Improving the Anchoring Vignette Methodology with Visual Vignettes

Mengyao Hu

Sunghee Lee

FCSM Conference 2018

Comparison between Different Population Groups

Q: Overall, in the last 30 days, how much pain or bodily aches did you have?

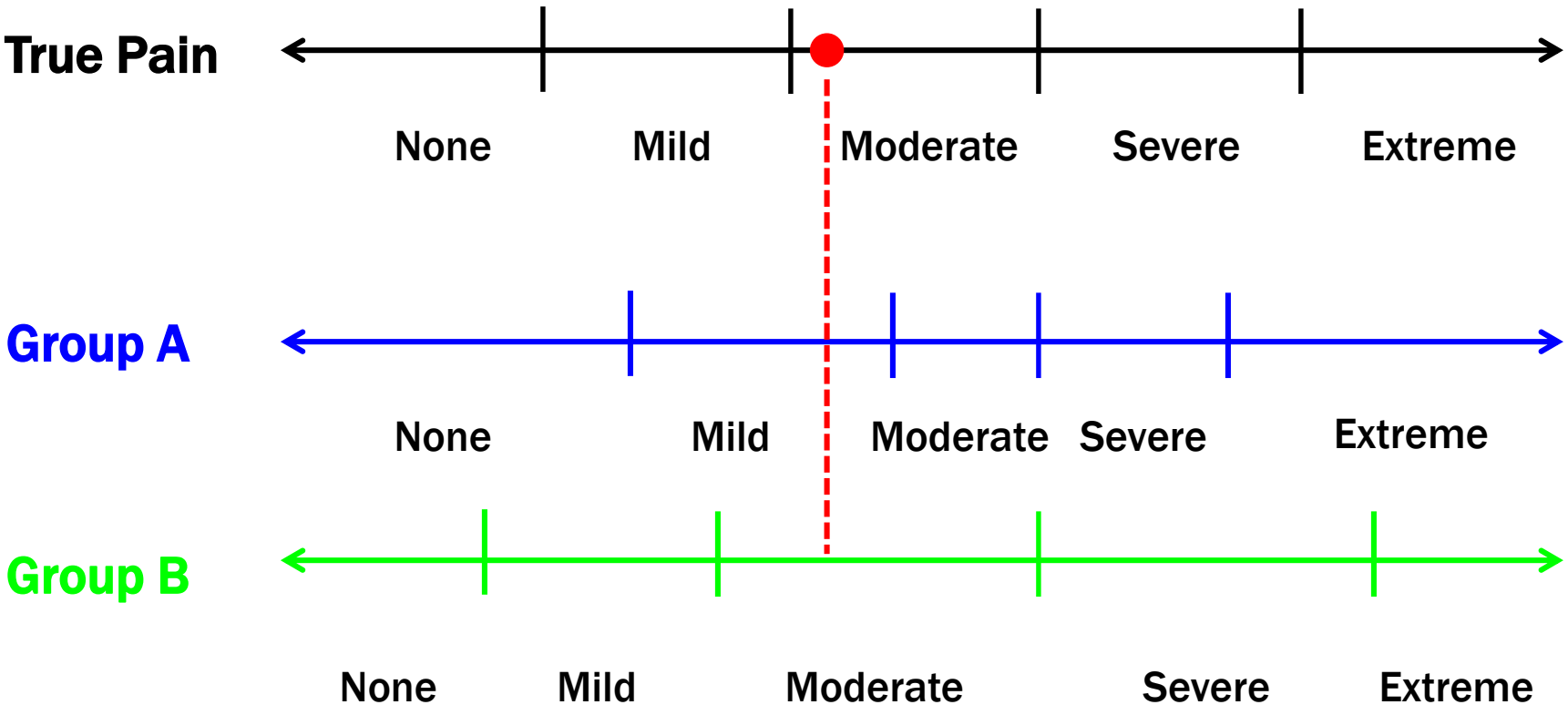
None **Mild** **Moderate** **Severe** **Extreme**

- Simple comparison (e.g., mean and distribution comparison)
- Adjustment for age, gender, and education differences through multivariate models (e.g., ordered probit)

IS THE COMPARISON VALID?

