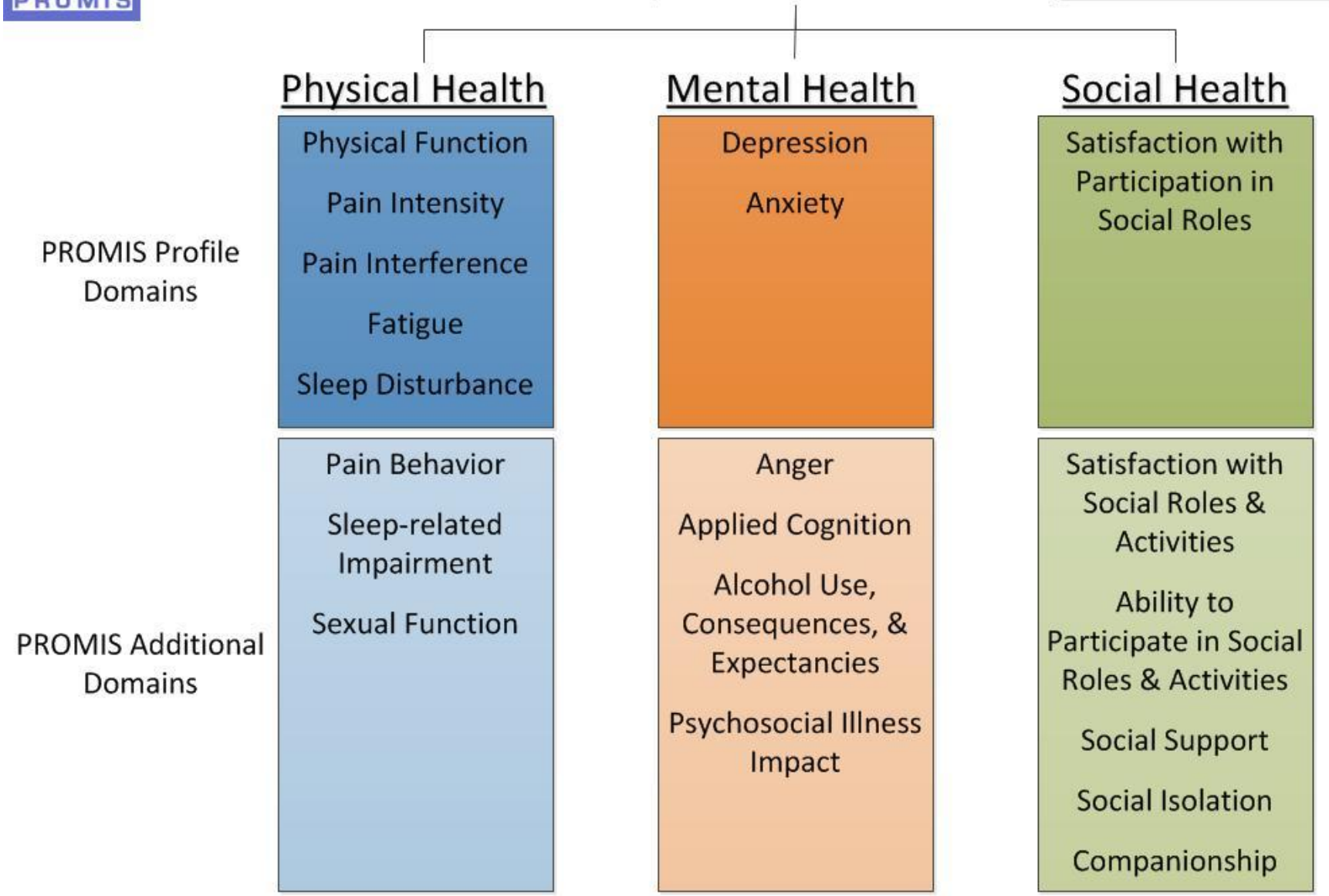
Sources of Comparability Between Probability Sample Estimates and Nonprobability Web Sample Estimates

William Riley, National Cancer Institute
Ron Hays, University of California Los Angeles (UCLA)
Robert Kaplan, National Institutes of Health
David Cella, Northwestern University of Feinberg School of Medicine
William Thompson, Centers for Disease Control
Catherine Okoro, Centers for Disease Control
Satvinder Dhingra, Northrop Grumman and Centers for Disease Control)

