1. a. Full Title: The SF-12 vs CES-D: A Comparison of Mental Health and Depression Scales in the ARIC Cohort Study

b. Abbreviated Title (Length 26 characters): Comparing the SF-12 and CES-D

2. Writing Group:
Writing group members: Jason Rotter, Sally Stearns, Bryce Reeve, Benjamin Capistrant, Kenneth Butler, Others welcome

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _____ [please confirm with your initials electronically or in writing]

First author: Jason Rotter
Address: 1104C McGavran-Greenberg Hall
UNC Gillings School of Global Public Health
135 Dauer Drive, Campus Box 7411
Chapel Hill, NC 27599-7411

Phone: 908-489-3792  Fax: 919-966-3671
E-mail: jrotter@live.unc.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
Name: Sally Stearns
Address: 1104C McGavran-Greenberg Hall
UNC Gillings School of Global Public Health
135 Dauer Drive, Campus Box 7411
Chapel Hill, NC 27599-7411
Phone: 919-843-2590  Fax: 919-966-3671
E-mail: sally_stearns@unc.edu

3. Timeline:
All data are currently available and the SF-12 measures have already been merged from Visit 5 and the GNC telephone interview. We anticipate beginning dataset construction and analyses in the Fall of 2016, with full co-author review by early 2017.
4. **Rationale:**

Depression is among the most common and least understood health conditions in the US today. Major depressive disorder (MDD) and the less severe persistent depressive disorder (PDD) (also called dysthymic disorder) account for the majority of clinically diagnosed chronic depression. Nearly 7% of the US adult population suffered from an episode of MDD in the last year (1.5% PDD).\(^1\) The lifetime risk of MDD is approximately 17%.\(^2\) The National Institutes for Mental Health (NIMH) has argued that the nature of the condition itself increases the likelihood of non-response (to surveys, clinical studies, or general practice), suggesting a substantial underreporting of underlying disease.\(^3\) People with depression are at increased risk of cognitive impairment, decreased functional status, and mortality from suicide and cardiovascular events.\(^4\) Direct and indirect economic costs of MDD alone are estimated north of $200 billion annually\(^5\), well above similar estimates for cancer ($125 billion\(^6\)) and just below diabetes ($245 billion\(^7\)).

Identifying patients with depression and mental illness remains a challenging target both in clinical care and investigator-driven research. Information on depression is most often collected through the use of patient-reported outcomes (PROs), commonly through pre-validated and standardized survey questionnaires. Among the many measures that can be used to assess mental health and depression are the general health Short Form-12 (SF-12) and the depression-specific CNAME (CES-D). Prior studies have validated both of these questionnaires across a range of different populations.

ARIC included both the SF-12 and CES-D questionnaires as part of its most recent follow-up visit, visit 5 (2011-2013). The SF-12 is a shortened version of the more
comprehensive SF-36, and ARIC used a ‘shortened’ 11 question version of the full CED-D. To our knowledge, the SF-12 and shortened CES-D have never been compared in a community-based cohort of older adults such as ARIC. Both for clinical research and clinical practice, it is important to know how these measures perform individually and collectively in a community setting. This study looks to compare the SF-12 and shortened CES-D to assess depression. The study will report descriptive differences in observed depression based on scale scores, will evaluate each scale’s psychometric properties for reliable use, and will assess the degree to which each scale is associated with diagnosed depression from hospital claims, outpatient ambulatory care, or medication use. Understanding the benefits and limitations of the SF-12 and CES-D as they relate to mental health and depression in a community-based sample will inform future research regarding the relative value of the scales.

5. **Main Hypothesis/Study Questions:**

1. What is the observed level of mental health as measured by the SF-12 and CES-D in the ARIC cohort, and how does their identification of depression differ?

2. What are the psychometric properties of the SF-12 and CES-D that make them valid and reliable estimates of the underlying depression construct in this sample?

