Diagnostic Validity of Patient Health Questionnaire-9 and Composite International Diagnostic Interview Assessment Scales for Depression in East Africa

Bizuayehu Gelaye

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Michelle A. Williams, Chair

Jesse Fann

Ann Vander Stoep

Xiao-Hua Andrew Zhou

Program Authorized to Offer Degree:
School of Public Health—Epidemiology

University of Washington

Abstract

Diagnostic Validity of Patient Health Questionnaire-9 and Composite International Diagnostic Interview Assessment Scales for Depression in East Africa

Bizuayehu Gelaye

Chair of the Supervisory Committee

Professor Michelle A. Williams
Department of Epidemiology

Background: Depression is one of the leading causes of disability and largest contributor to the global burden of disease. In sub-Saharan Africa, depression often goes undetected and untreated, in part, due to a lack of validated screening and diagnostic instruments. Therefore, we conducted a validation study of the Patient Health Questionnaire (PHQ-9) and fully structured Composite International Diagnostic Interview (CIDI) in diagnosing current major depressive disorder (MDD) among Ethiopian adults.

Methods: A total of 926 adults attending outpatient departments in a major referral hospital in Addis Ababa, Ethiopia participated in this study. We used a two stage-study design where participants were first interviewed using an Amharic versions of PHQ-9 and the CIDI instruments and then a stratified random sample underwent a follow-up semi-structured clinical interview conducted by a psychiatrist, blinded to the screening results, using Schedules for Clinical Assessment in Neuropsychiatry (SCAN) instrument. We tested construct validity using factor analysis and by examining associations of PHQ-9/CIDI with self-reported quality of life as assessed using the World Health Organization Quality of Life

(WHO-QOL) Questionnaire. We calculated the psychometric properties of the PHQ-9/CIDI using the SCAN diagnostic interview as a gold standard.

Results: Our study provided evidence for unidimensionality of core depression screening questions on both PHQ-9 and CIDI interviews with good factor loadings on a major core depressive factor. The PHQ-9 items showed good internal (Cronbach's alpha=0.85) and test re-test reliability (intraclass correlation coefficient=0.92). Similarly, we found that the CIDI diagnostic interview has good internal reliability (Cronbach's alpha= 0.97) among Ethiopian adults. Quality of life, as reflected in subscale scores for four WHO-QOL domains, was significantly lower among adults classified as depressed on the PHQ-9 and CIDI compared to those classified as not depressed. Receiver Operating Characteristics analysis showed that a PHQ-9 threshold score of 9 offered the optimal discriminatory power with respect to diagnosis of major depressive disorder via the clinical interview (sensitivity= 79% and specificity= 72%). Compared to the SCAN reference standard, the CIDI had fair specificity (72.2%) but low sensitivity (51.0%).

Conclusion: The PHQ-9 is a reliable and valid instrument that may be used as a screen for major depressive disorder among Ethiopian adults in outpatient health care settings. The CIDI had fair specificity and low sensitivity in detecting MDD compared with psychiatrist administered SCAN diagnosis. Our findings are generally consistent with prior studies. Use of fully structured interviews such as the CIDI for MDD diagnosis in clinical settings might lead to under detection of MDD.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Definition
AUC	Area Under the Curve
BC	Bayesian Credible Intervals
CHS	Cardiovascular Health Study
CI	Confidence Intervals
CIDI	Composite International Diagnostic Interview
CRF	Corticotrophin-Releasing Factor
CRP	C-reactive Protein
CVD	Cardiovascular Diseases
DBP	Diastolic Blood Pressure
DDE	Direct Data Entry
DSM-IV	Diagnostic Statistical manual of Mental Disorders, 4th Edition
EFA	Exploratory Factor Analysis
FBG	Fasting Blood Glucose
HDL-C	High-density Lipoprotein Cholesterol
HOMA-IR	Homeostatic Model Assessment of Insulin Resistance
HPA	Hypothalamic–Pituitary–Adrenal
ICC	Intraclass Correlation Coefficient
ICD-10	International Classification of Diseases 10th edition
ICL	International Clinical Laboratories
IDF	International Diabetes Federation
IQR	Interquartile Range
LDL-C	Low-density Lipoprotein Cholesterol (LDL-C)
LR+	Positive Likelihood Ratio
LR-	Negative Likelihood Ratio
MAR	Missing at Random

Abbreviation	Definition
MCAR	Missing Completely at Random
MDD	Major Depressive Disorders
MESA	Multi-Ethnic Study of Atherosclerosis
METS	Metabolic Syndrome
MINI	Mini International Neuropsychiatric Interview
NPV	Negative Predictive Value
NHANES	National Health and Nutrition Examination Survey
NIHM	National Institute of Mental Health
OR	Odds Ratios
PAPI	Paper and Pencil Interview
PHQ-9	Patient Health Questionnaire 9
PPS	Probability Proportional to Size
PPV	Positive Predictive Value
ROC	Receiver Operating Characteristics
SCID	Structured Clinical Interview for DSM-IV
SCAN	Schedule for Clinical Assessment in Neuropsychiatry
SBP	Systolic Blood Pressure
SD	Standard Deviation
SE	Standard Error
TC	Total Cholesterol
TG	Triglycerides
WHO	World Health Organization
WHO-QOL	World Health Organization Quality of Life
WHOQOL-BREF	World Health Organization Quality of Life Abbreviated Version
WMH	World Mental Health

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

I owe a debt of gratitude for my mentor Dr. Michelle A. Williams—for seeing the Epidemiologist in an Engineer, for all her support, inspiration, her unwavering commitment to excellence and encouraging me to be the best that I can be over the years.

I am thankful for my committee members– Drs. Jesse Fann, Ann Vander Stoep, and Andrew Zhou– for their important contributions to this project and their support.

Many thanks to my Graduate School Representative, Dr. Karen Thomas, for her leadership, wisdom, and constant support over the years.

I am thankful to the research staff members at Addis Continental Institute of Public Health and Saint Paul Hospital for their support to the project. Special thanks to Drs. Yemane Berhane, Yonas Bahretibeb, Negussie Deyessa, Teshome Shibre, Dawit Wondimagegn, Asnake Limenhe and Ms. Seblewengel Lemma.

This dissertation and my graduate school experience would never have been possible without the support and encouragement of my family and friends. And most of all, I deeply thank my loving wife Tersu (Hana) who encouraged and supported me every step of the way; Ruth and Nehemiah whose love, innocence and laughter makes everything possible; and my mom for being a constant source of inspiration in my life.

DEDICATION

To my uncle Gashe who did not live to see this day.

CHAPTER 1

Overview of Dissertation Research

Overview of Dissertation Research

Globally, mental health problems account for 13% of the total burden of disease and 31% of all years lived with disability [1]. By 2030, depression alone is likely to be the single highest contributor to burden of disease in the world –higher than heart disease, stroke, road traffic accidents, and HIV/AIDS [1]. Studies conducted in sub-Saharan Africa among primary health clinic attendees show that 20–30% of them present with depressive symptoms and other psychological disorders as the first or secondary reason for seeking medical care [2]. Depression is often underdiagnosed and undertreated in primary care settings, particularly in developing countries, despite its economic impact on families and communities, and its contribution to long-term disability [3-5]. This is, in part, due to challenges resulting from a lack of skilled mental health workers, stigma associated with mental illness, lack of cross-culturally validated screening instruments, and prominence of somatic presentations of mental disorders [5-7]. Effective treatment and prevention of depression requires accurate detection and diagnosis [8], which in turn involves the systematic use of valid measurement approaches. Below, we provide a rationale for evaluating the diagnostic validity of two standardized depression screening instruments that have been used extensively worldwide among adults in Ethiopia.

Patient Health Questionnaire (PHQ-9)

The PHQ-9, an instrument designed to screen for major depressive disorder, consists of one question for each of the nine criterion based up on the DSM-IV major depressive disorders [9, 10]. The questions refer to situations in the previous two weeks. Because of its brevity the PHQ-

9 has gained increased recognition as the preferred instrument for depression screenings in primary care settings among racially and ethnically diverse populations [11]. Additionally, it has been reported as a valuable tool for the diagnosis and management of depression as it can generate a diagnosis of major depression, as well a continuous score to monitor treatment [9]. Importantly, the PHQ-9 can be self-administered or interviewer-administered, and it has demonstrated validity in many countries as an instrument capable of measure probable major depressive disorders (MDD) [9]. To date, only three, investigative teams have published studies documenting the psychometric properties of the PHQ-9 when used in sub-Saharan African populations [11-14]; and none of these included Ethiopians. Of the three, only one study translated the PHQ-9 instrument into local dialect (Swahili) [9]. Despite differences in population characteristics, sample size, and study settings, the previous studies consistently demonstrate the good reliability and potential benefit of the PHQ-9 in sub-Saharan Africa. The current study will add to the sparse research findings specific to sub-Saharan Africa.

Composite International Diagnostic Interview (CIDI)

The CIDI was developed by the World Health Organization as a research diagnostic interview and to obtain valid information about the prevalence and correlates of mental disorders [10-12]. The first version of CIDI did not adequately and consistently measure risk factors, consequences, patterns and correlates of mental health problems [17]. It also had several methodological limitations including complexity of questions and instructions, which limited its application in cross cultural settings [17]. Cognizant of these limitations the WHO charged a multinational CIDI Editorial Committee to test the instrument in many different countries and

established the WHO World Mental Health (WMH) Survey Initiative in 1998. Since then, CIDI has gone through a number of revisions to align diagnostic algorithms with the DSM-IV and International Classification of Diseases 10th edition (ICD-10) coding systems. Although clinical appraisal studies have been conducted to assess the validity of CIDI in different countries [14], to date, there is no published literature reporting the validity of CIDI (version 3.0) when used in sub-Saharan African populations. We sought to address this gap in the literature by evaluating the validity and reliability of CIDI version 3.0 as a depression screening instrument among urban dwelling Ethiopians.

Translating Measures in Non-Western Setting

Although both PHQ-9 and CIDI, originally developed in Western countries, are extensively used around the world, concerns have been raised that standard translation and back-translation procedures may be inadequate to produce a culturally valid version of these instruments in non-Western countries [13, 14]. Indeed, some have argued that depression in Africa presents in somatic forms, culture specific idioms, interpersonal relationships or in spiritual phenomena that may obscure detection when standard instruments are used [15-18]. Somatization, the unconscious expression of psychological conflict as physical symptoms with no identifiable physical origin, is not unique to sub-Saharan Africa [19]. Investigators have made similar observations in Asia [8, 20, 21] and in Latin America [22, 23]. Similarly somatic symptoms were reportedly higher among African Americans in the United States compared to Whites [24]. However there is an accumulating body of objective evidence challenging the view of a cultural preponderance of somatization [25]. Investigators have reported that the measurement of

depression in international contexts may be undertaken with standard instruments in different cultural settings, provided care is taken to ensure an adequate translation and validation of the cutoff score [25]. Importantly, mental health researchers advocate the use of standardized instruments when there is evidence that major depressive disorder manifests in forms similar to those cultures for which there are standardized instruments [15].

To the best of our knowledge, only one study, thus far, has systematically evaluated how culture and somatic manifestations of depression affect standard diagnosis using PHQ-9 in sub-Saharan African context [26]. Okulate *et al* in their study among Nigerian military personnel evaluated somatic symptoms with diagnosis of depression using PHQ-9 applying a principal component analysis [26]. The authors noted that somatic symptoms were ubiquitous among their study subjects but had considerably less weight in the diagnosis of depression. The core depressive symptoms in PHQ-9 were better than somatic symptoms as predictors of depression [26]. To ensure proper expression and conceptualization of terminologies in local contexts, we used a standard approach of iterative back translation by panels of bilingual experts (please see Appendix 1).

Schedules for Clinical Assessment in Neuropsychiatry (SCAN) Diagnosis

The SCAN is a semi structured clinical interview developed by the WHO for use by trained clinicians to assess and diagnose psychiatric disorders among adults [27, 28]. SCAN was developed within the framework of the WHO and the National Institute of Mental Health (NIMH) Joint Project on Diagnosis and Classification of Mental Disorders, Alcohol and Related

Problems [27]. The use of SCAN gives flexibility in the diagnosis of mental disorders based on ICD-10 and DSM-IV [27]. The depression module of SCAN has been reported to work well in different languages and cultures including Ethiopia [29]. Alem *et al* conducted a clinical reappraisal of SCAN against a clinical diagnosis by a psychiatrist and found it a feasible and reliable tool in Ethiopia (percent agreement 93.0 with *kappa*=0.80)[29]. In the current study we used SCAN as a reference (or gold) standard.

Depression and Cardiovascular Disease Risk

The global prevalence of chronic non communicable diseases, particularly cardiovascular disease (CVD), is increasing at an alarming rate, with the majority of cases occurring in developing countries [1]. The rise in CVD is driven in part by significant changes in dietary habits, physical activity levels, and increased stress as a result of increase urbanization and economic development [30]. An expanding body of evidence now implicates depression as an important and prevalent co-morbid condition and risk factor for CVD [31-33]. Although causal associations between psychiatric disorders and CVD remain to be more firmly established, the influences of psychiatric disorders on CVD health promotion, disease control and prevention are far-reaching [31, 32, 34-36]. For example, investigators have reported that individuals with comorbid diabetes and depression are more likely than individuals with diabetes alone to have poor glycemic control and consequently to have more severe complications and lower quality of life [31, 32]. Other investigations have shown that depression is predictive of incident myocardial infarction and hypertension [22, 35].

Given (1) emerging evidence of chronic disease burden, particularly CVD, in low-income countries; (2) emerging evidence documenting the influence of psychiatric disorders on incident and prevalent CVD; (3) limited access to validated mental health screening instruments designed for use in Ethiopia; and (4) the scarcity of information concerning the prevalence and severity of depression and depressive symptoms among Ethiopians, we conducted this study to assess the psychometric properties of two widely used instruments among urban dwelling Ethiopians. The results of these analyses are presented in Chapters 2 and 3. As the gold standard might not always be available, we evaluated the psychometric properties of the PHQ-9 against a lay-interviewer administered CIDI reference standard. We also evaluated the extent to which use of suboptimal gold standard biases the estimates and demonstrated how statistical methods can be used to correct the bias and improve psychometric properties of the PHQ-9. The results of this study are presented in Chapter 4. Finally, we evaluated the extent to which, if at all, depression diagnosis is associated with CVD risk among a well characterized occupational cohort of bankers and teachers residing in Addis Ababa. The results of this study are provided in Chapter 5. Finally, in Chapter 6 we summarize the findings of the studies and discuss possible future directions.

REFERENCES

1. WHO: **The Global Burden of Disease 2004 Update.** . In. Geneva, Switzerland: World Health Organization; 2004.

2. Ngoma MC, Prince M, Mann A: **Common mental disorders among those attending primary health clinics and traditional healers in urban Tanzania**. *Br J Psychiatry* 2003, **183**:349-355.

3. Patel V, Garrison P, de Jesus Mari J, Minas H, Prince M, Saxena S: **The Lancet's series on global mental health: 1 year on**. *Lancet* 2008, **372**(9646):1354-1357.

4. Patel V, Prince M: **Global mental health: a new global health field comes of age**. *JAMA* 2010, **303**(19):1976-1977.

5. Wig NN: **WHO and mental health--a view from developing countries**. *Bull World Health Organ* 2000, **78**(4):502-503.

6. Patel V, Araya R, Chowdhary N, King M, Kirkwood B, Nayak S, Simon G, Weiss HA: **Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires**. *Psychol Med* 2008, **38**(2):221-228.

7. Patel V, Sartorius N: **From science to action: the Lancet Series on Global Mental Health**. *Curr Opin Psychiatry* 2008, **21**(2):109-113.

8. Liu SI, Yeh ZT, Huang HC, Sun FJ, Tjung JJ, Hwang LC, Shih YH, Yeh AW: **Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan**. *Compr Psychiatry* 2011, **52**(1):96-101.

9. Omoro SA, Fann JR, Weymuller EA, Macharia IM, Yueh B: **Swahili translation and validation of the Patient Health Questionnaire-9 depression scale in the Kenyan head and neck cancer patient population**. *Int J Psychiatry Med* 2006, **36**(3):367-381.

10. Kessler RC, Abelson J, Demler O, Escobar JI, Gibbon M, Guyer ME, Howes MJ, Jin R, Vega WA, Walters EE *et al*: **Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI)**. *Int J Methods Psychiatr Res* 2004, **13**(2):122-139.

11. Kessler RC, Angermeyer M, Anthony JC, R DEG, Demyttenaere K, Gasquet I, G DEG, Gluzman S, Gureje O, Haro JM *et al*: **Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative**. *World Psychiatry* 2007, **6**(3):168-176.

12. Kessler RC, Ustun TB: **The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI)**. *Int J Methods Psychiatr Res* 2004, **13**(2):93-121.

13. Kinder LS, Carnethon MR, Palaniappan LP, King AC, Fortmann SP: **Depression and the metabolic syndrome in young adults: findings from the Third National Health and Nutrition Examination Survey**. *Psychosom Med* 2004, **66**(3):316-322.

14. Kessler RC, Ustün TB, World Health O: **The WHO World Mental Health Surveys : global perspectives on the epidemiology of mental disorders**. Cambridge; New York; Geneva: Cambridge University Press ; Published in collaboration with the World Health Organization; 2008.

15. Bass JK, Bolton PA, Murray LK: **Do not forget culture when studying mental health**. *Lancet* 2007, **370**(9591):918-919.

16. Tomlinson M, Swartz L, Kruger LM, Gureje O: **Manifestations of affective disturbance in sub-Saharan Africa: key themes**. *J Affect Disord* 2007, **102**(1-3):191-198.

17. Halbreich U, Alarcon RD, Calil H, Douki S, Gaszner P, Jadresic E, Jasovic-Gasic M, Kadri N, Kerr-Correa F, Patel V *et al*: **Culturally-sensitive complaints of depressions and anxieties in women**. *J Affect Disord* 2007, **102**(1-3):159-176.

18. Johnson LR, Mayanja MK, Bangirana P, Kizito S: **Contrasting concepts of depression in Uganda: implications for service delivery in a multicultural context**. *Am J Orthopsychiatry* 2009, **79**(2):275-289.

19. Kleinman A: **Rethinking psychiatry : from cultural category to personal experience**. New York; London: Free Press ; Collier Macmillan; 1991.

20. Lotrakul M, Sumrithe S, Saipanish R: **Reliability and validity of the Thai version of the PHQ-9**. *BMC Psychiatry* 2008, **8**:46.

21. Dinh TQ, Yamada AM, Yee BW: **A culturally relevant conceptualization of depression: an empirical examination of the factorial structure of the Vietnamese Depression Scale**. *Int J Soc Psychiatry* 2009, **55**(6):496-505.

22. Gromova HA, Gafarov VV, Gagulin IV: **Depression and risk of cardiovascular diseases among males aged 25-64 (WHO MONICA--psychosocial)**. *Alaska Med* 2007, **49**(2 Suppl):255-258.

23. Slone LB, Norris FH, Murphy AD, Baker CK, Perilla JL, Diaz D, Rodriguez FG, Gutierrez Rodriguez Jde J: **Epidemiology of major depression in four cities in Mexico**. *Depress Anxiety* 2006, **23**(3):158-167.

24. Brown C, Schulberg HC, Madonia MJ: **Clinical presentations of major depression by African Americans and whites in primary medical care practice**. *J Affect Disord* 1996, **41**(3):181-191.

25. Patel V: **Cultural factors and international epidemiology**. *Br Med Bull* 2001, **57**:33-45.

26. Okulate GT, Olayinka MO, Jones OB: **Somatic symptoms in depression: evaluation of their diagnostic weight in an African setting**. *Br J Psychiatry* 2004, **184**:422-427.

27. Aboraya A, Tien A, Stevenson J, Crosby K: **Schedules for Clinical Assessment in Neuropsychiatry (SCAN): introduction to WV's mental health community**. *W V Med J* 1998, **94**(6):326-328.

28. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N: **SCAN. Schedules for Clinical Assessment in Neuropsychiatry**. *Arch Gen Psychiatry* 1990, **47**(6):589-593.

29. Alem A, Kebede D, Shibre T, Negash A, Deyassa N: **Comparison of computer assisted scan diagnoses and clinical diagnoses of major mental disorders in Butajira, rural Ethiopia**. *Ethiop Med J* 2004, **42**(2):137-143.

30. Vorster H: **The emergence of cardiovascular disease during urbanization of Africans**. *Public Health Nutr* 2002, **5**(1A):239-243.

31. Lloyd CE, Dyer PH, Barnett AH: **Prevalence of symptoms of depression and anxiety in a diabetes clinic population**. *Diabet Med* 2000, **17**(3):198-202.

32. Lustman PJ, Clouse RE: **Depression in diabetic patients: the relationship between mood and glycemic control**. *J Diabetes Complications* 2005, **19**(2):113-122.

33. Moller-Leimkuhler AM: **Higher comorbidity of depression and cardiovascular disease in women: a biopsychosocial perspective**. *World J Biol Psychiatry*, **11**(8):922-933.

34. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ: **The prevalence of comorbid depression in adults with diabetes: a meta-analysis**. *Diabetes Care* 2001, **24**(6):1069-1078.

35. Jonas BS, Franks P, Ingram DD: **Are symptoms of anxiety and depression risk factors for hypertension? Longitudinal evidence from the National Health and Nutrition Examination Survey I Epidemiologic Follow-up Study**. *Arch Fam Med* 1997, **6**(1):43-49.

36. Sher Y, Lolak S, Maldonado JR: **The impact of depression in heart disease**. *Curr Psychiatry Rep* 2010, **12**(3):255-264.

Chapter 2

Validity of the Patient Health Questionnaire-9 (PHQ-9) for Depression Screening in an East African Population

ABSTRACT

Background: Depression is one of the leading causes of disability and the largest contributor to the global burden of disease. In sub-Saharan Africa, depression often goes undetected and untreated, in part, due to a lack of validated screening and diagnostic instruments. The Patient Health Questionnaire-9 (PHQ-9) depression scale is a brief questionnaire that may be easily used by researchers and clinicians in resource limited settings to detect depression.

Objective: To evaluate the reliability and validity of the PHQ-9 depression questionnaire as a screen for detecting major depressive disorder among Ethiopian adults.

Methods: A total of 926 adults attending outpatient departments in a major referral hospital in Addis Ababa, Ethiopia participated in this study. We tested construct validity using factor analysis and by examining associations of PHQ-9 scores with self-reported quality of life as assessed using the World Health Organization Quality of Life (WHO-QOL) Questionnaire. We assessed criterion validity and performance characteristics against an independent, blinded, and psychiatrist administered semi-structured Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview (i.e., the diagnostic gold standard) using measures of sensitivity, specificity and receiver operating characteristics (ROC) curves. Finally, we used the Rasch analysis to examine the structure of the PHQ-9 questionnaire.

Results: Using the diagnostic gold standard, the prevalence of major depressive disorder (MDD) was 12.6% (95% CI 9.2-16.1%). The PHQ-9 items showed good internal (Cronbach's alpha=0.85) and test re-test reliability (intraclass correlation coefficient=0.92). A principal factor analysis confirmed a 1-factor structure (depression), which accounted for a total of 69.8% of the variance. Quality of life, as reflected in subscale scores for four WHO-QOL domains, was

significantly lower among adults classified as depressed on the PHQ-9 compared to those classified as not depressed. ROC analysis showed that a PHQ-9 threshold score of nine or higher offered optimal discriminatory power with respect to diagnosis of MDD via the clinical interview (sensitivity = 79% and specificity = 72%). Results of our Rasch analysis of the PHQ-9 showed that the data fit to the stringent Rasch model. In addition, the item and separation indices of the PHQ-9 were within acceptable range.

Conclusion: The PHQ-9 appears to be a reliable and valid instrument that may be used to detect major depressive disorders among Ethiopian adults in outpatient health care settings.

INTRODUCTION

Globally, mental health problems account for 13% of the total burden of disease and 31% of all years lived with disability [1]. Mental health problems in developing countries are largely overlooked despite their high prevalence, their economic impact on families and communities, and their contribution to long-term disability [2-4]. According to World Health Organization (WHO) estimates, major depressive disorder (MDD) is projected to become the leading cause of disability and the second leading contributor to the global burden of disease by 2020 [1]. MDD is associated with decreased quality of life, functional decline, health care utilization, and increased mortality [5-7]. However, its diagnosis and treatment remains low in developing countries due, in part, to the lack of skilled mental health workers, stigma associated with mental illness, lack of cross-culturally validated screening and diagnostic instruments, and the prominence of somatic presentations of mental disorders [4, 8, 9].

Studies conducted in sub-Saharan Africa among primary health clinic patients show that 20–30% of such patients present with MDD and other psychiatric disorders as the primary or secondary reason for seeking medical care [10]. Effective treatment of MDD requires accurate detection and diagnosis [11], which requires access to and systematic use of valid screening and diagnostic instruments.

A number of questionnaires are used to screen for MDD in both primary care and general population studies in North American and European settings [12-15]. The Patient Health Questionnaire-9 (PHQ-9), a brief 9-item questionnaire designed to detect MDD according to the

criteria from the Diagnostic and Statistical Manual for Mental Disorders (DSM-IV)[16], has been widely used in clinical and population-based studies across the globe as a screening instrument [14, 17]. Additionally, the PHQ-9 has been endorsed as a valuable tool for the screening and management of MDD [14]. The instrument can be used to detect MDD (dichotomous), as well as a continuous symptom score that may be used to monitor depression severity and patients' response to treatment [14]. Importantly, the PHQ-9 is a brief, freely available, easy to understand, simple to score questionnaire that can be useful in resource limited settings of sub-Saharan African countries where administering comprehensive structured or semi-structured screening instruments can be a challenge.

To date, only three investigative teams have published studies documenting the psychometric properties of the PHQ-9 used in sub-Saharan African populations [18-20]; and none of these studies were conducted in Ethiopia. Ethiopia, with a population of 84 million, is the second most populous country in sub-Saharan Africa. The country is undergoing social and economic changes that are predicted to have detrimental population level health effects including increased burden of chronic diseases, particularly among urban dwelling individuals. Although Ethiopia is one of the world's oldest civilizations with rich natural and mineral resources, it is also one of the poorest countries and has one of the fastest rates of urban growth in the world [21]. Ethiopia also represents one of the leading countries of origin for African-born immigrants in the United States [22]. Therefore, we conducted this study to evaluate the reliability and validity of an Amharic language version of the PHQ-9 against a psychiatrist-administered Schedules for Clinical Assessment in Neuropsychiatry (SCAN) [23] reference/gold standard

among urban dwelling Ethiopian adults. Specifically, we sought to evaluate the construct and criterion validity of the Amharic version of PHQ-9 for detecting an MDD diagnosis against a SCAN reference standard administered by a psychiatrist also in Amharic. We further sought to assess the optimal cut-off points for discrimination between participants with and without MDD using the PHQ-9 instrument. Additionally, we evaluated the structure of the PHQ-9 questionnaire using a contemporary item response theory, the Rasch Analysis. Finally, we conducted a sub-group analysis to examine the extent to which use of the PHQ-9 questionnaire promotes under or over detection of MDD according to participants' sociodemographic characteristics

METHODS AND MATERIALS

Study Participants

A total of 926 adults (≥ 18 years of age) attending outpatient departments in Saint Paul General Specialized Hospital in Addis Ababa, Ethiopia were invited to participate in the study. Saint Paul Hospital is a referral and teaching hospital under the Ethiopian Federal Ministry of Health. The hospital was established to serve the economically underprivileged population, providing services free of charge to about 75% of its patients. The hospital provides all basic services across different departments including: pediatrics, medical, gynecology, neurology, general surgery, psychiatry, ophthalmology, and emergency medicine. For the purposes of this study only medical, general surgery and gynecological outpatient departments were included.

Data collection was conducted between July and December, 2011. Study personnel included research nurses with public health training. Prior to the start of the study, research nurses were trained for four days on the contents of the questionnaire, ethical conduct of human subjects research, and data collection techniques. All study participants provided informed consent and all research protocols were approved by the Institutional Review Boards of Addis Continental Institute of Public Health, Addis Ababa, Ethiopia and the Human Subjects Division at the University of Washington, USA.

Study Procedures

We used a two-stage study design where participants were first interviewed by research nurses using the PHQ-9 questionnaire. On average it took five minutes to administer and score the PHQ-9 questionnaire. Then, on the same day, participants were interviewed using the SCAN diagnostic questionnaire by licensed psychiatrists who were blind to PHQ-9 results. On average it took 20 minutes to administer the SCAN interview. During the PHQ-9 interview, we collected other general information including demographic characteristics, behavioral risk factors, and self-reported health status[24]. Additionally, we collected information concerning self-reported quality of life using the WHO Quality of Life (WHO-QOL) questionnaire.

Screening Test

The PHQ-9 is a 9-item depression screening and diagnostic questionnaire for MDD based on DSM-IV criteria [14, 25, 26]. The PHQ-9, originally written in English, was translated into Amharic (the official language of Ethiopia). To ensure proper expression and conceptualization

of terminologies in local contexts, we used a standard approach of iterative back translation by panels of bilingual experts [27, 28]. The translated version was back-translated and modified until the back-translated version was comparable with the original English version (please see Appendix 1). Each question requires participants to rate the frequency of a depressive symptom experienced in the two weeks prior to evaluation. These items include: 1) anhedonia, 2) depressed mood, 3) insomnia or hypersomnia, 4) fatigue or loss of energy, 5) appetite disturbances, 6) guilt or worthlessness, 7) diminished ability to think or concentrate, 8) psychomotor agitation or retardation, and 9) suicidal thoughts. Scores for each item range from 0 ("not at all") to 3 "nearly every day" with a total score ranging from 0 to 27 [14, 25, 26]. The PHQ-9 contains one additional item (item 10) which assesses functional impairment, also based on a three point scale (not difficult at all, somewhat difficult, very difficult and extremely difficult [14, 25, 26]. We examined several methods of depression screening using the PHQ-9. First we examined the validity based on the total score of the nine items. Similar to other prior studies [29], we also examined the validity using a categorical algorithm. A positive outcome on categorical algorithm requires that at least 5 symptoms are present "more than half the days" (suicidal thoughts count if present at all), with at least one of these symptoms being depressed mood or anhedonia. We also examined the validity of lowering the symptom frequency threshold in the above scheme to "several days."

Diagnostic Interview

The SCAN is a semi-structured clinical interview developed by the WHO for use by trained clinicians to diagnose psychiatric disorders among adults [30, 31]. The SCAN diagnostic

interview, comprised of 28 modules, gives flexibility for diagnosing a number of mental disorders based on DSM-IV diagnostic criteria [30]. In this study, we used the instrument's depression module. The depression module has been reported to have good psychometric properties in diverse populations and in multiple languages [32-34]. Notably, the Amharic version of the SCAN depression module was shown to feasible and reliable tool in Ethiopia (percent agreement 93.0 with kappa=0.80) among Ethiopians [34]. On the basis of earlier reports, we elected to use the psychiatrist administered interview (using SCAN) as the reference (or gold) standard against which we would test the reliability and validity of the PHQ-9 as a screening questionnaire for major depressive disorder among Ethiopians.