PROMIS INTRODUCTION

PROMIS Adult Self-Reported Health

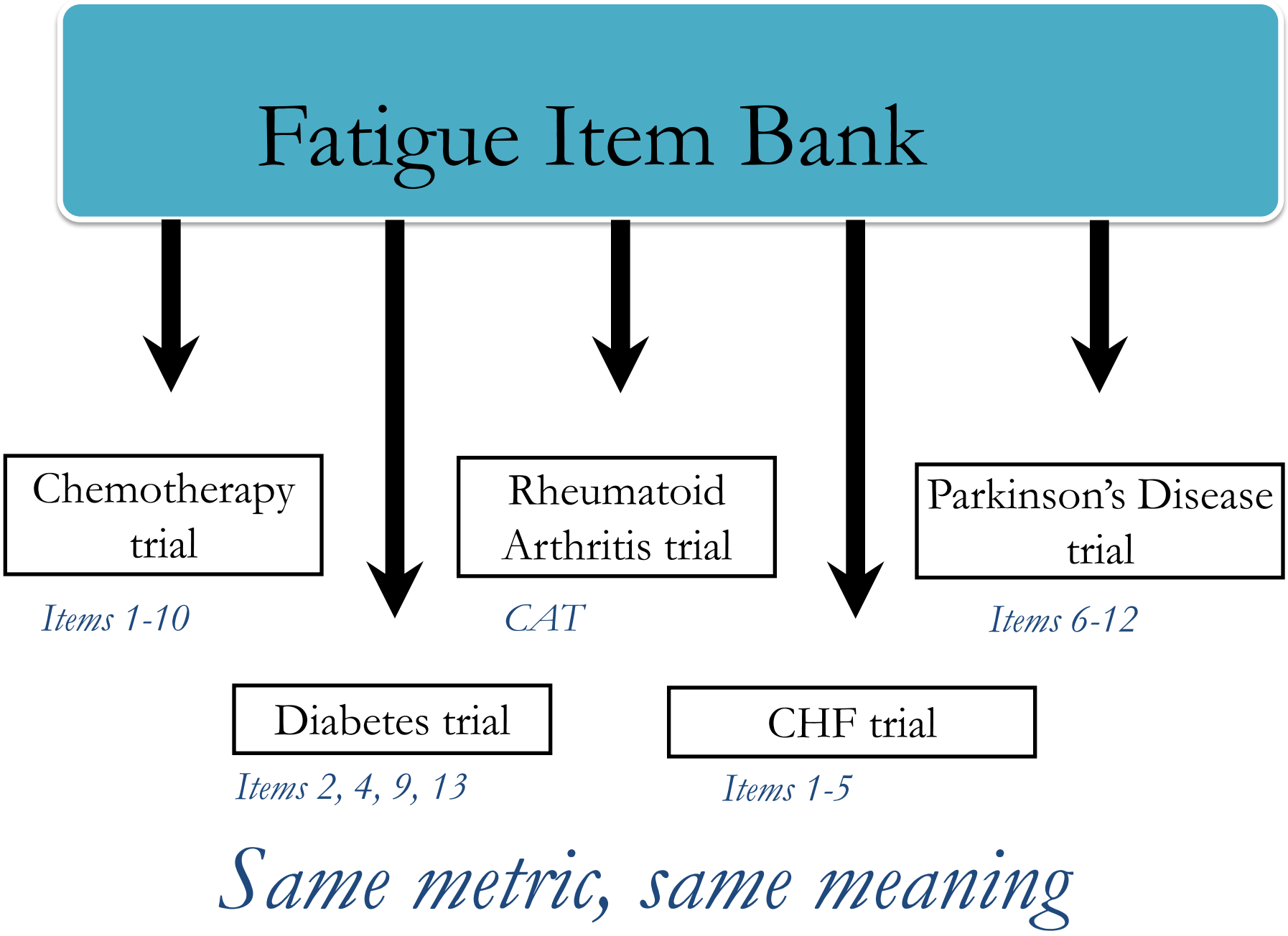
Global Health



Modern Psychometric Methodology

Item Response Theory and Computer Adaptive Testing

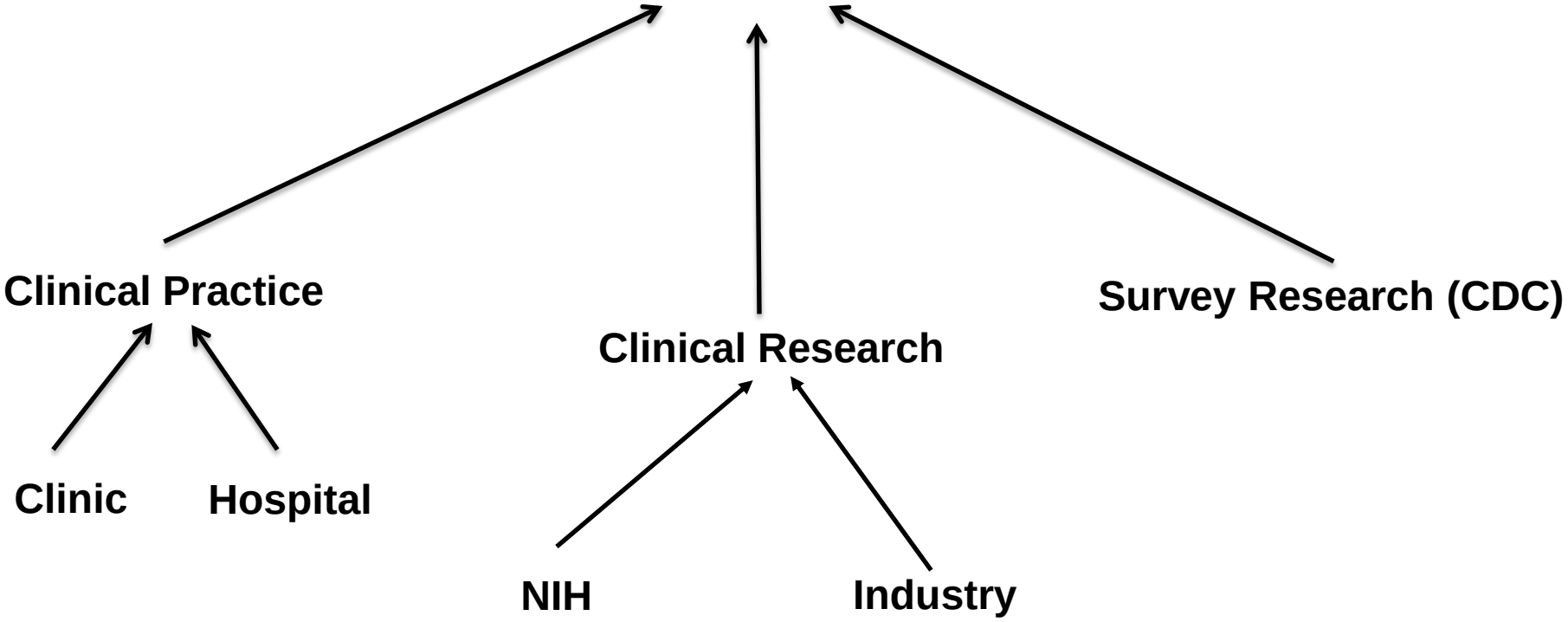
- Psychometric properties at the item level allow item banks that can be flexibly administered and further refined over time
- Improved efficiency (less respondent burden)
  - Automated administration, scoring and reporting
  - Tailored targeting of items
  - Item administration based on prior responses
- Improved precision
- Ability to crosswalk with existing instruments for comparability on the same metric



Same metric, same meaning

Advancing Patient-Centered Outcomes

PROMIS: A Common Source of PROs



Hays, R. D., Bjorner, J.B., Revicki, D. A., Spritzer, K. L., & Cella, D. (2009). Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. *Quality of Life Research, 18,* 873-880.

PROMIS GLOBAL ITEMS

PROMIS Global Items (10)

1. In general, would you say your health is . . .
2. In general, would you say your quality of life is . . .
3. In general, how would you rate your physical health?
4. In general, how would you rate your mental health?
5. In general, how would you rate your satisfaction with social activities and relationships?
6. To what extent are you able to carry out your everyday physical activities?
7. How would you rate your pain on average?
8. How would you rate your fatigue on average?
9. In general, please rate how well you carry out usual social activities and roles.
10. How often have you been bothered by emotional problems?