Differential Item Functioning (DIF)

Q: Overall, in the last 30 days, how much pain or bodily aches did you have?



Anchoring Vignettes

- Adjustment method for DIF
- Data requirement
  - Self-assessment (e.g., Self-assessed Pain)
  - Vignette items (typically more than one)
    - Hypothetical scenarios describing different pain levels

Anchoring Vignettes Example

Self-assessment:

- Overall, in the last 30 days, how much pain or bodily aches did you have?

None Mild Moderate Severe Extreme

Vignette Items:

- # 1: **Paul** has a headache once a month that is relieved after taking a pill. During the headache he can carry on with his day-to-day affairs. Overall, in the last 30 days, how much of bodily aches or pains did Henry have?
- # 2: **Henry** has pain that radiates down his right arm and wrist during his day at work. This is slightly relieved in the evenings when he is no longer working on his computer. Overall, in the last 30 days, how much of bodily aches or pains did Henry have?
- # 3:**Charles** has pain in his knees, elbows, wrists and fingers, and the pain is present almost all the time. Although medication helps, he feels uncomfortable when moving around, holding and lifting things. Overall, in the last 30 days, how much of bodily aches or pains did Charles have?

None Mild Moderate Severe Extreme

Anchoring Vignettes Example

Low

Medium

High

Vignette Items:

- # 1: Paul has a headache once a month that is relieved after taking a pill. During the headache he can carry on with his day-to-day affairs. Overall, in the last 30 days, how much of bodily aches or pains did Henry have?
- # 2: Henry has pain that radiates down his right arm and wrist during his day at work. This is slightly relieved in the evenings when he is no longer working on his computer. Overall, in the last 30 days, how much of bodily aches or pains did Henry have?
- # 3:Charles has pain in his knees, elbows, wrists and fingers, and the pain is present almost all the time. Although medication helps, he feels uncomfortable when moving around, holding and lifting things. Overall, in the last 30 days, how much of bodily aches or pains did Charles have?

Assumptions

- Assumptions
  - Vignette equivalence (VE)
  - Reporting consistency (RC)

Limitations of verbal vignettes

- Mixed results on the effectiveness of verbal anchoring vignettes
- RC and VE can be violated in a real population (e.g., Grol-Prokopczyk et al., 2015)
- Increased question difficulty
- Increased survey time
- Challenges in questionnaire translations

Motivations

- Visual anchoring vignettes:
  - Less cognitive effort
    - compared to texts, images are processed in a quicker and more automatic way (e.g., Paivio 2013)
  - Reduce survey time
  - Potentially help with assumption fulfillments
- No prior studies on the use of visual anchoring vignettes
- An open question whether this approach can remedy limitations of the current verbal vignettes

Research Objectives

- Evaluate the use of visual anchoring vignettes as an alternative to verbal vignettes
  - Response time
  - Realization of the assumptions
- Examine the effect of different vignette persons' characteristics

Design of Visual Vignettes

- Focus on four health domains:
  - sleep, affect, mobility and pain
- Steps to design visual vignettes:
  - examine critical elements of the four domains
  - identify common elements applicable across groups
  - select or design images with these elements at different intensity level for each domain
  - remove potential confounding effects of various image elements

Pain (Varying Age)

	No pain	Middle pain	High pain
Young Adults			
Older Adults			

Sleep (Varying Gender)

	No / low Difficulty	Middle Difficulty	High Difficulty
Male			
Female			

Mobility (Varying Body Size)

	No / low Difficulty	Middle Difficulty	High Difficulty
Fit			
Obese			

Affect (Varying Race / Ethnicity)

	No / low Depression	Middle Depression	High Depression
White			
Black			
Hispanic			

Experiment Design

- Web survey (programmed in Qualtrics)
- Sample: Qualtrics online panel
  - 760 White
  - 750 Black
  - 750 Hispanics speaking English
  - 889 Hispanics speaking Spanish
- 3 vignette type conditions
  - one verbal condition (adopted from HRS)
  - two visual conditions

Analysis Outline

- Response time
- Test of Vignette Equivalence (VE)
  - Rank-ordering test
  - Likelihood ratio test
- Test of Response Consistency (RC)
- Compare visual vignettes with different characteristics

Response Time (sec)

	Pain		Sleep		Mobility		Affect	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Verbal vignette	15.93	8.81	15.73	8.25	17.85	10.31	18.05	10.59
Visual vignette	7.95	3.58	8.38	3.61	8.33	3.79	7.42	3.17

Correct Rank-ordering

- Percentage of respondents ordering vignettes consistently with expected ordering.

