Online Supplement

Table of Contents

Final 10-item Version of the BASE© - Eating Disorder Screen Only

Copyright Note: The Eating Pathology Symptoms Inventory (EPSI) was developed by Dr. Kelsie Forbush and published in *Psychological Assessment* in 2013. Prior to publication, Dr. Forbush obtained a United States copyright to the EPSI and all derivative versions (which includes translations, shortened versions, and format adaptations). This means that Dr. Forbush, who has the authority to grant or deny any modifications to the original version of the EPSI, retains the copyright to the EPSI and any derivative or translated version of the EPSI, including the BASE. Please note that changing the instructions or deviations from the paper-and-pencil format (e.g., administering the survey over the Internet or through a mobile-health app) are considered modifications. Unauthorized translations or modifications would, therefore, represent copyright infringement. To request permission to modify the EPSI or BASE, please contact kforbush@ku.edu. Please note that a longer version of the BASE exists that includes screening items for depression, generalized anxiety disorder, and post-traumatic stress disorder.

BASE-10©

Below is a list of experiences and problems that people sometimes have. Read each item to determine how well it describes your recent experiences. Then select the option that best describes how frequently each statement applied to you during the PAST FOUR WEEKS, including today.

0	1	2	3	4
Never	**Rarely**	**Sometimes**	**Often**	**Very Often**

1. People told me that I do not eat very much 1._____
2. I felt that I needed to exercise nearly every day 2._____
3. I used muscle building supplements 3._____
4. I did not like how my body looked 4._____
5. I ate until I was uncomfortably full 5._____
6. I made myself vomit in order to lose weight 6._____
7. I engaged in strenuous exercise at least five days per week 7._____
8. I stuffed myself with food to the point of feeling sick 8._____
9. I used laxatives, diet pills, or diuretics to lose weight 9._____
10. I used medications or substances (such as nicotine) to reduce hunger or lose weight 10._____

Table s1. Summary of Strengths and Limitations of Common Eating Disorder Screening Measures

Measure	Items	Initial Development Sample	Psychometrics	Strengths	Limitations
SCOFF	5	N=212 women with AN (n=68) or BN (n=48) (Morgan et al., 1999).	.86 (sensitivity), .83 (specificity), and .91 (AUC) in women (Kutz et al., 2020).	Accurate at distinguishing individuals with EDs from those without. The SCOFF is brief and easy to score, making it ideal for healthcare settings.	Sensitivity is lower in samples that include men (Liu et al., 2015; Solmi et al., 2015). Samples with greater representation of BED found decreased accuracy (Kutz et al., 2020).
Weight Concerns Scale (WCS)	5	N=877 community girls and women with body image concerns or dieting history (Killen et al., 1994).	.72 (sensitivity), and .80 (specificity) for in college women (Graham et al., 2019). Moderate accuracy in college women (AUC=.64) (Jacobi et al., 2011).	Accurate at identifying girls and women at risk for EDs (Graham et al., 2019). Brief and easy to implement in healthcare settings.	Lower sensitivity for identifying OSFEDS, such as PD (.55) or sub-BN (.68) (Graham et al., 2019). Little coverage of male-based ED content. Scoring is relatively complex.
Eating Disorder Examination 7-item version (EDE-Q7)	7	N=801 young adults (71.5% female) with AN or BN (Grilo et al., 2015).	.67 (sensitivity), .88 (specificity), and .84 (AUC) in female adolescents and women (Machado et al., 2020).	Briefer than the full EDE-Q, while retaining specificity and high AUC values.	Lower sensitivity than the full EDE-Q. Sensitivity was .67 for the EDE-Q7 compared to .73 for the full EDE-Q (Machado et al., 2020).

