

### 1.5.1. Sampling and analysis framework<sup>1</sup>

From July 2006 to March 2007, data were collected from the U.S. general population and multiple disease populations. A sampling plan was developed for collecting responses to the candidate items from the targeted PROMIS domains. This plan was designed to accommodate multiple objectives: (1) obtain item calibrations for each domain; (2) estimate profile scores for various disease populations; (3) create linking metrics to legacy questionnaires (e.g., Short Form [SF]-36); (4) confirm the factor structure of the domains; and (5) conduct item and bank analyses.

Because of the large total number of items (> 1,000), it was unreasonable to ask participants to respond to the entire pool of items. We estimated that participants would respond to approximately four questions per minute and limited the maximum number of items administered to about 150, for an estimated average response time of 37 minutes.

Figure 1 outlines the two arms of the sampling design: “full-bank” and “block” administration. There were 14 candidate item banks (three physical functioning banks, anxiety, depression, anger, alcohol abuse, fatigue interference, fatigue experience, social role performance, social role satisfaction, pain interference, pain quality, and pain behavior). All 56 items for each of the two PROMIS candidate item banks (112 PROMIS items) were administered to a subset of individuals in the full-bank arm. They also completed appropriate “legacy” questionnaires (well-validated and widely used measures of the same concept). Another subset of the PROMIS wave 1 sample was administered blocks of seven items selected from each of the 14 candidate item banks (98 PROMIS items). All participants completed a clinical form consisting of approximately 25 auxiliary items measuring global health perceptions and sociodemographic variables, including age, income, number of hospitalizations, disability days, use of prescription medication, height, weight, gender, race/ethnicity, relationship status, educational attainment, and employment status. This clinical form also included a series of health questions about the presence and degree of limitations related to 25 chronic medical conditions: hypertension, angina, coronary artery disease, heart failure, heart attack, stroke or transient ischemic attack, liver disease, kidney disease, arthritis or rheumatism, osteoarthritis, migraines, asthma, chronic obstructive pulmonary disease, diabetes, cancer, depression, anxiety, alcohol or drug problems, sleep disorder, human immunodeficiency virus/acquired immunodeficiency syndrome, spinal cord injury, multiple sclerosis, Parkinson’s disease, epilepsy, and amyotrophic lateral sclerosis.

---

<sup>1</sup> from Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S, Cook K, Devellis R, DeWalt D, Fries JF, Gershon R, Hahn EA, Lai JS, Pilkonis P, Revicki D, Rose M, Weinfurt K, Hays R; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol.* 2010 Nov;63(11):1179–94. Epub 2010 Aug 4.

## PROMIS WAVE 1 DATA OVERVIEW (from promis\_w1\_final\_20100604 Rita Bode/Katy Wortman)

A. The final PROMIS Wave 1 data contains 8 variables used to describe a total of **21133** cases: **19601** from Polimetrix and **1532** from four PROMIS Network sites (UNC=540, Pittsburgh=252, Stanford=339 and Duke=401):

1. CaseID: subject unique ID
2. FORM: with forms A–W
3. FORMAT: with 1=full bank 2=block
4. SITE: polimetrix or site names for the clinical sites
5. SOURCE: with 1=polimetrix 2=clinical site
6. SAMPLE1: with 1=general population 2=clinical population
7. SAMPLE2:
  - 1=general population
  - 2=cancer (unknown type)
  - 3=heart disease
  - 4=OA
  - 5=RA
  - 6=psychiatric
  - 7=SCI
  - 8=COPD
  - 9=limited by disease (non-diabetes)
  - 10=limited by disease (diabetes)
  - 11=breast cancer
  - 12=colorectal cancer
  - 13=lung cancer
  - 14=prostate cancer
  - 15=medically capable with cancer

8. EXCLUDE: flag variable (exclude=1) to exclude subjects with:
  - a. Average response time per item < 1 sec (1000 milliseconds) (N=573).
  - b. 10 consecutive items with response time less than half a second (500 milliseconds) (N=192)

A total of **733** cases (32 overlapping with a and b) should be excluded from the final data analysis.

B. All item bank have been recoded with following directions:

- a. Global items: GLOBAL01-GLOBAL06, GLOBAL09—higher score, better health
- b. Global items: GLOBAL07 and GLOBAL08, GLOBAL10—higher score, more pain and fatigue
- c. Clinical items: higher score, more chronic conditions
- d. Physical functions (PFA, PFB, PFC): higher score, better physical function
- e. Pain (PAINQU, PAINBE, PAININ): higher score, more pain
- f. Fatigue (FATIMP, FATEXP): higher score, more fatigue
- g. Emotional distress (EDANX, EDDEP, EDALC, EDANG): higher score, more distress
- h. Social role function (SRPPER, SPRSAT): higher score, more social role function