3. What is the association between the SF-12 and CES-D, administered in-person at visit 5, with measures of clinical diagnosis of or treatment for depression?
   a. For hospital/ambulatory care depression diagnoses based on claims codes within 3, 12 months
   b. For patients taking anti-depressive medications as listed in their part D file

The focus of this work is descriptive and hypothesis generating.
6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Target Sample

For these analyses we will restrict our sample to participants who have completed both the SF-12 and the CES-D. We expect this sample to include approximately 5000-6000 individuals. For analyses using ambulatory and Part D claims, the subsample will be restricted to those not in HMO plans (~60%) and on Part D (~70%), respectively.

Outcomes and methods

The SF-12 consists of 12 questions, which sum to a Mental Component Score (MCS), which is a weighted, normalized composite score representing general mental health, with higher scores indicating better health. The SF-12 is comprised of 8 subdomains, 4 of which are considered structurally associated with the MCS: vitality, social functioning, role emotional, and mental health. The last sub-domain, mental health (MH), is comprised of questions that may be more strongly associated with depression. In the analyses, both the MCS and MH scales will be compared to the CES-D.

The shortened CES-D consists of 11 questions and is summed (unweighted) to provide a composite score, with higher scores indicating more severe forms of depression. It consists of 4 subdomains: Depression, Positive, Somatic, and Interpersonal.

The CES-D is used frequently in clinical practice to diagnose depression. The full-length version has established ‘cut-offs’ for further evaluation. For the shortened 11-item version, a score of >8 is suggested as optimal. The SF-12 is not as routinely used in clinical practice, but because of its widespread clinical research use, researchers have
modeled its sensitivity and specificity across various thresholds. For the purpose of this study, to allow for comparison with the CES-D, we use a suggested normalized SF-12 cut-off value of <45.11

Analyses

Scores from the SF-12 and CES-D will be reported as mean values and for individual subdomains. Using the suggested threshold values, prevalence estimates for depression will be compared between the scales.

Next, we will assess both scales’ psychometric properties to help understand the extent to which either is accurately measuring its purported underlying constructs. A confirmatory factor analysis will be completed to test the construct structure of the scales. Additionally, we will complete tests of reliability (internal consistency) using Chronbach’s coefficient alpha for overall component scores and subdomains, where applicable. Convergent, discriminant, and known groups validity will also be tested using simple correlation analysis and ANOVA for the known-group of low-SES, female, and comorbid diabetes (education will be used as a proxy for SES). Each of the tests above will be performed separately for both scales and compared.

For the primary analysis we will compare both scales against known cases of hospitalization, ambulatory care, and medication use. This helps us understand how well each measure correlates with clinical diagnoses (as presented in the claim), and may also provide a rough estimate of the ‘undiagnosed’ population in this setting. Using ARIC surveillance and CMS merged hospitalizations, individuals coded for depression according to the AHRQ comorbidity ICD-9 algorithm will be identified. We will look at those hospitalized with a diagnoses for depression (in any position) at 3, 6, and 12
months (prior to or following) Visit 5. We will report mean values and distributional plots, as well as threshold-based sensitivity and specificity estimates from both measures. The same procedure will be completed for ambulatory claims (with a reduced sample from non-HMO only participants [~60% of full sample]).

Separately, we will examine anti-depressive medication use in this sample (~70% on Part D) both prior to and immediately following visit 5. An indication of depressive symptoms from medication use alone, even more strongly than inpatient or outpatient claims requires careful interpretation, as the medication’s effect as well as its many indications can introduce conflicting estimates of a person’s underlying level of disease as represented by a scale, like the SF-12 or CES-D.

We will examine each of these claims-based indicators separately and together in comparing the SF-12 and CES-D. McNemar’s chi-squared test for paired data can be used to appropriately account for correlation within individuals.

**Limitations**

This proposed assessment has several limitations. The ARIC sample alive and present at visit 5 may suffer from a healthy participant bias, limiting generalizability to the community-based population. The appropriate ‘thresholds for depression’ for the two scales are also open to interpretation. Modeling has helped inform the ‘optimal’ sensitivity and specificity of some of these estimates, as well as experience in clinical practice, but it is not clear that one should consider a score of 46 on the SF-12 MCS as conclusively disease-free, nor a 44 conclusively as depression. Doctors and their patients make individualized decisions based on history and competing risk factors. This analysis
looks simply to describe how two scales, if used in cohort studies such as this or in clinical practice, may perform relative to each other, on average.