Eating Disorder Diagnostic Scale (EDDS)	22	N=367 female adolescents and women recruited from clinical trials (n=32), inpatient psychiatric unit (n=3), and two longitudinal studies of EDs (n=294), and a multi-site study on affect (n=38) (Stice et al., 2000).	AN: .93-.95 (sensitivity), and 1.00 (specificity). BN: .81-1.00 (sensitivity) and .97-.98 (specificity). BED: .57-.77 (sensitivity) and .96-.99 (specificity) in female youth and women. Excellent accuracy (AUC=.93-.99) in original validation sample and in a clinical sample Dutch women (Krabbenborg et al., 2012; Stice et al., 2000).	High predictive validity in assessing a likely ED diagnosis	Lengthy assessment. Somewhat complex to score. Less favorable psychometrics for identifying BED (Stice et al., 2000).
Eating Attitudes Test-26 (EAT-26)	26	N=300 of women (n=160 young adult women with AN and n=140 community-recruited young women; Garner et al., 1982).	.26 (sensitivity), .95 (specificity), and .70 (AUC) in a community sample of Spanish women (N=778) (Rivas et al., 2010).	Useful for identifying AN and BN. Substantial history of use with normative data available. Easy to score.	Less accurate for OSFED and BED (Orbitello et al., 2006; Scheinberg et al., 1993). Lower AUC (.62) in a sample of young adult women with BN or BED (Siervo et al., 2005).
Eating Disorder Examination-Questionnaire (EDE-Q)	28	N=36 women who had AN (n=23) or BN (n=13) (Fairburn & Beglin, 1994).	.83 (sensitivity) and .96 (specificity) in a community sample of women (Mond et al., 2004).	Performs well in samples of women. Long history of use with substantial normative data available.	Screening is less accurate in men (sensitivity of .77 and specificity of .77; Schaefer et al., 2018) and for some presentations of EDs (sensitivity of .73 and

					specificity of .81 for detecting BED; Vander Wal et al., 2011). Item instructions change across the measure, increasing administration length. Scoring is less straightforward than other screening tools.
Stanford-Washington Eating Disorder Screen (SWED)	31	N=549 college-aged women (Graham et al., 2019).	.72 (sensitivity), .80 (specificity), and .94 (AUC) in college-aged women.	Detects a broader range of ED diagnoses than most screeners.	Sensitivity is lower for sub-BN and PD; lack of data in male-identifying and minoritized groups (Graham et al., 2019). Lengthy and somewhat complex to score.

Note. This table provides a brief review of commonly used eating-disorder screening measures. For an excellent, comprehensive review of eating-disorder screening measures in college students, we refer the interested reader to Fitzsimmons-Craft, Karam et al., 2019. AN=anorexia nervosa; BN=bulimia nervosa; BED=binge eating disorder; ED=eating disorder; PD=purging disorder; OSFED=other specified feeding and eating disorder.

Table s2. Eating-Disorder Behavior Frequencies in Past 3 Months

	n (%)			
	Total (n = 596)	Men[g] (n = 146)	Women (n = 406)	Another gender identity (n = 32)
Objective Binge Episodes[a]				
0 times	389 (66.0)	97 (66.9)	262 (65.5)	22 (68.8)
1-11 times	78 (13.2)	16 (11.0)	58 (14.5)	3 (9.4)
12+ times	122 (20.7)	32 (22.1)	80 (20.0)	7 (21.9)
Subjective Binge Episodes[b]				
0 times	223 (38.3)	63 (44.1)	146 (36.8)	8 (25.8)
1-11 times	183 (31.4)	40 (28.0)	127 (32.0)	15 (48.4)
12+ times	177 (30.4)	40 (28.0)	124 (31.2)	8 (25.8)
Self-Induced Vomiting[c]				
0 times	523 (88.2)	137 (93.8)	360 (89.3)	18 (56.3)
1-11 times	42 (7.1)	7 (4.8)	24 (6.0)	9 (28.1)
12+ times	28 (4.7)	2 (1.4)	19 (4.7)	5 (15.6)
Laxatives or Diuretics[d]				
0 times	539 (91.0)	143 (97.9)	361 (89.6)	25 (80.6)
1-11 times	25 (4.2)	2 (1.4)	22 (5.5)	0 (0.0)
12+ times	28 (4.7)	1 (0.7)	20 (5.0)	6 (19.4)

Fasting (Skipping at least 2 meals in a row)[e]				
0 times	291 (49.1)	94 (64.4)	187 (46.4)	6 (18.8)
1-11 times	76 (12.8)	16 (11.0)	57 (14.1)	2 (6.3)
12+ times	226 (38.1)	36 (24.7)	159 (39.5)	24 (75.0)
Excessive Exercise[f]				
0 times	356 (60.1)	94 (64.4)	237 (58.8)	18 (58.1)
1-11 times	81 (13.7)	14 (9.6)	63 (15.6)	2 (6.5)
12+ times	155 (26.2)	38 (26.0)	103 (25.6)	11 (35.5)

Note. [a] Objective binge episodes was missing for seven participants. [b] Subjective binge episodes was missing for 13 participants. [c] Self-induced vomiting was missing for three participants. [d] Laxatives or diuretics was missing for four participants. [e] Fasting was missing for three participants. [f] Excessive exercise was missing for four participants.[g] Men includes both cisgender and transgender men.