Two-stage Sample Selection

As noted earlier, we used a two-stage design in which all participants who screened positive for depression on the PHQ-9 questionnaire (i.e., positive on PHQ-9), as well as a randomly selected sub-group of participants who screened negative for depression (i.e., screen negative on PHQ-9), were referred for clinical diagnostic interview. The SCAN diagnostic interview was conducted, on the same day, by a psychiatrist who was blinded to the PHQ-9 questionnaire outcome. For the purpose of this selection, based on prior literature [14], we defined positive screening for depression as having a score of 10 or higher on the PHQ-9. Following the PHQ-9 interview, a total of 384 participants were invited to participate in the SCAN diagnostic interview (178 who screened positive on the PHQ-9 and 276 who screened negative on the PHQ-9) and 363 of them agreed to do so (94% of selected positive screens and 95% of selected

negative screens). Those who refused to participate in the diagnostic interview cited reasons such as lack of time to do a follow-up interview with the psychiatrist.

Statistical Analyses

First, we assessed data collection instruments manually for quality and completeness using range, plausibility, and cross-validation checks confirming all data were logical. We used Epi-Info software (Version 3.3.2, public access software made available from the U.S. Centers for Disease Control and Prevention) for data entry. We performed double data entry for a sample of completed questionnaires. After data checking and cleaning, entered data were transferred to Stata 11.0 software (Statacorp, College Station, TX) for statistical analyses.

Reliability

We assessed reliability using a number of agreement and consistency indices. Specifically, Cronbach's alpha was computed to assess the internal consistency of items in the PHQ-9 scale. We also computed interclass correlation coefficients to assess the test-retest reliability of PHQ-9 total scores among 5% of the study participants who completed two questionnaires on the same day.

Construct Validity

We evaluated the construct validity, how the PHQ-9 instrument measures the underlying construct (depression), [35] using two approaches. First, we completed a factor analysis to assess whether a single-factor model could be generated among the nine items of PHQ-9 as

reported by others [36, 37]. Prior to performing factor analysis, we assessed the suitability of the data for performing factor analysis. This analysis showed that it is appropriate to proceed with factor analysis (Bartlett's test of sphericity ($p<0.001$), and the Kaiser-Meyer-Olkin measure of sampling adequacy=0.818). We then used the scree plot, presenting the eigenvalues associated with each factor (or the amount of variance accounted for by each factor), to identify the number of meaningful factors (shown in Figure 2.5). Factors with relatively large eigenvalues (eigenvalues greater than one) were assumed to be meaningful and were retained for rotation. Factors with small eigenvalues were not retained. Some investigators have noted that the PHQ-9 items may be inter-correlated with each other; hence we applied the promax oblique rotation procedure to estimate factor correlations.

Second, we used the WHO-QOL questionnaire to evaluate the construct validity of the PHQ-9 questionnaire. The WHO-QOL is a cross-cultural assessment tool that captures an individual's perception of their position in life in the context of culture and value systems in which they live and in relation to their goals, expectations, standards and concerns [38]. The instrument has been widely used globally including in sub-Saharan Africa [38]. In this study we used the abbreviated version of WHO-QOL (also known as WHOQOL-BREF) which has 26 items that cover four domains: physical health, psychological health, social relationships, and environment. The overall percentile score for each domain ranges from 0% (very poor) to 100% (very good). Since prior research has shown statistically significant associations between depression and quality of life [20, 39], we hypothesized that lower WHO-QOL scores would be associated with higher

PHQ-9 scores. We used Student's t-test to compare mean WHO-QOL scores for those classified as depressed and not depressed according to the PHQ-9 questionnaire.

Criterion Validity

We assessed the criterion validity by determining the concordance between the PHQ-9 score and a psychiatrist diagnosis of MDD using the SCAN. We computed the following parameters: sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, positive predictive values, and negative predictive values for the presence or absence of MDD [40]. Additionally, to identify the best PHQ-9 cut-off score to use in depression screening of Amharic-speaking adults, we completed receiver operating characteristic (ROC) curve analyses to identify optimal balance of sensitivity and specificity, area under the ROC curve (AUC) and its nonparametric 95% CI [40]. Additionally we calculated the Youden Index as an additional metric for cut off decision [41]. The Youden index is a function of sensitivity and specificity calculated as the (sensitivity + specificity – 1). The range of the index is 0 to 100 when converted to percentages. Although there are not established values of Youden Index, values above 50% are generally considered acceptable values of diagnostic accuracy [42].

Because of cost and time restrictions, we employed a two-stage design where a subset of participants screened with the PHQ-9 was assessed by a licensed psychiatrist using the SCAN depression module. Given the likelihood of referral or verification bias (i.e., participants with positive test results were preferentially referred for the gold standard procedures), we implemented two analytical strategies to assess and correct for bias introduced as a result of

our study design. First, we computed the psychometric properties of the PHQ-9 using missing completely at random (MCAR) assumptions. For this analysis, all participants with unknown true depression status (i.e., those not selected for a psychiatrist administered diagnostic interview using the SCAN depression module) were excluded from analyses. We presented the results from this analysis, often referred to as naive estimators [40], in **Table 2.6a**. Second, we evaluated the Begg and Greenes adjusted estimates of psychometric properties. Estimates generated from this analysis are corrected for verification bias using a Bayes Theorem approach [40, 43, 44] which applies a missing at random (MAR) assumption. Because it has been shown that the asymptotic confidence intervals might be unreliable in validation studies, we calculated the bootstrapped confidence intervals to render them closer to their expected values [45]. The results of this analysis are presented in **Tables 2.6b** and **2.6c**.

Rating Scale Analysis

We performed a Rasch analysis to evaluate the extent to which items from the Amharic version of the PHQ-9 are reliable and valid in detecting major depressive disorder. We used mean square infit and outfit values to determine how well individual PHQ-9 items fit the Rasch model. [46, 47]. The infit statistic is a weighted mean square residual value that is more sensitive to the unexpected response of an individual's ability level. The outfit statistic is the usual unweighted mean square residual and is more sensitive to unexpected observations or outliers. High infit and outfit values reflect under fit or lack of predictability of an item. As a general rule it has been suggested that mean square infit and outfit values between 0.5-1.5 are acceptable fit to the model [48]. Fit statistics higher than 1.5 and below 0.5 indicate too much and too little

variation in response patterns and should be considered for removal from the instrument to improve fit. Therefore, in our study, we used mean square infit and outfit criteria of between 0.7 and 1.40 (more stringent criterion) to test model misfit [46, 47]. In addition to the (mis)fit of the data and the model, we also evaluated the person and item or Wright map to evaluate the hierarchy of the item difficulties. The Rasch analysis was carried out using the Winsteps Software (version 3.73, Chicago, Illinois) [48].

Validity of PHQ-9 in Subgroups

Finally, we conducted a sub-group analysis in which we used logistic regression procedures to examine the extent to which use of the PHQ-9 questionnaire promotes under or over detection of major depressive disorder *vis a vis* the diagnostic gold standard according to participants' sociodemographic characteristics (e.g., sex, age, and educational attainment). We first compared the demographic characteristics of those misidentified as depressed according PHQ-9 (false positives) relative to those who were correctly classified as not depressed using the gold standard (true negative). We then compared the demographic characteristics of depressed patients misidentified as not depressed by PHQ-9 (false negatives) relative to those who were correctly classified as depressed by the gold standard (true positives).

RESULTS

Participant Characteristics

A summary of selected socio-demographic and lifestyle characteristics of study participants is presented in **Table 2.1**. A total of 926 participants between the ages of 18 and 69 years (mean

age=35 years, standard deviation=11 years) participated in the study. A majority of participants were women (61%), married (52.3%) and self-identified as Orthodox Christians (74%). Approximately 4% of participants reported that they are current cigarette smokers and 9.6% of participants reported consuming at least 1 alcoholic beverage per week. Khat consumption (a green plant with amphetamine-like effects commonly used as a mild stimulant for social recreation and to improve work performance in Ethiopia [49, 50]) was reported by 5.3% of participants. Approximately 44% of participants reported having a fair or poor physical health status, and 33% of them self-reported poor mental health status. Demographic characteristics, smoking behavior, alcohol consumption, and Khat consumption were similar in all participants and those who were referred for verification of psychiatric diagnoses.

Distributions of the socio-demographic and lifestyle characteristics according to participants' MDD status, based on a psychiatrist diagnosis made using the SCAN depression module, are presented in **Table 2.2**. A total of 46 patients fulfilled the DSM-IV criteria for MDD when interviewed by a psychiatrist using the SCAN depression module. Women were more likely to be diagnosed with MDD (14.4%; 95%CI 11.5-17.4%) than men (6.0%; 95%CI 5.4-6.5%). Overall, participants diagnosed with MDD, as compared with those not diagnosed, were more likely to have lower educational attainment, to be divorced or widowed, and to report poor physical and mental health status. The prevalence of MDD according to age and sex is presented in **Figure 2.1**. The prevalence of MDD varied with age among both women and men. Among women the prevalence of MDD was consistently high (15.7% to 16.4%) between the second and fourth decades of life (18-49 years) while among men, MDD prevalence was highest (10.0%) during the

third decade of life (30-39 years). As shown in **Figure 2.2**, PHQ-9 scores were higher among depressed individuals (median=13; interquartile range [51]=10-18) compared to the non-depressed individuals (median=8; IQR=4-12).

Reliability and Item Analysis

The reliability coefficient, Cronbach's alpha for the PHQ-9 total score was 0.81 (0.82 among women and 0.79 among men) (**Table 2.3**). The correlations between nine items of the PHQ-9 and the total scores ranged from 0.57 to 0.75, and all correlations were statistically significant (all 2-tailed p-values < 0.01). Depressed mood and feeling bad about oneself were the two most frequently endorsed items. Conversely, having trouble falling or staying asleep was the item least frequently endorsed by study participants. The test-retest reliability intraclass correlation coefficient of the PHQ-9 total score was 0.92 (0.91 among men and 0.93 among women).

Construct Validity

The results of the factor analysis showed that a rotated factor solution for the PHQ-9 (**Table 2. 4**) contained one factor with eigenvalues greater than 1.0, which accounted for 69.8% of the variance. Item loadings ranged from 0.39 to 0.78 (**Table 2.4**). These values suggest that depressed mood and feeling bad about self were most strongly related to the underlying construct. For both of these variables the correlation between the item and the construct was over 0.70. Suicidal thoughts, lethargy and anhedonia were the next set of items strongly related to the underlying construct. For all of these variables the correlations between PHQ-9 items

and the construct were all greater than 0.60. The items that loaded least strongly included trouble falling or staying asleep (0.39) and poor appetite or over eating (0.45).

WHO-QOL scores for each domain are indicated in **Table 2.5**. Overall, mean WHO-QOL scores for women and men with major depressive disorder were similar for each domain. Across all domains, mean WHO-QOL scores for those classified as depressed were statistically significantly lower than those not depressed in the total sample and within sex-specific comparisons. For instance, for psychological domain participants with MDD, compared to non-depressed individuals, were more likely to have lower mean WHO-QOL scores (42.3 (SD=15.8) versus 60.7 (SD=16.3), $p<0.001$).

Criterion Validity

The psychometric properties of PHQ-9 using MCAR assumptions where all patients with unknown true depression status were excluded from all analyses are presented in **Table 2.6a.** At a PHQ-9 score of ≥9 the sensitivity and specificity were 90.4% (95%CI: 82.6-95.5%) and 61.7% (95%CI: 55.6-67.5%), respectively.

We next evaluated the psychometric properties of PHQ-9 correcting for verification bias using the SCAN diagnosis as a gold standard (**Table 2.6b**). The optimal cut point for maximizing the sensitivity of the PHQ-9 without loss of specificity was a score of ≥9. At this cut point, the PHQ-9 had a sensitivity of 79.0% (95%CI: 68.2-89.1%) and a specificity of 72.0% (95%CI: 68.5-74.9%) for detecting MDD. The positive predictive value was 21.8% (95%CI: 17.5-26.8%) for detecting

major depression on SCAN, and the negative predictive value was 97.1% (95%CI: 95.5-98.2%) with positive likelihood ratio of 2.8 (95%CI: 2.4-3.3) and a negative likelihood ratio of 0.3 (95% CI: 0.2-0.4)

Using a screening criteria of at least five PHQ-9 symptoms present at least several days or more over the last two weeks, with at least one of the symptoms being a cardinal symptom (i.e., anhedonia or depressed mood), was found to be have a lower sensitivity (63.0%) but higher specificity (79.6%) while providing a positive predictive value of 23.9% and a negative predictive value of 95.4%. Using the standard PHQ-9 scoring—that is, at least five symptoms present at least half the days (suicidal ideation could be present only several days), with at least one being a cardinal symptom—provided good specificity (78.1) but poor sensitivity (63.9%). Using the 2-item screening method that uses the first two DSM-IV cardinal symptoms (i.e. anhedonia or depressed mood) present at least several days or more during the last week provided similar sensitivity (66.0) and specificity (70.0).

An ROC analysis was conducted to determine the PHQ-9 cutoff score that is optimal for detecting a diagnosis of major depression (**Figure 2.3**). The AUC under the ROC curve for detecting MDD at a PHQ-9 score of nine or higher was 0.82 (95% CI: 0.78– 0.86), and the SE was 0.023 ($P<0.0001$).

We also evaluated how using the interference with functioning question on the PHQ-9 impacted its sensitivity and specificity. We found that requiring endorsement of at least

"somewhat difficult" functioning as a result of the endorsed depressive symptoms decreased the sensitivity to 73%, but increased the specificity to 75% when using the optimal cutoff (a PHQ-9 score of nine or higher).

Rasch Scale Analysis

Table 2.7 indicates the item fit summary statistics for PHQ-9 using the Rasch analysis. The item difficulties ranged from 58.6 "feeling tired or low energy" (i.e., easier to endorse for participants) to 77.0 "thoughts of suicide" (i.e., difficult to endorse). None of the items misfit the model according to criteria set *a priori* with infit mean square values ranging from 0.81 to 1.24 and outfit mean square values ranging from 0.71 to 1.26. Finally, the separation index of PHQ-9, the ability to discriminate between participants who were depressed and those who were not, was within acceptable range.

Validity of PHQ-9 in Population Subgroups

Participants with higher educational status were less likely to be misdiagnosed using the PHQ-9 relative to SCAN diagnosis (OR=0.87; 95%CI: 0.79-0.97) (please see **Table 2.9**). No significant differences in under or over diagnosis of MDD using the PHQ-9 were noted according to other socio-demographic characteristics.

DISCUSSION

To our knowledge, this is the first validation of an Amharic version of the PHQ-9 questionnaire as a screening tool for major depressive disorder among Ethiopians. The estimated prevalence

of MDD (12.6%; 95%CI 9.2-16.1%) found in our study of clinic patients is consistent with those of previous epidemiologic studies conducted in sub-Saharan Africa [18, 19] and among clinic patients in the US [17]. When compared with a criterion gold standard (psychiatrist-administered SCAN diagnosis), the PHQ-9 has good reliability and fair validity (sensitivity (79%) and specificity (72%)) for diagnosing MDD among Ethiopian adults. The internal consistency reliability was also found to be excellent (ICC=0.92 and Cronbach's alpha=0.81). Our ROC analysis showed that a threshold of nine on the PHQ-9 was the most appropriate cutoff and offered the optimal discriminatory power in detecting MDD. In addition, our study provided strong evidence for the construct validity of the Amharic version of PHQ-9 questionnaire. Finally the results of our factor analysis revealed unidimensionality with acceptable factor loadings on a major core depressive factor and adequate item discrimination values. Our findings are consistent with reports from community based and hospital based studies conducted globally [14, 18, 19, 29, 36, 52-55].

A recent meta-analysis conducted by Manea *et al* [56] found the PHQ-9 to have acceptable screening properties for detecting major depressive disorder for cut-off scores between 8 and 11. In their pooled analysis, the authors reported that specificity estimates summarized across 11 published studies ranged from 73% (95% CI: 63%–82%) to 96% (95% CI 94%–97%) for a cut-off scores between 7 and 15. The authors also noted substantial variability in the sensitivity for cut-off scores between 7 and 15. The sensitivity and specificity of PHQ-9 reported in our study are consistent with those reported by Manea *et al* [56]. Notably, our study findings are comparable to what is reported in other developing countries. For instance, Lotrakul *et al* [54]

in their validation study of the Thai version of PHQ-9, found the optimal cut-off score of 9 resulted in a sensitivity of 84% and specificity of 77% when compared to a nurse administered Mini International Neuropsychiatric Interview (MINI) reference standard. To date, only three, investigative teams have published studies documenting the validity and reliability of the PHQ-9 when used in sub-Saharan African populations (**Table 2.8**) [18-20]; and none of these included Ethiopians. Omoro *et al* [20] in their study of cancer patients in Kenya reported that a Swahili language version of the PHQ-9 was a reliable instrument for depression screening. The authors reported a correlation coefficient of 0.71 and a Cronbach's alpha of 0.80. However, the investigators did not use a reference standard to evaluate criterion validity. Similarly, Adewuya *et al* [18], in their study among Nigerian college students reported good reliability of the English version of the PHQ-9 (correlation coefficient=0.89 and Cronbach's alpha=0.85). Of note, the Nigerian investigators evaluated the validity of the PHQ-9 against a clinician administered MINI which served as their diagnostic reference (gold) standard. The investigators reported high validity for the PHQ-9 compared with the DSM-IV diagnosis for MDD (sensitivity=0.84, specificity=0.99, AUC=0.99; 95% CI 0.98–1.00) [18]. It is important to note, however, that due to its highly structured nature, the MINI is typically regarded as a useful screening instrument instead of a diagnostic gold standard [57]. Thus it is possible that the psychometric properties reported by Adewuya *et al*, may be biased. Finally, results from a recent study of HIV/AIDS patients in Western Kenya [19] indicated that the PHQ-9 may be a reliable instrument for assessing DSM-IV depressive disorders. The investigators reported good internal consistency (Cronbach's alpha=0.78) and somewhat lower test-retest reliability (correlation coefficient=0.59) in comparison to previous studies. The investigators did not assess the

criterion validity of the PHQ-9. Of the three, only one study translated the PHQ-9 instrument into local dialect (Swahili) [20]. Our test-retest reliability of 0.92 for PHQ-9 total score is excellent. This is better than the three African studies and is comparable to what is reported among US outpatient samples [17]. Despite differences in population characteristics, sample size, and study settings, on balance, the findings of our study and those of others [18-20] consistently document the validity, reliability and potential benefits of using the PHQ-9 as a depression screening instrument among sub-Saharan Africans. By employing rigorous methodological approaches, including using a psychiatrist administered objective diagnosis for major depressive disorders, and implementation of multiple statistical analytical methods that accounted for verification bias and other potential limitations, our study adds important new information to the sparse research literature concerning the reliability and validity of screening that may be used to detect depression among sub-Saharan African populations, particularly Ethiopians.

Another important point that merits consideration is at cut-off score of 9 the PHQ-9 yielded low PPV (21.8%) and high NPV (97.1%). A low PPV may indicate lower specificity, lower disease prevalence in the population undergoing screening, or a combination of these factors [40]. In this study, results may come from both low specificity (72%) and low disease prevalence (9.0%). Notably, PPV and NPV are not measures of intrinsic accuracy of a test; rather they are functions of both the intrinsic accuracy and the prevalence of a condition. The positive likelihood ratio of PHQ-9 was 2.8 and the negative likelihood ratio was 0.3. This means that clinically in a similar outpatient settings, patients with MDD are 2.8-times as likely to have a positive test on PHQ-9

(cut-off score ≥ 9) compared to patients without MDD. Similarly, patients without MDD are 3.3 (1/0.3) times more likely to be screen negative on PHQ-9 as compared with patients without MDD.

Several investigators have noted that the PHQ-9 items may not accurately capture all components of MDD [46, 47] in particular cultural contexts. Hence, in this study, we used the Rasch analysis to corroborate the evidence suggesting the PHQ-9 questionnaire is a reliable and valid measure of MDD diagnosis. The results of our Rasch analysis did not detect item misfit using the mean infit and outfit square criteria set *a priori*. However, of the PHQ-9 items, we found that feeling tired or having little energy was easier to endorse (understand) while the question about suicidal thoughts was the most difficult to endorse. We do not have a clear explanation for this. It is possible that there may be cultural values and factors, such as stigma, that impact endorsement of certain depressive symptoms. This question warrants future research. Another important point is one of the double-barreled items (with polar-opposite symptoms) "sleep disturbance" had the highest infit index (1.29). As noted by Williams *et al* [46], in their study among spinal cord injury patients, including items that contain polar opposite symptom descriptions might be confusing for some subjects and could impact the psychometric properties of the PHQ-9 [46]. Investigators were able to improve the psychometric properties of the PHQ-9 by splitting items and removing those that misfit the Rasch model [46, 58]. Future studies need to evaluate misfitting among patients of different subgroups. Evaluating the extent to which, if at all, splitting items improves psychometric properties is also warranted.

The PHQ-9 has been reported to have a single factor in previous studies [36]. Our exploratory and confirmatory factor analyses showed that a single factor model exists among the nine items of the PHQ-9 among Ethiopian adults [37]. This finding is consistent with prior studies that showed a single factor structure of the PHQ-9 across participants among Chinese, African Americans, and Non- Hispanic Whites [36, 52]. Our study results showed that depressed mood and feeling bad about oneself were most strongly related to the underlying construct—meeting DSM-IV diagnostic criteria for depression. For both of these items the correlation between the item and the construct was over 0.70. This means that more than 50% of the variance of these items is related to the construct underlying these items (depression). It is important to note that depressed mood is one of the cardinal symptoms of depression, while feeling bad about self is a secondary diagnostic symptom. Our observations showing strong correlation between the cardinal symptom and the underlying construct is in agreement with prior reports [19]. A study conducted among Nigerian army personnel found that factor analysis of PHQ-9 to have acceptable loadings (ranging from 0.43 to 0.63) [59]. Similar findings were reported by Monahan *et al* [19]. Collectively, our study results and those of others [36, 37] support the thesis that a single-factor structure for the PHQ-9 depressive symptoms can be generalized to many different populations. Importantly, the PHQ-9 depression questionnaire appears to be a reliable screening instrument with fair validity for identifying DSM-IV MDD in sub-Saharan African populations.

Major strengths of our study include the use of a clinical diagnostic gold standard to assess validity, the large sample size, and execution of a rigorous analytic plan that assessed the magnitude and impact of biases on indices used to evaluate the psychometric properties of the Amharic version of the PHQ-9 questionnaire. Using rigorous methodological approaches and a psychiatrist administered objective diagnosis; we were able to overcome some of the previously noted methodological limitations of other studies conducted among African populations. Our study expands the literature by including assessment of an Amharic version of the PHQ-9 that may be used in one of the most populous countries in Africa. Some caveats, however, must be considered when interpreting the results of our study. First, our study was conducted in an urban hospital; therefore, the result may not be generalizable to populations in rural and remote areas. In addition, our study was limited to adults. There is increasing evidence that adolescents are particularly affected by depressive disorders and thus future studies must evaluate the psychometric properties of the PHQ-9 when used in these specific populations [60]. Second, since the PHQ-9 questionnaire and the SCAN diagnostic interview were conducted during the same day, it is possible that the short time interval between the two interviews administration created carryover or recall effects and increased the reliability [61]. The consistency of our study findings with those of other studies [20, 59] and the psychometric properties of other reliability measures, in part, mitigate this concern. Third, the PHQ-9 does not make exclusions for physically induced symptoms. Consequently, we cannot rule out the possibility that our study included some participants with depressive symptoms secondary to physical illnesses and/or medical effects. The inclusion of such participants would lead to spuriously elevated prevalence estimates of major depressive disorder. Finally, there is some

body of evidence that suggests the importance of incorporating a mixed method approach where an initial qualitative exploration is conducted to assess the existence and nature of depression-like syndrome in the local context and to use the data obtained from the qualitative approach to adapt the depression assessment tool [62-64]. This approach is typically used for validating mental health screening instrument in non-Western settings where research has not determined if Western models apply [62-64]. Prior validation studies conducted in Ethiopia that have assessed the use of DSM constructs [65], however, serve to mitigate these concerns.

Several mental health advocates have suggested the integration of mental health in to primary health care to adequately address the burden of mental disorders in low and middle income countries [66]. One of the primary challenges for such integration has been the lack of valid and easy to administer screening tools to detect the presence of mental disorders in primary care settings. Introducing brief screening instruments such as the PHQ-9 could be one solution to the problem. Indeed, improving the recognition of depression in clinical populations in low and middle income countries can be accomplished by the successful adaptation of depression screening instruments and diagnostic approaches from high income country settings into settings with few resources and weaker health systems [67]. The benefits of having well characterized, reliable and valid screening instruments like the PHQ-9 in low income and resource limited clinical and research settings are potentially far reaching [68, 69]. Though depression screening alone is insufficient for addressing growing mental health care needs in low and middle income settings, given depression is the leading contributor to global burden of diseases having a valid screening /diagnostic instrument is an important first step towards addressing a major public

health problem. There are cost effective intervention programs for depression in low income countries where valid screening instruments could be used to identify appropriate participants [70-72]. Importantly, early identification and proper treatment can significantly reduce the adverse impact of depression. Monitoring depressive symptoms using simple screening questionnaires like PHQ-9 that is brief, acceptable, and easy to administer (e.g., self-administration, nurse administration or interactive voice recording) is an important component of effective mental health treatment [73]. Recently, the Ethiopian Federal Ministry of Health released its first National Mental Health Strategy that outlines integrating mental health services into Primary Health Care, one of the most widely recommended approaches by the WHO [74]. As part of this national strategy, physicians, health officers and nurses are provided training to provide brief psychosocial interventions and prescribe psychotropic medications. The systematic use of PHQ-9 instruments in the context of outpatient hospitals allows for the early detection of MDD.

In conclusion, the PHQ-9 is very brief, as well as easy to administer and interpret questionnaire. These are advantageous for use in resource-constrained settings like Ethiopia. Our results demonstrated that PHQ-9 appears to be a reliable and valid instrument to identify DSM-IV major depressive disorder among Ethiopian adults. Future studies should assess the responsiveness and sensitivity of PHQ-9 to treatment in Ethiopian patients. In addition it would also be useful to determine what constitutes a minimal clinically important change in the PHQ-9 score.

REFERENCES

1. WHO: **The Global Burden of Disease 2004 Update.** Geneva, Switzerland.: World Health Organization; 2004.
2. Patel V, Garrison P, de Jesus Mari J, Minas H, Prince M, Saxena S: **The Lancet's series on global mental health: 1 year on**. *Lancet* 2008, **372**(9646):1354-1357.
3. Patel V, Prince M: **Global mental health: a new global health field comes of age**. *JAMA* 2010, **303**(19):1976-1977.
4. Wig NN: **WHO and mental health--a view from developing countries**. *Bull World Health Organ* 2000, **78**(4):502-503.
5. Simon G, Ormel J, VonKorff M, Barlow W: **Health care costs associated with depressive and anxiety disorders in primary care**. *Am J Psychiatry* 1995, **152**(3):352-357.
6. Broadhead WE, Blazer DG, George LK, Tse CK: **Depression, disability days, and days lost from work in a prospective epidemiologic survey**. *JAMA* 1990, **264**(19):2524-2528.
7. Lowe B, Spitzer RL, Grafe K, Kroenke K, Quenter A, Zipfel S, Buchholz C, Witte S, Herzog W: **Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses**. *J Affect Disord* 2004, **78**(2):131-140.
8. Patel V, Araya R, Chowdhary N, King M, Kirkwood B, Nayak S, Simon G, Weiss HA: **Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires**. *Psychol Med* 2008, **38**(2):221-228.
9. Patel V, Sartorius N: **From science to action: the Lancet Series on Global Mental Health**. *Curr Opin Psychiatry* 2008, **21**(2):109-113.

10. Ngoma MC, Prince M, Mann A: **Common mental disorders among those attending primary health clinics and traditional healers in urban Tanzania**. *Br J Psychiatry* 2003, **183**:349-355.

11. Liu SI, Yeh ZT, Huang HC, Sun FJ, Tjung JJ, Hwang LC, Shih YH, Yeh AW: **Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan**. *Compr Psychiatry* 2011, **52**(1):96-101.

12. Mastrogianni A, Bhugra D: **Globalization, cultural psychiatry and mental distress**. *Int J Soc Psychiatry* 2003, **49**(3):163-165.

13. Bhugra D: **Severe mental illness across cultures**. *Acta Psychiatr Scand Suppl* 2006(429):17-23.

14. Kroenke K, Spitzer RL, Williams JB: **The PHQ-9: validity of a brief depression severity measure**. *J Gen Intern Med* 2001, **16**(9):606-613.

15. Bhui K, Bhugra D: **Cross-cultural competencies in the psychiatric assessment**. *Br J Hosp Med* 1997, **57**(10):492-496.

17. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, Falloon K, Hatcher S: **Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population**. *Ann Fam Med* 2010, **8**(4):348-353.

17. Spitzer RL, Williams JB, Kroenke K, Hornyak R, McMurray J: **Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study**. *Am J Obstet Gynecol* 2000, **183**(3):759-769.

18. Adewuya AO, Ola BA, Afolabi OO: **Validity of the patient health questionnaire (PHQ-9) as a screening tool for depression amongst Nigerian university students**. *J Affect Disord* 2006, **96**(1-2):89-93.

19. Monahan PO, Shacham E, Reece M, Kroenke K, Ong'or WO, Omollo O, Yebei VN, Ojwang C: **Validity/reliability of PHQ-9 and PHQ-2 depression scales among adults living with HIV/AIDS in western Kenya**. *J Gen Intern Med* 2009, **24**(2):189-197.

20. Omoro SA, Fann JR, Weymuller EA, Macharia IM, Yueh B: **Swahili translation and validation of the Patient Health Questionnaire-9 depression scale in the Kenyan head and neck cancer patient population**. *Int J Psychiatry Med* 2006, **36**(3):367-381.

21. Bank TW: **The World Bank: Ethiopia Overview. Available at: www.worldbank.org/en/country/ethiopia, accessed on September 20, 2012**..

22. Capps R, McCabe K, Fix M: **New Streams: Black African Migration to the United States**. In. Washington, DC: Migration Policy Institute; 2011.

23. Rashid E, Kebede D, Alem A: **Evaluation of an Amharic version of the Composite International Diagnostic Interview (CIDI) in Ethiopia**. *Ethiop J Health Dev* 1996, **10**:69-77.

24. WHO: **STEPs manual**. Geneva: World Health Organization; 2008.

25. American Psychiatric A, American Psychiatric Association. Task Force on D-I: **Diagnostic and statistical manual of mental disorders : DSM-IV-TR**. Washington, DC: American Psychiatric Association; 2000.