	Pain		Sleep		Mobility		Affect	
	n	%	n	%	n	%	n	%
Verbal vignette	1051	47.6	1051	17.7	1051	19.8	1051	67.1
Visual vignette	2098	79.7	2098	74.0	2098	43.4	2098	81.8

Test of Vignette Equivalence (VE)

- Based on a likelihood ratio test (LRT) of two models

$$V_{ij}{}^{*} = \alpha_j + \varepsilon_{ij} \qquad (A)$$

  - ${V_{ij}}^{*}$: respondent i's perceived location of vignette j
  - α_j is a constant term; α for the reference vignette is set to be 0.
  - ε_{ij} is the random error term.

$$V_{ij}{}^{*} = \alpha_j + \boldsymbol{\lambda_j X_i} + \varepsilon_{ij} \qquad (B)$$

  - X_i, a vector of covariates including demographic variables.

- If VE is fulfilled, $\boldsymbol{\lambda_j}$ will be 0, and the LR test will be not significant.

Grol-Prokopczyk, H. (2017). In Pursuit of Anchoring Vignettes That Work: Evaluating Generality Versus Specificity in Vignette Texts. *The Journals of Gerontology: Series B*.

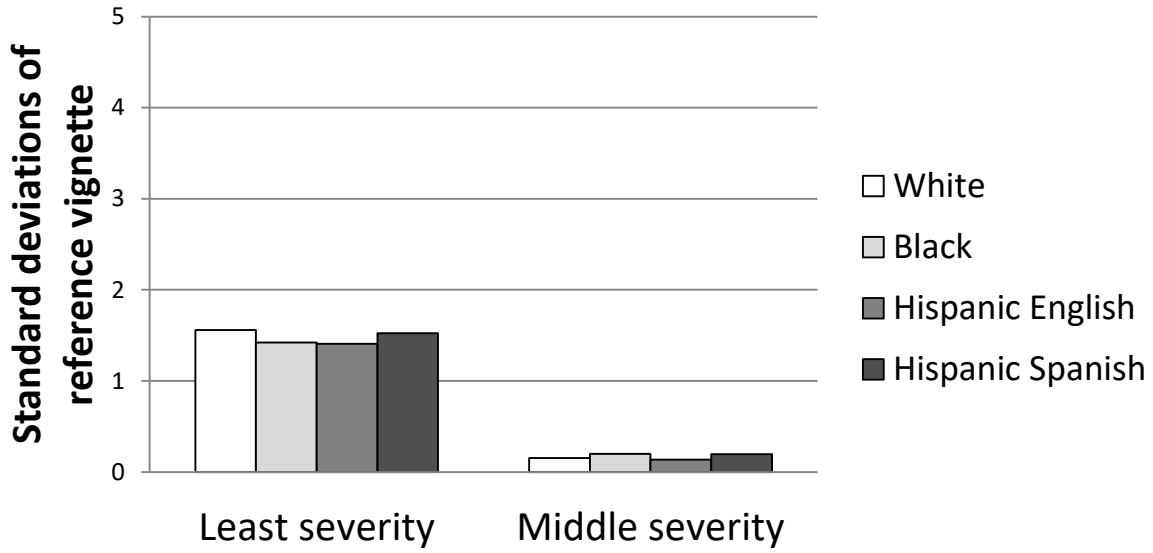
Results: Test of VE for Sleep

	Pain		Sleep		Mobility		Affect	
	df	LRT	df	LRT	df	LRT	df	LRT
Verbal vignettes	24	70.4***	24	24.4	24	55.1***	24	110.9***
Visual vignettes	24	137.4***	24	158.8***	24	67.1***	24	154.3***