C. Legacy items have not been recoded (to preserve their original forms). Thus, if you plan to analyze legacy items, you'll need to recode the following items (see Appendix B Legacy Scoring):

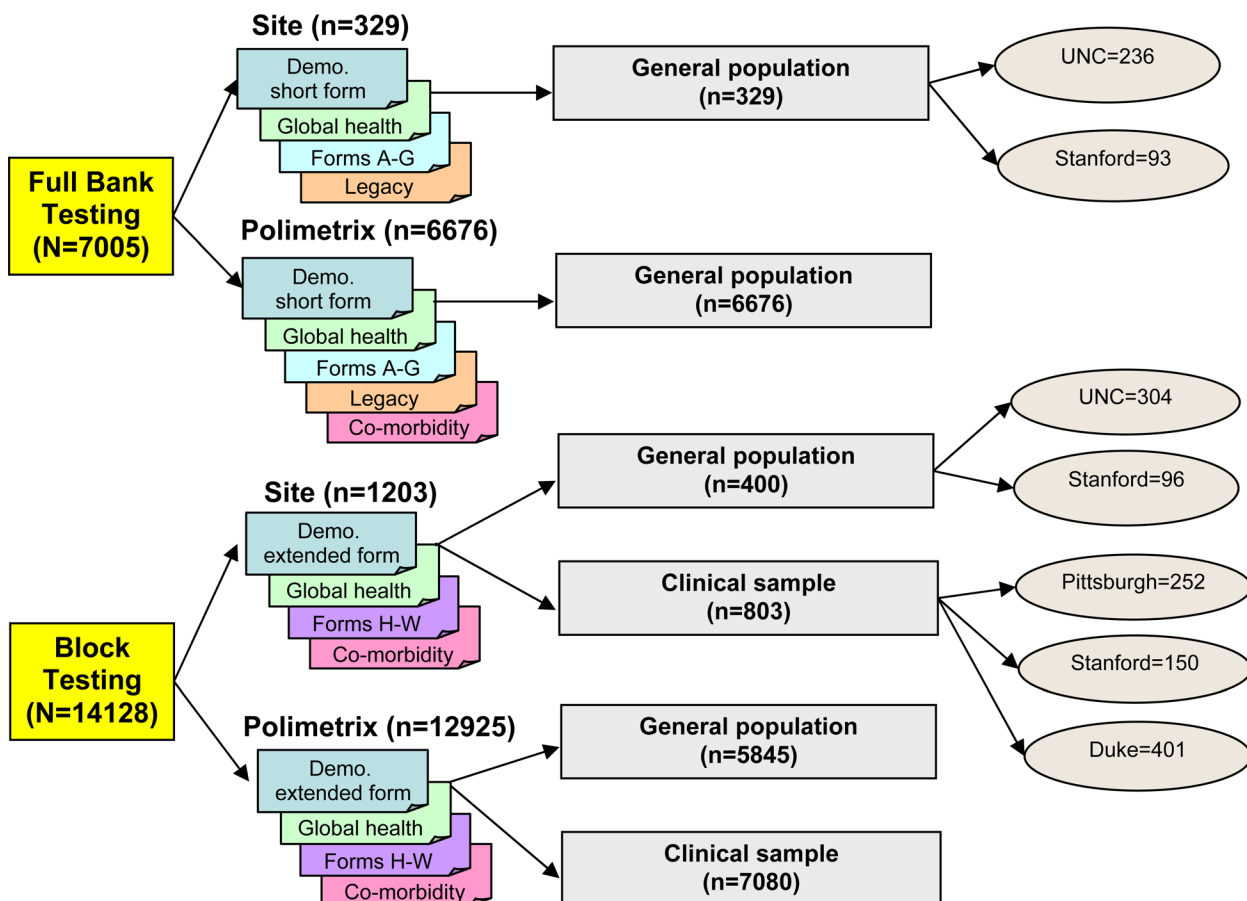
- a. MASQ (emotional distress—anxiety)—higher score, more distress (no recode needed)
- b. CESD-D (emotional distress—depression)—higher score, more distress (CESD4, CESD8, CESD12, CESD16 need to be recoded)

- c. AQ (emotional distress—anger)—higher score, more distress (AQ4 needs to be recoded)
- d. AUDIT (emotional distress—alcohol abuse)—higher score, more distress (no recode needed)
- e. HAQ (physical function—impairment)—higher score, lower physical function (no recode needed)
- f. BPI (pain)—higher score, more pain (no recode needed)
- g. SF-36 (health)—higher score, better health (HT, GH01, GH03, GH05, BP01, BP02, SF01, VT01, VT02, MH03, MH05 need to be recoded)
- h. FACIT-Fatigue (fatigue)—higher score, more fatigue (AN5, AN7 needed to be recoded)
- i. FACT-GP (social role function)—higher score, better social role function (no recode needed)

- D. Missing data have been coded as follows:
- a. .M = Missing (participant did not respond)
  - b. .S = Skipped due to branching
  - c. .N = Not administered due to study arm
  - d. .D = Don't know / not sure (participant response)
  - e. .O = Over age 89 (for socio02—age)

E. Final data include 7005 cases of full bank testing and 14128 of block testing. The sample summary is illustrated in Figure 1.