Results of the BASE-9

In this supplement, we presented results from the 9-item version of the BASE (BASE-9). We tested the BASE-9 and BASE-10 to inform whether the BASE-10 performed as well as or better than the BASE-9 using two-tailed tests, and whether the BASE-9 performed as well as or better than the SCOFF using two-tailed tests. We reported *p*-values unadjusted by multiple comparison correction methods.

BASE-9 Statistics

The BASE-9 had an ordinal alpha *r* of 0.81. Coefficient alpha for the BASE-9 was 0.72 [CI: 0.69-0.75]. Table s3 shows the means, medians, and standard deviations for the BASE-9 in comparison with the BASE-10 and the SCOFF. Table s4 shows predictive accuracy statistics (AUC, sensitivity, specificity, PPV, and NPV) from optimal ROC and PR curves for the BASE-9, BASE-10, and SCOFF in the full sample. Table s5 shows gender-specific statistics.

AUC Comparison Tests: BASE-9 and BASE-10

Using AUC comparison tests, we tested the AUC differences between BASE-9 and BASE-10 using a two-tailed test. In ROC curve analyses, we observed that the BASE-10 accounted for significantly more AUC than the BASE-9 in the full sample ($z = 2.07$, $p = .038$) before multiple comparison correction. There were no significant AUC differences between the BASE-9 and BASE-10 in women ($z = .935$, $p = .350$), but the BASE-10 accounted for marginally more AUC than the BASE-9 in men ($z = 1.85$, $p = .064$). Using the PR curve analyses, we observed that the BASE-10 accounted for more AUC than the BASE-9 in the full sample ($z = 2.037$, $p = .042$) before multiple comparison correction, but showed no differences in

women (z =1.325, p = .185) or men (z = 1.231, p = 0.218). Overall, the BASE-10 performed at least as well as the BASE-9, with some statistics indicating that the 10-item version performed better, particularly in men.

AUC Comparison Tests: BASE-9 and SCOFF

Using AUC comparison tests, we tested the AUC differences between the BASE-9 and SCOFF using a two-sided test. Using the ROC curve analyses, we observed that the BASE-9 and the SCOFF did not show evidence for AUC differences in the full sample (z = -.292, p =.770), nor in women (z = -1.139 p = .255) or men (z =1.534, p = .125). Using the PR curve analyses, we found that the BASE-9 and the SCOFF did not have significant AUC differences in the full sample (z = .095, p =.924) or in women (z =-1.539, p = .124), but the BASE-9 accounted for significantly more AUC than the SCOFF in men (z = 2.075, p = .038). Overall, we found that the BASE-9 performs at least as well as the SCOFF in the full sample and in women, but better than the SCOFF in men.

Table s3. Means, Medians, and Standard Deviations of Measures by Gender

Mean, Median (SD)

	Men (n = 146)		Women (n = 406)		Another gender identity (n = 32)	
	Mean (SD)	*Median*	*Mean (SD)*	*Median*	*Mean (SD)*	*Median*
SCOFF	1.06 (1.25)	1	1.45 (1.21)	1	2.03 (1.38)	2
BASE - 9-item	10.15 (5.73)	9.5	10.70 (5.46)	10	13.03 (6.67)	11.5
BASE - 10-item	10.42 (6.02)	10	11.21 (5.99)	10	14.03 (7.48)	13.5

Note. BASE = Brief Assessment of Stress and Eating

Table s4. Predictive Accuracy Statistics in Full Sample

	Statistics					
Full (n = 596)						
	*Cutoff-**PRC***	*AUC-**PRC***	*Sensitivity*	*Specificity*	*PPV*	*NPV*
SCOFF-optimal	0	0.626	0.937	0.425	0.431	0.935
SCOFF-1-cutoff	1	0.626	0.746	0.720	0.553	0.859
BASE - 9-item	7	0.630	0.952	0.425	0.435	0.951
BASE - 10-item	7	0.648	0.952	0.423	0.434	0.950
	*Cutoff-**ROC***	*AUC-**ROC***	*Sensitivity*	*Specificity*	*PPV*	*NPV*
SCOFF-optimal	1	0.797	0.746	0.720	0.553	0.859
BASE - 9-item	9	0.790	0.852	0.590	0.491	0.896
BASE - 10-item	9	0.799	0.868	0.577	0.488	0.904

Note: The cutoff score represents the highest value that an individual can score before being flagged as a likely ED case. For the rows representing the SCOFF with a cutoff score of 1, for instance, these statistics are associated with the assumption that anyone scoring a 2 or higher should be classified as an ED case.