26. Spitzer RL, Kroenke K, Williams JB: **Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire**. *JAMA* 1999, **282**(18):1737-1744.

27. Kinder LS, Carnethon MR, Palaniappan LP, King AC, Fortmann SP: **Depression and the metabolic syndrome in young adults: findings from the Third National Health and Nutrition Examination Survey**. *Psychosom Med* 2004, **66**(3):316-322.

28. Kessler RC, Ustün TB, World Health O: **The WHO World Mental Health Surveys : global perspectives on the epidemiology of mental disorders**. Cambridge; New York; Geneva: Cambridge University Press ; Published in collaboration with the World Health Organization; 2008.

29. Fann JR, Bombardier CH, Dikmen S, Esselman P, Warms CA, Pelzer E, Rau H, Temkin N: **Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury**. *J Head Trauma Rehabil* 2005, **20**(6):501-511.

30. Aboraya A, Tien A, Stevenson J, Crosby K: **Schedules for Clinical Assessment in Neuropsychiatry (SCAN): introduction to WV's mental health community**. *W V Med J* 1998, **94**(6):326-328.

31. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N: **SCAN. Schedules for Clinical Assessment in Neuropsychiatry**. *Arch Gen Psychiatry* 1990, **47**(6):589-593.

32. Cheng AT, Tien AY, Chang CJ, Brugha TS, Cooper JE, Lee CS, Compton W, Liu CY, Yu WY, Chen HM: **Cross-cultural implementation of a Chinese version of the Schedules for**

Clinical Assessment in Neuropsychiatry (SCAN) in Taiwan. *Br J Psychiatry* 2001, **178**:567-572.

33. Krisanaprakornkit T, Paholpak S, Piyavhatkul N: **The validity and reliability of the WHO Schedules for Clinical Assessment in Neuropsychiatry (SCAN Thai Version): Mood Disorders Section**. *Journal of the Medical Association of Thailand* 2006, **89**(2):205-211.

34. Alem A, Kebede D, Shibre T, Negash A, Deyassa N: **Comparison of computer assisted scan diagnoses and clinical diagnoses of major mental disorders in Butajira, rural Ethiopia**. *Ethiop Med J* 2004, **42**(2):137-143.

35. Moller-Leimkuhler AM: **Higher comorbidity of depression and cardiovascular disease in women: a biopsychosocial perspective**. *World J Biol Psychiatry*, **11**(8):922-933.

36. Yu X, Tam WW, Wong PT, Lam TH, Stewart SM: **The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong**. *Compr Psychiatry* 2012.

37. Thompson B: **Exploratory and confirmatory factor analysis : understanding concepts and applications**. Washington, DC: American Psychological Association; 2004.

38. Saxena S, Carlson D, Billington R, Life WGWHOQO: **The WHO quality of life assessment instrument (WHOQOL-Bref): the importance of its items for cross-cultural research**. *Qual Life Res* 2001, **10**(8):711-721.

39. Hyphantis T, Kotsis K, Voulgari PV, Tsifetaki N, Creed F, Drosos AA: **Diagnostic accuracy, internal consistency and convergent validity of the Greek version of PHQ-9 in diagnosing depression in rheumatological disorders**. *Arthritis Care Res (Hoboken)* 2011.

40. Zhou XH, McClish DK, Obuchowski NA: **Statistical methods in diagnostic medicine**, 2nd edn. Oxford: Wiley-Blackwell; 2011.

41. Youden WJ: **Index for rating diagnostic tests**. *Cancer* 1950, **3**(1):32-35.

42. Zhou H: **Statistical Inferences for the Youden Index**. *Dissertation.* Georgia State University; 2011.

43. Begg CB, Greenes RA: **Assessment of diagnostic tests when disease verification is subject to selection bias**. *Biometrics* 1983, **39**(1):207-215.

44. Pepe M: **The statistical evaluation of medical tests for classification and prediction**. Oxford: Oxford University Press; 2003.

45. Carpenter J, Bithell J: **Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians**. *Stat Med* 2000, **19**(9):1141-1164.

46. Williams RT, Heinemann AW, Bode RK, Wilson CS, Fann JR, Tate DG: **Improving measurement properties of the Patient Health Questionnaire-9 with rating scale analysis**. *Rehabil Psychol* 2009, **54**(2):198-203.

47. Cook KF, Bombardier CH, Bamer AM, Choi SW, Kroenke K, Fann JR: **Do somatic and cognitive symptoms of traumatic brain injury confound depression screening?** *Arch Phys Med Rehabil* 2011, **92**(5):818-823.

48. Linacre JM: **Winsteps Ministep Rasch-model computer programs**. 2007.

49. Belew M, Kebede D, Kassaye M, Enquoselassie F: **The magnitude of Khat use and its association with health, nutrition and socio-economic status**. *Ethiop Med J* 2000, **38**(1):11-26.

50. Kalix P: **Khat: scientific knowledge and policy issues**. *Br J Addict* 1987, **82**(1):47-53.

51. Haro JM, Arbabzadeh-Bouchez S, Brugha TS, de Girolamo G, Guyer ME, Jin R, Lepine JP, Mazzi F, Reneses B, Vilagut G *et al*: **Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys**. *Int J Methods Psychiatr Res* 2006, **15**(4):167-180.

52. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL: **Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients**. *J Gen Intern Med* 2006, **21**(6):547-552.

53. Kalpakjian CZ, Toussaint LL, Albright KJ, Bombardier CH, Krause JK, Tate DG: **Patient health Questionnaire-9 in spinal cord injury: an examination of factor structure as related to gender**. *J Spinal Cord Med* 2009, **32**(2):147-156.

54. Lotrakul M, Sumrithe S, Saipanish R: **Reliability and validity of the Thai version of the PHQ-9**. *BMC Psychiatry* 2008, **8**:46.

55. van Steenbergen-Weijenburg KM, de Vroege L, Ploeger RR, Brals JW, Vloedbeld MG, Veneman TF, Hakkaart-van Roijen L, Rutten FF, Beekman AT, van der Feltz-Cornelis CM: **Validation of the PHQ-9 as a screening instrument for depression in diabetes patients in specialized outpatient clinics**. *BMC Health Serv Res* 2010, **10**:235.

56. Manea L, Gilbody S, McMillan D: **Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis**. *CMAJ* 2012, **184**(3):E191-196.

57. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC: **The Mini-International Neuropsychiatric Interview (M.I.N.I.): the**

development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998, **59 Suppl 20**:22-33;quiz 34-57.

58. Smith AB, Rush R, Wright P, Stark D, Velikova G, Sharpe M: **Validation of an item bank for detecting and assessing psychological distress in cancer patients**. *Psycho-oncology* 2009, **18**(2):195-199.

59. Okulate GT, Olayinka MO, Jones OB: **Somatic symptoms in depression: evaluation of their diagnostic weight in an African setting**. *Br J Psychiatry* 2004, **184**:422-427.

60. Richardson LP, McCauley E, Grossman DC, McCarty CA, Richards J, Russo JE, Rockhill C, Katon W: **Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents**. *Pediatrics* 2010, **126**(6):1117-1123.

61. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF: **A comparison of two time intervals for test-retest reliability of health status instruments**. *J Clin Epidemiol* 2003, **56**(8):730-735.

62. Betancourt TS, Bass J, Borisova I, Neugebauer R, Speelman L, Onyango G, Bolton P: **Assessing local instrument reliability and validity: a field-based example from northern Uganda**. *Soc Psychiatry Psychiatr Epidemiol* 2009, **44**(8):685-692.

63. Bass JK, Bolton PA, Murray LK: **Do not forget culture when studying mental health**. *Lancet* 2007, **370**(9591):918-919.

64. Bass JK, Ryder RW, Lammers MC, Mukaba TN, Bolton PA: **Post-partum depression in Kinshasa, Democratic Republic of Congo: validation of a concept using a mixed-methods cross-cultural approach**. *Trop Med Int Health* 2008, **13**(12):1534-1542.

65. Kortmann F: **DSM-III in Ethiopia: a feasibility study**. *Eur Arch Psychiatry Neurol Sci* 1988, **237**(2):101-105.

66. Patel V: **Mental health in low- and middle-income countries**. *Br Med Bull* 2007, **81-82**:81-96.

67. Patel V, Simon G, Chowdhary N, Kaaya S, Araya R: **Packages of care for depression in low- and middle-income countries**. *PLoS Med* 2009, **6**(10):e1000159.

68. Alem A: **Mental health services and epidemiology of mental health problems in Ethiopia**. *Ethiop Med J* 2001, **39**(2):153-165.

69. Giel R: **The prehistory of psychiatry in Ethiopia**. *Acta Psychiatr Scand Suppl* 1999, **397**:2-4.

70. Araya R, Rojas G, Fritsch R, Gaete J, Rojas M, Simon G, Peters TJ: **Treating depression in primary care in low-income women in Santiago, Chile: a randomised controlled trial**. *Lancet* 2003, **361**(9362):995-1000.

71. Bolton P, Bass J, Neugebauer R, Verdeli H, Clougherty KF, Wickramaratne P, Speelman L, Ndogoni L, Weissman M: **Group interpersonal psychotherapy for depression in rural Uganda: a randomized controlled trial**. *JAMA* 2003, **289**(23):3117-3124.

72. Patel V, Chisholm D, Rabe-Hesketh S, Dias-Saxena F, Andrew G, Mann A: **Efficacy and cost-effectiveness of drug and psychological treatments for common mental disorders in general health care in Goa, India: a randomised, controlled trial**. *Lancet* 2003, **361**(9351):33-39.

73. Kroenke K, Spitzer R: **The PHQ-9: A new depression diagnostic and severity measure.** . *Psychiatric Annals* 2002, **32**(9):509-515.

74. EMOH: **Federal Democratic Republic of Ethiopia Ministry of Health. National Mental Health Strategy. Addis Ababa, Ethiopia.** . Geneva: World Health Organization; 2012.

75. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, Falloon K, Hatcher S: **Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population**. *Ann Fam Med* 2010, **8**(4):348-353.

Table 2.1—Characteristics of the entire study population (N=926), and those selected for psychiatrist diagnostic interview (N=363)

Characteristic	All N=926		Diagnostic Interview N=363	
	n	%	n	%
Mean age (years)*	35.1±1.7		34.9±11.6	
Sex				
Women	568	61.3	229	63.1
Men	358	38.7	134	36.9
Marital status				
Married	486	52.5	186	51.3
Never married	293	31.6	109	30.0
Other	147	15.9	68	18.7
Education				
≤ Primary (1-6 years)	400	43.2	169	46.6
Secondary (7-12 years)	322	34.8	124	34.2
College graduate	204	22.0	70	19.3
Religion				
Orthodox Christian	692	74.3	278	76.6
Protestant	122	13.2	48	13.2
Muslim	98	10.6	31	8.5
Other	14	1.5	6	1.6
Smoking status				
Never	802	86.6	310	85.4
Former	88	9.5	41	11.3
Current	36	3.9	12	3.3
Alcohol consumption past year				
Non-drinker	528	57.0	209	57.6
Less than once a month	309	33.4	119	32.8
≥ 1 day a week	89	9.6	35	9.6
Khat consumption				
None	679	73.7	261	71.9
Former	198	21.4	89	24.5
Current	49	5.3	13	3.6

*Mean ± standard deviation (SD)

Table 2.2—Characteristics of the study population according psychiatrist diagnosed major depressive disorder status

Characteristic	Depressed N= 46 %	Non-depressed N=317 %	P-value
Age (years), Mean ± SD	33.7 ± 9.6	35.1 ± 11.9	0.448
Sex			
Women	80.4	60.6	0.001
Men	19.6	39.4	
Marital status			
Married	36.9	53.3	<0.001
Never married	28.3	30.3	
Other	34.8	16.4	
Education			
≤ Primary (1-6)	52.2	45.7	0.425
Secondary (7-12)	30.4	34.7	
College graduate	17.4	19.6	
Smoking status			
Never	78.3	86.4	0.329
Former	17.4	10.4	
Current	4.3	3.2	
Alcohol consumption past year			
Non-drinker	73.9	55.2	0.060
Less than once a month	21.7	34.4	
≥ 1 day a week	4.4	10.4	
Khat chewing			
None	69.6	72.3	0.623
Former	2.2	3.8	
Current	28.2	23.9	
Self-reported physical health			
Excellent/very good/good	36.9	47.9	0.002
Poor/fair	63.1	52.1	
Self-reported mental health			
Excellent/very good/good	34.8	56.2	0.007
Poor/fair	65.2	43.8	

Table 2.3—PHQ-9 item level values and item-total correlations

PHQ-9 item	Corrected item-total correlation	Alpha if item deleted
Little interest or pleasure in doing things	0.62	0.77
Feeling down, depressed, or hopeless	0.74	0.75
Trouble falling or staying asleep	0.52	0.79
Feeling tired or having little energy	0.64	0.78
Poor appetite or overeating	0.54	0.79
Feeling bad about self	0.72	0.75
Trouble concentrating	0.56	0.78
Moving or speaking slowly	0.55	0.78
Thoughts of being better off dead	0.64	0.77

***Overall Cronbach's alpha = 0.81**

Table 2.4—Factor analysis of depressive symptoms using the PHQ-9 questionnaire

	Factor Loadings
Depressive symptoms	**Factor 1**
Little interest or pleasure in doing things	0.55
Feeling down, depressed, or hopeless	0.78
Trouble falling or staying asleep	0.39
Feeling tired or having little energy	0.57
Poor appetite or overeating	0.45
Feeling bad about self	0.72
Trouble concentrating	0.48
Moving or speaking slowly	0.51
Thoughts of being better off dead	0.58
Eigenvalue	2.96
% Variance	69.8

Table 2.5—Mean WHO quality of life scores according to PHQ-9 questionnaire determined MDD disorder status by domain

Quality of life assessed by Domain	Scored high for depression on PHQ-9				P-value
	Yes (n=300)		No (n=626)		
	Mean score	SD	Mean score	SD	
Physical	44.6	12.7	54.0	12.9	<0.001
Psychological	42.2	15.8	60.7	16.3	<0.001
Social relationships	52.5	23.7	67.4	21.4	<0.001
Environmental	34.3	14.8	44.9	15.3	<0.001

* A score of ≥ 9 on PHQ-9 questionnaire was used as cutoff to indicate major depressive disorder (MDD)

Table 2.6a—Sensitivity and specificity for MDD diagnosis across various cut-off points of the PHQ-9 by sex—naïve estimates*

	Sensitivity (95% CI)	Specificity (95% CI)	Youden Index	LR+ (95% CI)	LR- (95% CI)	PPV (95% CI)	NPV (95% CI)
All participants							
Score≥9	90.4 (82.6-95.5)	61.7 (55.6-67.5)	52.1	2.6 (2.0-2.8)	0.2 (0.1-0.3)	45.2 (38.0-52.6)	94.9 (90.5-97.6)
Score≥10	86.2 (77.5-92.4)	67.3 (61.3-72.9)	53.5	2.6 (2.2-3.2)	0.2 (0.1-0.3)	47.9 (40.2-55.7)	93.3 (88.8-96.4)
Score≥11	78.7 (69.1-86.5)	74.0 (68.3-79.1)	52.7	3.4 (3.0-3.9)	0.3 (0.2-0.5)	54.3 (45.3-63.2)	89.4 (84.8-93.0)
Women							
Score≥9	90.1 (80.7-95.9)	62.0 (54.0-69.6)	52.1	2.4 (1.9-2.9)	0.2 (0.1-0.3)	51.0 (42.5-60.7)	93.3 (86.7-97.3)
Score≥10	85.9 (75.6-93.0)	68.4 (60.5-75.5)	54.3	2.9 (2.5-3.5)	0.2 (0.1-0.4)	55.0 (45.2-64.4)	91.5 (85.0-95.9)
Score≥11	81.7 (70.7-89.9)	72.2 (64.5-79.0)	53.9	2.9 (2.2-3.9)	0.3 (0.1-0.4)	56.9 (46.7-66.6)	89.8 (83.1-94.4)
Men							
Score≥9	91.3 (72.0-98.9)	61.3 (51.5-70.4)	52.6	2.4 (1.8-3.1)	0.2 (0.1-0.5)	32.8 (21.6-45.7)	97.1 (90.1-99.7)
Score≥10	87.0 (66.4-97.2)	65.8 (56.2-74.5)	52.8	2.5 (1.9-3.4)	0.2 (0.1-0.6)	34.5 (22.5-48.1)	96.1 (88.9-99.2)
Score≥11	69.6 (47.1-86.8)	76.6 (67.6-84.1)	46.1	2.9 (1.9-4.6)	0.4 (0.2-0.7)	38.1 (23.6-54.4)	92.4 (84.9-96.9)

*Naive estimates calculated using missing completely random assumption; LR+: positive likelihood ratio, LR-: negative likelihood ratio, PPV: positive predictive value, NPV: negative predictive value

Table 2.6b—Verification bias corrected* sensitivity and specificity for MDD diagnosis across various cut-off points of the PHQ- 9 by sex

	Sensitivity (95% CI)	Specificity (95% CI)	Youden Index	LR+ (95% CI)	LR- (95% CI)	PPV (95% CI)	NPV (95% CI)
All participants							
Score≥9	79.0 (68.2-89.1)	72.0 (68.5-74.9)	51.0	2.8 (2.4-3.3)	0.3 (0.2-0.4)	21.8 (17.5-26.8)	97.1 (95.5-98.2)
Score≥10	71.1 (61.2-83.9)	76.6 (74.8-79.5)	47.7	3.0 (2.5-3.6)	0.4 (0.3-0.)	23.0 (18.3-28.6)	96.4 (94.7-97.6)
Score≥11	57.5 (41.6-72.0)	79.8 (78.0-81.5)	37.3	2.8 (2.3-3.5)	0.5 (0.4-0.7)	23.5 (18.4-29.5)	94.5 (92.5-95.9)
Women							
Score≥9	78.7 (67.75-86.9)	71.5 (67.4-75.2)	50.2	2.7 (2.3-3.3)	0.3 (0.2-0.5)	26.5 (20.8-33.1)	96.3 (93.8-97.6)
Score≥10	70.2 (58.0-79.2)	75.6 (71.1-79.2)	45.8	2.9 (2.3-3.6)	0.4 (0.3-0.6)	28.1 (21.9-35.2)	94.8 (92.2-96.6)
Score≥11	59.1(47.5-69.8)	77.6(73.6-81.4)	36.7	2.6 (2.0-3.4)	0.5 (0.4-0.7)	27.6 (21.1-35.2)	92.9 (90.0-95.0)
Men							
Score≥9	78.4 (52.9-90.4)	73.7 (68.9-78.1)	52.1	2.9 (2.1-3.9)	0.3 (0.1-0.7)	12.4 (7.4-20.0)	98.5 (96.1-99.4)
Score≥10	76.9 (50.5-89.8)	78.2 (73.5-82.2)	55.1	3.4 (2.4-4.8)	0.3 (0.2-0.7)	13.8 (8.1-22.6)	98.5 (96.3-99.4)
Score≥11	52.6 (31.7-72.7)	82.6 (78.1-86.2)	35.2	3.0 (1.8-4.9)	0.6 (0.4-0.9)	14.5 (8.1-24.7)	96.9 (94.1-98.3)

* Begg and Greenes adjusted estimates; ** LR+: positive likelihood ratio, LR-: negative likelihood ratio, PPV: positive predictive value, NPV: negative predictive value

Table 2.6c—Sensitivity and specificity for MDD diagnosis using different PHQ-9 screening criteria

	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	PPV (95% CI)	NPV (95% CI)
Naive estimates						
PHQ-9 score ≥9, at least one cardinal symptom scored ≥2	78.3 (63.6-89.1)	64.0 (58.5-69.3)	2.1 (1.8-2.7)	0.3 (0.2-0.6)	24.0 (17.4-31.6)	95.3 (91.5-97.7)
At least 5 symptoms scored ≥ 2, at least 1 cardinal symptom	50.0 (34.9-65.1)	84.2 (79.7-88.1)	3.2 (2.2-4.7)	0.6 (0.4-0.8)	31.5 (21.1-43.4)	92.1 (88.3-94.9)
At least 5 symptoms scored ≥ 1, at least 1 cardinal symptom	89.1 (76.4-96.4)	42.9 (37.4-48.6)	1.6 (1.4-1.8)	0.2 (0.1-0.6)	18.5 (13.6-24.2)	96.5 (91.9-98.8)
At least one cardinal symptom scored ≥ 2	78.3 (63.6-89.1)	54.9 (49.2-60.5)	1.7 (1.4-2.1)	0.4 (0.2-0.6)	20.1 (14.5-26.7)	94.6 (90.2-97.4)
Verification bias corrected estimates						
PHQ-9 score ≥9, at least one cardinal symptom scored ≥2	63.0 (45.9-77.2)	79.6 (76.0-81.8)	3.0 (2.5-3.7)	0.5 (0.3-0.6)	23.9 (18.9-29.8)	95.4 (93.5-96.7)
At least 5 symptoms scored ≥ 2, at least 1 cardinal symptom	35.8 (24.2-49.3)	90.5 (88.1-92.5)	3.8 (2.7-5.3)	0.7 (0.6-0.8)	31.6 (23.8-40.6)	90.7 (88.6-92.5)
At least 5 symptoms scored ≥ 1, at least 1 cardinal symptom	78.1 (58.6-89.9)	63.9 (59.8-66.7)	2.2 (1.8-2.5)	0.3 (02-0.5)	18.4 (14.8-22.6)	96.5 (94.6-97.7)
At least one cardinal symptom scored ≥ 2	66.0 (50.0-79.0)	70.0 (66.3-72.9)	2.1 (1.8-2.6)	0.5 (0.3-0.6)	20.1 (16.1-24.8)	94.6 (92.5-96.1)

*LR+: positive likelihood ratio, LR-: negative likelihood ratio, PPV: positive predictive value, NPV: negative predictive value

Table 2.7—Item fit statistics and misfit order for the four category of the PHQ-9 questionnaire using the Rasch Analysis

PHQ-9 item	Measure	Fit statistics	
		Infit mean square	Outfit mean square
Low energy	-0.83	0.98	1.04
Anhedonia	-0.37	0.93	1.03
Depressed mood	-0.34	0.81	0.74
Feelings of worthlessness or guilt	-0.04	0.83	0.71
Appetite disturbances	-0.02	1.17	1.23
Sleep disturbances	0.06	1.29	1.26
Trouble concentrating	0.18	1.23	1.15
Psychomotor agitation/retardation	0.65	1.15	0.96
Suicidal thoughts	0.72	1.04	0.78

Table 2.8—Summaries of epidemiologic studies examining validity of the PHQ-9 in sub-Saharan Africa

Author (Year)	Country	Participants	Language used	Diagnostic Interview	Study Findings
Omoro *et al* (2006)	Kenya	Cancer patients	Swahili	None	The Swahili language of PHQ-9 was found to be a reliable instrument for depression screening (correlation coefficient=0.71 and Cronbach's alpha=0.80). Criterion validity was not established using a gold standard
Monahan *et al* (2010)	Kenya	HIV/AIDS patients	English	None	The PHQ-9 was found to be reliable (Cronbach's alpha=0.78) instrument but with low test retest reliability (correlation coefficient=0.59). Criterion validity was not established using a gold standard
Adewuya *et al* (2006)	Nigeria	College students	English	Mini International Neuropsychiatric Interview (MINI)	The PHQ-9 was noted to have good reliability (Cronbach's alpha=0.85) and psychometric properties (sensitivity=0.84, specificity=0.99) pared to MINI Definitive clinical diagnostic interview such as SCAN or SCID was not used as reference standard

Table 2.9—Sensitivity and specificity for MDD diagnosis by educational status

	Sensitivity	Specificity
Niave Estimates		
≤ High School	92.1 (78.6-98.3)	50.6 (44.3-56.9)
≥ College	75.0 (34.9-96.8)	66.1 (53.0-77.7)
Verification Bias Corrected		
≤ High School	75.4 (56.9-87.6)	78.6 (76.4-82.5)
≥ College	72.7 (50.5-89.8)	80.6 (73.8-85.9)

Table 2.10—Results from hypothetical screening program with 100 patients, prevalence of depression 15%

	SCAN Diagnosis	
	True positive	**True negative**
PHQ-9 positive	12	24
PHQ-9 negative	3	61
Total	15	85

Figure 2.1—Prevalence of major depressive disorder according to sex and age

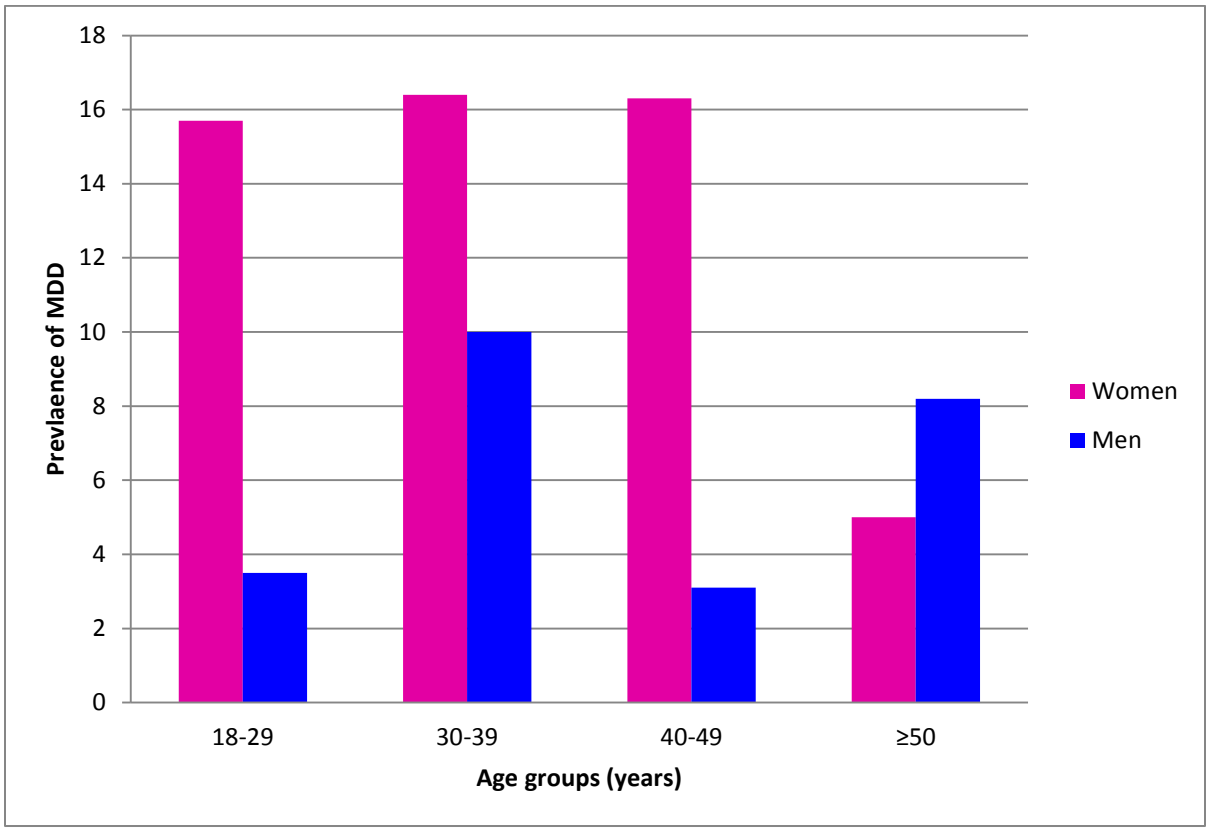
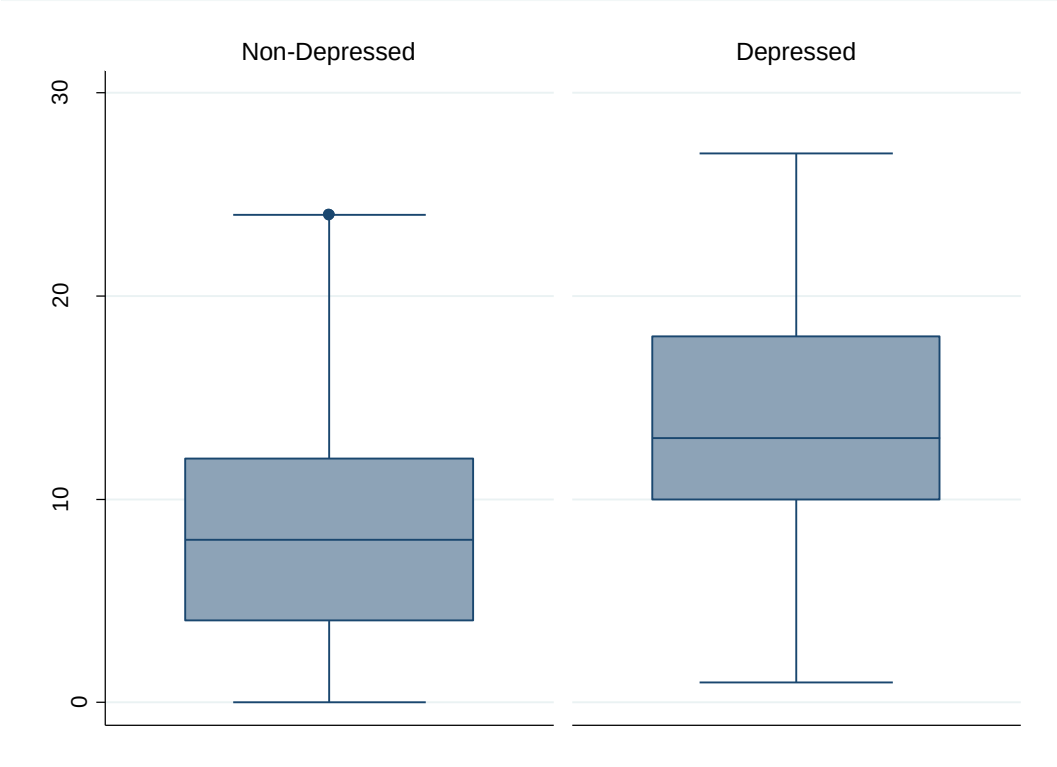
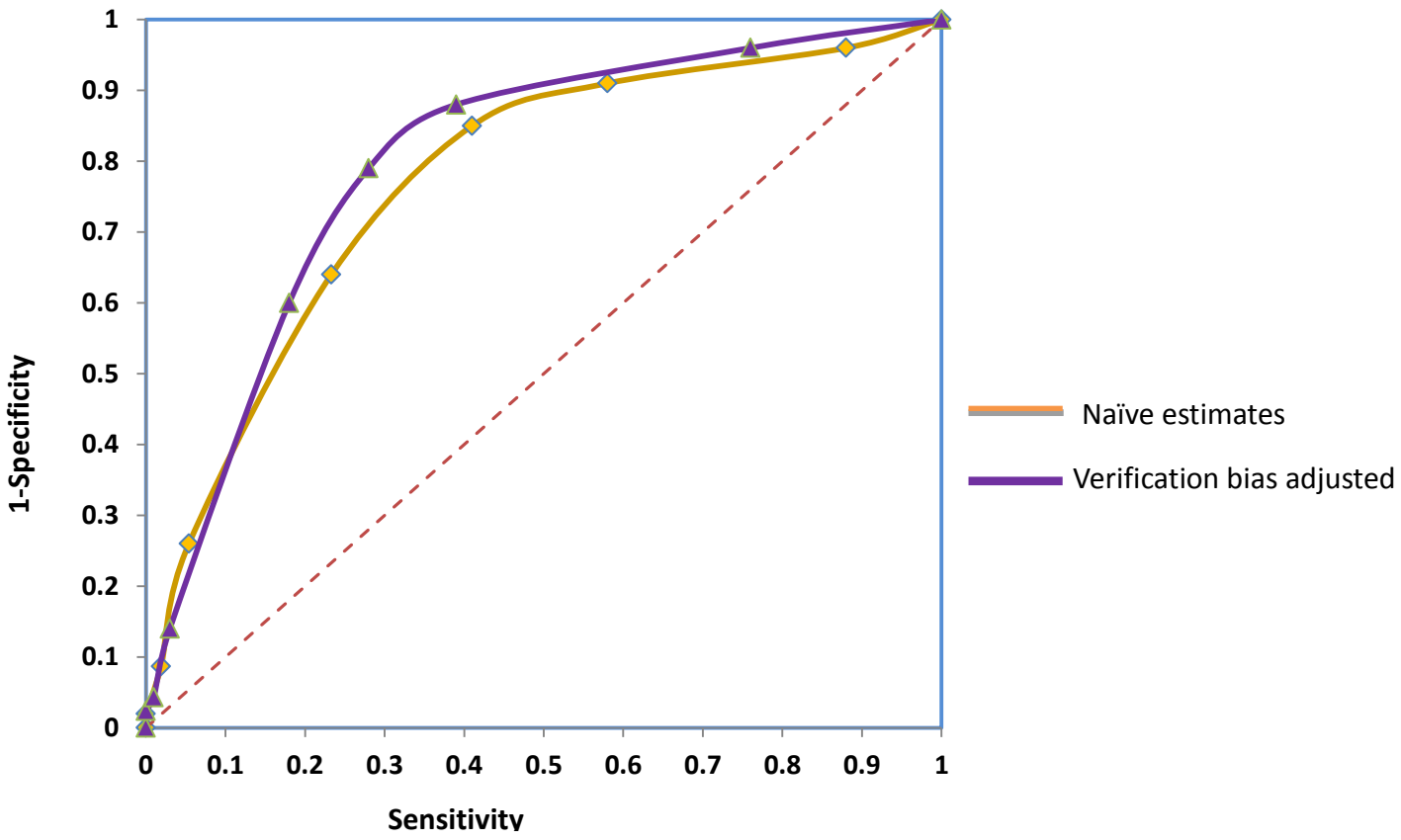


Figure 2.2—-Distribution of PHQ-9 scores according to major depressive disorder status



Box plots comparing PHQ-9 scores among those classified as having major depressive disorder (right side) and those without depression (left side) based on psychiatrist diagnosed major depressive disorder using the SCAN depression module. The central box shows the data between the upper and lower quartiles, with the median represented by the middle line. The "whiskers" (lines on either side of central box) extend from the upper and lower quartiles to a distance of 1.5xIQR (interquartile range) away or the most extreme data point within that range, whichever is smaller.