The claims-based analysis also has limitations. Depression may be specifically associated with increased hospitalizations, for example, but we may not conclude that a lack of hospitalization, or missed code indicates a lack of disease. This ‘missingness’ which represents the true underlying levels of disease for those that did not present or have coded depression limits the interpretability of the ‘specificity’ measurements from both scales. This is true as well for ambulatory or medication use, though we expect these measures to be less ‘severe’ cases, and therefore more comprehensive. The temporality and episodic nature of depression is also a limitation. Depressive symptoms can last hours, days, or months. Associations we find between these scales and other clinical diagnosis measures will not be perfect across any meaningful length of time, but we expect to see stronger effects (correlations) as we shorten the time frame (12 vs 3 months).

With cautions about the limitations, we believe this study can contribute meaningful insight into the performance of two commonly used questionnaires in a sample of older adults. We believe this analysis is the first to systematically compare these two specific measures of depression and, as such, can provide valuable information to future clinical research and health services research studies.

7.a. Will the data be used for non-CVD analysis in this manuscript? ___X__ Yes  ____ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and
for DNA analysis RES_DNA = “CVD Research” would be used? __X__ Yes  
____ No  
(This file ICTDER has been distributed to ARIC PIs, and contains  
the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript?  
____ Yes  ___X__ No

8.b. If yes, is the author aware that either DNA data distributed by the  
Coordinating Center must be used, or the file ICTDER03 must be used to  
exclude those with value RES_DNA = “No use/storage DNA”?  
____ Yes  ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing  
ARIC Study manuscript proposals and has found no overlap between this  
proposal and previously approved manuscript proposals either published or still  
in active status. ARIC Investigators have access to the publications lists under the  
Study Members Area of the web site at:  http://www.csc.unc.edu/ARIC/search.php

___X___ Yes  _______ No

10. What are the most related manuscript proposals in ARIC (authors are  
encouraged to contact lead authors of these proposals for comments on the new  
proposal or collaboration)?

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use  
any ancillary study data? ____ Yes  ___X__ No

11.b. If yes, is the proposal
____  A. primarily the result of an ancillary study (list number* ______________)  
____  B. primarily based on ARIC data with ancillary data playing a minor  
role (usually control variables; list number(s)* ______________ ______________  
____________________)

*ancillary studies are listed by number at  http://www.csc.unc.edu/aric/forms/

12a. Manuscript preparation is expected to be completed in one to three years. If a  
manuscript is not submitted for ARIC review at the end of the 3-years from the date  
of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the  
public has access to the published results of NIH funded research. It is your  
responsibility to upload manuscripts to PUBMED Central whenever the journal does  
not and be in compliance with this policy. Four files about the public access policy  
Publications, Policies & Forms. [http://publicaccess.nih.gov/submit_process_journals.htm](http://publicaccess.nih.gov/submit_process_journals.htm) shows you which journals automatically upload articles to Pubmed central.

13. Per Data Use Agreement Addendum for the Use of Linked ARIC CMS Data, approved manuscripts using linked ARIC CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript _X___ Yes _____ No.
REFERENCES


Lungenhausen, Margitta, et al. Randomised controlled comparison of the Health Survey Short Form (SF-12) and the Graded Chronic Pain Scale (GCPS) in telephone interviews versus self-administered questionnaires. Are the results equivalent? BMC medical research methodology 7.1 2007: 50.


Ware, John E., Mark Kosinski, and Susan D. Keller. SF-12: How to score the SF-12 physical and mental health summary scales. Health Institute, New England Medical Center, 1995.
2Anxiety and Depression Association of America. Depression. October 2015. Available at: http://www.adaa.org/understanding-anxiety/depression