Sample (n = 21,133)

- You Gov Polimetrix nonprobability Internet panel, augmented by clinical samples
- Age: 18-100 (mean = 53)
- 52% Female
- 9% Latino/Hispanic, 9% black, 2% other
- 3% < high school, 16% high school only
- 59% Married
- 39% Working full-time

Physical Health
Item Parameters

Item	A	B1	B2	B3	b4
Global03	2.31	-2.11	-0.89	0.29	1.54
Global06	2.99	-2.80	-1.78	-1.04	-0.40
Global07	1.74	-3.87	-1.81	-0.67	1.00
Global08	1.90	-3.24	-1.88	-0.36	1.17

3. In general, how would you rate your physical health?
6. To what extent are you able to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries or moving a chair?
7. How would you rate your pain on average?
8. How would you rate your fatigue on average?

3: Poor, Fair, Good, Very Good, **Excellent**
6: Not at all, A Little, Moderately, Mostly, **Completely**
7: **No pain** (0) – Worse pain imaginable (10)
8: **None,** Mild, Moderate, Severe, Very Severe

Mental Health
Item Parameters

Item	A	B1	B2	B3	b4
Global02	2.41	-2.45	-1.32	-0.19	1.07
Global04	3.67	-2.31	-1.26	-0.33	0.67
Global05	2.98	-1.78	-0.90	-0.01	1.07
Global10	1.89	-2.82	-1.51	-0.25	0.99

2. In general, would you say your quality of life is …?
4. In general, how would you rate your mental health, including your mood and your ability to think?
5. In general, how would you rate your satisfaction with social activities and relationships?
10. How often have you been bothered by emotional problems such as feeling anxious, depressed or irritable?

2, 4, 5: Poor, Fair, Good, Very Good, **Excellent**
10: **Never,** Rarely, Sometimes, Often, Always

Physical and Mental Health (r = 0.63)

- Physical (α = 0.81)
  - r = 0.82 with EQ-5D
  - r = -0.75 (pain impact), -0.73 (fatigue), 0.71 (physical functioning), & -0.67 (pain behavior)

- Mental (α = 0.86)
  - r = 0.61 with EQ-5D
  - r = -0.71 (depression), - 0.65 (anxiety), & 0.60 (satisfaction with discretionary social activities)

General Health Rating Across Samples

Table 2

General Health Item Comparisons

Variable	N	Mean	Std Error of Mean	95% CL for Mean		90% confidence interval from equivalence testing vs. MEPS	vs. NHANES	vs. BRFSS
General Health (*5-Excellent, 4-Very Good, 3-Good, 2-Fair, 1-Poor*)								
2004 MEPS	20777	3.56	0.012	3.54	3.59			
2001-2002 NHANES	6873	3.50	0.017	3.47	3.54			
2005 BRFSS	352036	3.52	0.004	3.52	3.53			
PROMIS: General population (Unweighted)	11794	3.50	0.009	3.48	3.52	**−0.085, −0.035**	**−0.032, 0.032**	**−0.036, −0.004**
PROMIS: General population (weighted)	218022053	3.42	0.016	3.39	3.45	**−0.175, −0.108**	**−0.120, −0.043**	**−0.129, −0.074**
PROMIS: General population sub-sample	2196	3.53	0.020	3.49	3.57	**−0.068, 0.008**	**−0.013, 0.073**	**−0.024, 0.044**

Note: equivalencies are printed in boldface type.

Liu et al. Representativeness of PROMIS Internet Panel, J Clin Epi 2010;63: 1169-78

PROMIS GLOBAL AMONG DIFFERENT SAMPLING METHODOLOGIES

Sampling Methodologies

- **National Health Interview Survey** (2010 - a probability sample of households)
  - The National Health Interview Survey is a cross-sectional household interview survey.
  - Data are collected through a personal household interview conducted by interviewers employed and trained by the U.S. Bureau of the Census according to procedures specified by the NCHS.
  - Sampling and interviewing are continuous throughout each year.
  - The sampling plan is redesigned after every decennial U.S. census.

- **Health Styles** (2010 - a non-probability mail panel)
  - Porter Novelli contracted with Synovate
  - The sampling and data collection are conducted by Synovate, Inc.
  - The Synovate, Inc. consumer mail panel contains approximately 200,000 potential respondents.

