Published in final edited form as:

J Clin Epidemiol. 2014 January ; 67(1): 108–113. doi:10.1016/j.jclinepi.2013.07.016.

Method of Administration of PROMIS Scales Did Not Significantly Impact Score Level, Reliability or Validity

Jakob B. Bjorner^{1,2,3}, Matthias Rose^{4,5}, Barbara Gandek⁴, Arthur A. Stone⁶, Doerte U. Junghaenel⁶, and John E. Ware Jr.^{4,7}

¹National Research Centre for the Working Environment, Copenhagen, Denmark

²QualityMetric, Optum PatientInsight, Lincoln, RI, USA

³Department of Public Health, University of Copenhagen, Copenhagen, Denmark

⁴Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA

⁵Department of Psychosomatic Medicine and Psychotherapy, Charité, Universitätsmedizin, Berlin, Germany

⁶Department of Psychiatry and Behavioral Science, Stony Brook University, Stony Brook, NY, USA

⁷John Ware Research Group, Worcester, MA, USA

Abstract

Objective—To test the impact of method of administration (MOA) on score level, reliability, and validity of scales developed in the Patient Reported Outcomes Measurement Information System (PROMIS).

Study Design and Setting—Two non-overlapping parallel forms each containing 8 items from each of three PROMIS item banks (Physical Function, Fatigue and Depression) were completed by 923 adults with COPD, depression, or rheumatoid arthritis. In a randomized cross-over design, subjects answered one form by interactive voice response (IVR) technology, paper questionnaire (PQ), personal digital assistant (PDA), or personal computer (PC) and a second form by PC, in the same administration. Method equivalence was evaluated through analyses of difference scores, intraclass correlations (ICC), and convergent/discriminant validity.

Results—In difference score analyses, no significant mode differences were found and all confidence intervals were within the pre-specified MID of 0.2 SD. Parallel forms reliabilities were very high (ICC=0.85-0.93). Only one across mode ICC was significantly lower than the same

---

© 2013 Elsevier Inc. All rights reserved.

Corresponding author: Jakob Bue Bjorner, MD, PhD, QualityMetric / Optum PatientInsight, 24 Albion Road, Lincoln 02865 RI, USA, Telephone: +45 39 16 52 00, Fax: +45 39 16 52 01, jbjorner@qualitymetric.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

mode ICC. Tests of validity showed no differential effect by MOA. Participants preferred screen interface over PQ and IVR.

Conclusion—We found no statistically or clinically significant differences in score levels or psychometric properties of IVR, PQ or PDA administration as compared to PC.

Keywords

Patient Reported Outcomes; Quality of Life; Questionnaire; Mode of Administration; Method of Administration; Item Response Theory

---

Introduction

Measurement of patient-reported outcomes (PRO) is facilitated through electronic data capture (EDC) technologies like telephone interactive voice response (IVR), personal computer (PC), and hand-held devices (e.g., personal digital assistant (PDA)). These EDC technologies are quite flexible, allowing for use of static (pre-selected) or dynamic (matched to the respondent) selection of survey items. Having multiple methods of administration of PRO surveys improves response rates in population studies [1]. However, transition to EDC and the combination of methods assume that the method of administration (MOA) does not influence item response.

Several studies have found equivalence between paper questionnaire (PQ) and computerized (PC or PDA) methods of administration [2-5]. Fewer studies have assessed IVR technology for health outcomes measurement; these have found only non-significant [6] or small score differences [7,8]. However, many previous studies were small and did not include explicit calculations of statistical power. Few studies have performed a comprehensive evaluation of equivalence across the most frequently used modes.

The Patient Reported Outcomes Measurement Information System (PROMIS®) is a National Institutes of Health (NIH) Roadmap initiative to generate new PRO measures for clinical research and practice (<http://www.nihpromis.org>). Using patient input through focus groups and cognitive interviews, PROMIS has created large item banks to measure many domains, including physical function, fatigue, pain, social role, and emotional distress [9]. These banks have been tested rigorously using modern psychometric methods such as confirmatory factor analysis and item response theory [10]. The present PROMIS study was designed to examine how differences in MOA affect psychometric properties and score differences and to evaluate -with large statistical power-the consistency of any differences between MOA across health domains. Four MOA were compared: PQ, IVR, PC, and PDA. A previous paper evaluated psychometric equivalence across these modes of administration through confirmatory factor analyses (CFA) and item response theory (IRT) and found that CFA and IRT models were robust across MOA [11]. The present paper will evaluate the equivalence between MOA in terms of mean score differences, intraclass correlations (ICC), and analyses of convergent/discriminant validity. Finally, we evaluate respondent preference for different MOA.