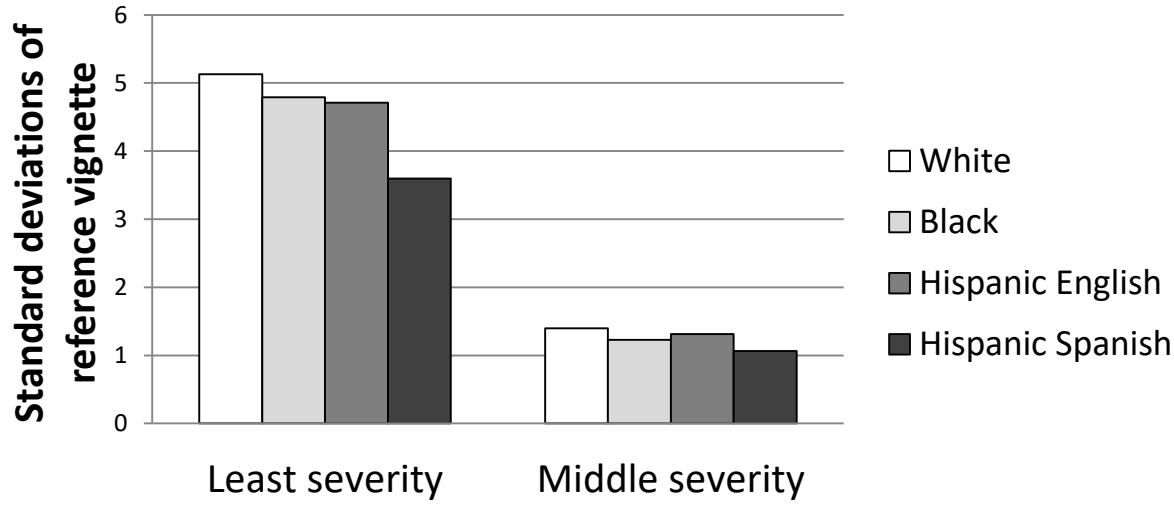
Note: LRT = $2(\max \ell - \max_{\lambda=0} \ell) \sim \chi^2(p(m-1))$, *where* ℓ *is* the log-likelihood function.

Results: Test of VE for Sleep

Verbal Vignettes:



Visual Vignettes:



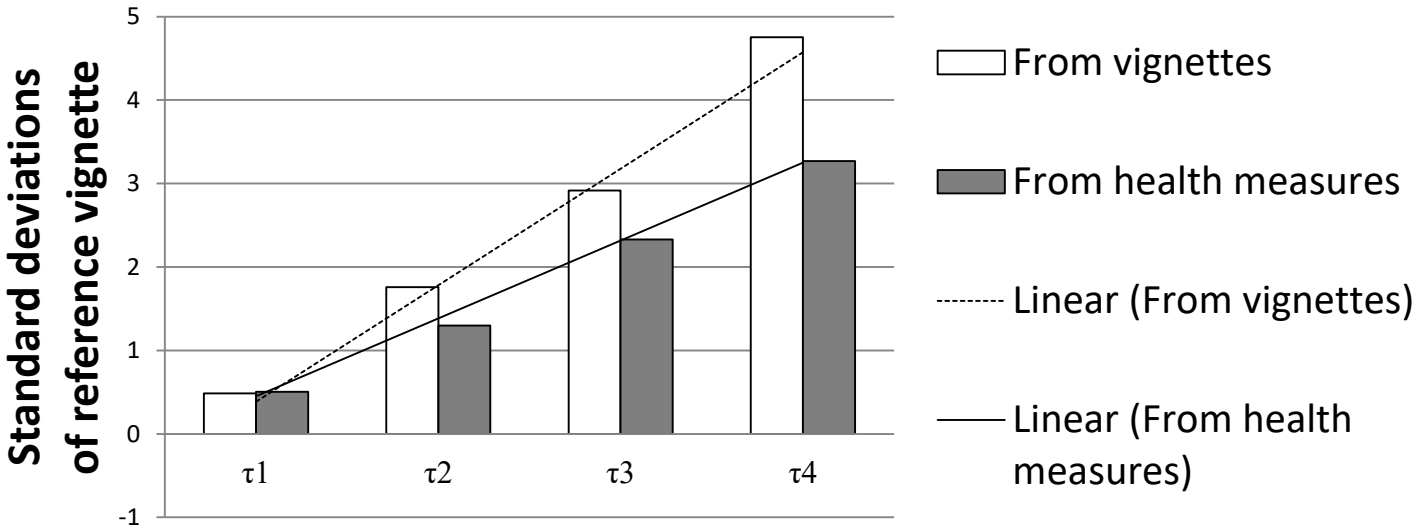
Test of Response Consistency (RC)