**Figure 1—PROMIS Sample Summary (N=21133)**



F. PROMIS data consist of Full Bank Testing and Block Testing, administered in 23 different forms with slight variation between versions administered by Polimetrix and PROMIS Network sites. Items included in each form are summarized in Tables 1 and 2.

**Table 1—PROMIS items by form (N=21133)**

DATA	FORM	ITEM	NO. ITEM	TOTAL SAMPLE
Full Bank (N=7005)	A	56 Anxiety and 56 Depression	128	925
	B	56 Alcohol Abuse and 56 Anger	128	1049
	C	56 Physical Function Part I and 56 PF Part II	128	942
	D	56 Fatigue Experience and 56 Fatigue Impact	128	1052
	E	56 SRC Performance and 56 SRC Satisfaction	128	1010
	F	56 Pain Interference and 56 Pain Quality	128	1032
	G	56 Physical Function Part III and 56 Pain Behavior	128	995
Block (N=14128)	H	Items 1–7 from each subdomain	7	978
	I	Items 4–10 from each subdomain	7	866
	J	Items 8–14 from each subdomain	7	1004
	K	Items 11–17 from each subdomain	7	809
	L	Items 15–21 from each subdomain	7	638
	M	Items 18–24 from each subdomain	7	821
	N	Items 22–28 from each subdomain	7	689
	O	Items 25–31 from each subdomain	7	784
	P	Items 29–35 from each subdomain	7	1110
	Q	Items 32–38 from each subdomain	7	974
	R	Items 36–42 from each subdomain	7	1122
	S	Items 39–45 from each subdomain	7	992
	T	Items 43–49 from each subdomain	7	653
	U	Items 46–52 from each subdomain	7	998
	V	Items 50–56 from each subdomain	7	647
W	Items 53–56 and 1–3 from each subdomain	7	1043	

**Table 2—Non-PROMIS items by study arm**

Study Arm	Full Bank: Forms A–G— PROMIS Network	Full Bank: Forms A–G— Polimetrix	Block: Forms H–W
Demographic—Short Form	11	11	11
Demographics—Extended (EQ-5D, criterion)			8
Global Health Status	10	10	10
Primary Co-Morbidity		21	21
Co-Morbidity Limitation		1–3	1–3
Average 3 Disease Specific Items/Disease		3–9	3–9
Max Legacy (range 15–36)	13–36	13–36	
TOTAL ITEMS	146–169	171–202	152–160

G. Demographics including all PROMIS and clinical sample are shown in Tables 3 and 4.

**Table 3 -PROMIS Sample Demographics (N=21133)**

<b>Demographic</b>	<b>ALL Sample (N=21133)</b>	<b>General Population (N=13250)</b>	<b>Clinical Population (N=7883)</b>
Age (Mean[S.E.])	53.2 (17.2)	50.6 (18.3)	57.4 (13.8)
Age group (%)			
18–29	11.8	16.5	4.1
30–39	12.3	15.2	7.6
40–49	15.8	17.0	13.9
50–59	21.4	17.9	27.2
60–64	10.4	7.5	15.3
65–84	27.1	24.7	31.2
85+	1.1	1.3	0.7
Gender (%)			
Male	48.0	44.8	53.3
Female	52.0	55.2	46.7
Ethnicity (%)			
Spanish/Hispanic/Latino origin*	8.8	11.8	3.8
White	82.3	78.1	88.6
Black	8.6	10.7	5.0
American Indian/Alaska Native	0.6	0.6	0.6
Asian	0.5	0.6	0.5
Native Hawaiian/Pacific Islanders	0.1	0.1	0.1
Multi-racial	8.2	9.9	5.2

\*Spanish/Hispanic/Latino origin can be of any race

**Table 4—PROMIS Clinical Sample Summary (N=7883)**

<b>Clinical Sample</b>	<b>N</b>	<b>%</b>
Cancer (unknown type)	1308	16.6
Heart disease	1156	14.7
OA	918	11.7
Rheumatoid Arthritis	557	7.1
Psychiatric	1193	15.1
SCI	531	6.7
COPD	1214	15.4
Limited by disease (non-diabetes)	497	6.3
Limited by disease (diabetes)	63	0.8
Breast cancer	106	1.3
Colorectal cancer	30	0.4
Lung cancer	10	0.1
Prostate cancer	155	2.0
Medically capable with cancer	145	1.8