Methods

Two sub-studies were conducted using a randomized cross-over design, in which two non-overlapping parallel forms (Form A and Form B) consisting of eight items from each of three PROMIS domains (Physical Function (PF), Fatigue (FAT), Depression (DEP)) were administered. In study 1, participants were randomized to complete one form by IVR, PQ, or PC, and the second form by PC. In study 2, all participants were assessed by PDA and PC. Study 2 was performed separately, because the PDA administration required in-person contact. The order of forms and the combination of form and MOA were randomized in a balanced design (see [11] for details on study design). Since the administration of the two parallel forms was done in one session, all participants completed both forms. Except for participants randomized to the PC-PC arm, all participants answered two MOA.

Measures

Parallel Static Forms—From each of three PROMIS item banks (PF, FAT, DEP), two parallel static short forms containing eight non-overlapping items per domain were developed [11]. Following standard PROMIS scoring [10], scores were normed to a mean of 50 and SD of 10 in the US general population with high scores indicating good physical function and high levels of fatigue and depression, respectively. Each pair of short forms was developed to provide equivalent scores. However, due to the balanced design, the study was robust to departures from perfectly parallel forms.

Other measures included: (1) *PROMIS global items* measuring general health, physical function, fatigue, and emotional distress [12]; (2) *Number of restricted-activity days and days off from work* in the past four weeks for health reasons; and (3) *Demographics, health care utilization, previous computer use, and health literacy* [13]. Also, participants completed a *user experience questionnaire* based on a survey developed by Buxton et al. [14] for each MOA.

Sample and Procedures

Study 1—Data for the IVR-PC, PQ-PC and PC-PC arms was collected from the online panel used for the initial calibration of the PROMIS item banks [15]. Panelists were age 18 and older, fluent in English and had previously indicated that they had rheumatoid arthritis, chronic obstructive pulmonary disease (COPD), or depression. To confirm the diagnosis, subjects verified that they were diagnosed by a treating physician and taking one or more diagnosis-specific medications; subjects with depression also needed to verify current treatment by a mental health professional. To ensure a sufficient distribution of impairment within each diagnostic group, a quota was imposed aiming to achieve equal representation of low impact, medium impact, and severe disease impact. This study included a total of 723 persons, well above the target sample of 600.

All patients answered screener questions, followed by the first set of PROMIS items (either Form A or Form B) and user experience questionnaire one, literacy and demographic items. Then respondents answered the second set of PROMIS items (if previously Form B - then Form A), user experience questionnaire two, and other health items. Except for the PROMIS

items, which were answered by either PQ, IVR, or PC, all items were answered on the PC. Subjects randomized to the PQ arm were instructed to complete a previously mailed paper-pencil questionnaire, while subjects randomized to the IVR arm were instructed call a toll-free number for the IVR assessment. Following PROMIS conventions; the PC administration displayed one item per screen, while the PQ layout grouped items with the same response category together. IVR recordings were developed for the PROMIS initiative using a female voice.

Study 2—For the PDA-PC arms, 200 rheumatology patients, 18 years or older, who were fluent in English, able to hold a pen and had no visual impairment, were recruited through a rheumatology practice on Long Island. The order of survey elements was similar to study 1. Randomization to one of four arms (order of MOA and order of form) was accomplished by opening the next envelope in a numbered sequence that had one of the four administration orders. The two assessments took place on site and were separated by a short interval (e.g., 5-10 minutes), to allow participants to switch from one MOA to another. Both PC and PDA displayed one item per screen. Participants were compensated with up to \$50 for participation.

Studies were approved by the New England Institutional Review Board (#09-107, Study 1 and 2) and the Stony Brook University Institutional Review Board (#2008-0280, Study 2).

Analyses

A detailed analytic protocol was approved by the PROMIS Steering Committee. The current paper evaluates equivalence across MOA with respect to these key parameters: (1) score levels; (2) reliability (internal consistency and alternate forms); (3) convergent and discriminant validity; and (4) respondent acceptability. The same analytic approach was applied to the comparison of PQ, IVR, and PC MOA and the comparison of the PDA and PC MOA. However, because selection criteria for the PDA arms were different than for the other MOA, the PDA arms were analyzed separately.

For each form (A and B), scores were calculated based on the standard PROMIS IRT scoring approach. The impact of MOA on score level was tested through analysis of difference scores, because simulation studies showed that difference score analysis in this setting would provide more statistical power than mixed linear models for repeated observations. Score differences between Form B and Form A were regressed on MOA controlling for sequence of administration. The study was powered to evaluate equivalence across MOA within 2 points, corresponding to the 0.2 effect size suggested by Cohen [16] as the smallest relevant effect. This difference was chosen since potential MOA effects should be smaller than score differences that would be considered clinically relevant.

In addition to evaluation of the impact of MOA on score level, we evaluated *reliability* for each domain and MOA by estimating alternate forms reliability (using the ICC). Confidence intervals for the ICC were estimated through empirical bootstrap. Also, internal consistency reliability was estimated by domain, separately for each MOA and form, using Cronbach's coefficient alpha.