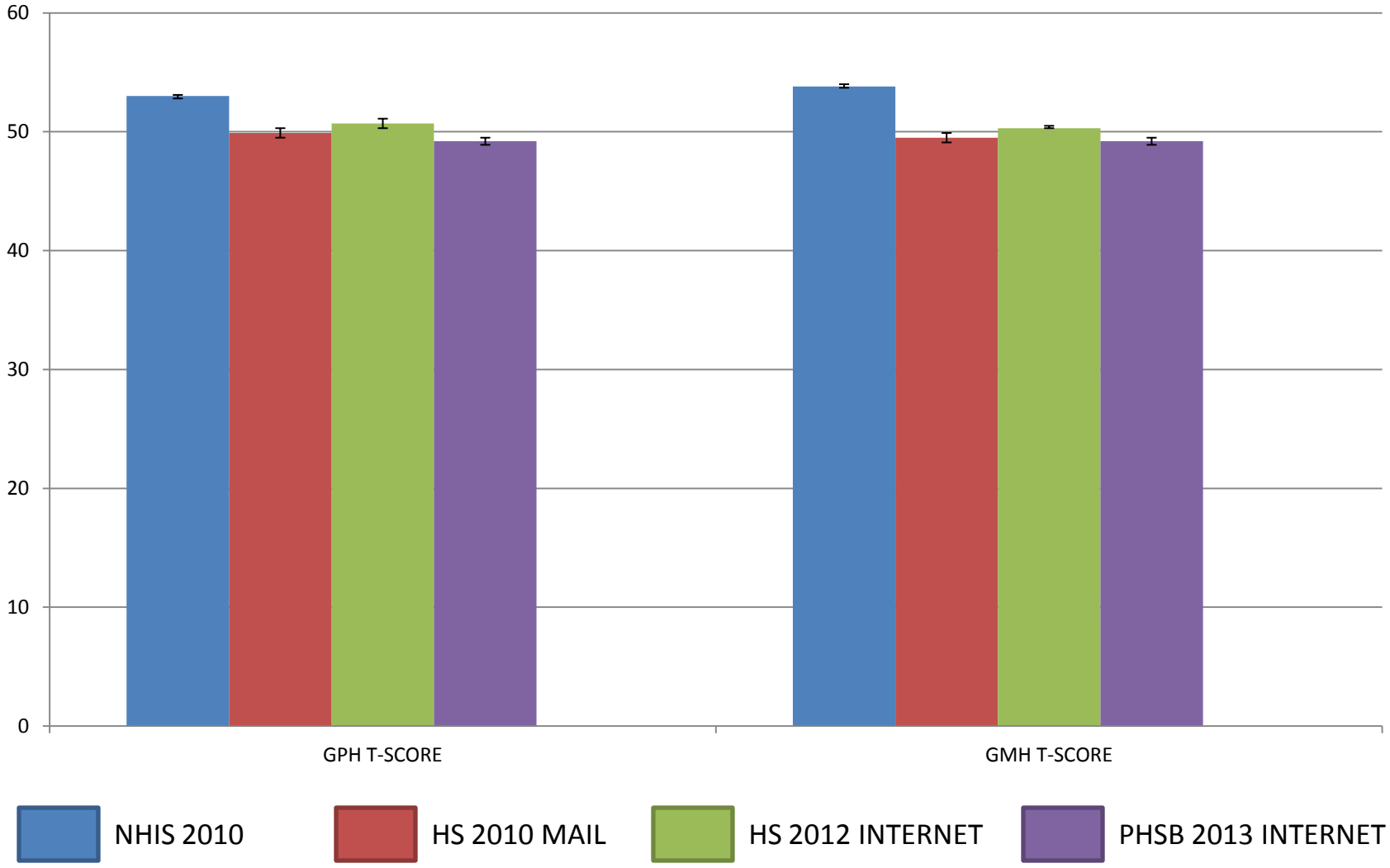
- **Health Styles** (2012 - a probability Internet panel)
  - Porter Novelli contracted with Knowledge Networks
  - Knowledge Networks maintains an online panel of 50,000 representative of the entire U.S. population.
  - Panel members randomly recruited by probability-based sampling (using both random-digit dial and address-based sampling methods).
  - If needed, households provided with a laptop computer and access to the Internet.