Figure 2.3—Receiver operating characteristic (ROC) curve of PHQ-9 compared with gold standard SCAN for MDD diagnosis.



Verification Bias Adjusted	Naïve Estimates
AUC= 0.82(95%CI: 0.78-0.86)	AUC=0.77 (95%CI: 0.68-0.85)
SE= 0.0237	SE= 0.043
P (Area=0.5)= <0.0001	P (Area=0.5)<0.001

The receiver operation characteristics (ROC) curve—a plot of sensitivity and (1-specificity) across all possible PHQ-9 score threshold values for diagnosis of MDD in comparison with the gold standard. The better the diagnostic accuracy is, the closer an ROC curve is to the upper left corner of the square. The area under the ROC curve (AUC) is a single summary diagnostic accuracy number stating that the probability that a randomly selected patient with MDD has a greater test result than a randomly selected patient without MDD. The AUC ranges from 0.5 to 1.0

Figure 2.4—Study flow chart

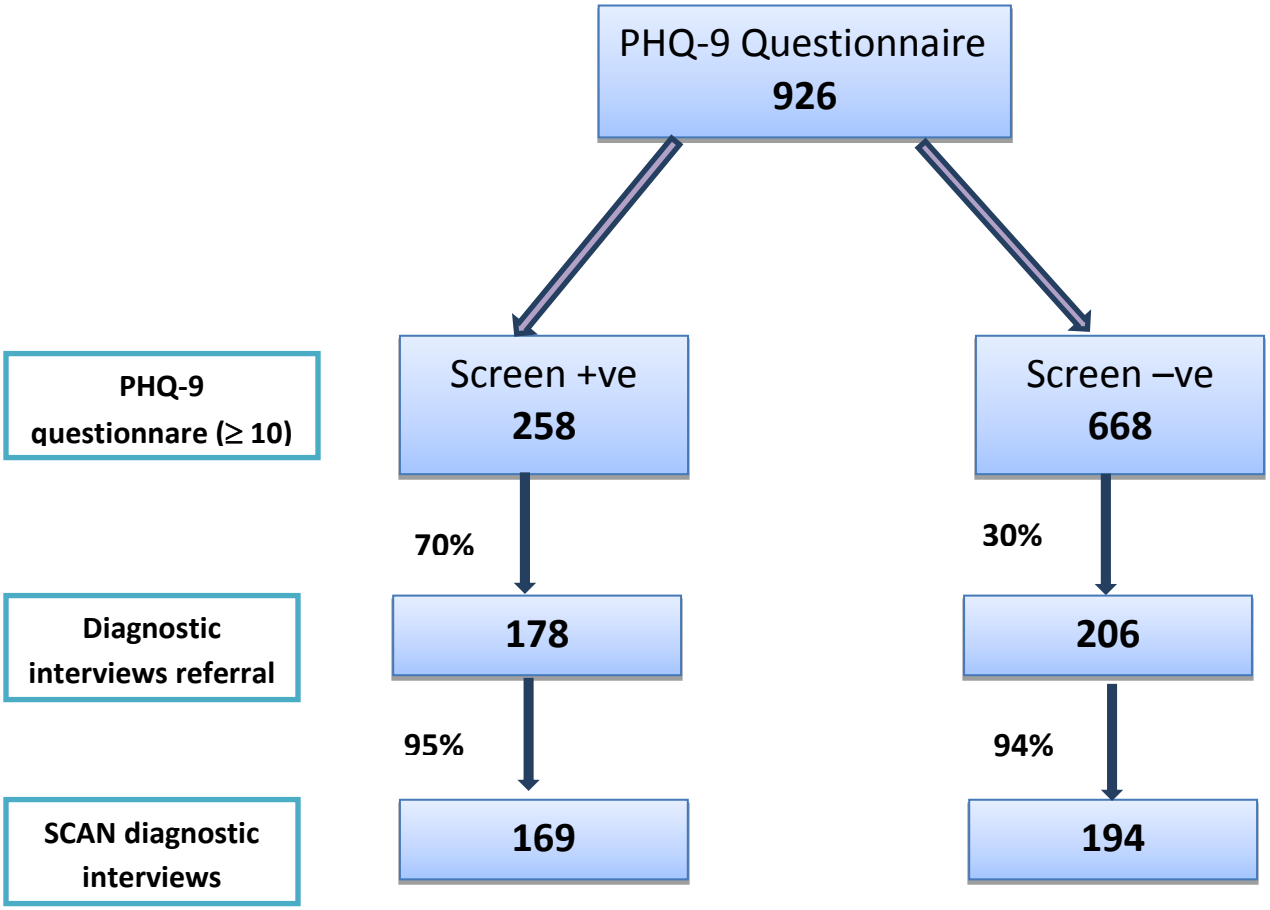
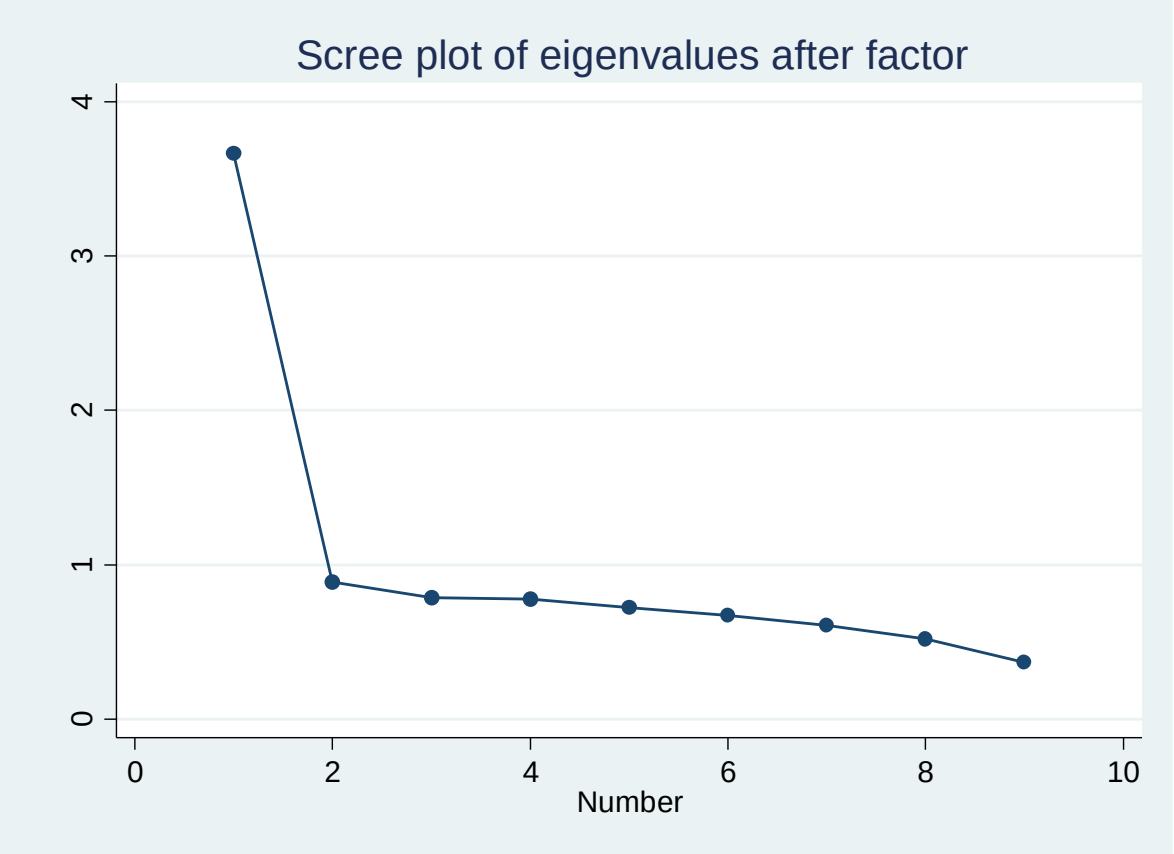
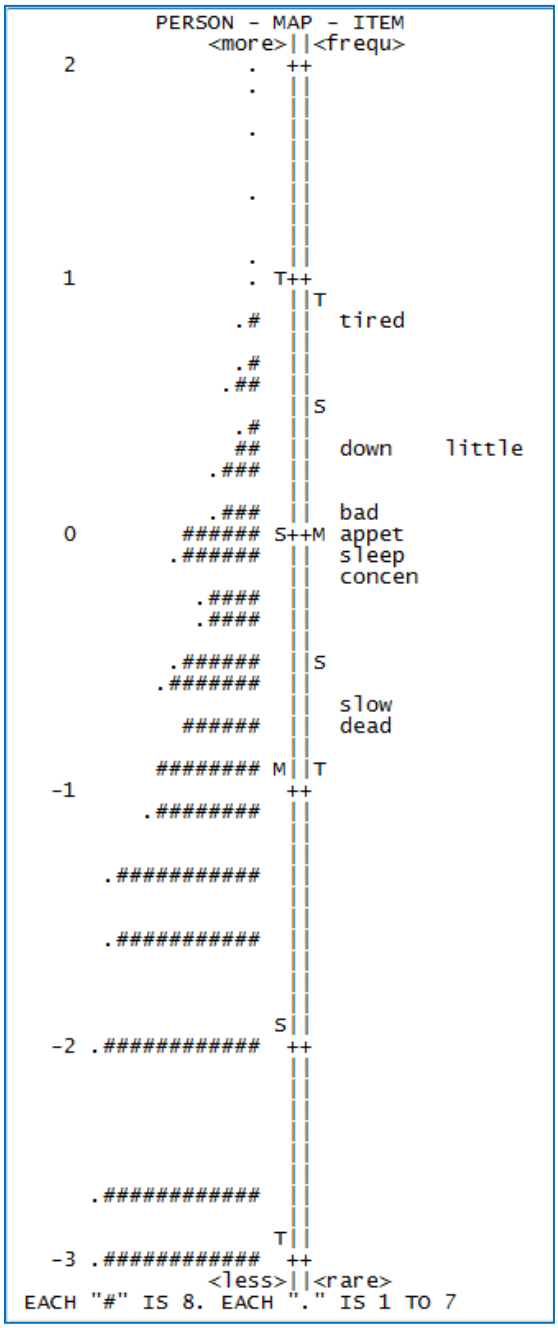


Figure 2.5—Scree plot of PHQ-9 questionnaire items



A screeplot is a graph plotting each eigenvalue associated with the factor(s). Factors with relatively large eigenvalues (greater than one) are assumed to be meaningful and retained for rotation. Factors with small eigenvalues(less than one) are not retained.

Figure 2.6—Wright map for Amharic version of PHQ-9



Wright map for Amharic version of PHQ-9 with a person (n = 926) and item (9items)

Items from the scale are shown on the right side of the figure, and person measures are highlighted by "#" or "."The items at the top are those items the patients easily endorsed. For example, feeling tired or having little energy was easier to endorse than suicidal thoughts.

Abbreviations: M, mean; s, one standard deviation from the mean; T, two standard deviations from the mean

Chapter 3

Diagnostic Validity of the Composite International Diagnostic Interview (CIDI) Depression Module in an East African Population

ABSTRACT

Background: Unipolar depression is one of the leading causes of disability worldwide. However, its recognition and treatment remains low in developing countries, in part, due to lack of cross-culturally validated measures. Therefore, we conducted a validation study of the structured Composite International Diagnostic Interview (CIDI) in diagnosing current major depressive disorder (MDD) among Ethiopian adults.

Methods: A sample of 926 patients attending a major referral hospital participated in this diagnostic assessment study. We used a two stage-study design where participants were first interviewed using an Amharic version of the CIDI and a stratified random sample underwent a follow-up semi-structured clinical interview conducted by a psychiatrist, blinded to the screening results, using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) instrument. We tested construct validity by examining the association of the CIDI and World Health Organization Quality of Life (WHO-QOL) questionnaire. We calculated the psychometric properties of the CIDI using the SCAN diagnostic interview as a gold standard.

Results: We found that the Amharic version of the CIDI diagnostic interview has good internal reliability (Cronbach's alpha= 0.97) among Ethiopian adults. Compared to the SCAN reference standard, the CIDI had fair specificity (72.2%) but low sensitivity (51.0%). Our study provided evidence for unidimensionality of core depression screening questions on the CIDI interview with good factor loadings on a major core depressive factor.

Conclusion: The Amharic language version of the CIDI had fair specificity and low sensitivity in detecting MDD compared with psychiatrist administered SCAN diagnosis. Our findings are

generally consistent with prior studies. Use of fully structured interviews such as the CIDI for MDD diagnosis in clinical settings may lead to under detection of DSM-IV MDD.

INTRODUCTION

Mental health problems in developing countries are largely overlooked despite their high prevalence, their adverse economic impact on families and communities, and their contribution to long-term disability [1-3]. By 2020 major depressive disorder (MDD) is projected to become the leading cause of disability and the second leading contributor to the global burden of disease [4]. However, its recognition and treatment remain low in developing countries due to challenges such as a shortage of skilled mental health workers, stigma associated with mental illness, lack of cross-culturally validated measurement tools, and the prominence of somatic presentations of mental disorders [3, 5, 6].

The Composite International Diagnostic Interview (CIDI) is a fully standardized structured diagnostic interview designed for assessment of mental disorders by trained lay interviewers. The CIDI, developed by the World Health Organization (WHO) , is widely used as a research diagnostic interview to obtain information about the prevalence and correlates of mental disorders from large community samples in different countries [7-10]. A number of studies have demonstrated the validity of the CIDI diagnostic assessment against a trained clinical interviewer [7, 9]. Since its development, the instrument has been widely used in community based and health care settings worldwide. The first version of the CIDI did not adequately and consistently measure risk factors, consequences, patterns and correlates of mental health problems [9]. It also had some methodological limitations including complexity of questions and instructions, which hindered its application in cross cultural settings [9]. Cognizant of these limitations the WHO charged a multinational CIDI Editorial Committee to test the instrument in

many different countries and established the WHO World Mental Health (WMH) Survey Initiative in 1998. Since then the CIDI has gone through a number of revisions to make the diagnostic sections in accordance with the DSM-IV and International Classification of Diseases 10th edition (ICD-10) systems. The CIDI is currently the most widely used fully structured diagnostic interview in psychiatric epidemiologic research [9]. In 1993 a group of investigators assessed the feasibility and acceptability of an Amharic language translated the CIDI version 1.0 among residents of Addis Ababa, Ethiopia [11]. The primary purpose of their study was to assess feasibility and acceptability of the instrument. However, the diagnostic validity (clinical appraisal) of the instrument against a gold standard has not been established [11]. The new version of the CIDI (version 3.0) has been used in Nigeria and South Africa as part of the WMH survey [12-14]. To date, however, there is no published literature reporting the validity of the CIDI version 3.0 in sub-Saharan Africa. This study aims to address this gap in the literature by evaluating the validity and reliability of the CIDI version 3.0 for detecting depression among adults in Ethiopia, the second most populous African country, using a psychiatrist administered Schedules for Clinical Assessment in Neuropsychiatry (SCAN) reference (or gold) standard.

METHODS AND MATERIALS

Study population

Adults (≥ 18 years of age) attending outpatient departments in St. Paul General Specialized Hospital in Addis Ababa, Ethiopia were invited to participate in the study. The hospital is the second largest public hospital in Ethiopia and serves as a referral and teaching hospital. The hospital was established to serve the economically underprivileged population, providing

services free of charge to about 75% of its patients. The hospital provides all basic services across different departments including pediatrics, medical, gynecology, neurology, general surgery, psychiatry, ophthalmology, and emergency medicine. For the purposes of this study only the medical, general surgery and gynecological outpatient departments were invited to complete an interview where a trained research nurse interviewer administered the CIDI instrument in a private room. During this interview other general information were collected including socio-demographic characteristics, quality of life, and self-reported health status. Participants were also asked nine depression screening questions from the Patient Health Questionnaire (PHQ-9) [15]. The items were selected correspond to the criteria for the diagnosis of MDD based on DSM-IV [16]. Those who screened positive for depression on initial interview (i.e., positive on PHQ-9), as well as a randomly selected sub-group of participants who screened negative for depression (i.e., screen negative on PHQ-9) were invited for a diagnostic interview with a psychiatrist who was blinded to the CIDI and PHQ-9 screening outcome. This psychiatrist-administered SCAN diagnostic interview was completed after the initial CIDI interview within the same day.

Study Design

The study followed a two-stage design in which study participants first underwent a structured interview using the CIDI. Subsequently, participants' MDD status was verified by a psychiatrist-administered diagnostic interview using the SCAN instrument.

Study Procedures

The data collection was conducted between June and December 2011. Research nurses were recruited and received structured training on administration of the CIDI at Addis Continental Institute of Public Health. The training program was similar to the one that the principal investigator attended at the Social Survey Institute at the University of Michigan (WHO Training Center). In addition to the structured training course for the interviewers, item-by-item description of questionnaires and role plays were used further by two days of debriefing and review after each interviewer had done at least three practice interviews. To ensure the highest possible quality of collected data, interviewers were provided strict on-site supportive supervision while in the field. All paper and pencil recorded questionnaires, collected manually, were entered using Blaise version 4.6 (Statistics Netherlands), which contained the entire WMH-CIDI algorithm along with an automatic checking mechanism to identify item omissions and unusual responses.

Composite International Diagnostic Interview (CIDI version 3.0)

We used the depression and mania modules of the CIDI 3.0 to diagnose MDD. The CIDI interview, designed for administration by trained lay interviewers, includes three screening questions (known as STEM) about sadness/depressed mood, feelings of discouragement, and loss of interest lasting several days or longer (**Table 3.6**). Participants who endorsed any of the three questions were administered the depression module. Those who failed to endorse any of the three STEM questions were skipped out of the depression module. On average, it took 25 minutes to administer the CIDI interview. In accordance with DSM-IV criteria, we defined MDD

as the presence of five out of nine depressive symptoms that persist for two weeks or longer, are present for most of the day nearly every day, and cause significant distress or impairment. These symptoms include dysphoric mood or anhedonia (cardinal symptoms) that persist most of the day, and clinically significant weight gain/loss or appetite disturbance, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue or loss of energy, feelings of worthlessness or excessive guilt, diminished ability to concentrate or think clearly, and recurrent thoughts of death or suicide. For the purposes of this validation study, we defined current MDD as experiencing MDD in the past 12 months without plausible organic causes and without history of mania or hypomania [7, 9].

Schedules for Clinical Assessment in Neuropsychiatry (SCAN) Diagnosis

The SCAN is a semi structured clinical interview developed by the WHO for use by trained clinicians to assess and diagnose psychiatric disorders among adults [17, 18]. The SCAN was developed within the framework of the WHO and the National Institute of Mental Health (NIMH) Joint Project on Diagnosis and Classification of Mental Disorders, Alcohol and Related Problems [17]. The use of the SCAN gives flexibility in the diagnosis of mental disorders based on ICD-10 and DSM-IV [17]. The depressed mood and ideation module of SCAN has been reported to work well in different languages and cultures including Ethiopia [19]. For the current study, we used SCAN administered in Amharic as a reference (or gold) standard. On average it took 20 minutes to administer the SCAN interview. All SCAN interviews (and diagnoses) were conducted without knowledge of the results of initial CIDI interviews.

World Health Organization Quality of Life Questionnaire

The WHO Quality of Life questionnaire (WHO-QOL) [20] is a cross-cultural assessment tool that captures an individual's perception of their position in life in the context of culture and value systems in which they live and in relation to their goals, expectation, standards and concerns [20]. In this study we used the abbreviated version of WHO-QOL (also known as WHOQOL-BREF) which has 26 items across four domains. The physical domain consists of questions about daily activities, dependence on medication and treatment, energy and exhaustion, mobility, pain and discomfort, sleep and rest, and capacity to work. The psychological domain consists of questions about positive and negative feelings, self-esteem, body image and external image, personal beliefs, and attention. The social relationships domain consists of questions about relationships with others, social support, and sex life. The environmental domain of the scale consists of questions about the home environment, physical security and safety, financial resources, availability of health services, leisure activities, physical environment, and transportation. The overall percentile score for each domain ranges from 0% (very poor) to 100% (very good).

Statistical Analysis

Completed data collection instruments were assessed for quality and completeness. Data from CIDI 3.0 paper and pencil interview (PAPI) were entered using direct data entry (DDE) software. After data checking and cleaning, the data were transferred to Stata 11.0 software (Statacorp, College Station, TX) for analyses. Standard descriptive statistics was performed. Participants' characteristics were summarized using means (± standard deviation) for continuous variables

with symmetric distribution and median (interquartile range) for variables with non-normal distribution and counts and percentages for categorical variables. Variables with skewed distributions were transformed using the natural logarithm and summarized using medians (interquartile range).

Diagnostic Validity Measures

We assessed the construct validity, which is defined as how a test measures the underlying construct of depression [21], on the total sample (N=926) using two approaches. First, we used exploratory factor analysis (EFA) to assess the factor structure of the symptoms of depression. Promax oblique rotation was performed because depression symptoms may be inter-correlated with each other. The number of factors to extract was determined by using the criterion of eigenvalue greater than one and the scree plot (shown in Appendix 2). Next, we used the WHO-QOL questionnaire to assess the associations between depression and quality of life. To test this hypothesis, we used a Student's t-test to compare mean WHO-QOL scores between those classified as depressed (yes/no) according to results from our administration of the CIDI depression screening instrument.

We evaluated the criterion validity—the results of an instrument with unknown merit against a gold standard—by determining the concordance between the CIDI and SCAN clinical diagnosis. We computed the following parameters: sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, positive predictive values, and negative predictive values for the presence or absence of MDD [22]. First, we evaluated the naïve psychometric properties of the

CIDI using missing completely at random (MCAR) assumptions where all patients with unknown SCAN depression status were discarded. We then reported Begg and Greenes adjusted estimates of psychometric properties of the CIDI, where the estimates were corrected for verification bias using inverse probability weighting approach [22-24] under missing at random (MAR) assumptions.

RESULTS

A summary of selected socio-demographic and lifestyle characteristics of study participants is presented in **Table 3.1**. A total of 926 participants with an average age of 35 ± 11 years, ranging from 18 to 69 years were included in the study. The majority of participants were women (61%), married (52.3%) and more likely to be Orthodox Christians (74%). Approximately 4% of participants reported that they were current smokers and 9.6% of participants reported drinking at least one alcoholic beverage per week. Khat chewing was reported by 5.3 % of participants. Approximately 44% of participants reported having a fair or poor physical health status, while 33% reported poor mental health status.

Lifetime prevalence of depressive symptoms, derived from response to the initial CIDI STEM questions for depression, are presented in **Figure 3.1**. The most commonly endorsed depressive symptom was feeling sad, empty or depressed (54.6%) followed by feeling discouraged (42.6%). Distributions of the socio-demographic and lifestyle characteristics according to participants' current CIDI MDD status are shown in **Table 3.2**.

Overall, the 12 month prevalence estimate of the CIDI MDD in the total 926 sample was 6.5% (95% CI, 4.9- 8.1) (**Table 3.2**). Distributions of the socio-demographic and lifestyle characteristics according to participants' CIDI MDD status are also presented in **Table 3.2**. Those who were classified as depressed were less likely to be married, more likely to be smokers, and more likely to report poor physical and mental health.

The exploratory factor analysis of the CIDI depressive symptoms questions showed that a rotated factor solution contained one factor with eigenvalues greater than 1.0, which accounted for 84.3% of the variance (**Table 3.3**). Item loadings ranged from 0.54 to 0.95. Depressed mood, diminished ability to concentrate or think, and change in appetite were most strongly related to the underlying construct (loading values over 0.90). Anhedonia and suicidal thoughts were the next items strongly related to the underlying construct. For these two variables the correlation between the item and the construct was over 0.85. The items that loaded least strongly included psychomotor agitation or retardation (0.54) worthlessness or excessive or inappropriate guilt (0.75).

The WHO-QOL scores for physical, psychological, social relationship and environmental domains are summarized in **Tables 3.4a** and **3.4b.** Since the scores have fairly symmetrical distributions, we provide mean (SD) summaries. The mean WHO-QOL scores for depressed women were significantly lower in each domain as compared to men. Across all domains for both men and women, the mean scores for those classified as depressed were significantly lower than those not depressed. For instance, for psychological domain participants with MDD,

compared to non-depressed, were more likely to have lower mean WHO-QOL scores (46.9 (SD=18.8) versus 57.6 (SD=17.5), p<0.001). Similar significantly lower mean WHO-QOL scores were noted for social relationship, physical and environmental domains. Furthermore, similar significant differences were noted with sex specific comparisons.

Table 3.5 shows the psychometric characteristics of CIDI using SCAN as a gold standard. First, we evaluated the psychometric properties of CIDI using MCAR assumptions where all patients with unknown true MDD status were discarded. For instance, the sensitivity and specificity of the CIDI under naïve estimators for all participants were 43.5% (95%CI: 28.9-58.9%) and 77.9% (95%CI: 72.9-82.4%), respectively. The positive predictive value for detecting MDD with the CIDI was 22.4%, and the negative predictive value was 90.5%. After adjusting for verification bias, the sensitivity was 51% (95%CI: 41.3-60.6%) and specificity was 72.2% (95%CI: 68.7-75.6%). The positive predictive value for detecting MDD with the CIDI assessment was 22.4%, and the negative predictive value was 90.5%. The sex stratified analysis showed that sensitivity was higher in women (sensitivity = 56.1% and specificity = 69.3%) compared with men (sensitivity = 45.0% and specificity = 69.6%).

DISCUSSION

Our results provided evidence that the Amharic version of the CIDI is a tool with high internal reliability (Cronbach's alpha = 0.97) among Ethiopian adults. Our study also provided strong evidence for the construct validity of the Amharic version of the CIDI in diagnosing MDD. The factor analysis revealed unidimensionality of core depression screening questions with good

factor loadings on a major core depressive factor. Adults classified with MDD by the CIDI had lower quality of life scores across physical, psychological, social relationship and environmental domains. Compared to SCAN reference standard, the CIDI is modestly effective in MDD diagnosis with a moderate specificity (72.2%) but low sensitivity (51.0%). Finally, the CIDI was most successful at identifying a person without MDD (with a good NPV of 90.5%). In other words, a person who was identified as non-depressed using CIDI had more than 90% probability of not having MDD diagnosis using SCAN. However, a person who identified as depressed using CIDI had 22% (PPV) chance of having MDD diagnosis using SCAN. The low PPV could be due to the low prevalence of MDD in our study. The positive likelihood ratio commonly considered to "rule in disease" of the CIDI was 1.8. The negative likelihood ratio commonly considered to "rule out of disease" was 0.7. This means that clinically in a similar outpatient settings, patients with MDD are 1.8 times more likely to have an MDD diagnosis using CIDI compared to patients without MDD. Similarly patients without MDD are 1.4 (1/0.7) times more likely to have negative test results using CIDI compared to those with MDD. The high NPV and low negative LR show the potential for CIDI in ruling out MDD in a similar outpatient setting.

Findings from our study are generally consistent with prior validation studies conducted in other countries. In their study among UK residents, Brugha *et al* [25] compared the performance of the CIDI in diagnosing depression using SCAN diagnostic interview as a reference standard [25]. The authors employed a two phased study design where participants were initially screened using the CIDI instrument and then evaluated using the SCAN clinical interview. They reported a sensitivity of 50% (95%CI: 12.0-88%) and a specificity of 87% (95%CI:

81-91%). Similarly, a recent reappraisal study conducted by Haro *et al* compared the current version of CIDI (CIDI 3.0) with a clinician-administered Structured Clinical Interview for DSM-IV (SCID) in a probability subsamples of the WHO World Mental Health (WMH) surveys in France, Italy, Spain, and the US [26]. The investigators noted a sensitivity of 55.3 % and specificity of 93.7% for MDD. In contrast, Jordanova and colleagues [27] in their study among 105 primary care attendees in UK evaluated the diagnostic properties of CIDI with reference to a clinician administered SCAN. The authors used ICD-10 diagnostic categories to compare the two instruments. Although the concordance for depressive disorders was fair (kappa=54%), the sensitivity was 100% and specificity of 88%. CIDI prevalence estimates of depression (18.1%) were more than two-fold higher compared with SCAN prevalence estimates (7.6%) [27] although the CIDI estimates are generally conservative compared with semi-structured interviews [28].

