known depression, diagnoses were made at higher symptom severity levels and using more criteria. This seems to suggest that PCPs exercise greater caution when diagnosing new cases. As mental illness remains highly stigmatizing in Asian cultures, one possibility is that PCPs in our setting are wary about excessively burdening patients experiencing distress and may only diagnose depression when they feel active treatment is absolutely necessary. Alternatively, our patients may be reluctant to disclose the extent of their depressive symptoms, requiring the doctor to 'fish' for symptoms before coming to a decision about whether or not the patient has depression. One potential risk for patients with no history of depression is that if diagnostic thresholds are too high, many patients who may benefit from treatment may be missed.

In keeping with previous studies, a range of psychological, somatic, and other symptoms was found to be associated with the diagnosis of depression (Malhi et al. 2014), however, the model found that physical symptoms (sleep, fatigue) and functional impairment occurred higher in the tree. The PHQ-9 symptom model demonstrated a hierarchy of symptoms associated with the diagnosis of depression, with low mood, sense of worthlessness, fatigue, sleep disturbance, and functional impairment branching earlier. Assuming that such symptoms were elicited during the consultation, this suggests that these criteria may be of greater diagnostic value to PCPs. One possibility is that some depressive symptoms such as appetite change or poor concentration may be of limited value because of their relatively low specificity in generalist settings, where clinicians often prefer to rule out physical disease first (Dew et al. 2005). This finding was supported by a British study comparing PCPs to psychiatrists which found that PCPs tend to use a more limited number of constructs to establish psychological diagnoses, more commonly basing their judgment on low mood, sleep disturbance and functional impairment (Armstrong & Earnshaw, 2004). The findings of our model were also supported by earlier studies that found physician sensitivity to depression is higher when the functional impairment is present and patient presentations with sleep disturbance promote depression assessment and diagnosis in primary care (Ani et al. 2008).

Two key diagnostic criteria for depression were notably absent from our PHQ-9 symptom model namely anhedonia and suicidal ideation. Although anhedonia is traditionally considered a cardinal diagnostic feature of depression, it appears that PCPs in our setting either do not elicit this symptom or find it less useful for diagnosing depression. In a fast-paced and workaholic environment such as Hong Kong, loss of motivation or loss of pleasure from activities may well be masked or considered less significant. Doing things even when they are not pleasurable is often considered a positive attribute in Chinese culture (Shek et al. 2003).

Whilst many studies have found high concordance rates between the presence of suicidality and depression recognition, this was not observed in our study sample (Ani et al. 2008; Henriques et al. 2009). One potential reason may be that the spectrum of severity in our primary care sample was relatively mild, and there may have been too few patients with significant suicidal symptoms for the model to identify this criterion as a classifier. Another possibility

is that our study doctors may not be sufficiently sensitive to their patients' suicidal communications or did not encourage them to talk about it (Riihimäki et al. 2014; Younes et al. 2015), raising potential concerns regarding the quality of suicide prevention in this setting. An Australian study found that PCPs are more likely to use suicidality as a measure of depression severity (rather than for diagnosing depression), and may not enquire about suicidal thoughts unless they perceive the patient is severely depressed (Malhi et al. 2014). Thirdly, it is possible that many people in our setting will have suicidal ideas but do not appear depressed. Although suicidal ideation is most often associated with depression, other social circumstances such as marital discord, financial problems or sense of hopelessness may also be associated with suicidal thoughts (Cheung et al. 2006). Nevertheless, what our findings show is that patients with suicidal thoughts was not an associated factor with PCP diagnosis of depression, which may have potential implications on whether the quality of care regarding suicide detection is adequate in our setting.

Limitations

This study has several limitations. Firstly, data on patient characteristics including PHQ-9 symptoms of depression were all self-reported, and are susceptible to self-report bias. It is not known how many symptoms the doctor elicited during the consultation and it is entirely possible that some of the patient's self-reported symptoms did not get conveyed. It must be stressed that the findings of our models only represent the statistical relationship between the presence of symptoms (as reported by the patient) and PCP diagnosis of depression, and these have been used to simulate the diagnostic pathways. A further study where the symptoms elicited by the doctor are collected independently would help to validate our findings.

Secondly, we only examined the factors associated with PCP diagnosis of depression without considering diagnostic accuracy and there was no gold standard assessment to validate the doctor's clinical diagnosis. The scope of this current study was limited to exploring PCP behavior patterns in relation to diagnosing depression in a real-world setting. Further studies where the PCP's performance is benchmarked against a gold standard are needed to compare the differences in pathways leading to accurate and inaccurate recognition to determine whether clinical assessments are adequately performed, particularly in those with past depression.