Table s5. Predictive Accuracy Statistics by Gender

	Statistics					
Women (n = 406)						
	*Cutoff-**PRC***	*AUC-**PRC***	*Sensitivity*	*Specificity*	*PPV*	*NPV*
SCOFF-optimal	0	0.684	0.956	0.395	0.440	0.947
SCOFF-1-cutoff	1	0.684	0.793	0.701	0.569	0.872
BASE - 9-item	7	0.616	0.948	0.406	0.443	0.940
BASE - 10-item	7	0.633	0.948	0.402	0.441	0.940
	*Cutoff-**ROC***	*AUC-**ROC***	*Sensitivity*	*Specificity*	*PPV*	*NPV*
SCOFF-optimal	1	0.810	0.793	0.701	0.569	0.872
BASE - 9-item	9	0.783	0.844	0.587	0.504	0.883
BASE - 10-item	11	0.787	0.733	0.697	0.547	0.840
Men (n = 140)						
	*Cutoff-**PRC***	*AUC-**PRC***	*Sensitivity*	*Specificity*	*PPV*	*NPV*
SCOFF-optimal	Undefined	0.354	1.000	0.000	0.207	0.000
SCOFF-1-cutoff	1	0.354	0.517	0.775	0.375	0.860
BASE - 9-item	6	0.561	1.000	0.405	0.305	1.000
BASE - 10-item	6	0.605	1.000	0.405	0.305	1.000
	*Cutoff-**ROC***	*AUC-**ROC***	*Sensitivity*	*Specificity*	*PPV*	*NPV*
SCOFF-optimal	0	0.710	0.828	0.514	0.308	0.919

			Statistics			
SCOFF-1-cutoff	1	0.710	0.517	0.775	0.375	0.860
BASE - 9-item	7	0.808	0.966	0.495	0.333	0.982
BASE - 10-item	7	0.821	0.966	0.495	0.333	0.982

Note: The cutoff score represents the highest value that an individual can score before being flagged as a likely ED case. For the rows representing the SCOFF with a cutoff score of 1, for instance, these statistics are associated with the assumption that anyone scoring a 2 or higher should be classified as an ED case.