Reasons for the differences in test characteristics across studies are unclear. We speculate that the lower sensitivity of the CIDI noted in our study and those of others [25, 26] may stem from the fact that the CIDI is a fully structured interview that employs precisely worded questions that cannot be rephrased or reworded [9, 25, 28]. In addition, the CIDI is designed to be used by lay interviewers, thus does not allow use of clinical judgment. Whereas the gold standard for our comparison was the semi-structured psychiatrist administered SCAN interview with freedom for probing and rephrasing questions [18]. In addition, given that a large proportion of patients endorsed the STEM questions of depression (75%), it is possible that the strict skip structure of CIDI led to poor sensitivity. The diagnostic STEM (core) questions used at the

beginning of the CIDI and SCAN interviews are similar. However, SCAN allows clinicians to elicit more information based upon open ended probing whereas CIDI follows a strict Yes/No structure [25]. It is important to recognize that relaxing the threshold of the symptoms might be something that merits consideration in future studies.

Investigators have discussed the importance of using factor structure when an instrument is applied in a new context or cultural group[29]. However, to date, we are not aware of previous studies that examined the factor structure of CIDI depression screening instrument. Our study provided the first evidence of the unidimensional nature of the depression symptom questions included in CIDI. As shown in our factor analysis, the correlations between the item and the construct for depressed mood, diminished ability to concentrate or think and change in appetite were over 0.90 showing strong correlation to the construct underlying these items. This is important in that depressed mood is one of the cardinal symptoms of depression, diminished ability to concentrate or think and changes in appetite are secondary diagnostic symptoms. Anhedonia and suicidal thoughts were the next set of items strongly related to the underlying construct. This finding is consistent with what is reported with other depression screening instruments such as the PHQ-9 [30].

Several potential limitations must be considered when interpreting results from our study. First, our study was conducted in a clinical setting. Some investigators have noted that in community based studies of CIDI, respondents were more comfortable admitting personal or socially unacceptable feelings and behaviors to lay interviewers than to clinical interviewers [7, 31]. The

concordances of our results with those from other studies that have included community based samples, however, serve to attenuate some of these concerns. It will be important for future studies to evaluate the diagnostic validity of the Amharic version of the CIDI within the general population. Second, some investigators have argued that semi-structured diagnostic interviews such as SCAN or SCID do not represent a valid gold standard, as neither perfectly reflects the DSM diagnosis [28]. Hence, they suggest evaluating the standard psychometric properties with performance measures such as sensitivity or specificity may not be appropriate. Furthermore, it is important to recognize that the self-report of subjective symptoms is an inherent problem in measurement of mental disorders in clinical and research settings that cannot be solved by use of any standardized instruments such as SCAN or SCID [32]. Notwithstanding the noted limitations, our study has several strengths. First, we used strict protocols of a two-stage study design and appropriate statistical analysis ensuring there is no verification bias. Second, psychiatrists who administered the criterion reference standard SCAN were blinded to the results of the CIDI interviews.

In conclusion, this study provided fundamental information concerning the reliability and validity of an Amharic language translated the CIDI for depression diagnosis among Ethiopian adults. Knowledge gained from this study will facilitate efforts in identifying unmet mental health needs and improving mental health services. There is epidemiologic evidence suggesting that locally validated screening instruments can be used by non-physicians and community health workers [33]. The benefits of having well characterized instruments like the CIDI in low income and resource limited clinical and research settings are far reaching [34, 35]. First, they

have the ability to provide reliable local data on specific disorders. Having local data compels health personnel and policy makers in specific countries or regions and to evaluate unmet needs and to evaluate implementation of health programs and policies. Second, they can be used to evaluate modifiable risk factors associated with depression in the local context. Lastly, they provide a basis for designing and evaluating the impact of intervention programs [36-38]. Given that semi-structured interviews such as SCAN are time-demanding, expensive and require a trained clinician, the use of fully structured interviews like CIDI is critical. However, in light of our findings and those of others, researchers should interpret the findings of these surveys in the general population cautiously.

REFERENCES

1. Patel V, Garrison P, de Jesus Mari J, Minas H, Prince M, Saxena S: **The Lancet's series on global mental health: 1 year on**. *Lancet* 2008, **372**(9646):1354-1357.

2. Patel V, Prince M: **Global mental health: a new global health field comes of age**. *JAMA* 2010, **303**(19):1976-1977.

3. Wig NN: **WHO and mental health--a view from developing countries**. *Bull World Health Organ* 2000, **78**(4):502-503.

4. WHO: **The Global Burden of Disease 2004 Update.** . In. Geneva, Switzerland.: World Health Organization; 2004.

5. Patel V, Araya R, Chowdhary N, King M, Kirkwood B, Nayak S, Simon G, Weiss HA: **Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires**. *Psychol Med* 2008, **38**(2):221-228.

6. Patel V, Sartorius N: **From science to action: the Lancet Series on Global Mental Health**. *Curr Opin Psychiatry* 2008, **21**(2):109-113.

7. Kessler RC, Abelson J, Demler O, Escobar JI, Gibbon M, Guyer ME, Howes MJ, Jin R, Vega WA, Walters EE *et al*: **Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI)**. *Int J Methods Psychiatr Res* 2004, **13**(2):122-139.

8. Kessler RC, Angermeyer M, Anthony JC, R DEG, Demyttenaere K, Gasquet I, G DEG, Gluzman S, Gureje O, Haro JM *et al*: **Lifetime prevalence and age-of-onset distributions**

of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry* 2007, **6**(3):168-176.

9. Kessler RC, Ustun TB: **The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI)**. *Int J Methods Psychiatr Res* 2004, **13**(2):93-121.

10. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA *et al*: **The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures**. *Arch Gen Psychiatry* 1988, **45**(12):1069-1077.

11. Rashid E, Kebede D, Alem A: **Evaluation of an Amharic version of the Composite International Diagnostic Interview (CIDI) in Ethiopia**. *Ethiop J Health Dev* 1996, **10**:69-77.

12. Bass JK, Bolton PA, Murray LK: **Do not forget culture when studying mental health**. *Lancet* 2007, **370**(9591):918-919.

13. Johnson LR, Mayanja MK, Bangirana P, Kizito S: **Contrasting concepts of depression in Uganda: implications for service delivery in a multicultural context**. *Am J Orthopsychiatry* 2009, **79**(2):275-289.

14. Lloyd CE, Dyer PH, Barnett AH: **Prevalence of symptoms of depression and anxiety in a diabetes clinic population**. *Diabet Med* 2000, **17**(3):198-202.

15. Kroenke K, Spitzer RL, Williams JB: **The PHQ-9: validity of a brief depression severity measure**. *J Gen Intern Med* 2001, **16**(9):606-613.

16. American Psychiatric A, American Psychiatric Association. Task Force on D-I: **Diagnostic and statistical manual of mental disorders : DSM-IV-TR**. Washington, DC: American Psychiatric Association; 2000.

17. Aboraya A, Tien A, Stevenson J, Crosby K: **Schedules for Clinical Assessment in Neuropsychiatry (SCAN): introduction to WV's mental health community**. *W V Med J* 1998, **94**(6):326-328.

18. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N: **SCAN. Schedules for Clinical Assessment in Neuropsychiatry**. *Arch Gen Psychiatry* 1990, **47**(6):589-593.

19. Alem A, Kebede D, Shibre T, Negash A, Deyassa N: **Comparison of computer assisted scan diagnoses and clinical diagnoses of major mental disorders in Butajira, rural Ethiopia**. *Ethiop Med J* 2004, **42**(2):137-143.

20. Saxena S, Carlson D, Billington R, Life WGWHOQO: **The WHO quality of life assessment instrument (WHOQOL-Bref): the importance of its items for cross-cultural research**. *Qual Life Res* 2001, **10**(8):711-721.

21. Moller-Leimkuhler AM: **Higher comorbidity of depression and cardiovascular disease in women: a biopsychosocial perspective**. *World J Biol Psychiatry*, **11**(8):922-933.

22. Zhou XH, McClish DK, Obuchowski NA: **Statistical methods in diagnostic medicine**, 2nd edn. Oxford: Wiley-Blackwell; 2011.

23. Begg CB, Greenes RA: **Assessment of diagnostic tests when disease verification is subject to selection bias**. *Biometrics* 1983, **39**(1):207-215.

24. Pepe M: **The statistical evaluation of medical tests for classification and prediction**. Oxford: Oxford University Press; 2003.

25. Brugha TS, Jenkins R, Taub N, Meltzer H, Bebbington PE: **A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN)**. *Psychol Med* 2001, **31**(6):1001-1013.

26. Haro JM, Arbabzadeh-Bouchez S, Brugha TS, de Girolamo G, Guyer ME, Jin R, Lepine JP, Mazzi F, Reneses B, Vilagut G *et al*: **Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys**. *Int J Methods Psychiatr Res* 2006, **15**(4):167-180.

27. Jordanova V, Wickramesinghe C, Gerada C, Prince M: **Validation of two survey diagnostic interviews among primary care attendees: a comparison of CIS-R and CIDI with SCAN ICD-10 diagnostic categories**. *Psychol Med* 2004, **34**(6):1013-1024.

28. Kessler RC, Ustün TB, World Health O: **The WHO World Mental Health Surveys : global perspectives on the epidemiology of mental disorders**. Cambridge; New York; Geneva: Cambridge University Press ; Published in collaboration with the World Health Organization; 2008.

29. Reise SP, Widaman KF, Pugh RH: **Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance**. *Psychol Bull* 1993, **114**(3):552-566.

30. Yu X, Tam WW, Wong PT, Lam TH, Stewart SM: **The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong**. *Compr Psychiatry* 2012.

31. Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen H-U: **The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF)**. *International Journal of Methods in Psychiatric Research* 1998, **7**(4):171-185.

32. Tsuang MT, Tohen M, Jones PB: **Textbook of psychiatric epidemiology**. Chichester, West Sussex; Hoboken, NJ: Wiley-Blackwell; 2011.

33. Patel V, Simon G, Chowdhary N, Kaaya S, Araya R: **Packages of care for depression in low- and middle-income countries**. *PLoS Med* 2009, **6**(10):e1000159.

34. Alem A: **Mental health services and epidemiology of mental health problems in Ethiopia**. *Ethiop Med J* 2001, **39**(2):153-165.

35. Giel R: **The prehistory of psychiatry in Ethiopia**. *Acta Psychiatr Scand Suppl* 1999, **397**:2-4.

36. Araya R, Rojas G, Fritsch R, Gaete J, Rojas M, Simon G, Peters TJ: **Treating depression in primary care in low-income women in Santiago, Chile: a randomised controlled trial**. *Lancet* 2003, **361**(9362):995-1000.

37. Bolton P, Bass J, Neugebauer R, Verdeli H, Clougherty KF, Wickramaratne P, Speelman L, Ndogoni L, Weissman M: **Group interpersonal psychotherapy for depression in rural Uganda: a randomized controlled trial**. *JAMA* 2003, **289**(23):3117-3124.

38. Patel V, Chisholm D, Rabe-Hesketh S, Dias-Saxena F, Andrew G, Mann A: **Efficacy and cost-effectiveness of drug and psychological treatments for common mental disorders**

in general health care in Goa, India: a randomised, controlled trial. *Lancet* 2003, **361**(9351):33-39.

Table 3.1—Characteristics of the entire study population (N=926)

Characteristic	N=926	
	N	%
Mean age (years)*	35.1±1.7	
Sex		
Women	568	61.3
Men	358	38.7
Marital status		
Married	486	52.5
Never married	293	31.6
Other	147	15.9
Education		
≤ Primary (1-6)	400	43.2
Secondary (7-12)	322	34.8
College graduate	204	22.0
Smoking status		
Never	802	86.6
Former	88	9.5
Current	36	3.9
Alcohol consumption past year		
Non-drinker	528	57.0
Less than once a month	309	33.4
≥ 1 day a week	89	9.6
Khat chewing		
None	679	73.7
Former	198	21.4
Current	49	5.3
Self-reported physical health		
Excellent/very good/good	522	56.4
Poor/fair	404	43.6
Self-reported mental health		
Excellent/very good/good	616	66.5
Poor/fair	310	33.5

* Mean ± standard deviation (SD)

Table 3.2—Characteristics of subjects according to CIDI and SCAN major depressive disorder status

	CIDI		SCAN	
	Depressed N=60	Non-depressed N=866	Depressed N= 46	Non-depressed N=317
Characteristic	%	%	%	%
Mean age (years)	35.1 ± 11.5	35.6 ± 12.9	33.7 ± 9.6	35.1 ± 11.9
Sex				
Women	60.0	61.4	80.4	60.6
Men	40.0	38.6	19.6	39.4
Marital status				
Married	40.0	53.1	36.9	53.3
Never married	36.7	31.4	28.3	30.3
Other	23.3	15.5	34.8	16.4
Education				
≤ Primary (1-6)	33.3	43.9	52.2	45.7
Secondary (7-12)	46.7	33.9	30.4	34.7
College graduate	20.0	22.2	17.4	19.6
Smoking status				
Never	81.7	86.9	78.3	86.4
Former	5.0	3.8	17.4	10.4
Current	13.3	9.2	4.3	3.2
Alcohol consumption past year				
Never	58.3	56.9	73.9	55.2
Former	31.7	33.5	21.7	34.4
Current	10.0	9.6	4.4	10.4
Khat consumption				
None	76.7	73.1	69.6	72.3
Former	1.7	5.5	2.2	3.8
Current	21.6	21.4	28.2	23.9
Self-reported physical health				
Excellent/very good/good	33.0	58.0	36.9	47.9
Poor/fair	66.0	42.0	63.1	52.1
Self-reported mental health				
Excellent/very good/good	45.0	68.0	34.8	56.2
Poor/fair	55.0	32.0	65.2	43.8

*Mean ± standard deviation (SD)

Table 3.3—Exploratory factor analysis of depressive symptoms using CIDI

	Factor Loadings
Depressive symptoms	**Factor 1**
Anhedonia	**0.894**
Depressed mood	**0.950**
Insomnia or hypersomnia	**0.838**
Fatigue or loss of energy	**0.878**
Significant weight loss when not dieting or weight gain	**0.908**
Worthlessness or excessive or inappropriate guilt	**0.750**
Diminished ability to think or concentrate	**0.951**
Psychomotor agitation or retardation	**0.544**
Recurrent suicidal thoughts	**0.880**
Eigenvalue	**7.95**
% Variance	**84.3**

Table 3.4a—Mean WHH-QOL scores according to CIDI determined major depressive disorder status by domain

Quality of life assessed by Domain	Depressed		Not Depressed		P-value
	Mean score	SD	Mean score	SD	
Physical	46.1	13.0	52.4	13.4	<0.001
Psychological	46.9	18.8	57.6	17.5	<0.001
Social relationships	54.4	24.9	65.2	22.1	<0.001
Environmental	35.7	15.0	43.6	15.6	<0.001

Table 3.4b—Sex specific mean WHO-QOL scores according to CIDI determined major depressive disorder status by domain

Quality of life	Women					Men				
	Depressed		**Not Depressed**		P-value	**Depressed**		**Not Depressed**		P-value
	Mean score	SD	Mean score	SD		Mean score	SD	Mean score	SD	
Physical	45.6	13.4	52.2	13.4	<0.001	47.3	11.6	52.8	13.5	0.004
Psychological	46.1	18.0	57.4	17.6	<0.001	48.6	20.3	57.8	17.2	<0.001
Social relationships	52.9	26.2	66.1	21.8	<0.001	57.4	22.0	63.8	22.6	0.045
Environmental	35.4	15.8	43.5	15.5	<0.001	36.4	15.1	43.6	15.8	0.001

Table 3.5—Sensitivity and Specificity for the detection of major depressive disorder of CIDI by sex

	Sensitivity	Specificity	LR+	LR-	PPV	NPV-
All subjects						
Naïve	43.5 (28.9-58.9)	77.9(72.9-82.4)	2.0(1.3-2.9)	0.7(0.6-0.9)	22.2(14.1-32.2)	90.5(86.4-93.7)
Adjusted	**51.0 (41.3-60.6)**	**72.2 (68.7-75.6)**	**1.8(1.4-2.3)**	**0.7(0.5-0.8)**	**22.4(17.3-28.0)**	**90.5(88.3-92.4)**
Women						
Naïve	45.9 (31.9-56.0)	77.1(70.5-82.8)	2.0(1.3-3.1)	0.7(0.5-0.9)	27.9(17.1-40.8)	88.1(82.2-92.6)
Adjusted	**56.1 (45.3-66.3)**	**69.3 (64.6-73.3)**	**1.8(1.4-2.3)**	**0.6(0.5-0.8)**	**27.9(21.6-35.2)**	**88.2(84.1-91.4)**
Men						
Naïve	33.1(7.5-70.1)	79.2(71.0-85.9)	1.6(0.6-4.3)	0.8(0.5-1.3)	10.3(2.2-27.4)	94.3(88.0-97.9)
Adjusted	**45.0 (26.0-65.8)**	**69.6(63.6-75.0)**	**1.5(0.9-2.5)**	**0.8(0.5-1.2)**	**10.6(5.7-18.9)**	**94.1(89.7-96.6)**

*Naive estimates are using missing completely random assumption and adjusted estimates are Begg and Greenes estimates; **LR+: positive likelihood ratio, LR-: negative likelihood ratio, PPV: positive predictive value, NPV: negative predictive value

Table 3.6—Questions used in the CIDI depression module

STEM Questions

1. Have you ever in your life had a period lasting several days or longer when most of the day you felt sad, empty or depressed?

2. Have you ever had a period lasting several days or longer when most of the day you were very discouraged about how things were going in your life?

3. Have you ever had a period lasting several days or longer when you lost interest in most things you usually enjoy like work, hobbies, and personal relationships?

Major Depressive Disorder algorithm

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

Part 1 AND Part 2.

Part 1. Symptoms have been present during the same 2 week period and at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

Part 2. At least five of the following symptoms must be present and represent a change from previous functioning:

Note: "change from previous functioning" is implicit in the item corresponding to each symptom (e.g. "more than usual", "less than usual").

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report(e.g., feels sad or empty) or observation made by others.

2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day(as indicated by either subjective account or observation made by others)

3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day.

4. Insomnia or hypersomnia nearly every day.

5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).

6. Fatigue or loss of energy nearly every day.

7. Feelings of worthlessness or excessive or inappropriate guilt(which may be delusional) nearly every day(not merely self-reproach or guilt about being sick)

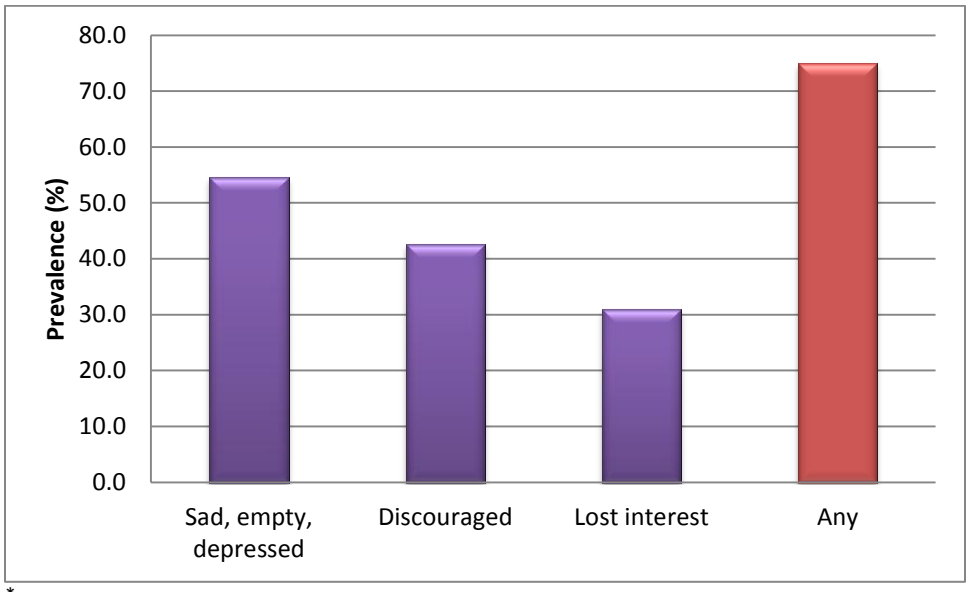
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day(either by subjective account or as observed by others)

9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

Table 3.7—Results from hypothetical CIDI screening program with 100 patients, prevalence of depression 15%

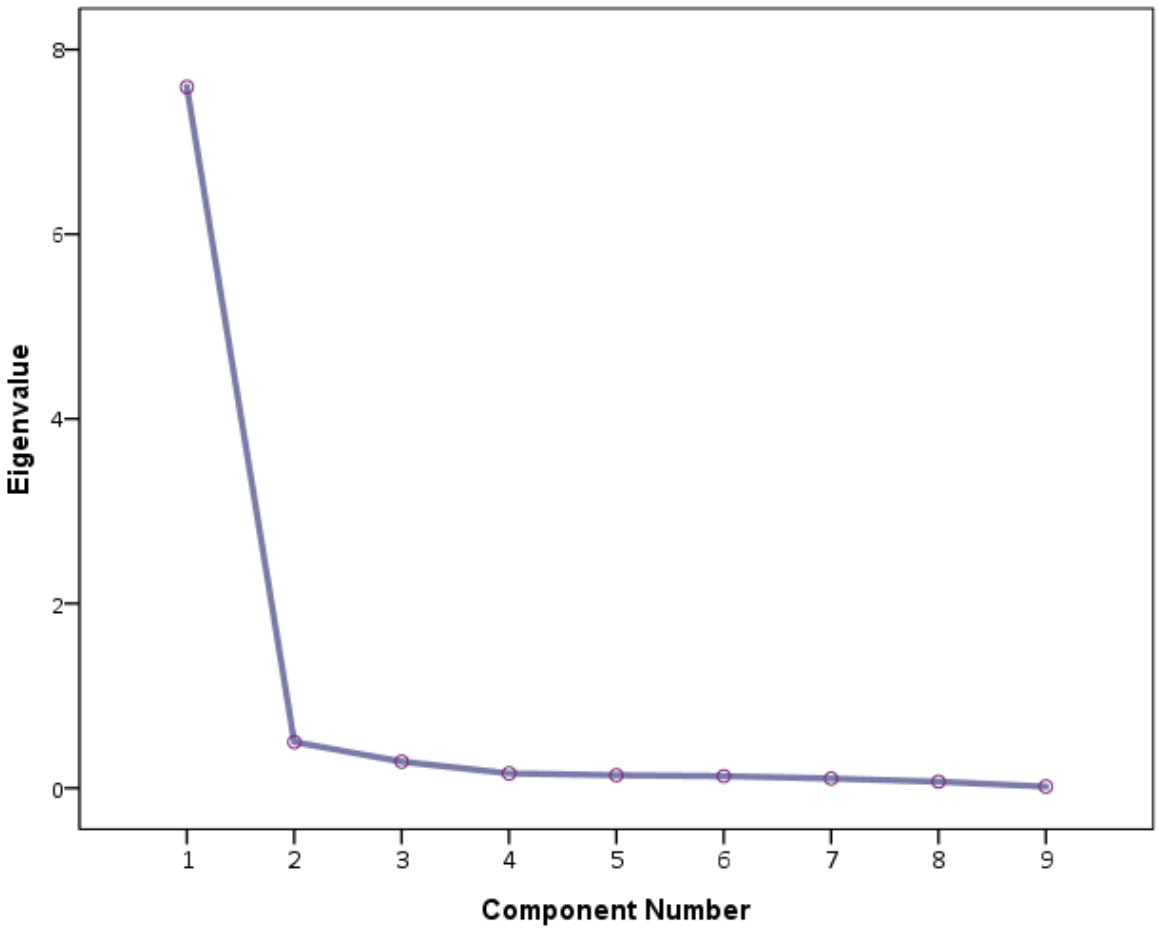
	SCAN Diagnosis	
	True positive	**True negative**
CIDI positive	8	24
CIDI negative	7	61
Total	15	85

Figure 3.1—Depressive symptoms endorsed using CIDI STEM questions for depression



*STEM questions are core questions used at the beginning of CIDI interviews

Figure 3.2—Scree plot of CIDI depressive symptoms



Chapter 4

Assessing Validity of a Depression Screening Instrument in the Absence of a Gold Standard

ABSTRACT

Background and Objective: Screening for mental health problems such as depression in clinical and community based settings is an important activity in an overall public health disease prevention and health promotion strategy. The success of screening, however, is largely dependent on the accuracy of the diagnostic procedure that is the gold standard used to evaluate the effectiveness of the screening instrument. When the gold standard is subject to error, the perceived accuracy of the screening test can be affected. In this study, we evaluated the extent to which use of a hypothesized imperfect gold standard, the Composite International Diagnostic Interview (CIDI), biases the estimates of diagnostic accuracy of the Patient Health Questionnaire-9 (PHQ-9) depression screening instrument. We also evaluated how statistical correction methods can be used to address this bias.

Methods: The study was conducted among 926 adults attending an outpatient department in a major referral hospital in Ethiopia. Structured interviews were conducted to collect information about participants' current major depressive disorder using PHQ-9 and CIDI instruments. First, we evaluated the relative psychometric properties of PHQ-9 using CIDI as a gold standard. Next, we employed a Bayesian latent class model analysis to correct for bias resulting from using an imperfect gold standard.

Results: In comparison with CIDI, the relative sensitivity and specificity of the PHQ-9 for detecting major depressive disorder at a cut point of ≥9 were 61.7% (95% confidence interval (CI): 54.2-69.2%) and 73.8% (95%CI: 70.6-76.9%), respectively. Correcting for bias resulting from imperfect gold standard, using Bayesian latent class modeling analysis, resulted in a modest increase in sensitivity and specificity of PHQ-9 82.5% (95% Bayesian credible interval (BCI): 79.9-

85.3%) and 78.3% (95% BCI: 62.0-89.4%), respectively. Overall, the psychometric properties calculated using Bayesian modeling approaches were closer to the values obtained using a true gold or criterion standard.

Conclusion: Our results provided evidence that assessing diagnostic validity of mental health screening instrument in resource limited settings, where application of a gold standard might not be available, can be accomplished by using appropriate statistical methods.

INTRODUCTION

Screening for mental health problems in clinical and community based settings is an important component in the overall public health disease prevention and health promotion strategy. The success of screening procedures, however, is largely dependent on the accuracy of the diagnostic procedure and the gold standard, or criterion standard, used to evaluate screening instruments [1-3]. Psychometric properties such as sensitivity and specificity are common measures used to evaluate the quality of a screening test. These psychometric properties are unbiased if the screening tests results are compared with a gold standard measure [2, 3]. We recognize that there is no perfect gold standard. Hence, we use the term gold standard to refer to the best available method that determines the presence or absence of the condition or disease of interest. However, verification of the true status using the gold standard may be impossible to obtain due to cost and human resources. In some instances the gold standard may be invasive, impractical to obtain or unethical to conduct. For example, gold standard diagnosis of Alzheimer's disease cannot be ascertained until a patient dies and an autopsy is performed. In epidemiologic studies where a comparison with a gold standard is not possible, validation studies often compare screening instruments with instruments that are imperfect but more precise than the screening instrument. The key assumption is the measurement error for the imperfect reference or gold standard is unlikely to be correlated with the screening instrument [4]. If an imperfect standard is used as if it were a gold standard, the estimated accuracy of the tests would be biased due to misclassification [3, 5] . Zhou *et al* [3] call this type of bias "imperfect gold standard bias". A number of authors have proposed model-based

estimates or estimates that make use of prior information, to reduce or correct this imperfect gold standard bias without retesting any subjects [1].

The Patient Health Questionnaire-9 (PHQ-9) is a very brief, easy to administer and interpret depression screening instrument [6]. Because of its brevity, the PHQ-9 is widely used as a depression screening instrument in primary care settings among racially and ethnically diverse populations. Additionally, it has been reported to be a valuable tool for the detection and management of depression [6]. Use of the PHQ-9 instrument in new clinical and research settings requires evaluating the validity of the instrument in comparison with a diagnostic 'gold standard'. The Schedules for Clinical Assessment in Neuropsychiatry (SCAN), a semi structured clinical interview widely considered as gold standard, is used to assess and diagnose psychiatric disorders including depression among adults [7]. The instrument offers flexibility for clinicians to phrase questions about particular symptoms taking into account local context. However, it requires that clinicians make their clinical decisions following the definitions and criteria provided in the Diagnostic and Statistical Manual (DSM) [8, 9]. Although the SCAN has been reported to have excellent accuracy in depression diagnosis, it is time-demanding, expensive and requires a trained clinician, thus limiting its use in resource limited clinical settings [10].

To address these limitations, alternative diagnostic tools for the measurement of depression have been developed. One of these tools is the Composite International Diagnostic Interview (CIDI). The CIDI is a fully structured lay-administered diagnostic interview that can be used to diagnose major depressive disorder according to DSM criteria, though it is regarded as a less

optimal gold standard compared to the SCAN [11, 12]. In an effort to employ a less burdensome assessment, some investigators have used the CIDI as the criterion standard to evaluate the diagnostic accuracy of the PHQ-9 [13, 14]. Use of CIDI as gold standard (an imperfect but more precise than PHQ-9) is expected to bias the validity of the PHQ-9. However, no study has systematically evaluated the impact of using the CIDI as a gold standard. Therefore, we conducted this study to assess the extent to which the use of what we hypothesized as an imperfect gold standard [3] biases the estimates of diagnostic accuracy of a screening instrument (the PHQ-9). We further sought to demonstrate how statistical methods can be used to correct the bias and improve psychometric properties of the PHQ-9.

METHODS AND MATERIALS

Participants

The study was conducted at Saint Paul General Specialized Hospital, a major referral and teaching hospital, in Addis Ababa, Ethiopia between the months of July and December, 2011. We used a two-stage study design where a total of 926 adults (≥ 18 years of age) attending outpatient departments were first interviewed by research nurses using the PHQ-9 and CIDI depression scales. Then those who screened positive for depression on the PHQ-9 questionnaire (PHQ-9 score ≥10), as well as a randomly selected sub-group of participants who screened negative for depression (PHQ-9 score ≤ 10), were interviewed using the SCAN diagnostic questionnaire by licensed psychiatrists who were blind to PHQ-9 results. The questionnaire was originally written in English and then translated into Amharic. The translated version was back-translated and modified until the back-translated version was comparable

with the original English version. All study participants provided informed consent, and all research protocols were approved by the institutional review boards of Addis Continental Institute of Public Health, Addis Ababa, Ethiopia and the Human Subjects Division at the University of Washington, Seattle, WA.