- Based on visual comparisons of two sets of cut-points:
  - Cut-points for vignettes
  - Cut-points for self-assessment
- Two sets of cutpoints were graphed in a figure for visual comparisons.
- Suggestive but not definitive

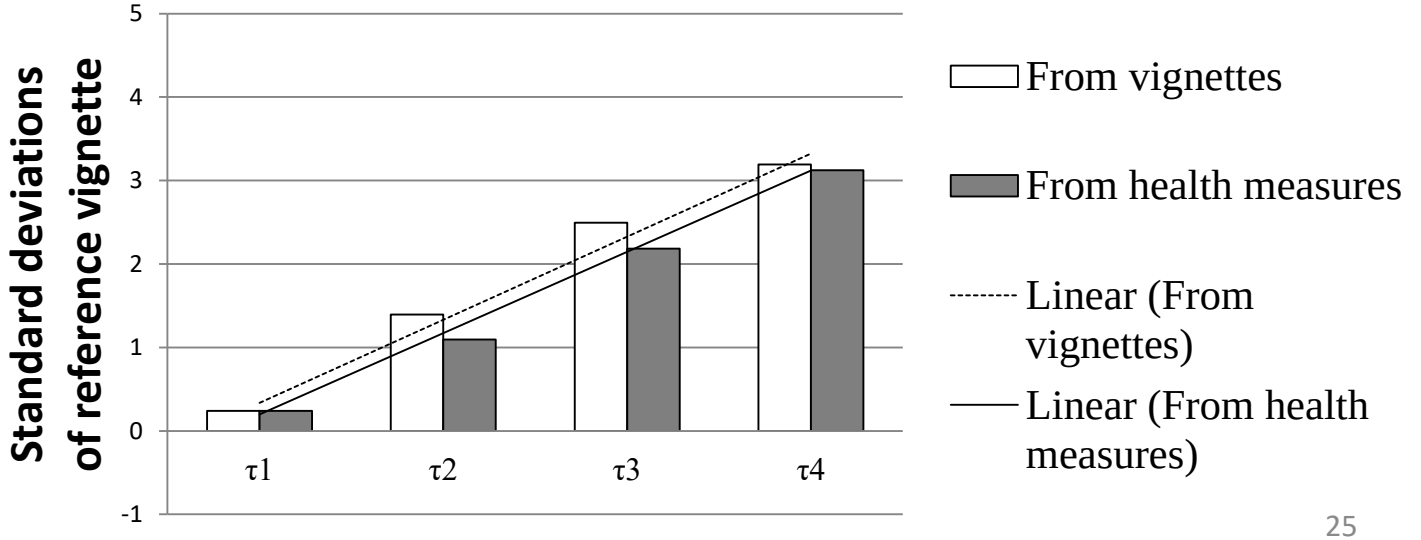
Grol-Prokopczyk, H. (2017). In Pursuit of Anchoring Vignettes That Work: Evaluating Generality Versus Specificity in Vignette Texts. *The Journals of Gerontology: Series B*.