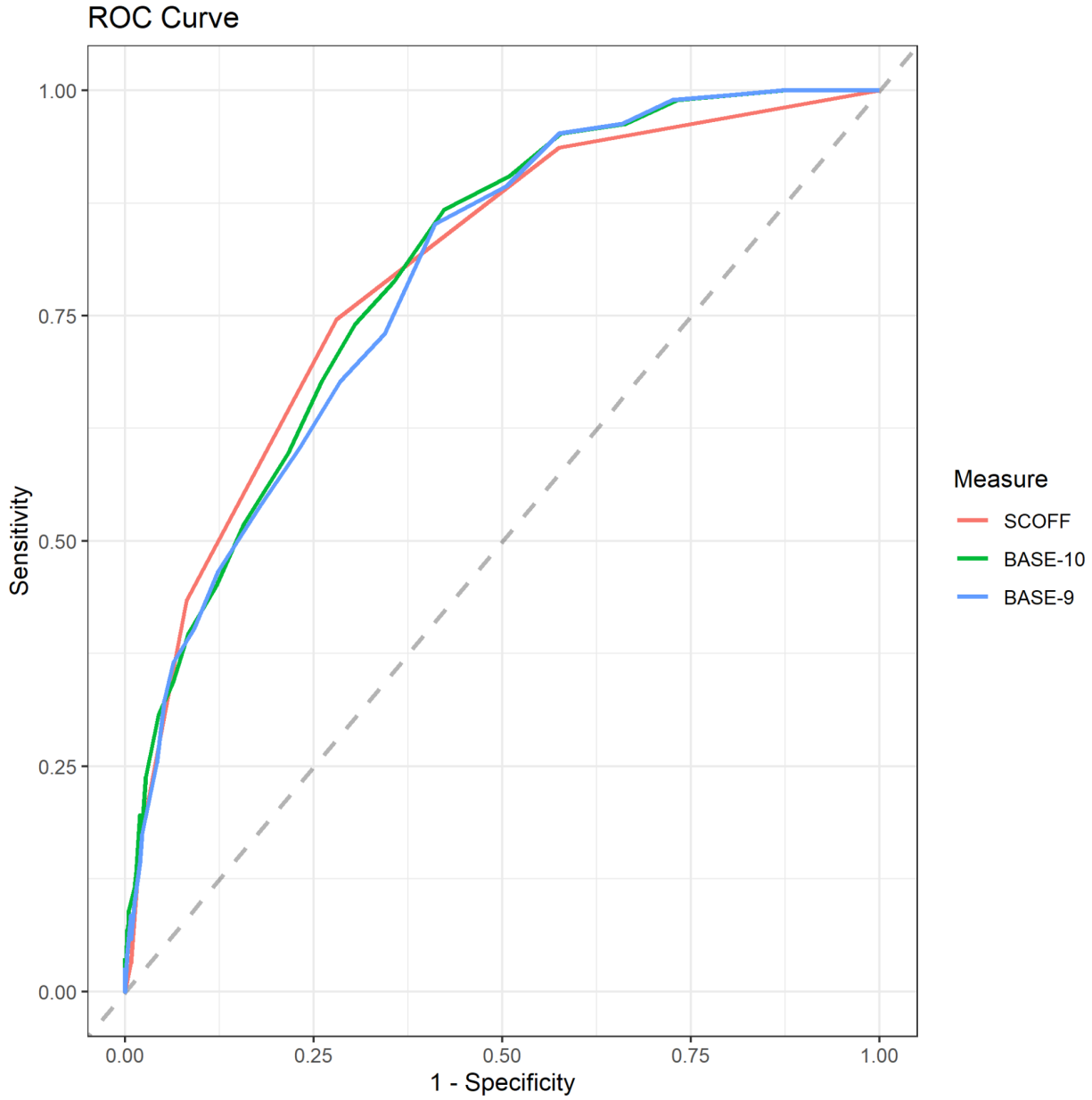

Figure s1. ROCs for each Measure in the Full Sample.