Study Measurements

Patient Health Questionnaire-9 (PHQ-9)

All study participants were evaluated for depressive symptoms using the PHQ-9 depression scale [6]. The PHQ-9 queries participants about the frequency with which they experienced nine depressive symptoms. Each item requires participants to rate the frequency of a depressive symptom experienced in the two weeks prior to evaluation. The items include: 1) anhedonia, 2) depressed mood, 3) insomnia or hypersomnia, 4) fatigue or loss of energy, 5) appetite disturbances, 6) guilt or worthlessness, 7) diminished ability to think or concentrate, 8) psychomotor agitation or retardation, and 9) suicidal thoughts. Scores for each item range from 0 (“not at all”) to 3 (“nearly every day”) with a total score ranging from 0 to 27 [6, 15, 16]

Composite International Diagnostic Interview (CIDI version 3.0)

Following PHQ-9 administration, participants were interviewed using the CIDI instrument. The CIDI is a fully structured diagnostic interview designed to be administered by lay interviewers was developed by the World Health Organization (WHO) to estimate the prevalence and identify correlates of mental disorders from large numbers of subjects across different countries. We used the depression and mania modules of the CIDI 3.0 to diagnose major

depressive disorder (MDD). The mania module was used to rule out history of bipolar disorders (manic, mixed or hypomanic episodes). The CIDI interview includes three screening questions about sadness/depressed mood, feelings of discouragement, and loss of interest lasting several days or longer. Participants who endorsed any of these three questions were given the depression module. Those who failed to endorse any of the three core questions were skipped out of the depression module. In accordance with DSM-IV criteria, we defined MDD as the presence of five out of nine cardinal symptoms that persist for two weeks or longer, are present for most of the day nearly every day, and cause significant distress or impairment. These symptoms include dysphoric mood or anhedonia that persist most of the day, and clinically significant weight gain/loss or appetite disturbance, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue or loss of energy, feelings of worthlessness or excessive guilt, diminished ability to concentrate or think clearly, and recurrent thoughts of death or suicide. For the purposes of this validation study, we defined current MDD as experiencing MDD in the past 12 months without plausible organic causes and without history of mania or hypomania [17, 18]. In the current study we considered the CIDI interview as an imperfect reference (or gold) standard [11, 12].

Schedules for Clinical Assessment in Neuropsychiatry (SCAN) Diagnosis

The SCAN is a semi structured clinical interview developed by the WHO for use by trained clinicians to assess and diagnose psychiatric disorders among adults [7, 19]. The use of the SCAN gives flexibility in the diagnosis of mental disorders based on ICD-10 and DSM-IV [19]. The depressed mood and ideation module of SCAN has been reported to work well in different

languages and cultures including Ethiopia [20]. For the current study, we used SCAN administered in Amharic as a perfect reference (or gold) standard. All SCAN interviews (and diagnoses) were conducted without knowledge of the results of initial PHQ9 and CIDI interviews.

Statistical Analysis

We first computed the psychometric properties of PHQ-9 using the CIDI as a reference or gold standard. Considering the CIDI as a reference standard subject to error, or imperfect gold standard, we used the terms "relative sensitivity" and "relative specificity" to report test characteristics of the PHQ-9 [2]. Relative sensitivity is the percent of time that the PHQ-9 detected depression when the CIDI detected it. Similarly, relative specificity is the percent of time that the PHQ-9 did not detect depression when the CIDI did not detect it. We used a Bayesian diagnostic test modeling approach in the absence of a gold standard developed by Branscum *et at* [21]. The Bayesian approach combines the likelihood of the data and an assumed prior distribution of parameters through Bayes' theorem to generate a posterior distribution of the parameters of interest [3]. We followed this approach because i) prior scientific information about the psychometric properties of PHQ-9 can be incorporated in our estimation, ii) the ability to use a publicly available open source WinBUGS software and iii) a freely available code that can be readily used in our study setting. Briefly, prior uncertainties about the true values of the parameters assessed here (sensitivity and specificity) were modeled using prior probability distributions. The first important step in Bayesian analysis is to obtain a prior distribution over all model parameters using past data, if available, or by drawing

upon expert knowledge, or a combination of both[22]. As pointed by Greenland [23], priors may vary considerably across individuals; a reasonable approach should reflect results from prior studies or reviews. Since there was no prior study conducted in the study setting, we chose priors based a recent diagnostic meta-analysis conducted by Manea *et al* [24]. We reasoned this will provide a realistic prior distribution than an expert opinion. In their study, Manea and colleagues summarized the psychometric properties of the PHQ-9 across 18 validation studies (n=7,180) conducted in various clinical setting around the world. The authors provided psychometric prosperities of PHQ-9 for detecting MDD for cut-off scores between 8 and 11. For instance, the prior values for sensitivity and specificity for diagnosing MDD using a cut-off score of 10 and above have modal values of 85% and 89%, respectively with 95% of their areas above 75%. These prior values were then used to generate the corresponding Beta distributions using the BetaBuster software (http://www.epi.ucdavis.edu./diagnostictests/betabuster.html). The prior densities were assumed to be independent beta distributions [25]. Then, a WinBUGS code that utilizes Bayesian latent class modeling approach was used in these analyses [21] (code available at http://www.epi.ucdavis.edu/diagnostictests). Because both the PHQ-9 and CIDI are based on DSM-IV criteria and are measuring the same psychological phenomenon (depression) and symptoms, the results are expected to be correlated [26]. Hence, a Bayesian model to assess test accuracy for two conditionally dependent tests was used [21] to estimate the sensitivity and specificity of the instruments. The model was checked for convergence using two chains with different initial values. Mixing of the chains was assessed by the lack of autocorrelation among the sampled values. The hyperparameters for the prior distributions were intentionally

specified to lead to vague priors that accommodate the large variability expected in the field for the diagnostic measures. Monte Carlo methods (a class of computational algorithms that rely on repeated random sampling to compute values) were used to estimate the lower and upper limits of 95% Bayesian credible intervals (95% BCI). All parameters and associated 95% BCIs were computed using the software WinBUGS [27]. The posterior distributions were computed based on 5,000 iterations after discarding the initial 500 iterations as burn-in.

RESULTS

The median PHQ-9 score was 5 (range 0-27). A total of 258 participants fulfilled DSM-IV criteria for MDD on the PHQ-9 (27.8%; 95% CI 24.9-30.7%) using a score of $\geq$ 10, while 668 were classified as non-depressed (**Table 4.1**). Using CIDI as a reference standard, a total of 86 subjects were classified as depressed and 564 as non-depressed on both instruments. Distributions for other PHQ-9 score cut-offs are also displayed in **Table 4.1**. As shown in **Figure 4.1**, PHQ-9 scores were higher among individuals classified as depressed on CIDI (median=10; 1^{st} quartile=6; 3^{rd} quartile=14; interquartile range =8, while for non-depressed participants the corresponding values were median=4; 1^{st} quartile =2; 3^{rd} quartile =9; interquartile range =7).

Relative psychometric properties of PHQ-9 using CIDI as a gold standard are shown in **Table 4.2**. In comparison with CIDI the relative sensitivity and specificity of PHQ-9 for a cut point of $\geq$9 were 61.7% (95%CI: 54.2-69.2%) and 73.8% (95%CI: 70.6-76.9%), respectively. As expected, lower sensitivities and improved specificities were noted for increased PHQ-9 score cut points. For instance the sensitivity and specificity of the PHQ-9 for MDD diagnosis at a cut point of $\geq$10

were 53.1% (95%CI: 45.4-60.8%) and 77.5% (95%CI: 74.5-80.5%), respectively. After applying a Bayesian latent class modeling analysis, the sensitivity and specificity of PHQ-9 for a cut point of ≥9 increased to 82.5% (95% BCI: 79.9-85.3%) and 78.3% (95% BCI: 62.0-89.4%), respectively (**Table 4.2**). Using a cut point of ≥10 the sensitivity and specificity were estimated to be 80.7% (95% BCI: 78.2-83.1%) and 84.4% (95% BCI: 73.5-90.2%), respectively. The psychometric properties remained stable for PHQ-9 cut-point of ≥11 and different prior prevalence assumptions. Based on the Youden Index, the diagnostic accuracy of the PHQ-9 was optimized at a cutoff of 10 or more (Youden index=0.65).

For comparison psychometric properties from our validation study of the PHQ-9 that used psychiatrist administered SCAN diagnosis (gold standard) are also presented in **Table 4.2** (submitted manuscript). For instance, a score of ≥ 10 on the PHQ-9 is associated with 71 % sensitivity (61.2-83.9) and 77% specificity (74.8-79.5) in diagnosing MDD compared with psychiatrist administered SCAN diagnosis.

DISCUSSION

Overall, improvements in psychometric properties of the PHQ-9 (when compared against CIDI, an imperfect gold standard) were noted after employing Bayesian latent class modeling approaches compared to the relative estimates. Overall, sensitivity and specificity estimates calculated using Bayesian modeling approaches were closer to the values obtained using the SCAN gold standard (regarded the true or perfect gold standard in this study).

Brief, reliable and valid screening instruments help health care professionals determine the magnitude of mental health problems such as depression in populations. As noted earlier, because of its brevity, the PHQ-9 [6, 16] has gained increased recognition as a preferred instrument for depression screening in primary care settings among racially and ethnically diverse populations [28, 29]. Use of screening and diagnostic instruments, such as the PHQ-9, in new settings requires evaluating their psychometric properties in comparison with a gold or criterion standard. In many developing countries such evaluations of the diagnostic validity of instruments remain a main challenge due to cost, time, and human resources limitation of using the gold or reference standard. One strategy to overcome this problem will be using structured interviews that do not require a specially trained clinician to administer them. Indeed, some investigators have used the CIDI as a gold standard to evaluate the diagnostic accuracy of the PHQ-9 [13]. Use of the CIDI as a gold standard results in errors due to the fact that it is recognized as an imperfect gold standard [11, 12]. To our knowledge, this is the first study to examine the impact of using an imperfect gold standard (CIDI interview to determine the validity of the PHQ-9 and to use statistical methods to correct for the bias. After applying Bayesian statistical modeling, we note the psychometric proprieties of the PHQ-9 were somewhat higher to what we would have obtained if we used the SCAN gold standard.

Some caveats should be considered when interpreting the results of our study. Due to the self-report nature of psychological conditions, it is possible that the SCAN clinical diagnosis may not be the best gold standard or reference test. Furthermore, although we used prior estimates

based on a previously conducted meta-analysis, one might arrive at slightly different estimates using different priors. As stated by Zou *et al* [3], estimates can be greatly affected by a change in prior distribution regardless of how large the sample size is.

In conclusion, our findings underscore the importance of evaluating the feasibility and validity of using an imperfect gold standard. In addition, using Bayesian statistical approaches in the absence of a gold standard may result in improvements of psychometric properties of mental health screening instruments. This analytical approach may thus represent an important tool for adjusting the bias that might result from the use of an imperfect gold standard.

REFERENCES

1. Bouma A, Stegeman JA, Engel B, de Kluijver EP, Elbers AR, De Jong MC: **Evaluation of diagnostic tests for the detection of classical swine fever in the field without a gold standard**. *J Vet Diagn Invest* 2001, **13**(5):383-388.
2. Pepe M: **The statistical evaluation of medical tests for classification and prediction**. Oxford: Oxford University Press; 2003.
3. Zhou XH, McClish DK, Obuchowski NA: **Statistical methods in diagnostic medicine**, 2nd edn. Oxford: Wiley-Blackwell; 2011.
4. White E, Armstrong BK, Saracci R, Armstrong BK: **Principles of exposure measurement in epidemiology : collecting, evaluating, and improving measures of disease risk factors**. Oxford; New York: Oxford University Press; 2008.
5. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM: **Evaluation of diagnostic tests when there is no gold standard. A review of methods**. *Health Technol Assess* 2007, **11**(50):iii, ix-51.
6. Kroenke K, Spitzer RL, Williams JB: **The PHQ-9: validity of a brief depression severity measure**. *J Gen Intern Med* 2001, **16**(9):606-613.
7. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N: **SCAN. Schedules for Clinical Assessment in Neuropsychiatry**. *Arch Gen Psychiatry* 1990, **47**(6):589-593.
8. Janca A, Ustun TB, Sartorius N: **New versions of World Health Organization instruments for the assessment of mental disorders**. *Acta Psychiatr Scand* 1994, **90**(2):73-83.

9. Andrews G, Peters L, Guzman AM, Bird K: **A comparison of two structured diagnostic interviews: CIDI and SCAN**. *Aust N Z J Psychiatry* 1995, **29**(1):124-132.

10. Hirsch SR, Weinberger DR: **Schizophrenia**. Oxford; Cambridge, Mass., USA: Blackwell Science; 1995.

11. Brugha TS, Jenkins R, Taub N, Meltzer H, Bebbington PE: **A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN)**. *Psychol Med* 2001, **31**(6):1001-1013.

12. Wittchen HU: **Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review**. *J Psychiatr Res* 1994, **28**(1):57-84.

13. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, Falloon K, Hatcher S: **Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population**. *Ann Fam Med* 2010, **8**(4):348-353.

14. Sherina MS, Arroll B, Goodyear-Smith F: **Criterion validity of the PHQ-9 (Malay version) in a primary care clinic in Malaysia**. *Med Jf Malaysia* 2012, **67**(3):309-315.

15. **American Psychiatric Association. Diagnostic and statistical manual of mental disorders DSM-IV** [http://www.psychiatryonline.com/DSMPDF/dsm-iv.pdf]

16. Spitzer RL, Kroenke K, Williams JB: **Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire**. *JAMA* 1999, **282**(18):1737-1744.

17. Kessler RC, Abelson J, Demler O, Escobar JI, Gibbon M, Guyer ME, Howes MJ, Jin R, Vega WA, Walters EE *et al*: **Clinical calibration of DSM-IV diagnoses in the World Mental**

Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). *Int J Methods Psychiatr Res* 2004, **13**(2):122-139.

18. Kessler RC, Ustun TB: **The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI)**. *Int J Methods Psychiatr Res* 2004, **13**(2):93-121.

19. Aboraya A, Tien A, Stevenson J, Crosby K: **Schedules for Clinical Assessment in Neuropsychiatry (SCAN): introduction to WV's mental health community**. *W V Med J* 1998, **94**(6):326-328.

20. Alem A, Kebede D, Shibre T, Negash A, Deyassa N: **Comparison of computer assisted scan diagnoses and clinical diagnoses of major mental disorders in Butajira, rural Ethiopia**. *Ethiop Med J* 2004, **42**(2):137-143.

21. Branscum AJ, Gardner IA, Johnson WO: **Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling**. *Prev Vet Med* 2005, **68**(2-4):145-163.

22. Joseph L, Gyorkos TW, Coupal L: **Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard**. *Am J Epidemiol* 1995, **141**(3):263-272.

23. Greenland S: **Bayesian perspectives for epidemiological research: I. Foundations and basic methods**. *Int J Epidemiol* 2006, **35**(3):765-775.

24. Manea L, Gilbody S, McMillan D: **Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis**. *CMAJ* 2012, **184**(3):E191-196.

25. Johnson N, Kotz S: **Continuous univariate distributions**. New York, NY: J. Wiley & Sons; 1970.

26. Dendukuri N, Joseph L: **Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests**. *Biometrics* 2001, **57**(1):158-167.

27. Lunn DJ, Thomas A, Best N, Spiegelhalter D: **WinBUGS \– A Bayesian modelling framework: Concepts, structure, and extensibility**. *Statistics and Computing* 2000, **10**(4):325-337.

28. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL: **Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients**. *J Gen Intern Med* 2006, **21**(6):547-552.

29. Lichtman JH, Bigger JT, Jr., Blumenthal JA, Frasure-Smith N, Kaufmann PG, Lesperance F, Mark DB, Sheps DS, Taylor CB, Froelicher ES: **Depression and coronary heart disease: recommendations for screening, referral, and treatment: a science advisory from the American Heart Association Prevention Committee of the Council on Cardiovascular Nursing, Council on Clinical Cardiology, Council on Epidemiology and Prevention, and Interdisciplinary Council on Quality of Care and Outcomes Research: endorsed by the American Psychiatric Association**. *Circulation* 2008, **118**(17):1768-1775.

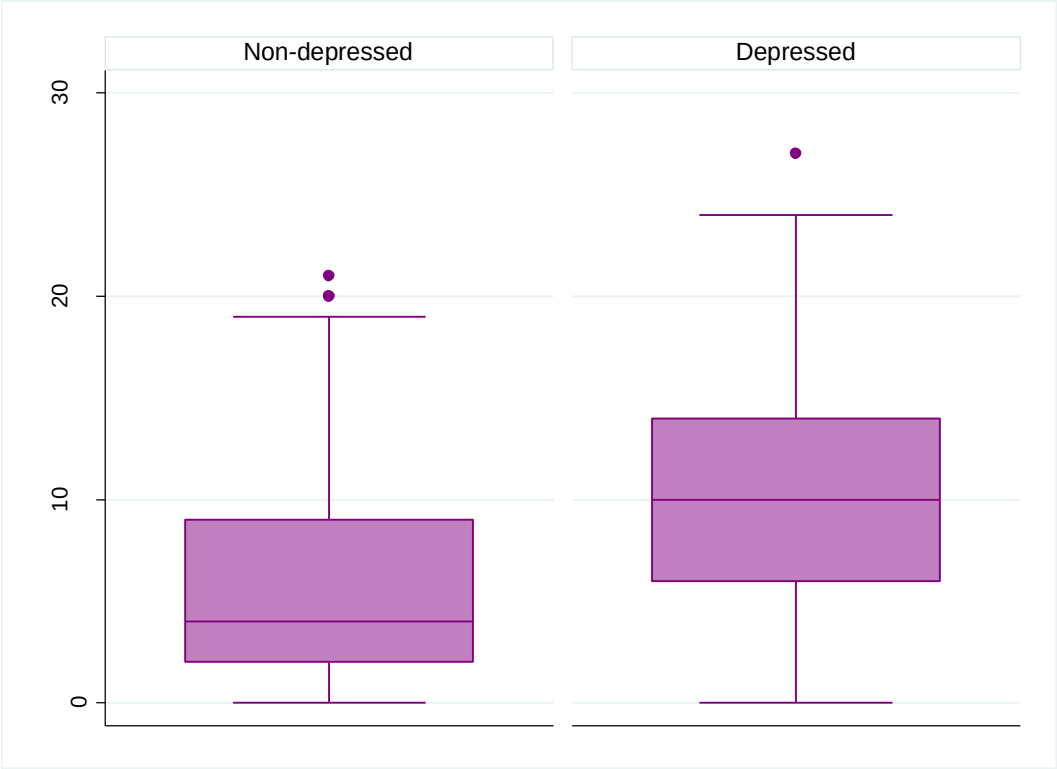
Table 4.1—Distribution of Patient Health Questionnaire-9(PHQ-9) and Composite International Diagnostic Interview (CIDI) results

	PHQ-9 Score cut-off		
	9	**10**	**11**
CIDI positive			
PHQ-9 positive	100	86	76
PHQ-9 negative	62	76	86
CIDI positive			
PHQ-9 positive	200	172	146
PHQ-9 negative	564	592	618
Total	926	926	926

Table 4.2—Relative and adjusted sensitivity and specificity values and associated 95% confidence intervals (CIs) for major depressive disorder diagnosis across various cut-off points of the PHQ-9 using Bayesian approach

	Sensitivity	Specificity	PPV	NPV
Relative estimates				
Score≥9	61.7 (54.2-69.2)	73.8 (70.6-76.9)	33.3	90.1
Score≥10	53.1 (45.4-60.8)	77.5 (74.5-80.5)	33.3	88.6
Score≥11	46.9 (39.2-54.6)	80.9 (78.1-83.7)	34.2	87.8
Bayesian estimates				
Score≥9	82.5(79.9-85.3)	78.3 (62.0-89.4)	26.9	97.7
Score≥10	80.7 (78.2-83.1)	84.4 (73.5-90.2)	33.8	97.4
Score≥11	74.3 (76.9-79.6)	88.7 (85.4-92.0)	39.4	96.5
Using SCAN gold standard				
Score≥9	79.0 (68.2-89.1)	72.0 (68.5-74.9)	23.0	96.4
Score≥10	71.1 (61.2-83.9)	76.6 (74.8-79.5)	23.5	94.5
Score≥11	57.5 (41.6-72.0)	79.8 (78.0-81.5)	33.3	90.1

Figure 4.1—Distribution of PHQ-9 scores according to major depressive disorder status



Box plots comparing PHQ-9 scores among those classified as having major depressive disorder (right side) and those without depression (left side) based on the CIDI depression module. The central box shows the data between the upper and lower quartiles, with the median represented by the middle line. The "whiskers" (lines on either side of central box) extend from the upper and lower quartiles to a distance of 1.5xIQR (interquartile range) away or the most extreme data point within that range, whichever is smaller.

Chapter 5

Major Depressive Disorder and Cardiometabolic Disease Risk among Sub-Saharan African Adults

ABSTRACT

Background: Research conducted in developed countries has documented an association between major depressive disorder (MDD) and cardiometabolic disease risk, but this association has not been studied in sub-Saharan Africa where both morbidities are increasing dramatically.

Objective: We sought to evaluate the extent to which MDD is associated with cardiometabolic diseases and risk factors among a cohort of urban dwelling Ethiopian adults.

Methods: This was a cross-sectional epidemiologic study of 1,924 employed adults (1,165 men and 759 women) in Addis Ababa, Ethiopia. The World Health Organization STEPwise approach was used to collect sociodemographic data, anthropometric measurements, and blood samples among study subjects. Fasting blood glucose (FBG), insulin, C-reactive protein, and lipid concentrations were measured using standard approaches. Metabolic syndrome was defined according to the International Diabetes Federation (IDF) criteria. Insulin resistance was assessed using the homeostatic model assessment (HOMA-IR). We ascertained the presence of MDD using the validated screening questionnaire, the Patient Health Questionnaire-9 (PHQ-9) depression scale. Multivariate logistic regression models were fitted to estimate odds ratios (OR) and 95% confidence intervals (95% CI) for the association between MDD and cardiometabolic disease risk while adjusting for potential confounders.

Results: A total of 154 participants screened positive for MDD on PHQ-9 (8.0%; 95% CI 6.7-9.2%). Among women, MDD was associated with more than 4-fold increased odds of diabetes (OR=4.14; 95% CI: 1.03-16.62). Among men the association was not significant (OR=1.12; 95% CI: 0.63-1.99). The odds for insulin resistance among women and men were (OR=2.89; 95% CI:

0.73-11.35) and (OR=0.66; 95% CI: 0.08-11.35), respectively. After controlling for confounders, there was no clear evidence for a positive association between MDD and hypertension status. Similarly, MDD was not associated with metabolic syndrome among women (OR=1.51; 95% CI: 0.69-3.29) and men (OR=0.61; 95% CI: 0.28-1.34). Lastly, MDD was not associated with increased odds of systemic inflammation as measured by C-reactive protein among women (OR=0.87; 95% CI: 0.12-1.49) or men (OR=1.51: 95% CI: 0.50-4.49).

Conclusion: The results of our study do not provide convincing evidence that MDD is associated with cardiometabolic diseases among Ethiopian adults. Future studies need to evaluate the effect of other psychiatric disorders on cardiometabolic disease risk.

INTRODUCTION

The global prevalence of cardiometabolic risk such as cardiovascular disease (CVD), diabetes and obesity is increasing at an alarming rate, with the majority of cases occurring in low and middle income countries [1]. A growing body of epidemiologic evidence shows that incidence of cardiometabolic diseases are increasing in sub-Saharan Africa [2-6]. Kearney at *al.* reported that in the year 2000 an estimated 639 million individuals had hypertension in low and middle income countries and this number is expected to rise to 1.15 billion by 2025 [7]. In 2006, 10.8 million sub-Saharan Africans were estimated to have diabetes. This number is expected to rise to 18.7 million by 2025 [8]. The rise in cardiometabolic disease prevalence is driven, in part, by significant changes in dietary habits, physical activity levels, and increased stress as a result of increased urbanization and economic development [1]. An expanding body of evidence now implicates unipolar major depressive disorder (MDD) as one of the major risk factors for and conditions co-occurring with cardiometabolic disease [9-11]. Several studies, primarily conducted in developed countries, have documented associations of MDD with cardiometabolic disease including hypertension, diabetes, metabolic syndrome, myocardial infarction, sudden death, and other cardiac events [12-15]. Some investigators, however, have found no significant associations between cardiometabolic disease risk and MDD [16-18]. Reasons for these inconsistent findings are unclear.

Although causal relationships, and biological mechanisms underlying associations of MDD with cardiometabolic diseases have yet to be clearly established, understanding the epidemiological characteristic of these disorders (e.g., assessment of comorbidity) may help inform health

promotion and disease control efforts [11]. For example, investigators have reported that individuals with comorbid diabetes and MDD are more likely than individuals with diabetes alone to have poor glycemic control and consequently to have more severe complications and lower quality of life [11, 19]. Given the increased burden of cardiometabolic disease risk and the available body of evidence documenting the association between MDD and cardiometabolic disease risk, we sought to evaluate the extent to which MDD is associated with cardiometabolic disease risk factors among an epidemiologically well characterized occupational cohort of bankers and teachers residing in Addis Ababa, Ethiopia.

METHODS AND MATERIALS

Design and Participants

This study was conducted in Addis Ababa, Ethiopia, during the months of December 2009 and January 2010. Study participants were permanent employees of the Commercial Bank of Ethiopia and teachers in government and public schools of Addis Ababa. These workplaces were selected based on their high stability of workforce and willingness to participate in the study. Multistage sampling was done by means of probability proportional to size (PPS) sampling [20]. This was performed for both institutions, and all individuals at selected locations were invited to participate. The original study population was comprised of 2,207 individuals. Subjects were excluded due to missing anthropometric information (n=35), pregnancy (n=21), and incomplete laboratory measures (n=227), the final analytical sample included 1,924 (1,165 men and 759 women) participants. Participants who were excluded were similar in sociodemographic and lifestyle characteristics to those who were included in the analysis.

Data Collection and Variable Specification

Each participant was interviewed by a trained interviewer in accordance with the WHO STEPwise approach for non-communicable diseases surveillance in developing countries [21]. The approach had three levels: (1) questionnaire to ascertain demographic and behavioral characteristics, (2) simple physical measurements, and (3) biochemical tests. Some questions were added to supplement the WHO questionnaire reflect on the local context. Questions were also included regarding behavioral risk factors such as tobacco, alcohol, and Khat consumption. Khat is an evergreen plant with amphetamine-like effects commonly used as a mild stimulant for social recreation and to improve work performance in Ethiopia [22, 23]. The modified questionnaire was first written in English and then translated into Amharic by experts and was translated back in to English. The questionnaire was pre-tested before the initiation of the study and contained information regarding socio-demographic characteristics, tobacco and alcohol use, nutritional status, and physical activity. A five-day training of the contents of the STEPs questionnaire, data collection techniques, and ethical conduct of human research was provided to research interviewers prior to the commencement of the study. . Details regarding data collection methods and study procedures have been previously described in detail [5, 24]

Cardiometabolic Disease Risk Factors

Blood pressure was digitally measured (Microlife BP A50, Microlife AG, Switzerland) after individuals had been resting for five minutes. Two additional blood pressure measurements were taken with three minutes elapsing between successive measurements. In accordance with

the WHO recommendation the mean systolic and diastolic BP from the second and third measurements were considered for analyses. For the collection of blood samples, individuals were advised to skip meals for 12 hours. Blood samples of 12 mL were obtained, using proper sanitation and infection prevention techniques. The collected aliquots of blood were used to determine participants' fasting blood glucose (FBG) concentrations and lipid profiles. Serum was used for the measurement of triglycerides (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), glucose concentrations, insulin and C-reactive protein (CRP). These were measured at the International Clinical Laboratories (ICL) in Addis Ababa, Ethiopia. ICL is the only clinical laboratory in East Africa accredited by the Joint Commission International of USA. TG concentrations were determined by standardized enzymatic procedures using glycerol phosphate oxidase assay. HDL-C was measured using the Ultra HDL assay which is a homogeneous method for directly measuring HDL-C concentrations in serum or plasma without the need for off-line pretreatment or centrifugation steps. Participants' FBG was determined using the standardized glucose oxidase method. Serum CRP concentrations were measured by an ultrasensitive competitive immunoassays. All laboratory assays were completed without knowledge of participants' medical history. Lipid, lipoprotein and FBG concentrations were reported as mg/dL and CRP as mg/L.

Height and weight were measured with participants wearing light clothing and no shoes[21]. Waist circumference measurements were performed with a fixed tension tape, at the midpoint between the lower margin of the last palpable rib and the top of the iliac crest. This was done in

a private place over light clothing. Hip circumference measurements were conducted in a similar manner, at the point of the maximum circumference of the buttocks.

Analytical Variable Specification

Metabolic Syndrome

We defined abdominal obesity using the International Diabetes Federation (IDF) criteria [25] where those having a waist circumference of ≥94 cm for men and ≥80 cm for women. Low HDL-C was defined to be <40 mg/dL in men and <50 mg/dL in women. We defined elevated blood pressure as having systolic blood pressure (SBP) of ≥135 mmHg or a diastolic blood pressure (DBP) ≥85 mmHg. Impaired fasting glucose was defined to be ≥100 mg/dL (5.6 mmol/L) or with a previous history of diabetes. Elevated TG was defined as ≥150 mg/dL. Metabolic syndrome was defined in accordance with the IDF as presence of abdominal obesity and presence of two or more metabolic syndrome components described above [25].

According to the definitions of the American Heart Association and the National Cholesterol Education Program [26] we grouped fasting blood glucose in to normal (<100 mg/dL), impaired fasting glucose (100-125 mg/dL), and diabetes (≥126 mg/dL or a previous history of diabetes or currently on medication). LDL concentrations were classified as: optimal (<100 mg/dL); near or above optimal (100-129 mg/dL); and high (≥130 mg/dL). Total cholesterol concentrations were classified as: desirable (<200 mg/dL), borderline high (200-239 mg/dL), and high (≥ 240 mg/dL). HDL concentrations levels were grouped as: low (<40 mg/dL), normal, (40-59 mg/dL), and high

(≥60 mg/dL). We grouped triglyceride concentrations as: desirable (<200 mg/dL), borderline high (200-239 mg/dL), and high (≥ 240 mg/dL).