- **PHSB** (2013 - a non-probability sample constructed to be representative of the U.S. population)
  - YouGov utilizes different modes of recruitment continuously over time ensuring hard-to reach populations adequately represented in survey samples
  - Sampling targets set based on gender, age, race, and education of this group using information from the 2010 American Community Survey
  - Weighted the matched set of survey respondents to known characteristics in the U.S. using propensity score weighting

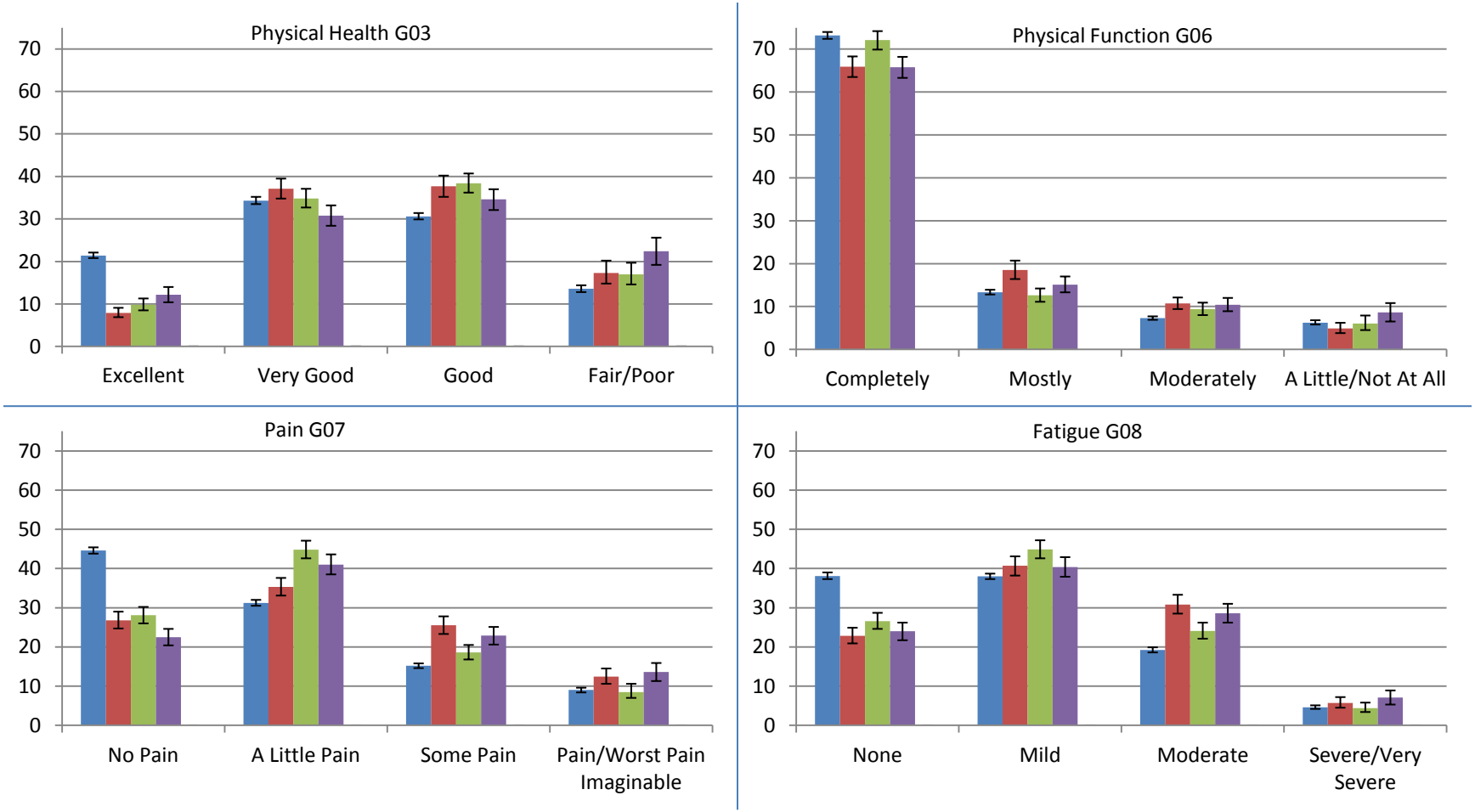
Population Characteristics

	PHSB 2013 (n=3,500)			HealthStyles 2012 (n=3,503)			HealthStyles 2010 (n=4,184)			NHIS 2010 (n=27,157)		
	Unweighted		Weighted	Unweighted		Weighted	Unweighted		Weighted	Unweighted		Weighted
	n	%	%	n	%	%	n	%	%	n	%	%
Race or Ethnicity												
Non-Hispanic White	2,635	75.3	67.4	2,641	75.4	67.0	2,842	68.0	69.0	15,510	57.2	68.4
Non-Hispanic Black	326	9.3	11.4	334	9.5	11.5	477	11.0	12.0	**4,394**	**16.2**	11.65
Hispanic	311	8.9	14.1	116	9.5	14.4	495	12.0	14.0	**5,054**	**18.6**	13.7
Other	228	6.5	7.2	412	5.6	7.1	370	9.0	6.0	2,171	8.0	6.3
Gender												
Female	1,968	56.2	52.0	1,770	50.5	51.7	2,181	52.0	52.0	15,171	**55.9**	51.7
Male	1,532	43.8	48.0	1,733	49.5	48.3	2,003	48.0	48.0	11,986	44.1	48.3
Age of Respondent												
18-24	283	8.1	13.1	317	9.0	12.6	60	**1.0**	13.0	2,801	10.3	12.8
25-34	565	16.1	15.7	418	11.9	17.2	414	10.0	18.0	4,974	18.3	17.9
35-44	786	22.5	18.9	518	14.8	17.2	707	17.0	18.0	4,805	17.7	17.4
45-54	533	15.2	16.8	718	20.5	18.9	1,269	**30.0**	20.0	4,855	17.9	19.4
55-64	751	21.5	18.2	706	20.2	16.2	806	19.0	15.0	4,272	15.7	15.6
65+	582	16.6	17.3	826	23.6	17.9	928	22.0	17.0	5,450	20.1	16.9