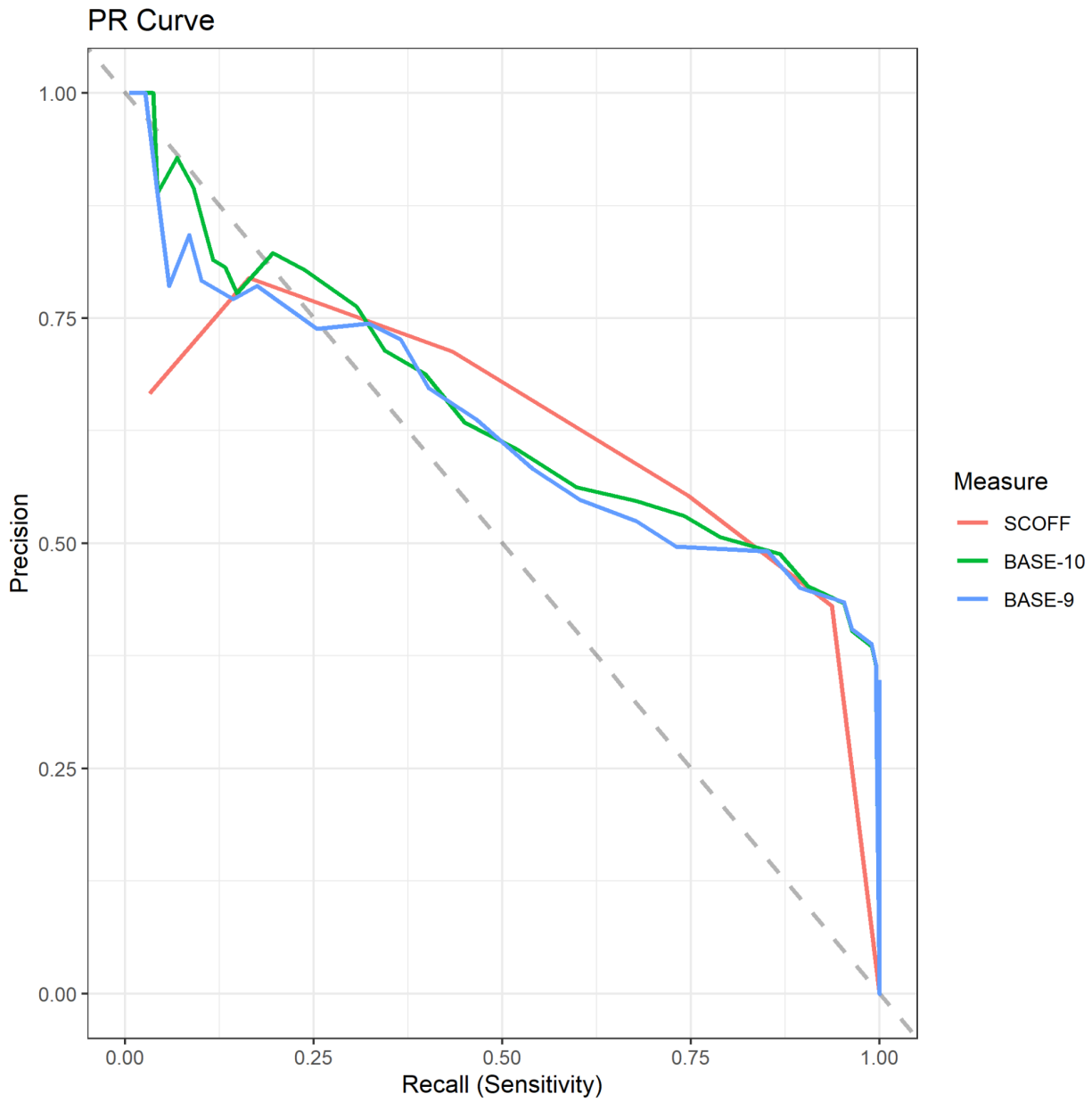

Figure s2. Precision-Recall Curves for each Measure in the Full Sample.