Results – Test of RC for Sleep

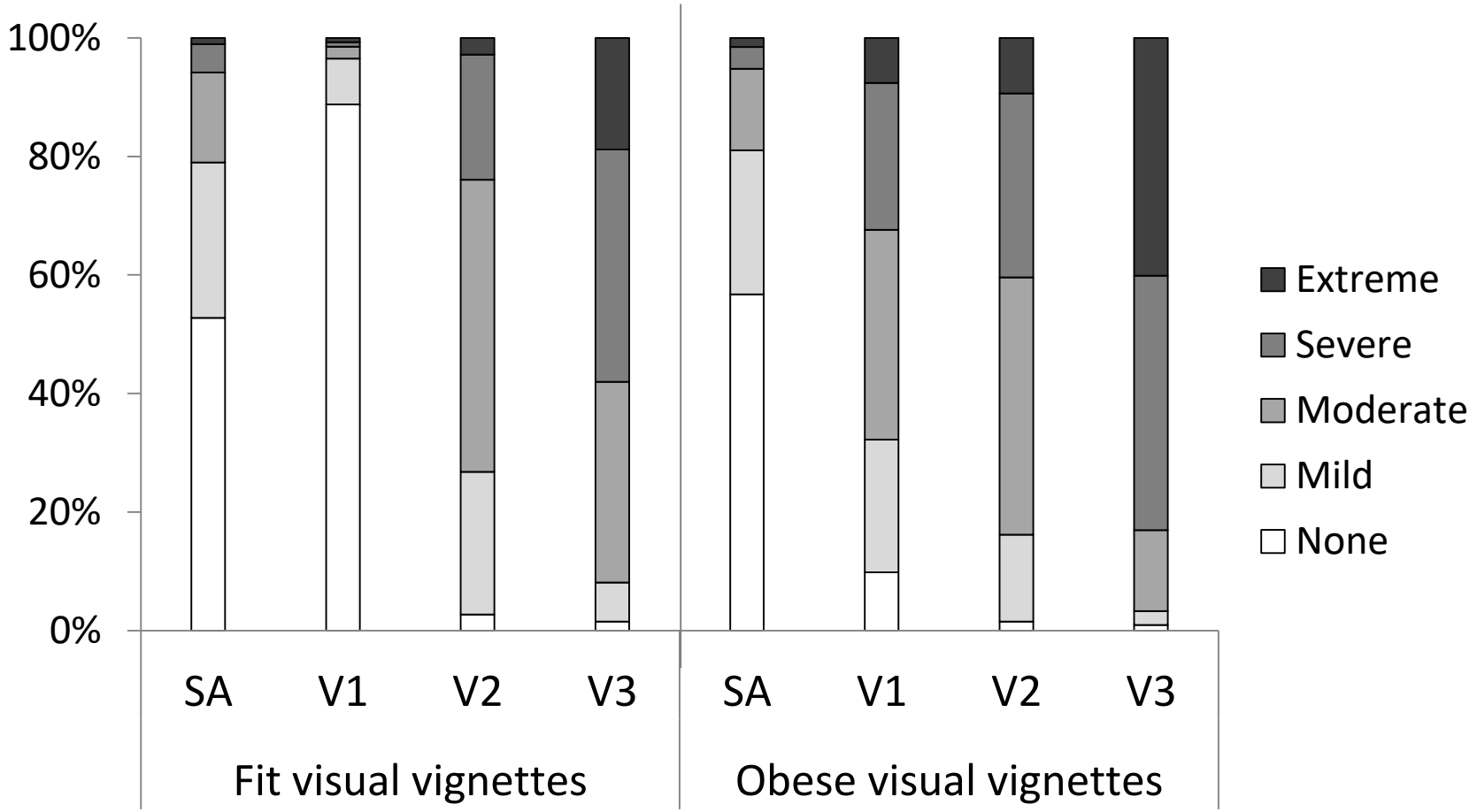
Verbal Vignettes:



Visual Vignettes:



Results – Vignette Characteristics (Mobility)



Discussion

- Visual vignettes can reduce survey time.
- Respondents can better distinguish the different intensity levels in visual vignettes.
- VE violations found in both verbal and visual vignettes.
- Similar DIF-adjusting results as verbal vignettes.
- A great potential in the use of visual anchoring vignettes in adjusting DIF.