Third, our models were based on the diagnostic behaviors of 59 PCPs on approximately 10 000 predominantly Chinese primary care patients in Hong Kong and may not be representative of depression diagnosed in other settings. Differences in age, gender, education, culture, and health systems of both patients and doctors will affect diagnostic behavior. Our findings provide valuable information on the direction of effects, the relative importance of classifiers and the utility of the decision tree methodology, however, further studies in other settings and using other instruments are needed to help build further knowledge in this field and confirm, amend or refute our findings. This study needs to be replicated in other settings for cross-cultural comparison and to examine how diagnostic behaviors differ among other types of mental health providers.

Conclusions

We found that decision tree analysis is an informative method to explore the factors associated with a clinical diagnosis of

depression. Our findings provide new insights regarding the associations between PHQ-9 symptoms, symptom severity and depression diagnoses. Data on these associations were used to model how primary care doctors diagnose depression and the potential factors influencing their decisions. Our tree diagrams illustrated how primary care doctors use a hypothetical-deductive problem-solving approach incorporating pre-test probability with different criteria and diagnostic pathways for patients with and without past depression. The models suggested that there is a hierarchical structure to the PHQ-9 criteria, demonstrating that some depressive symptoms may have greater diagnostic value in primary care. We found that many patients experiencing suicidal ideation are not diagnosed as having depression, which raises concerns regarding the quality of suicide detection and prevention in our setting. Our methods could be replicated in other settings or on other data to compare if different types of mental health providers diagnose depression differently to build further knowledge on how depression is diagnosed. Future studies using decision tree analyses to compare the decision-making pathways between accurate and inaccurate diagnoses are needed to inform better quality of care.

References

Akhtar-Danesh N and Landeen J (2007) Relation between depression and sociodemographic factors. *International Journal of Mental Health Systems* **1**, 4.

Ani C, et al. (2008) Depression symptomatology and diagnosis: discordance between patients and physicians in primary care settings. *BMC Family Practice* **9**, 1.

Armstrong D and Earnshaw G (2004) What constructs do GPs use when diagnosing psychological problems? *British Journal of General Practice* **54**, 580–583.

Barbui C and Tansella M (2006) Identification and management of depression in primary care settings. A meta-review of evidence. *Epidemiology and Psychiatric Sciences* **15**, 276–283.

Bekhuis E, et al. (2015) Differential associations of specific depressive and anxiety disorders with somatic symptoms. *Journal of Psychosomatic Research* **78**, 116–122.

Bermejo I, et al. (2005) Patients' and physicians' evaluation of the PHQ-D for depression screening. *Primary Care and Community Psychiatry* **10**, 125.

Carey M, et al. (2014) Accuracy of general practitioner unassisted detection of depression. *Australian & New Zealand Journal of Psychiatry* **48**, 571–578.

Cepoiu M, et al. (2008) Recognition of depression by non-psychiatric physicians—a systematic literature review and meta-analysis. *Journal of General Internal Medicine* **23**, 25–36.

Cheung YB, et al. (2006) Suicidal ideation and suicidal attempts in a population-based study of Chinese people: risk attributable to hopelessness, depression, and social factors. *Journal of Affective Disorders* **90**, 193–199.

Chin W, et al. (2012) The epidemiology and natural history of depressive disorders in Hong Kong's primary care. *BMC Family Practice* **12**, 129.

Chin WY, et al. (2014) Detection and management of depression in adult primary care patients in Hong Kong: a cross-sectional survey conducted by a primary care practice-based research network. *BMC Family Practice* **15**, 30.

Chin WY, et al. (2015a) 12-Month naturalistic outcomes of depressive disorders in Hong Kong's primary care. *Family Practice* **32**, 288–296.

Chin WY, et al. (2015b) Help-seeking intentions and subsequent 12-month mental health service use in Chinese primary care patients with depressive symptoms. *BMJ Open* **5**, e006730.

Chin W-Y, et al. (2016a) The 12-month incidence and predictors of PHQ-9–screened depressive symptoms in Chinese primary care patients. *The Annals of Family Medicine* **14**, 47–53.

Chin WY, Choi EP and Wan EY (2016b) Trajectory pathways for depressive symptoms and their associated factors in a Chinese primary care cohort by growth mixture modelling. *PLoS ONE* **11**, e0147775.