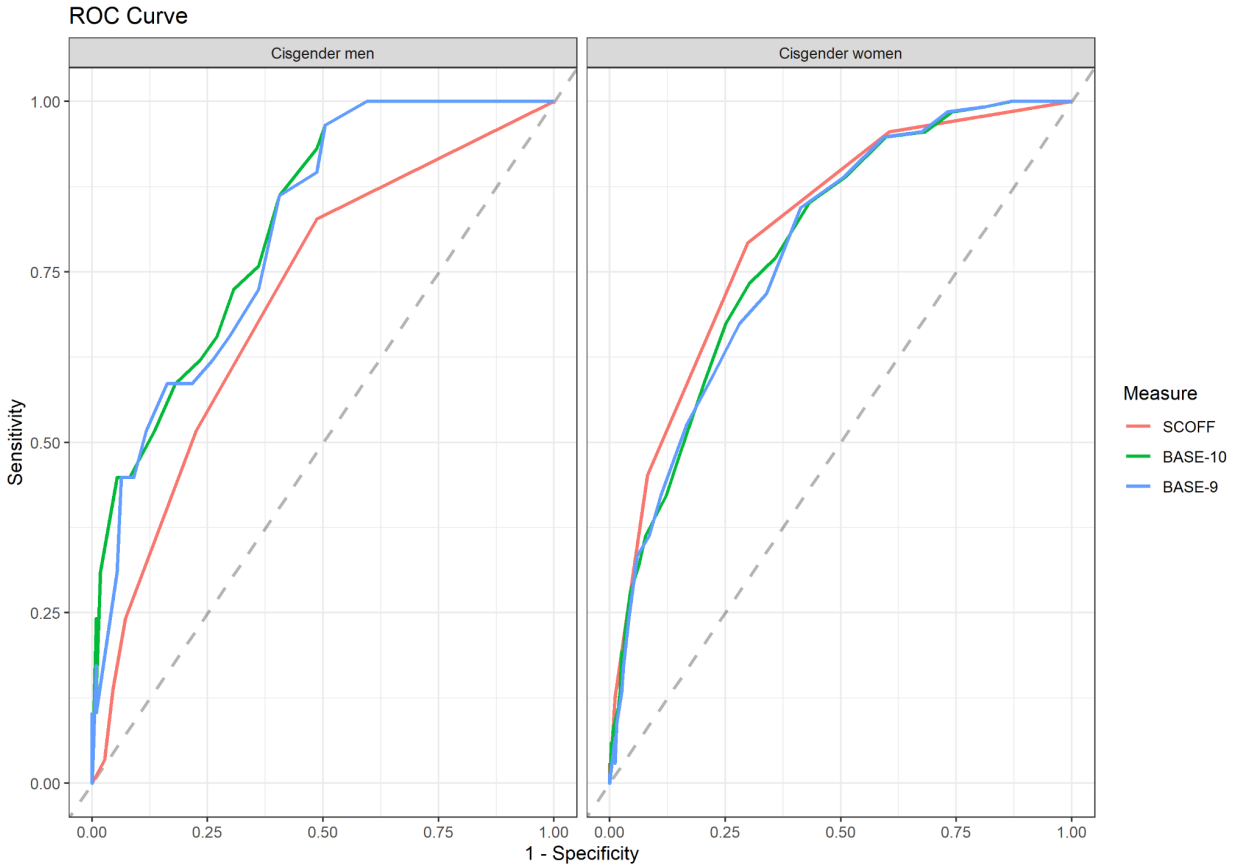

Figure s3. ROCs for each Measure in the Gendered Subsamples.

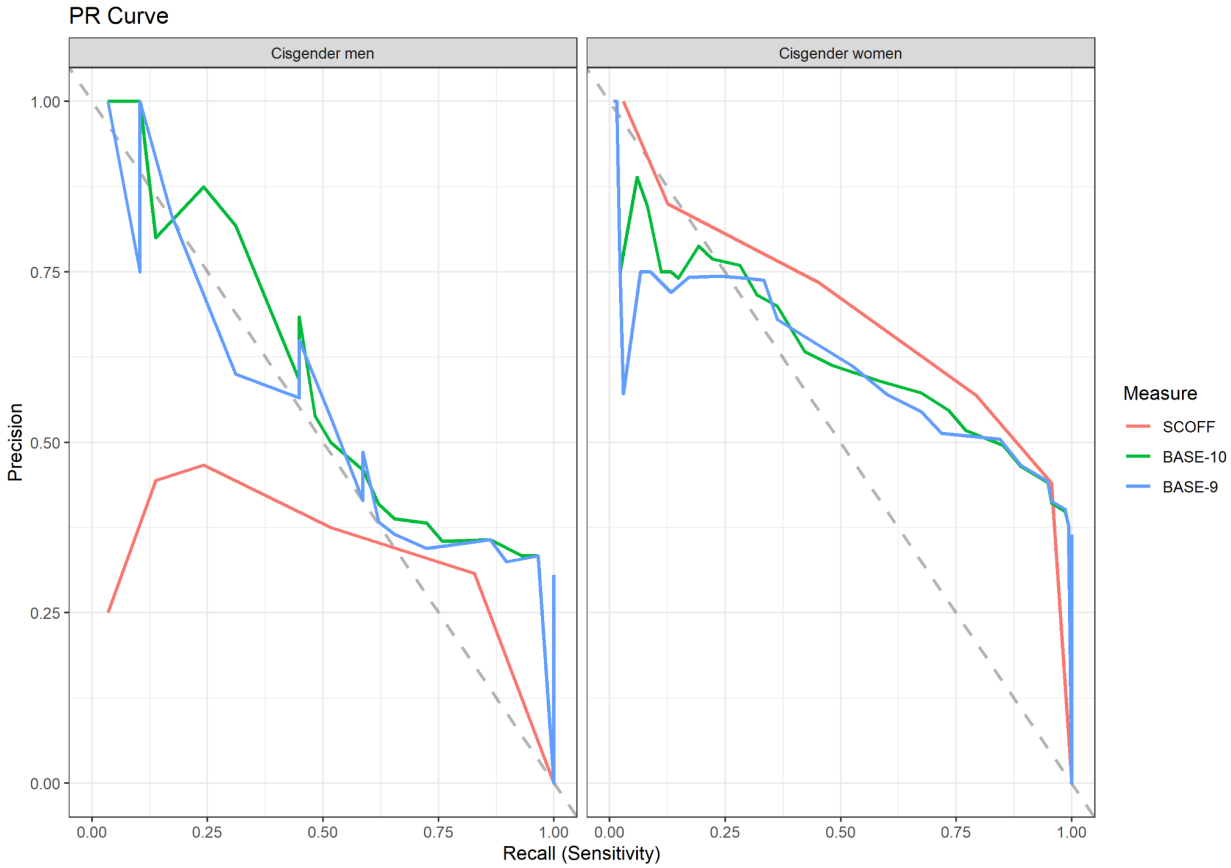

Figure s4. Precision-Recall Curves for each Measure in the Gendered Subsamples.