Future Research

- Further evaluate the effects of various vignette characteristics on visual vignette performances.
- Use more stringent tests to test RC.
- Compare visual and verbal vignettes using the multidimensional IRT model.
- Extend this study to cross-nation surveys and / or to a wide variety of other racial / ethnic groups.
- Explore the use of short video vignettes.

Thank you!

(maggiehu@umich.edu)

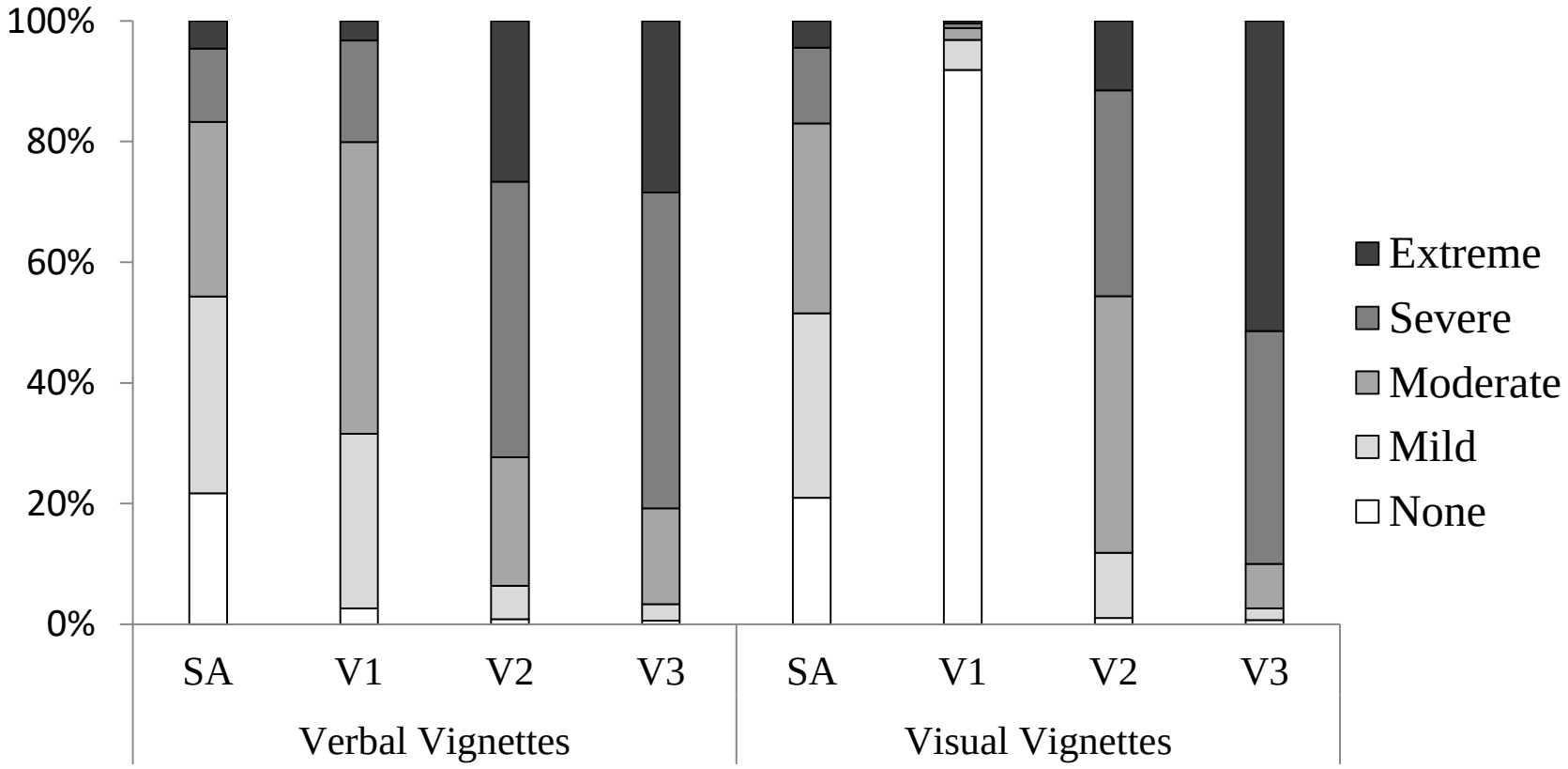
Study 2: Visual vignette question display

Overall, how much pain or bodily aches does this person have?