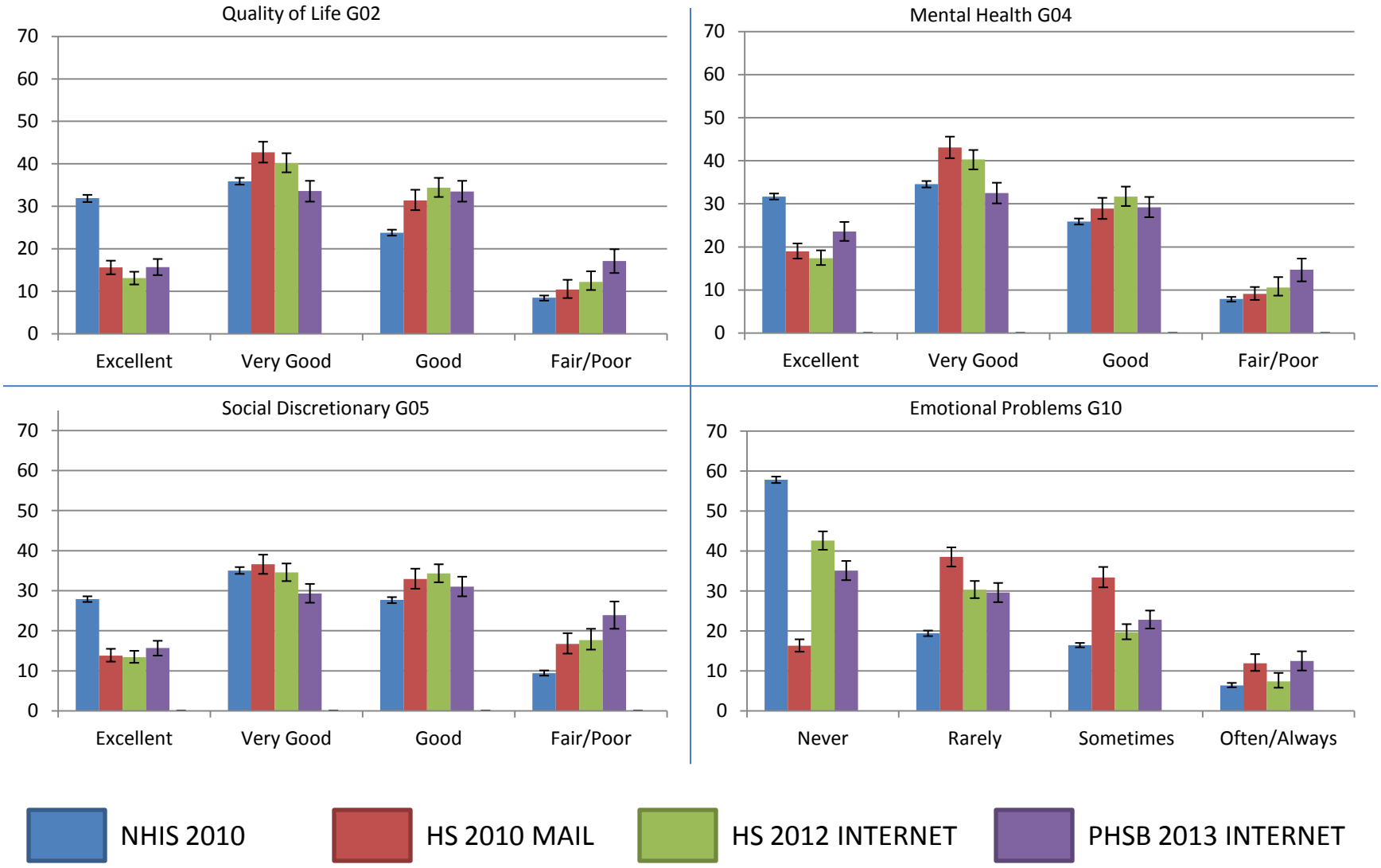
T-Scores


60
50
40
30
20
10
0
GPH T-SCORE
GMH T-SCORE
NHIS 2010
HS 2010 MAIL
HS 2012 INTERNET
PHSB 2013 INTERNET

PROMIS Global Physical Health



PROMIS Global Mental Health



Discussion

- Comparable results from probability and nonprobability samples

- NHIS difference potentially due to mode of administration, not probability vs. nonprobability sampling
  - Interviewer interaction produces a 0.2 – 0.5 SD increase in HRQL responses compared to mail/Internet responses

    Hays et al., Values in Health, 2009, 12: 1035-1039

- Limitations of comparing probability and nonprobability sampling
  - Other survey confounds
  - If different, was it the sampling approach?
  - If the same, is it generalizable?

Pros and Cons of Internet Panels

- Pros
  - Rapid and efficient data collection
  - Ability to recruit targeted samples
  - Ease of replication
- Cons
  - Selection biases
  - Difficulty ensuring integrity, security of the data
  - Higher rates of loss to follow-up
  - Considerable variability among Internet opt-in panels

But Are Nonprobability Samples Really That Different?

- Little practical distinction between opting in vs. opting out (Gotway Crawford, 2013)
  - Growing limits on RDD (no landline, call screening) introduce a form of selection bias
  - Weighting by the inverse of the selection probability (design-based) only works if the nonresponse rate is small (Rivers, 2013)
    - 10% nonresponse rate upper limit (Cochran, 1977)
    - 30% nonresponse rate with flawed results (Lohr, 2010)