Cornelius B, et al. (2014) Under-recognition and under-treatment of DSM-IV classified mood and anxiety disorders among disability claimants. *Disability and Rehabilitation* **36**, 1161–1168.

Dew K, et al. (2005) "This glorious twilight zone of uncertainty": mental health consultations in general practice in New Zealand. *Social Science & Medicine* **61**, 1189–1200.

Dowrick C, et al. (2011) Estimating probability of sustained recovery from mild to moderate depression in primary care: evidence from the THREAD study. *Psychological Medicine* **41**, 141–150.

Gask L, et al. (2008) Capturing complexity: the case for a new classification system for mental disorders in primary care. *European Psychiatry* **23**, 469–476.

Goldberg DP, et al. (2017) Screening for anxiety, depression, and anxious depression in primary care: a field study for ICD-11 PHC. *Journal of Affective Disorders* **213**, 199–206.

Guo B, et al. (2017) Factor structure and longitudinal measurement invariance of PHQ-9 for specialist mental health care patients with persistent major depressive disorder: exploratory structural equation modelling. *Journal of Affective Disorders* **219**, 1–8.

Henriques SG, et al. (2009) Recognition of depressive symptoms by physicians. *Clinics* **64**, 629–635.

Höfler M and Wittchen HU (2000) Why do primary care doctors diagnose depression when diagnostic criteria are not met? *International Journal of Methods in Psychiatric Research* **9**, 110–120.

Hothorn T, Hornik K and Zeileis A (2006) Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics* **15**, 651–674.

Hothorn T, et al. (2010) Party: a laboratory for recursive partitioning. R package version. 2015:1–0.

Kroenke K, Spitzer RL and Williams JBW (2001) The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* **16**, 606–613.

Kuhn L, et al. (2014) The process and utility of classification and regression tree methodology in nursing research. *Journal of Advanced Nursing* **70**, 1276–1286.

Lam CLK, et al. (2011) Utilization pattern of primary care services in Hong Kong – does having a family doctor make a difference? *Hong Kong Medical Journal* **17**, S28–S32.

Lemon SC, et al. (2003) Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Annals of Behavioral Medicine* **26**, 172–181.

Malhi GS, et al. (2014) Unlocking the diagnosis of depression in primary care: which key symptoms are GPs using to determine diagnosis and severity? *Australian and New Zealand Journal of Psychiatry* **48**, 542–547.

McGrady A, et al. (2010) Coherence between physician diagnosis and patient self reports of anxiety and depression in primary care. *Journal of Nervous and Mental Disease* **198**, 420–424.

Mitchell AJ, Rao S and Vaze A (2011) International comparison of clinicians' ability to identify depression in primary care: meta-analysis and meta-regression of predictors. *British Journal of General Practice* **61**, e72–e80.

Mitchell AJ, Vaze A and Rao S (2009) Clinical diagnosis of depression in primary care: a meta-analysis. *The Lancet* **374**, 609–619.

Paul P, Pennell ML and Lemeshow S (2013) Standardizing the power of the Hosmer–Lemeshow goodness of fit test in large data sets. *Statistics in Medicine* **32**, 67–80.

Podgorelec V, *et al.* (2002) Decision trees: an overview and their use in medicine. *Journal of Medical Systems* **26**, 445–463.

Riihimäki K, *et al.* (2014) Incidence and predictors of suicide attempts among primary-care patients with depressive disorders: a 5-year prospective study. *Psychological Medicine* **44**, 291–302.

Shek DT, *et al.* (2003) The relationship between Chinese cultural beliefs about adversity and psychological adjustment in Chinese families with economic disadvantage. *The American Journal of Family Therapy* **31**, 427–443.

Simon GE, *et al.* (1999) An international study of the relation between somatic symptoms and depression. *New England Journal of Medicine* **341**, 1329–1335.

Spitzer RL, *et al.* (1999) Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* **282**, 1737–1744.

Tylee A and Gandhi P (2005) The importance of somatic symptoms in depression in primary care. *Primary Care Companion to the Journal of Clinical Psychiatry* **7**, 167.

Üstün TB, *et al.* (2004) Global burden of depressive disorders in the year 2000. *The British Journal of Psychiatry* **184**, 386–392.

Younes N, *et al.* (2015) Attempted and completed suicide in primary care: not what we expected? *Journal of Affective Disorders* **170**, 150–154.

Yu X, *et al.* (2012) The patient health questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. *Comprehensive Psychiatry* **53**, 95–102.