- None
- Mild
- Moderate
- Severe
- Extreme

Sleep Domain Descriptive

- Responses to sleep self-assessment (SA) and the three vignettes difficulty/intensity questions (V1=none/mild; V2 = moderate; V3= severe/extreme) by vignette types



HOPIT model

Model for Self-Assessment:

The self-assessment component starts with the distribution of true state for person *i*, ${Y_i}^*$, where $i = 1, \ldots, N$

$$ {Y_i}^* \sim N(\mu_i, \sigma^2). $$

Note that ${Y_i}^*$ reflects an unobserved continuum (e.g., pain level). The mean of ${Y_i}^*$, μ_i, is a latent variable further modeled using linear regression on covariates, $\boldsymbol{X}_i$, such as cultural groups, age and gender, as $\mu_i = \boldsymbol{X}_i\beta + \eta_i$, where η_i is an independent random error, which follows a normal distribution as:

$$ \eta_i \sim N(0, \omega^2). $$

The actual response in the self-assessment question is denoted as Y_i, where the question is asked with a response scale with *K* response categories, $k = 1, \ldots, K$. Y_i reflects ${Y_i}^*$ as follows:

$$ Y_i = k, \text{ if } {\tau_i}^{k-1} < {Y_i}^* \leq {\tau_i}^k. $$

Here, $\tau_i = \left({\tau_i}^0, {\tau_i}^1, \ldots, {\tau_i}^k, \ldots, {\tau_i}^K\right)$ is a vector of thresholds (i.e., cutpoints) respondent *i* uses to answer the question. For instance, ${\tau_i}^1$ is where response categories of "None" and "Mild" in Figure 1.1 are differentiated. Naturally, ${\tau_i}^0 = 0$ and ${\tau_i}^K = \infty$. These person-specific thresholds are modeled as:

$$ {\tau_i}^1 = \gamma^1 \boldsymbol{Z}_i \text{ for } k = 1 \text{ and } {\tau_i}^k = {\tau_i}^{k-1} + \mathrm{e}^{\gamma^k \boldsymbol{Z}_i} \text{ for } k = 2, \ldots, K-1, $$

where $\boldsymbol{Z}_i$ is a vector of covariates relevant to the thresholds, which may or may not be the same as $\boldsymbol{X}_i$.

HOPIT model

Model for Vignette Responses:

Vignette responses are modeled in a similar way as the self-assessment. Let ${V_{ij}}^*$ be the unobserved state of the person in the j^{th} vignette item perceived by respondent i, which follows a normal distribution with a random error denoted as:

$$ {V_{ij}}^* \sim N(\alpha_j, \sigma_j^2), $$

where $i = 1, \ldots, N$ (the respondents can be different from those who answered for self-assessments, if vignettes are only asked to a subset of the whole sample) and $j = 1, \ldots, J$. Note that α_j is assumed to be the same across respondents, which implies the VE assumption.

Respondents' response to the j^{th} vignette is denoted as v_{ij} , which is obtained by mapping the latent ${V_{ij}}^*$ onto the response scale and modeled as:

$$ v_{ij} = k, \text{ if } {\tau_l}^{k-1} < {V_{ij}}^* \leq {\tau_l}^k. $$

The thresholds of vignettes are modeled exactly the same as the self-assessment model.

$$ {\tau_i}^1 = \gamma^1 \boldsymbol{Z}_i \text{ for } k = 1 \text{ and } {\tau_i}^k = {\tau_i}^{k-1} + \mathrm{e}^{\gamma^k \boldsymbol{Z}_i} \text{ for } k = 2, \ldots, K-1. $$

Imposing the thresholds to be modelled the same for the vignettes and the self-assessment implies the RC assumption.