Insulin Resistance Syndrome and C-reactive Protein

Furthermore, we implemented a nested, case–control design based on metabolic syndrome status within the entire cohort to assess the extent to which insulin resistance syndrome (a condition in which peripheral tissues become nonresponsive to the effects of insulin) and C-reactive protein (marker of systemic inflammation) are associated with MDD. This was done first by stratifying the study cohort according to the presence or absence of metabolic syndrome. Next, we selected all participants with metabolic syndrome (case group). Then, we randomly selected those without metabolic syndrome (control group). A total of 516 participants (156 metabolic syndrome cases and 360 controls) were selected for this nested case-control analysis. The age and sex distribution of those included in the nested analyses were similar to those not included. As an index of insulin resistance, we determined the Homeostasis Model Assessment of Insulin Resistance (HOMA-IR) index by dividing the product of fasting glucose (mg/dl) and fasting insulin (μIU/ml) by 405 [27]. The HOMA-IR is widely used in clinical trials and has been shown to be correlated with the results obtained from the "gold standard" euglycemic glucose clamp method [28]. We evaluated the presence of insulin resistance syndrome based on the upper 10^{th} percentile of HOMA-IR values for normal-weight male and female subjects with normal fasting blood glucose in this study. This is an approach that has been used by other investigators [29]. Categories of CRP were defined using tertiles

(based on the distribution among controls). The groups with the lowest two tertiles were defined as low CRP (reference group) and the group with the highest tertile as high CRP.

Major Depressive Disorder (MDD)

We used the PHQ-9 to classify participants with regard to MDD. Because of its brevity, the PHQ-9 has gained increased recognition as a preferred instrument for depression screenings and diagnosis in research and primary care settings among racially and ethnically diverse populations [30, 31]. The PHQ-9 items queries participants about the frequency of nine depressive symptoms experienced. Scores for each question range from 0 ("not at all") to 3 ("nearly every day"). The PHQ-9 total score is the sum of scores for the nine items for each participant, and ranged from 0-27. In our validation study of PHQ-9 (submitted manuscript), we have shown that in a sample of Ethiopian adults a score of ≥9 on the PHQ-9 is associated with 79% sensitivity and 72% specificity in diagnosing MDD using *the Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) criteria. Therefore, we defined presence of MDD using a score ≥9 on PHQ-9.

Covariates

Participants were classified according to their alcohol consumption habits: nondrinker (< 1 alcoholic beverage a week), moderate (1–21 alcoholic beverages a week), and high to excessive consumption (> 21 alcoholic beverages a week) according to the WHO classification [32]. Other variables were categorized as follows: age (years), sex (male, female), education (≤ high school, technical School, ≥ bachelors), smoking history (current, former, never), and current Khat

consumption (yes, no). Participants were also asked the following question about their self-reported health status: "Would you say your health in general is excellent, very good, good, fair, or poor?" We classified those who reported fair or poor health and those who reported excellent, very good, or good health. All participants provided informed consent, and the research protocol was approved by the Institutional Review Boards of Addis Continental Institute of Public Health, Addis Ababa, Ethiopia, and the Human Subjects Division at the University of Washington, USA.

Statistical Analysis

Subjects' characteristics were summarized using means (± standard deviation) for continuous variables and counts and percentages for categorical variables. For skewed variables median [interquartile range] were provided. Differences in categorical variables were evaluated using Chi-square test. For continuous variables with normal distributions, Student's t tests were used to evaluate differences in mean values by depression status. We used unadjusted and multivariable-adjusted logistic regression models to calculate odds ratios (ORs) and 95% confidence intervals (CIs) of the association between cardiometabolic risk and presence of MDD. Confounding was assessed by entering potential confounders into a logistic model one at a time, and then comparing the unadjusted and adjusted ORs. Final logistic regression models included covariates that altered unadjusted ORs by at least 10% [33]. Finally, we assessed for any evidence of modification in the magnitude of effect in participants with and without cardiometabolic risk by including an interaction term between MDD and potential effect modifiers. Evidence for effect modification was assessed in Likelihood Ratio Tests comparing the goodness of fit of models with and without the interaction term. We considered the

following variables, *a priori*, as potential confounders and/or effect modifiers in these analyses: age, sex, occupation, self-reported health status, physical activity, education, cigarette smoking, Khat use and alcohol consumption [34, 35]. All analyses were performed using STATA 11.0 statistical software for Windows (Statacorp, College Station, TX, USA). All reported p-values are two-sided and deemed statistically significant at α=0.05.

RESULTS

A summary of selected socio-demographic and lifestyle characteristics of study participants is presented in Table 1. A total of 1,924 participants between the ages of 18 and 67 years (mean age=35 years, standard deviation=11 years) participated in the study. The majority of participants were men (60.6%), unmarried (51%) and more likely to have a college diploma, bachelor's degree, or higher education (70.7%). Approximately 7% of participants reported that they were heavy drinkers and 9% reported that they were current smokers. Khat consumption was reported by 8.8% of participants. Approximately 40% of participants reported having a fair or poor health status.

Distributions of the socio-demographic and lifestyle characteristics according to participants' depression status are also presented in Table 1. A total of 154 participants screened positive for MDD on PHQ-9 (8.0%; 95% CI 6.7-9.2). No significant difference in MDD prevalence was noted between men and women. Depressed participants were more likely to be younger, to have a lower level of educational attainment, to be heavy drinkers, and to report a poor health status.

The prevalence of MDD according to age groups is presented in Figure 1. Depression prevalence varied with age with the highest being in early adult life (18-29 years), and declining thereafter.

As shown in Table 2, we calculated the odds ratio for diabetes and hypertension status in relation to MDD. Among women, MDD was associated with more than 4-fold increased odds of diabetes (OR =4.14; 95% CI: 1.03-16.62) after adjusting for confounders. Among men, we observed no evidence of an association of MDD with diabetes (OR=1.04; 95% CI: 0.35-3.05). After controlling for confounders, there was no clear evidence for a positive association between MDD and hypertension status in women (OR=1.05; 95% CI: 0.37-2.99) or men (OR=0.76; 95% CI: 0.37-1.57).

As presented in Table 3, the odds of lipid abnormalities in relation to MDD were evaluated. After adjusting for possible confounding by age, alcohol consumption, Khat consumption, smoking status, body mass index, and occupation, there was no statistically significant association of MDD with lipid abnormalities.

As shown in Table 4, there was no statistically significant association of MDD with metabolic syndrome among women (OR=1.51; 95% CI: 0.68-3.29) or men (OR=0.61; 95% CI: 0.28-1.34). In addition, there was no evidence of statistically significant associations of MDD with increasing number of MetS components (p for trend=0.189) (Figure 2). The odds of insulin resistance among women and men were (OR=2.89; 95% CI: 0.73-11.35) and (OR=0.66; 95% CI: 0.08-5.61), respectively. Lastly, MDD was not statistically significantly associated with the odds of chronic

systemic inflammation as measured with high CRP concentrations among both women (OR=0.87; 95% CI: 0.12-1.49) and men (OR=1.51: 95% CI: 0.50-4.49).

DISCUSSION

Given the body of epidemiologic evidence showing that depression is associated with and increased risk of cardiometabolic disease, we expected that MDD would be associated with cardiometabolic disease risk factors. However, we found little evidence of such an association among Ethiopian adults after controlling for confounders. Specifically, MDD was not associated with increased odds of hypertension, metabolic syndrome, insulin resistance or inflammation. However, among women, MDD was associated with more than 4-fold increased odds of diabetes.

There is a growing body of epidemiologic evidence that shows MDD as risk factor for diabetes [19]. Our study results showing significant association between MDD and diabetes among women are in agreement with some prior studies although inferences from this analysis are limited by our relatively small size as reflected by the wide 95% CI. A longitudinal study conducted among participants in the Multi-Ethnic Study of Atherosclerosis (MESA) found that depressive symptoms at baseline were associated with an increased incidence of type 2 diabetes after adjusting for confounding factors. However, baseline impaired fasting glucose was associated with reduced risk of depression. Other investigators have reported that depression is a risk factor for diabetes. For instance, Carnethon *et al* [36] in their study among Cardiovascular Health Study (CHS) participants found that older adults who reported higher

depressive symptoms were more likely to develop diabetes than their counterparts. Wagner *et al* [37] also found higher HbA1c (marker of diabetes status) and more diabetes complications among African Americans with higher depressive symptoms after controlling for confounders. In the current study, similar increased odds in insulin resistance were observed for women, although statistical significance was not achieved. The lack of association observed in our study is, in part, due to small sample size as reflected by the wider 95% CIs. Collectively the results of our study and those of others underscore the importance of evaluating depression among diabetics as it is associated with poor diabetes outcomes such as glycemic control and need of insulin therapy [19].

In the current study, we found no evidence of an association between MDD and CRP after adjustment for possible confounders. There is an accumulating epidemiologic evidence suggesting an association between depression and elevated levels of inflammatory markers such as CRP [15, 38-41]. For instance, Ford *et al* using the National Health and Nutrition Examination Survey (NHANES) reported that MDD was strongly associated with increased CRP concentrations among men (OR=3.00;95% CI: 1.39-6.48). There was, however, no evidence of an association of elevated CRP concentrations and depression among women [42]. Similar observation was noted by Danner *et al* using NHANES data [43]. Others have also noted that depression was independently associated with elevated CRP concentrations [44, 45]. Recently Wium-Andersen *et al* in Denmark using population based studies noted that elevated levels of CRP were associated with increased risk for and depression in the general population [46]. On the other hand, Tiemeier *et al* [17] in their population-based Rotterdam Study found no

significant association between depression and CRP after multivariate adjustment. Although available evidence suggests an association between depression and inflammation, it remains unclear whether the inflammation seen in depressed patients is a result of stress response or whether cardiometabolic disease risk related inflammation contributes to the pathogenesis of depression.

Most investigators, though not all, have shown previously that a bi-directional association between depression and hypertension in the US and European populations [47]. Very little is known, however, about this association among sub-Saharan Africans. To the best of our knowledge, to date, only one group of investigators has evaluated the relationship. Using a nationally representative sample from South Africa, as part of the World Mental Health Survey, Grimsrud *et al* found that MDD was not associated with previous medical diagnosis of hypertension [18]. Our study results documenting lack of association between MDD and hypertension are in agreement with their finding. Notably, the South African study [18] showed that anxiety disorders were associated with hypertension status. Perhaps, future studies among sub-Saharan Africans that evaluate other comorbid psychiatric disorders and with longitudinal cardiometabolic disease risk measures might shed light on this issue.

Similarly, we found no evidence of associations of MDD with lipid abnormalities. Some investigators have shown that total cholesterol, in particular, low density lipoprotein (LDL) cholesterol is associated with increased risks of depression; while high density lipoprotein is

inversely related to depressive symptoms [48]. This finding is however in contrast with other reports. Tedders *et al* [49] using the NHANES survey reported no significant association between depression and lipid concentrations. However, the authors noted a U shaped association between LDL and depression among men [49].

The relationship between MDD and metabolic syndrome has been widely studied. However, results have been inconclusive, and few studies have examined sub-Saharan African population [47, 50-52]. Our study results are consistent with some prior studies that showed no significant association between depression and metabolic syndrome (Figures 3 and 4) [50, 52]. Our results, however, are not in agreement with findings from studies by other investigators (supplemental Figures 1 and 2) [53, 54]. Recently, a meta-analysis conducted by Pan *et al* [51] reviewed cohort and cross-sectional studies that examined the association between depression and metabolic syndrome. Within cohort studies, the authors noted a bi-directional relationship where the pooled adjusted OR among studies that used depression as the outcome was (OR= 1.49; 95% CI: 1.19–1.87) and among studies that used metabolic syndrome as an outcome (OR=1.52; 95% CI: 1.20–1.91]. Similar bi-directional associations were found in cross-sectional studies. The lack of association observed in most of the prior studies, in part, may be due to small sample size. Our findings documenting gender difference in the magnitude of association between MDD and metabolic syndrome are in agreement with those reported by others. We do not have clear explanation for these findings but we speculate that difference in physical activity and other lifestyle characteristics might have contributed to the observed differences [55]. It is also

possible that that there could be underlying biological differences in metabolic syndrome risk among men and women that requires further study [56].

The biological mechanisms linking depression and cardiometabolic disease risk are plausible but not fully established. Most investigators have suggested that alterations in the hypothalamic–pituitary–adrenal (HPA) axis might be playing an important role in the pathophysiology of depression and cardiometabolic metabolic disease [57]. A large proportion of depressed subjects have autonomic imbalance characterized by increased sympathetic activation, decreased vagal tone, and abnormal HPA activity [12, 41, 58, 59]. In addition activation of the HPA axis results in increased secretion of corticotrophin-releasing factor (CRF) [15] resulting in excess cortisol secretion. Cortisol is a counter-regulatory hormone known to be associated with type 2 diabetes, insulin resistance, dyslipidemia, and hypertension [11, 13]. There is also an expanding literature suggesting endothelial dysfunction as a potential link between cardiometabolic disease risk and depression [14, 60]. Finally, some investigators have speculated that depression may be associated with unhealthy lifestyle habits, such as smoking, non-compliance with medical recommendation, unhealthy diet, and physical inactivity, which in turn increase the risk of cardiometabolic disease [61]. While conclusive answers about the mechanistic link between cardiometabolic disease risk and depression remain elusive, an increasing body of research has begun to shed light on these important topics. Clearly, however, future research is needed to more definitively establish a causal link and to help understand the role of depression in the pathogenesis of cardiometabolic disease.

The results of the current study have several potential limitations that bear on the interpretation of the study findings. First, our analyses are based on cross-sectionally collected data which leaves ambiguity concerning the temporal relation between depression and cardiometabolic disease. Longitudinal studies are needed to examine the temporal association between depression and cardiometabolic disease risk. Second, although we adjusted for several potential confounders, we cannot exclude the possibility of residual confounding due to misclassification of adjusted variables or confounding by other unmeasured variables. Third, the classification of MDD was done using the PHQ-9 questionnaire that does not give definitive clinical diagnosis of depression. However, use of validated instruments such as PHQ-9 remains the most feasible method of data collection for large-scale epidemiological studies [62]. Fourth, given the imperfect sensitivity and specificity of PHQ-9 in our study, there is a possibility of misclassification of MDD. The impact of this non-differential misclassification would generally underestimate the true magnitude of associations detected in our study [63]. Findings from other studies that used screening and diagnostic instruments with better psychometric properties attenuate concerns about misclassification of MDD status. Fifth, though we consider it very unlikely, we cannot rule out the possibility of missing weaker associations or association that may only be present among sub-specific subject groups. However the consistencies of our findings with prior studies, in part, provide some assurance that misclassification and other biases are probably not responsible for the observed association. Finally, our study findings may not be generalized to the broader Ethiopian population since our study was limited to a largely well-educated, urban dwelling, occupational cohort comprised of white-collar professionals in banking and academic sectors. The concordance of our results with those from other studies

that have included various socio-economic status and geographically diverse populations, however, serve to attenuate some concerns about the generalizability of our findings.

In conclusion, our study results do not provide convincing evidence of associations of MDD with cardiometabolic disease among Ethiopian adults. Future studies need to evaluate the associations of other psychiatric disorders on cardiometabolic risk and shed further light on these relationships among sub-Saharan Africans.

REFERENCES

1. Reddy K, Yusuf S: **Emerging epidemic of cardiovascular disease in developing countries**. *Circulation* 1998, **97**(6):596-601.

2. Njelekela MA, Mpembeni R, Muhihi A, Mligiliche NL, Spiegelman D, Hertzmark E, Liu E, Finkelstein JL, Fawzi WW, Willett WC *et al*: **Gender-related differences in the prevalence of cardiovascular disease risk factors and their correlates in urban Tanzania**. *BMC Cardiovasc Disord* 2009, **9**:30.

3. Oladapo OO, Salako L, Sodiq O, Shoyinka K, Adedapo K, Falase AO: **A prevalence of cardiometabolic risk factors among a rural Yoruba south-western Nigerian population: a population-based survey**. *Cardiovasc J Afr* 2010, **21**(1):26-31.

4. Tesfaye F, Byass P, Wall S: **Population based prevalence of high blood pressure among adults in Addis Ababa: uncovering a silent epidemic**. *BMC Cardiovasc Disord* 2009, **9**:39.

5. Tran A, Gelaye B, Girma B, Lemma S, Berhane Y, Bekele T, Khali A, Williams MA: **Prevalence of Metabolic Syndrome among Working Adults in Ethiopia**. *Int J Hypertens* 2011, **2011**:193719.

6. Wai WS, Dhami RS, Gelaye B, Girma B, Lemma S, Berhane Y, Bekele T, Khali A, Williams MA: **Comparison of measures of adiposity in identifying cardiovascular disease risk among Ethiopian adults**. *Obesity (Silver Spring)* 2012, **20**(9):1887-1895.

7. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J: **Global burden of hypertension: analysis of worldwide data**. *Lancet* 2005, **365**(9455):217-223.

8. Roglic G, Unwin N, Bennett PH, Mathers C, Tuomilehto J, Nag S, Connolly V, King H: **The burden of mortality attributable to diabetes: realistic estimates for the year 2000**. *Diabetes Care* 2005, **28**(9):2130-2135.

9. Tomlinson M, Swartz L, Kruger LM, Gureje O: **Manifestations of affective disturbance in sub-Saharan Africa: key themes**. *J Affect Disord* 2007, **102**(1-3):191-198.

10. Kinder LS, Carnethon MR, Palaniappan LP, King AC, Fortmann SP: **Depression and the metabolic syndrome in young adults: findings from the Third National Health and Nutrition Examination Survey**. *Psychosom Med* 2004, **66**(3):316-322.

11. Katon W, Maj M, Sartorius N: **Depression and diabetes**. Chichester, West Sussex; Hoboken, NJ: Wiley-Blackwell; 2010.

12. Carney RM, Freedland KE, Sheline YI, Weiss ES: **Depression and coronary heart disease: a review for cardiologists**. *Clin Cardiol* 1997, **20**(3):196-200.

13. Celano CM, Huffman JC: **Depression and cardiac disease: a review**. *Cardiol Rev* 2011, **19**(3):130-142.

14. Glassman AH, Maj M, Sartorius N: **Depression and heart disease**. Chichester, West Sussex; Hoboken, NJ: Wiley-Blackwell; 2011.

15. Carney RM, Freedland KE, Miller GE, Jaffe AS: **Depression as a risk factor for cardiac mortality and morbidity: a review of potential mechanisms**. *J Psychosom Res* 2002, **53**(4):897-902.

16. Foley DL, Morley KI, Madden PA, Heath AC, Whitfield JB, Martin NG: **Major depression and the metabolic syndrome**. *Twin Res Hum Genet* 2010, **13**(4):347-358.

17. Tiemeier H, Hofman A, van Tuijl HR, Kiliaan AJ, Meijer J, Breteler MM: **Inflammatory proteins and depression in the elderly**. *Epidemiology* 2003, **14**(1):103-107.

18. Grimsrud A, Stein DJ, Seedat S, Williams D, Myer L: **The association between hypertension and depression and anxiety disorders: results from a nationally-representative sample of South African adults**. *PLoS One* 2009, **4**(5):e5552.

19. Katon WJ: **The comorbidity of diabetes mellitus and depression**. *Am J Med* 2008, **121**(11 Suppl 2):S8-15.

20. World Health Organization: **Primary Health Care Management of Trachoma**. Geneva: World Health Organization; 1993.

21. WHO: **STEPs manual**. Geneva: World Health Organization; 2008.

22. Belew M, Kebede D, Kassaye M, Enquoselassie F: **The magnitude of Khat use and its association with health, nutrition and socio-economic status**. *Ethiop Med J* 2000, **38**(1):11-26.

23. Kalix P: **Khat: scientific knowledge and policy issues**. *Br J Addict* 1987, **82**(1):47-53.

24. Wai WS, Dhami RS, Gelaye B, Girma B, Lemma S, Berhane Y, Bekele T, Khali A, Williams MA: **Comparison of Measures of Adiposity in Identifying Cardiovascular Disease Risk Among Ethiopian Adults**. *Obesity (Silver Spring)* 2011.

25. Alberti KG, Zimmet P, Shaw J: **Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation**. *Diabet Med* 2006, **23**(5):469-480.

26. Expert Panel on Detection E, Treatment of High Blood Cholesterol in A: **Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP)**

Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA* 2001, **285**(19):2486-2497.

27. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC: **Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man**. *Diabetologia* 1985, **28**(7):412-419.

28. Deveci E, Yesil M, Akinci B, Yesil S, Postaci N, Arikan E, Koseoglu M: **Evaluation of insulin resistance in normoglycemic patients with coronary artery disease**. *Clin Cardiol* 2009, **32**(1):32-36.

29. Hillman AJ, Lohsoonthorn V, Hanvivatvong O, Jiamjarasrangsi W, Lertmaharit S, Williams MA: **Association of High Sensitivity C-Reactive Protein Concentrations and Metabolic Syndrome among Thai Adults**. *Asian Biomed* 2010, **4**(3):385-393.

30. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL: **Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients**. *J Gen Intern Med* 2006, **21**(6):547-552.

31. Lichtman JH, Bigger JT, Jr., Blumenthal JA, Frasure-Smith N, Kaufmann PG, Lesperance F, Mark DB, Sheps DS, Taylor CB, Froelicher ES: **Depression and coronary heart disease: recommendations for screening, referral, and treatment: a science advisory from the American Heart Association Prevention Committee of the Council on Cardiovascular Nursing, Council on Clinical Cardiology, Council on Epidemiology and Prevention, and Interdisciplinary Council on Quality of Care and Outcomes Research: endorsed by the American Psychiatric Association**. *Circulation* 2008, **118**(17):1768-1775.

32. WHO: **Global status report on alcohol**. Geneva, Switzerland: World Health Organization. Department of Mental Health and Substance Abuse; 2004.

33. Rothman KJ, Greenland S, Lash TL: **Modern epidemiology**. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.

34. Alem A, Kebede D, Woldesemiat G, Jacobsson L, Kullgren G: **The prevalence and socio-demographic correlates of mental distress in Butajira, Ethiopia**. *Acta Psychiatr Scand Suppl* 1999, **397**:48-55.

35. Gelaye B, Lemma S, Deyassa N, Bahretibeb Y, Tesfaye M, Berhane Y, Williams MA: **Prevalence and correlates of mental distress among working adults in ethiopia**. *Clin Pract Epidemiol Ment Health* 2012, **8**:126-133.

36. Carnethon MR, Biggs ML, Barzilay JI, Smith NL, Vaccarino V, Bertoni AG, Arnold A, Siscovick D: **Longitudinal association between depressive symptoms and incident type 2 diabetes mellitus in older adults: the cardiovascular health study**. *Arch Intern Med* 2007, **167**(8):802-807.

37. Wagner JA, Abbott GL, Heapy A, Yong L: **Depressive symptoms and diabetes control in African Americans**. *J Immigr Minor Health* 2009, **11**(1):66-70.

38. Smith PJ, Blumenthal JA: **Psychiatric and Behavioral Aspects of Cardiovascular Disease: Epidemiology, Mechanisms, and Treatment**. *Rev Esp Cardiol* 2011.

39. Blumberger DM, Daskalakis ZJ, Mulsant BH: **Biomarkers in geriatric psychiatry: searching for the holy grail?** *Curr Opin Psychiatry* 2008, **21**(6):533-539.

40. Howren MB, Lamkin DM, Suls J: **Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis**. *Psychosom Med* 2009, **71**(2):171-186.

41. Frasure-Smith N, Lesperance F, Irwin MR, Talajic M, Pollock BG: **The relationships among heart rate variability, inflammatory markers and depression in coronary heart disease patients**. *Brain Behav Immun* 2009, **23**(8):1140-1147.

42. Ford DE, Erlinger TP: **Depression and C-reactive protein in US adults: data from the Third National Health and Nutrition Examination Survey**. *Arch Intern Med* 2004, **164**(9):1010-1014.

43. Danner M, Kasl SV, Abramson JL, Vaccarino V: **Association between depression and elevated C-reactive protein**. *Psychosom Med* 2003, **65**(3):347-356.

44. Panagiotakos DB, Pitsavos C, Chrysohoou C, Tsetsekou E, Papageorgiou C, Christodoulou G, Stefanadis C, study A: **Inflammation, coagulation, and depressive symptomatology in cardiovascular disease-free people; the ATTICA study**. *Eur Heart J* 2004, **25**(6):492-499.

45. Kop WJ, Gottdiener JS, Tangen CM, Fried LP, McBurnie MA, Walston J, Newman A, Hirsch C, Tracy RP: **Inflammation and coagulation factors in persons > 65 years of age with symptoms of depression but without evidence of myocardial ischemia**. *Am J Cardiol* 2002, **89**(4):419-424.

46. Wium-Andersen MK, Orsted DD, Nielsen SF, Nordestgaard BG: **Elevated C-reactive protein levels, psychological distress, and depression in 73 131 individuals**. *JAMA Psychiatry* 2013, **70**(2):176-184.

47. Meng L, Chen D, Yang Y, Zheng Y, Hui R: **Depression increases the risk of hypertension incidence: a meta-analysis of prospective cohort studies**. *J Hypertens* 2012, **30**(5):842-851.

48. Manfredini R, Caracciolo S, Salmi R, Boari B, Tomelli A, Gallerani M: **The association of low serum cholesterol with depression and suicidal behaviours: new hypotheses for the missing link**. *J Int Med Res* 2000, **28**(6):247-257.

49. Tedders SH, Fokong KD, McKenzie LE, Wesley C, Yu L, Zhang J: **Low cholesterol is associated with depression among US household population**. *J Affect Disord* 2011, **135**(1-3):115-121.

50. Pulkki-Raback L, Elovainio M, Kivimaki M, Mattsson N, Raitakari OT, Puttonen S, Marniemi J, Viikari JS, Keltikangas-Jarvinen L: **Depressive symptoms and the metabolic syndrome in childhood and adulthood: a prospective cohort study**. *Health Psychol* 2009, **28**(1):108-116.

51. Pan A, Keum N, Okereke OI, Sun Q, Kivimaki M, Rubin RR, Hu FB: **Bidirectional association between depression and metabolic syndrome: a systematic review and meta-analysis of epidemiological studies**. *Diabetes Care* 2012, **35**(5):1171-1180.

52. Vogelzangs N, Beekman AT, Boelhouwer IG, Bandinelli S, Milaneschi Y, Ferrucci L, Penninx BW: **Metabolic depression: a chronic depressive subtype? Findings from the InCHIANTI study of older persons**. *J Clin Psychiatry* 2011, **72**(5):598-604.

53. Mast BT, Miles T, Penninx BW, Yaffe K, Rosano C, Satterfield S, Ayonayon HN, Harris T, Simonsick EM: **Vascular disease and future risk of depressive symptomatology in older adults: findings from the Health, Aging, and Body Composition study**. *Biol Psychiatry* 2008, **64**(4):320-326.

54. Almeida OP, Calver J, Jamrozik K, Hankey GJ, Flicker L: **Obesity and metabolic syndrome increase the risk of incident depression in older men: the health in men study**. *Am J Geriatr Psychiatry* 2009, **17**(10):889-898.

55. Workalemahu T, Gelaye B, Berhane Y, Williams MA: **Physical Activity and Metabolic Syndrome among Ethiopian Adults**. *Am J Hypertens* 2013; 26(4):535-40

56. Muhtz C, Zyriax BC, Klahn T, Windler E, Otte C: **Depressive symptoms and metabolic risk: effects of cortisol and gender**. *Psychoneuroendocrinology* 2009, **34**(7):1004-1011.

57. Grippo AJ, Johnson AK: **Biological mechanisms in the relationship between depression and heart disease**. *Neurosci Biobehav Rev* 2002, **26**(8):941-962.

58. Esler M, Turbott J, Schwarz R, Leonard P, Bobik A, Skews H, Jackman G: **The peripheral kinetics of norepinephrine in depressive illness**. *Arch Gen Psychiatry* 1982, **39**(3):295-300.

59. Siever LJ, Davis KL: **Overview: toward a dysregulation hypothesis of depression**. *Am J Psychiatry* 1985, **142**(9):1017-1031.

60. Pariante CM: **Understanding depression : a translational approach**. London; New York: Oxford University Press; 2009.

61. Empana JP, Jouven X, Lemaitre RN, Sotoodehnia N, Rea T, Raghunathan TE, Simon G, Siscovick DS: **Clinical depression and risk of out-of-hospital cardiac arrest**. *Arch Intern Med* 2006, **166**(2):195-200.

62. Gelaye B, Peterlin BL, Lemma S, Tesfaye M, Berhane Y, Williams MA: **Migraine and Psychiatric Comorbidities Among Sub-Saharan African Adults**. *Headache* 2012.

63. White E, Armstrong BK, Saracci R, Armstrong BK: **Principles of exposure measurement in epidemiology : collecting, evaluating, and improving measures of disease risk factors**. Oxford; New York: Oxford University Press; 2008.