    - What are we to expect of >95% nonresponse rates?
  - Modeling based inferences required for self-selection vs. nonresponse are mathematically indistinguishable (Rivers, 2013)

From JSSAM (2013) comments on the Summary Report of the AAPOR Task Force on Nonprobability Sampling

Weighting Adjustments for Opt-In Internet Panels

- Sample weighting
  - Cell weighting – adjust the sample distributions so they conform to the population distributions on a cell by cell basis
    - Assumes missing at random
    - Less stable aggregated estimates with large numbers of cells
  - Raking – iterative matching of cell counts to marginal distributions of the grouping variable
  - Propensity score weighting – correct selection bias in internet surveys but limited if only a few variables can be used to generate propensity scores

One Approach for Addressing the Problem of Estimate Precision from for Non Probability Samples: Bayesian Credibility Intervals

If you have a uniform(0,1) prior (the usual diffuse prior), then the posterior is:

$$\beta(np+1,n(1-p)+1)$$

where p is the sample fraction

Roshwalb et al. (2012)Towards the use of Bayesian credibility intervals in online survey results. NY: Ipsos Public Affairs

Conclusion

- Nonprobability Internet Panel Surveys are here to stay
- Focus on improving these methodologies and addressing their limitations
- Probability samples will remain the standard by which nonprobability samples are compared, and data from them remain critical for modeling and weighting nonprobability samples