Table 5.1-Characteristics of the study population according to major depressive disorder status

Characteristic	All N=1,924 n (%)	Depression N=154 n (%)	No Depression N=1,770 n (%)	P-value**
Mean Age*	35.9 ± 11.8	33.1 ± 10.9	36.1 ± 11.9	
Sex				
Women	759 (39.4)	60 (39.0)	699 (39.5)	0.897
Men	1,165 (60.6)	94 (61.0)	1071 (60.5)	
Age (years)				
18-29	799 (41.6)	83 (54.3)	716 (39.8)	0.017
30-39	441 (22.9)	32 (20.9)	409 (23.6)	
40-49	307 (16.0)	17 (11.1)	290 (16.5)	
50-59	353 (18.4)	20 (13.1)	333 (18.9)	
≥60	21 (1.1)	1 (0.6)	20 (1.2)	
Education				
≤ High school	564 (29.3)	55 (35.7)	509 (28.8)	0.069
≥ College education	1,360 (70.7)	99 (64.3)	1,261 (71.2)	
Smoking status				
Never	1,670 (86.8)	126 (81.8)	1,544 (87.2)	0.153
Former smoker	84 (4.4)	10 (6.5)	74 (4.2)	
Current smoker	170 (8.8)	18 (11.7)	152 (8.6)	
Alcohol consumption in				
Non drinker	435 (22.6)	25 (16.2)	410 (23.2)	0.032
Moderate	1,353 (70.3)	112 (72.7)	1,241 (70.1)	
Heavy	136 (7.1)	17 (11.1)	119 (6.7)	
Khat use				
No	1,758 (91.4)	137 (89.0)	1,621 (91.6)	0.256
Yes	165 (8.6)	17 (11.0)	148 (8.4)	
Self-reported health status				
Excellent/very good/good	1,161(60.3)	61 (39.6)	1,100 (62.2)	<0.001
Poor/fair	763 (39.7)	93 (60.4)	670 (37.8)	

*Mean ± standard deviation (SD); **P-value from Chi-Square test or Student's t test

Table 5.2—Odds ratios (OR) and 95% confidence intervals (CI) of diabetes status and hypertension status in relation to major depressive disorder among men and women

	Normal Fasting Glucose	Impaired Fasting Glucose	Diabetes Mellitus	Normotensive	Hypertensive
Women					
Unadjusted	1.0 (Reference)	1.05 (0.56-1.98)	1.28 (0.37-4.37)	1.0 (Reference)	0.93 (0.38-2.25)
Model 1	1.0 (Reference)	1.24 (0.66-2.37)	2.77 (0.73-10.65)	1.0 (Reference)	0.94 (0.33-2.65
Model 2	1.0 (Reference)	1.35 (0.93-3.86)	3.37 (0.86-13.22)	1.0 (Reference)	0.98 (0.35-2.77)
Model 3	1.0 (Reference)	1.44 (0.74-2.78)	4.14 (1.03-16.62)	1.0 (Reference)	1.05 (0.37-2.99)
Men					
Unadjusted	1.0 (Reference)	0.92 (0.55-1.55)	0.76 (0.27-2.16)	1.0 (Reference)	0.68 (0.35-1.32)
Model 1	1.0 (Reference)	0.99 (0.58-1.67)	0.93 (0.32-2.70)	1.0 (Reference)	0.75 (0.36-1.53)
Model 2	1.0 (Reference)	0.94 (0.54-1.62)	0.98 (0.33-2.89)	1.0 (Reference)	0.73 (0.35-1.47)
Model 3	1.0 (Reference)	0.98 (0.57-1.71)	1.04 (0.35-3.05)	1.0 (Reference)	0.76 (0.37-1.57)

Model 1 (age and BMI): Adjusted for age and body mass index

Model 2 (lifestyle): Adjusted for age, Khat consumption, alcohol drinking, smoking and physical activity

Model 3 (fully adjusted): Adjusted for age, body mass index, occupation, Khat consumption, alcohol consumption, smoking and physical activity

Table 5.3—Odds ratios (OR) and 95% confidence intervals (CI) of lipid abnormalities in relation to major depressive disorder among men and women

	Women		Men	
Lipids (mg/dL)	**Unadjusted OR (95% CI)**	**Adjusted* OR (95% CI)**	**Unadjusted OR (95% CI)**	**Adjusted* OR (95% CI)**
HDL-C				
Low (<40)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Normal (40-60)	0.47 (0.18-1.18)	0.54 (0.21-1.38)	0.73 (0.42-1.30)	0.69 (0.39-1.25)
High (≥ 60)	0.67 (0.23-1.94)	0.80 (0.26-2.41)	0.45 (0.13-1.62)	0.39 (0.11-1.41)
LDL-C				
Optimal (<100)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Above optimal (100-129)	1.05 (0.56-1.96)	1.25 (0.65-2.42)	0.97 (0.60-1.59)	1.01 (0.62-1.68)
High (>130)	0.67 (0.33-1.36)	0.94 (0.42-2.06)	0.77 (0.45-1.33)	0.84 (0.47-1.52)
TC				
Desirable (<200)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Borderline high (200-239)	1.05 (0.58-1.88)	1.27 (0.68-2.36)	0.87 (0.51-1.48)	0.86 (0.49-1.51)
High (≥ 240)	0.36 (0.11-1.29)	0.41 (0.09-1.84)	0.62 (0.29-1.32)	0.67 (0.31-1.50)
Triglycerides				
Normal(<150)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Borderline-high (150-199)	0.85 (0.33-2.20)	1.32 (0.48-3.61)	1.01 (0.56-1.81)	1.16 (0.61-2.19)
High (≥ 200)	1.75 (0.66-4.67)	3.08 (0.92-10.35)	0.64 (0.33-1.24)	0.70 (0.34-1.43)

*Adjusted for alcohol consumption, Khat consumption, smoking status, age, body mass index, occupation, and physical activity

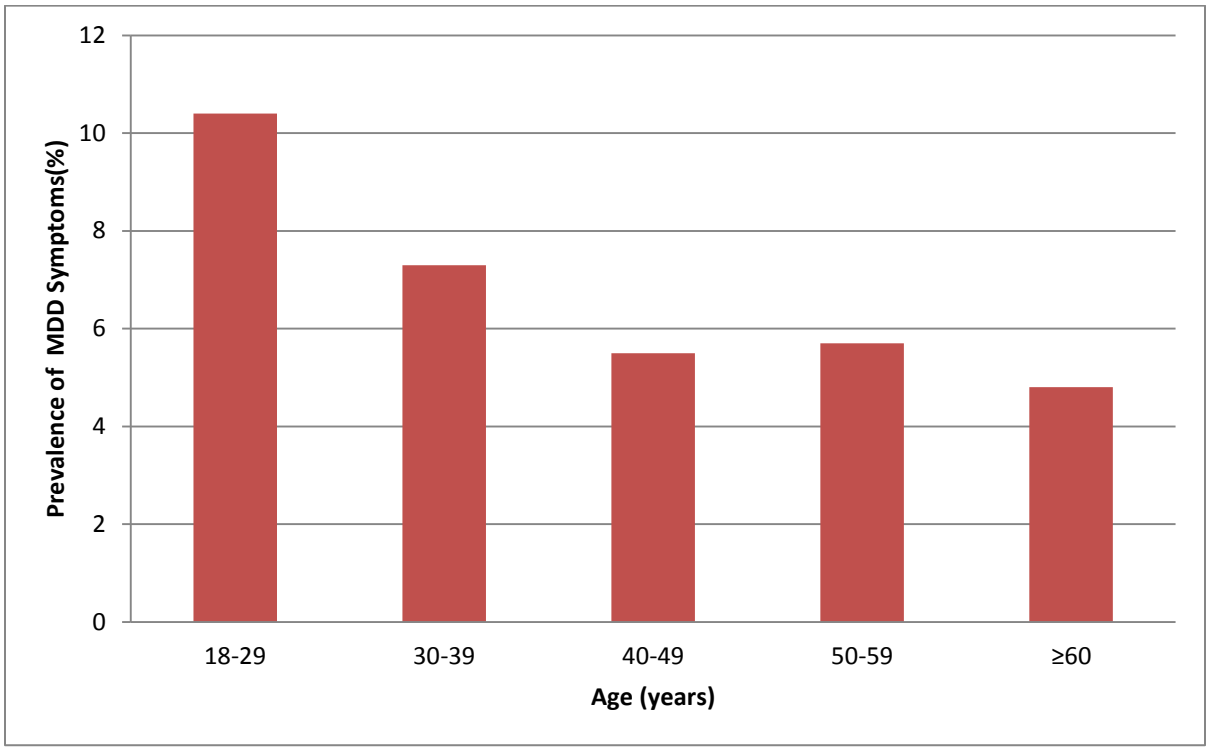
Table 5.4—Odds ratios (OR) and 95% confidence intervals (CI) of metabolic syndrome, C-reactive protein, and insulin resistance in relation to major depressive disorder among men and women – Nested case control (156 cases and 360 controls)

	All		Women		Men	
	Unadjusted OR (95% CI)	Adjusted* OR (95% CI)	Unadjusted OR (95% CI)	Adjusted* OR (95% CI)	Unadjusted OR (95% CI)	Adjusted** OR (95% CI)
Metabolic Syndrome						
No	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Yes	0.70 (0.43-1.17)	0.93 (0.52-1.64)	0.54 (0.26-1.13)	1.51 (0.68-3.29)	1.24 (0.43-3.59)	0.61 (0.28-1.34)
C-reactive Protein						
Low	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
High	0.71 (0.32-1.54)	0.75 (0.34-1.66)	0.41 (0.12-1.43)	0.42 (0.12-1.49)	1.24 (0.43-3.59)	1.51 (0.50-4.49)
Insulin Resistance						
No	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Yes	1.13 (0.38-3.32)	1.32 (0.43-3.99)	2.58 (0.68-9.83)	2.89 (0.73-11.35)	0.45 (0.06-3.52)	0.66 (0.08-5.61)

*Adjusted for age, work place, alcohol consumption, and physical activity

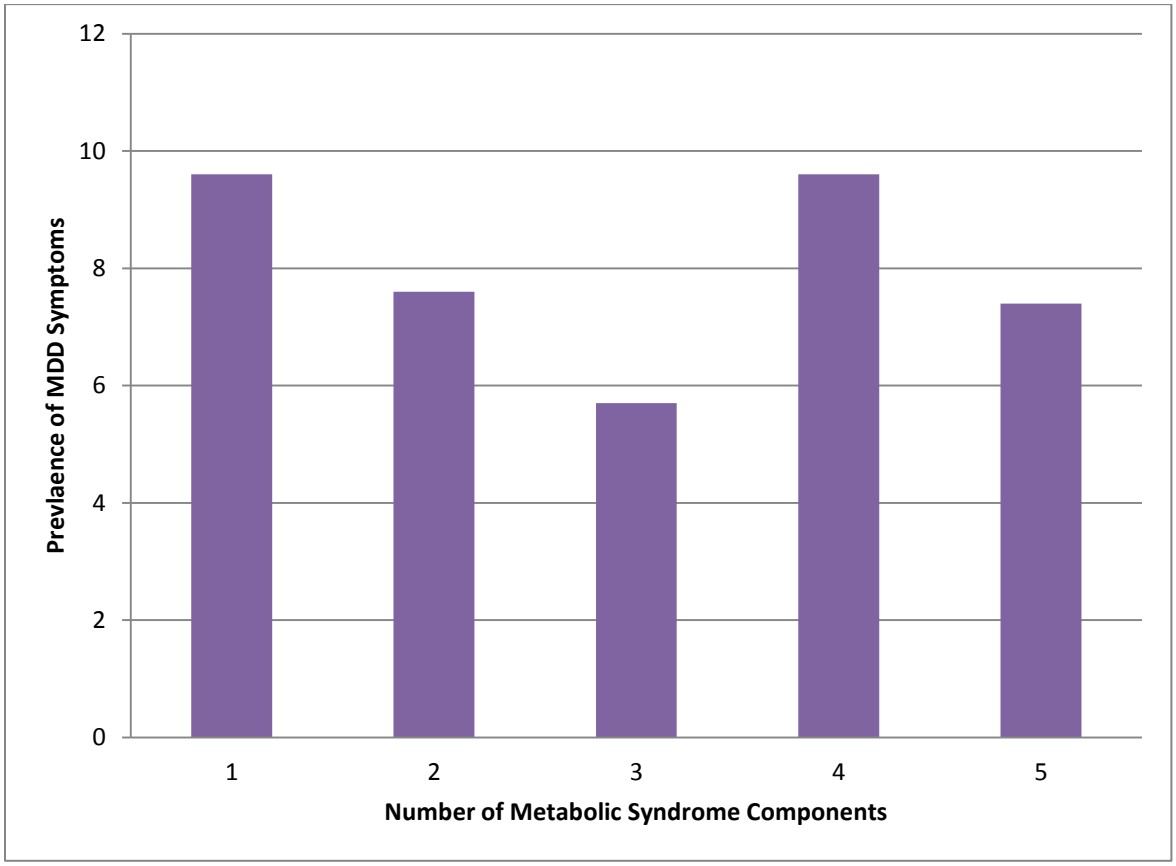
**Adjusted for age, work place, BMI, Khat consumption, and physical activity

Figure 5.1—Prevalence of MDD symptoms according age



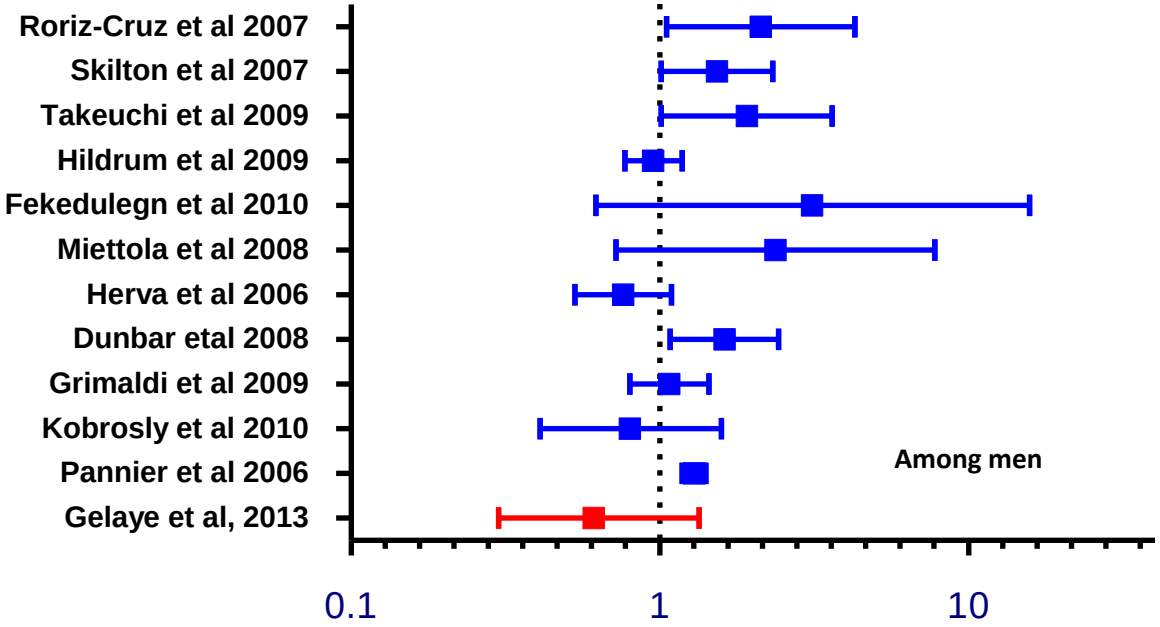
The prevalence of MDD varied with age with the highest being in early adult life (18-29 years), and declining thereafter.

Figure 5.2— Prevalence of MDD symptoms according to number of MetS components



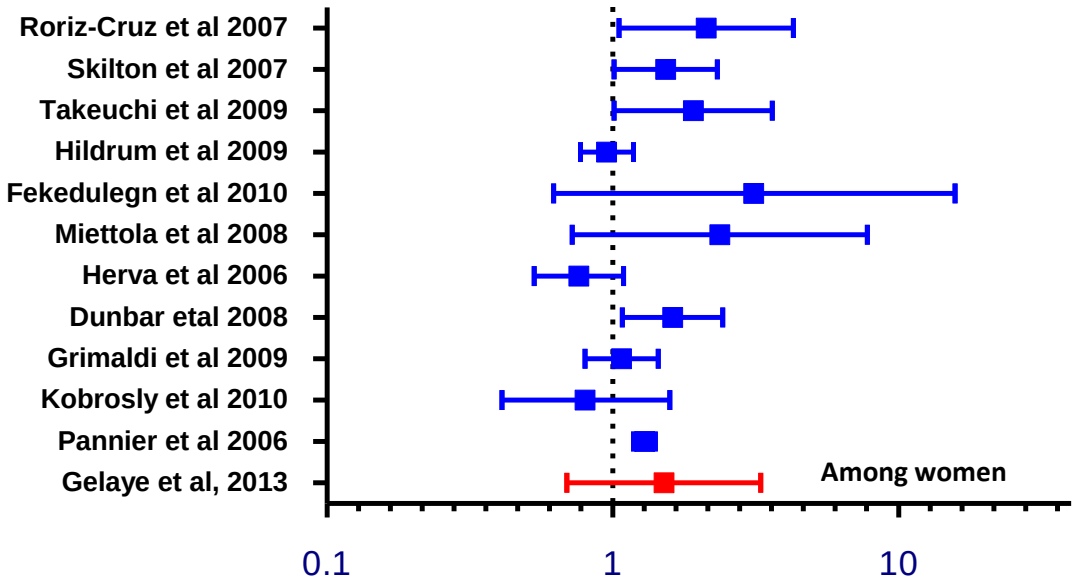
There was no evidence of statistically significant associations of MDD with increasing number of MetS components (p for trend=0.189).

Figure 5.3—Forest plot of cross-sectional studies evaluating the odds of metabolic syndrome in relation to depression



Among men, the magnitude and direction of associations between metabolic syndrome and depression was similar to some prior studies

Figure 5.4—Forest plot of cross-sectional studies evaluating the odds of metabolic syndrome in relation to depression



Among women, the magnitude and direction of associations between metabolic syndrome and depression was largely similar to prior studies.

Chapter 6

Conclusions

Conclusions

Our overall objective in this thesis was to examine the reliability and validity of the PHQ-9 and CIDI instruments among adults in Ethiopia against a psychiatrist administered SCAN reference standard. As the gold standard might not always be available, we also evaluated the psychometric properties of the PHQ-9 against a lay-interviewer administered CIDI reference standard and utilized advanced statistical approaches to improve psychometric properties of the PHQ-9. Additionally, we evaluated the extent to which symptoms of depression are associated with cardiometabolic risk and risk factors among an occupational cohort of Ethiopians residing in Addis Ababa.

As presented in chapter two, the PHQ-9 has good reliability and validity (sensitivity (79%) and specificity (72%)) for diagnosing MDD among Ethiopian adults. The internal consistency reliability was also found to be excellent (ICC=0.92 and Cronbach's alpha=0.81). Our ROC analysis showed that a threshold of nine on the PHQ-9 was the most appropriate cutoff and offered the optimal discriminatory power in detecting MDD. In addition, our study provided strong evidence for the construct validity of the Amharic version of PHQ-9 questionnaire. The results of our factor analysis revealed unidimensionality with acceptable factor loadings on a major core depressive factor and adequate item discrimination values.

In chapter three, our results provided evidence that the Amharic version of the CIDI is a tool with high internal reliability (Cronbach's alpha = 0.97) among Ethiopian adults and excellent

the construct validity of the Amharic version of the CIDI in diagnosing MDD. The factor analysis revealed unidimensionality of core depression screening questions with good factor loadings on a major core depressive factor. Adults classified with MDD by the CIDI had lower quality of life scores across physical, psychological, social relationship and environmental domains. Compared to SCAN reference standard, the CIDI is modestly effective in MDD diagnosis with a moderate specificity (72.2%) but low sensitivity (51.0%). Finally, the CIDI had high NPV and low negative LR indicated the potential for CIDI in ruling out MDD in a similar outpatient setting.

As shown in chapter four, we demonstrated how psychometric properties can be biased when one uses a suboptimal gold standard. Furthermore, improvements in psychometric properties of the PHQ-9 (when compared against CIDI, an imperfect gold standard) were noted after employing Bayesian latent class modeling approaches.

In chapter five, we examined the extent to which MDD is associated with cardiometabolic disease risk factors. We found little evidence of such an association among Ethiopian adults after controlling for confounders. Specifically, MDD was not associated with increased odds of hypertension, metabolic syndrome, insulin resistance or inflammation. However, among women, MDD was associated with a more than 4-fold increased odds of diabetes.

In conclusion, our study is an important contribution to research focused on guiding screening and treatment of depression in sub-Saharan Africa. We provided the first objective evidence for the use of PHQ-9 and CIDI among Ethiopian adults. Integration of mental health in primary health care settings has been one of the suggested approaches to adequately address the burden of mental disorders in low and middle income countries [1]. To this end, recently, the Ethiopian Federal Ministry of Health released its first National Mental Health Strategy that outlines integrating mental health services into Primary Health Care, one of the most widely recommended approaches by the WHO [2]. As part of this national strategy, physicians, health officers and nurses are provided training to provide brief psychosocial interventions and prescribe psychotropic medications. The first step in the management provision of treatments for depression is its effective detection. The systematic use of PHQ-9 instruments allows for the early detection of MDD [3]. However, as with use of any depression assessment tools, screening in routine clinical practice need to be combined with detailed clinical assessment of screen positives to reduce the proportion of false positives identified [4].

Future studies should assess the responsiveness and sensitivity of PHQ-9 derived symptom scores to treatment among Ethiopian patients. Although we have evaluated the validity of CIDI in a clinical setting, it will be important for future studies to evaluate its diagnostic validity within the general population. Furthermore future longitudinal prospective studies are warranted to confirm our findings and elucidate the causal pathways between depression and cardiometabolic risk. Finally, further studies are also needed to evaluate the effect of

other comorbid psychiatric disorders (e.g., generalized anxiety disorders) on cardiometabolic risk among sub-Saharan Africans.

REFERENCES

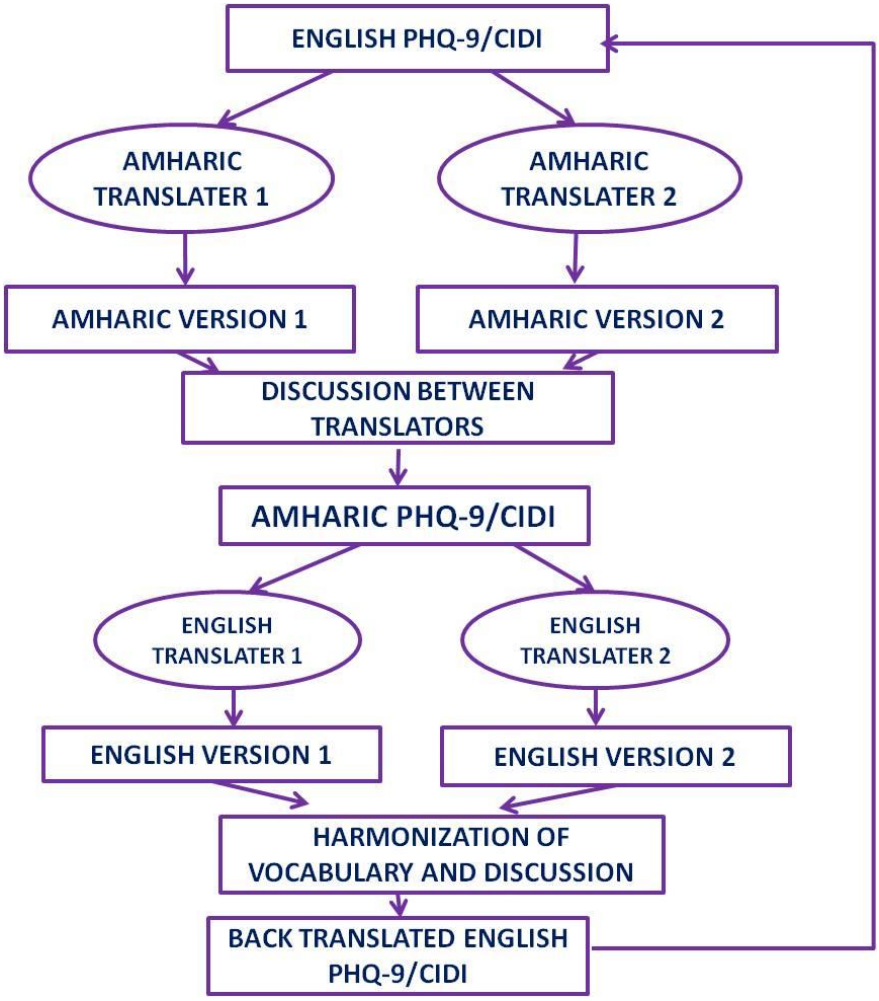
1. Patel V: **Mental health in low- and middle-income countries**. *Br Med Bull* 2007, **81-82**:81-96.

2. EMOH: **Federal Democratic Republic of Ethiopia Ministry of Health. National Mental Health Strategy. Addis Ababa, Ethiopia.** . Geneva: World Health Organization; 2012.

3. Kroenke K, Spitzer RL, Williams JB: **The PHQ-9: validity of a brief depression severity measure**. *J Gen Intern Med* 2001, **16**(9):606-613.

4. Patel V, Araya R, Chowdhary N, King M, Kirkwood B, Nayak S, Simon G, Weiss HA: **Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires**. *Psychol Med* 2008, **38**(2):221-228.

Appendix I—Translation process (Adopted from Omoro *et al*)



Appendix II –Amharic Version of PHQ-9 Questionnaire

		በፍፁም	ከሰባት ቀናት ያነሰ	ከሰባት ቀናት በላይ	ከሞላ ጎደል በየቀኑ	መልስ
1	ነገሮችን ሲሰሩ ፍላጎትም ወይም የሚያገኙት ደስታ በጣም ትንሽ(አምብዛም) ነበር?	0	1	2	3	[___]
2	የትካዜ፤ የበታችነት፤የጭንቀት፤የመደበር ወይንም ተስፋ የመቁረጥ ስሜት ነበረዎት?	0	1	2	3	[___]
3	እንቅልፍ የመተኛት ፤ተኝቶ የመቆየት ችግር ወይንም ከመጠን በላይ የመተኛት ችግር ነበረዎት?	0	1	2	3	[___]
4	ድካም የመሰማት ወይንም አቅም የማነስ ሁኔታ ነበረዎት?	0	1	2	3	[___]
5	የምግብ ፍላጐት አለመኖር ወይንም በጣም ብዙ የመብላት ችግር ነበረብዎት?	0	1	2	3	[___]
6	ስለራስዎ መጥፎ ስሜት ተስምቶዎት ወይም አልተሳካልኝም ብለው አስበው፤ ወይንም ቤተሰቤን አሳፈረኩ ብለው አስበው ነበር?	0	1	2	3	[___]
7	ነገሮች ላይ ሀሳብዎትን መስብሰብ ወይም ልብ የማለትችግር ነበረብዎት፡ ለምሳሌ ጋዜጣ ሲያነቡ ወይንም ቴሌቭዥን ሲመለከቱ?	0	1	2	3	[___]
8	ከተለመደው ውጪ እረፍት የማጣት፤ ወዲያና ወዲህ የማለት ወይንም በተቃራኒው ሌሎች ሰዎች ሊገነዘቡት በሚችሉት ሁኔታ ቀስ ብሎ የመናገር ወይም የመንቀሳቀስ ችግር ነበረብዎት?	0	1	2	3	[___]
9	ብሞት ይሻላል ወይንም እራሴን በሆነ መንገድ ብጉዳ ይሻላል ብለው ያስቡበት ጊዜ ነበር?	0	1	2	3	[___]
		ምንም ችግር አልፈጠሩም	በመጠኑ ችግር ፈጥረዋል	በጣም ተቸግሬያለሁ	እጅግ በጣም ተቸግረያለሁ	መልስ
10	ከነዚህ ከላይ የጠቀስናቸው ችግሮች አጋጥሞዎት ከነበሩ፡ ችግሮቹ ስራዎትን እንዳይሰሩ፤ የቤተሰብ ኃላፊነትዎን እንዳይወጡ፤ ራስዎን ለመጠበቅ ወይንም ከሌሎች ሰዎች ጋር ባለዎት ግንኙነት ለእርስዎ ምን ያህል አስቸጋሪ (አዳጋች) ነበሩ?	0	1	2	3	[___]

VITA

Bizu Gelaye

Education:	B.Sc. in Mechanical Engineering Addis Ababa University	July, 2002
	MPH in Epidemiology University of Washington	March, 2011
	PhD in Epidemiology University of Washington	March 2013

Employment

- **Researcher and Program Manager**, Department of Epidemiology, Harvard School of Public Health, Boston, MA– November 2011 to present
- **Research Scientist**, Department of Epidemiology, University of Washington Seattle, WA – August 2010 to November 2011
- **Program Operations Specialist/Manager**, Department of Epidemiology , University of Washington Seattle, WA – January 2006 to August 2010
- **Research Coordinator**, Department of Epidemiology, University of Washington Seattle, WA – September 2004 to January 2006
- **Program Assistant**, Department of Epidemiology, University of Washington Seattle, WA – April 2004 to September 2004
- **Mechanical Engineer and Section Head**, Addis Ababa Bottle and Glass Share Company, Addis Ababa, Ethiopia – February 2003 to January 2004
- **Instructor,** Delta Computer Systems
  Addis Ababa, Ethiopia – July 2001 to January 2004
- **Research Assistant**, Global Vision Consultants International
  Addis Ababa, Ethiopia – June 2000 to October 2002

Other Experience and Professional Membership

2006 – 2009	Member and Reviewer, Human Subjects Division, UW
2005 – 2011	Advisory Board Member, Global Health Resource Center, UW
2005 – 2011	Advisory Board Member, Michael Ealy Social Justice Foundation
2007 – 2011	Reviewer, Violence and Victims Journal
2007 – 2011	Reviewer, Scandinavian Journal of Public Health
2008 – present	Reviewer, Scientific Research and Essays
2010 – present	Reviewer, Diabetology and Metabolic Syndrome
2010 – present	Reviewer, Social Psychiatry and Psychiatric Epidemiology
2012 -- present	Reviewer, Journal of Affective Disorders
2012 – present	Reviewer, European Journal of Psychiatry
2012 -- present	Reviewer, Journal of Health, Population and Nutrition

Honors and Awards

1996	Five-year Scholarship, Engineering, Addis Ababa University
2007	Outstanding Staff Award, Department of Epidemiology, UW
2007	Kenneth Anderson Outstanding Staff Award, School of Public Health, UW
2007	Proposal writing training scholarship, Verizon Foundation
2011	Outstanding Masters Student Award, Department of Epidemiology, UW

Training

- Total Quality Management (Management Institute of Ethiopia)
- Institutional Review Board (IRB) reviewer training (University of Washington)
- World Health Organization Composite International Diagnostic Interview Certified Trainer (University of Michigan, Ann Arbor, MI)
- Office of Regulatory Affairs and Research Compliance Quality Assurance/Quality Improvement (Harvard School of Public Health, Boston, MA)

First Author Research Papers

1. **Gelaye B**, Philpart M, Goshu M, Berhane Y, Fitzpatrick AL, Williams MA. Anger expression, negative life events and violent behavior among male college students in Ethiopia. *Scand J Public Health*. 2008 ;36(5):538-45.
2. **Gelaye B**, Arnold D, Williams MA, Goshu M, Berhane Y. Depressive symptoms among female college students experiencing gender-based violence in Awassa, Ethiopia. *J Interpers Violence*. 2009 ;24(3):464-81.
3. **Gelaye B**, Lam N, Cripe S, Sanchez SE, Williams MA. Correlates of violent response among women exposed to intimate partner violence. *J Interpers Violence*. 2010; 25(1):136-51.
4. **Gelaye B**, Revilla L, Lopez T, Sanchez SE, Williams MA. Prevalence of metabolic syndrome and its relationship with leisure time physical activity among Peruvian adults. *Eur J Clin Invest.* 2009 ;39(10):891-8.
5. **Gelaye B,** Revilla L, Lopez T, Suarez L, Sanchez SE, Hevner K, Fitzpatrick AL, Williams MA. Association between insulin resistance and c-reactive protein among Peruvian adults. *Diabetol Metab Syndr*. 2010 May 18;2(1):30.
6. **Gelaye B**, Bekele T, Khali A, Haddis Y, Lemma S, Berhane Y, Williams MA. Laboratory reference values of complete blood count for apparently healthy adults in Addis Ababa, Ethiopia. C*linical Lab*. 2011 2011;57:635-640
7. **Gelaye B**, Peterlin BL, Lemma S, Tesfaye M, Berhane Y, Williams MA. Migraine and psychiatric aomorbidities among Sub-Saharan African Adults. Headache. 2012; ;53(2):310-21.
8. **Gelaye B,** Lemma S, Deyassa N, Bahretibebe Y, Tesfaye M, Berhane Y, Williams MA. Prevalence and correlates of mental distress among working adults in Ethiopia Clin Pract Epidemiol Ment Health 2012; 8:126-33.