Convergent and discriminant validity across MOA was evaluated by a mixed linear model to conduct separate analyses of each of the following exogenous variables as fixed effects: the four PROMIS global measures [12], limited activity days and sick days (for people who work), and diagnostic group. We hypothesized that the global general health item would be most strongly associated with Physical Function and Fatigue; that the global items for physical function, fatigue, and emotional distress would be most strongly associated with their respective scales; that bed days and activity limitation days would be significantly associated with all three scales; that a diagnosis of arthritis/rheumatism would be associated with Physical Function and Fatigue; that COPD would be strongly associated with Physical Function, and that a diagnosis of depression would be strongly associated with the Depression scale. Further, we hypothesized that the strength of these associations would not depend on MOA. These latter hypotheses were evaluated through tests of the interaction between MOA and each exogenous variable. Due to the large number of tests (27), we adjusted the significance level for evaluation of the interaction terms [17].

Respondent acceptability was evaluated in terms of subjective user evaluation through simple analyses of survey experience by method.

We performed *subgroup* analyses of score difference and reliability focusing on three groups for which electronic data capture might be problematic: persons 60 years or older (n=275), persons with at most a high school or GED education (n=142), and persons with low health literacy (n=235).

Results

Participants ranged from 18 to 89 years of age (mean age 56 years). Patients with rheumatoid arthritis and depression were predominantly (65-80%) women, while the COPD group had equal gender distribution. More than 90% of participants were White and a little more than half were married. Most participants had some post-high school education and 38% were college graduates. In study 1, RA and COPD patients had low PF (poorer physical function) scores, high FAT (more fatigue) scores and somewhat increased DEP (more depression) scores. Patients with depression had somewhat decreased PF scores and high FAT and DEP scores. In study 2 rheumatology patients had low PF scores, somewhat increased FAT scores and normal DEP scores. Across all comparisons, mean scores across Forms A and B were very similar.

Table 1 shows the effect of PQ, IVR and PDA MOA compared to PC. No MOA effects were significant, and all confidence intervals were within the bound of +/- 2 points that we consider an ignorable effect. Table 1 also reports the effect of a scale being placed in the last half of the form, and the effect of being assessed by Form B rather than Form A. For the purpose of MOA evaluation, these parameters can be regarded as nuisance parameters.

Parallel forms reliability was generally very high (Table 2). Not surprisingly, the highest ICCs were found for the PC-PC arm (0.93 on average), but results for other MOA were also high (0.90 on average), as were all internal consistency reliabilities. In one instance, Fatigue, the ICC for the IVR-PC MOA (0.89) was significantly lower than the ICC for PC-PC (0.94).

Note that the reliability results for the PDA-PC MOA in Study 2 should not be compared directly to the results from Study 1, since they are based on a different sample.

Analyses of convergent and discriminant validity showed a pattern of associations that fits well with our hypotheses (table 3). Tests of differential validity by MOA found no significant interaction effects after controlling for multiple comparisons. The strongest effect estimated, concerning a slightly stronger discrimination for the PQ MOA compared to the PC MOA within the depression domain, was approximately 1 point, thus far below our MID threshold.

Analyses of study 1 participants showed that 81% preferred PC over IVR, 18% were indifferent, and 1% preferred IVR, while 48% preferred PC over PQ, 54% were indifferent, and 3% preferred PQ. In study 2, 25% of participants preferred PC over PDA, 62% had no preference, and 13% preferred PDA.

The results of subgroup analyses agreed with the analyses of the total sample. We found no significant MOA difference in mean scores. While parallel forms reliability was slightly lower in subgroup analyses (0.91 on average for the PC-PC method and 0.89 on average for other MOA), the patterns were similar to the total sample. Patient preferences were virtually unchanged compared to the total sample.

Discussion

In this study of the effects of different MOA, we found neither statistically nor clinically significant effects on mean scale score levels of PQ, IVR or PDA administration as compared to PC. In parallel forms reliability analyses, ICCs were lower for IVR-PC than for PC-PC for the Fatigue domain, suggesting a slightly lower score precision for IVR administration, similar to results observed from analyses of item properties [11]. However, the differences seem minor and we found no decrement in convergent validity for IVR MOA compared to other MOA. Therefore, we conclude that a possible lower score precision with IVR is of limited practical importance. Thus, in line with the rather substantial number of studies comparing PQ with PC or PDA [3] and with the more limited number of studies comparing IVR and PQ [18], our results provide strong support for the equivalence of scores from PC, PQ, IVR and PDA MOA.

While study 2 used site-based sampling, study 1 used an internet sample that may be more similar to sampling strategies used in effectiveness studies than in efficacy trials. Both samples had higher education than the US average. However, subgroup analyses in the elderly and the low education and health literacy subgroups gave the same results as the overall analysis.

The parallel forms design – made possible by the PROMIS item banks - combined with the large sample size provided greater statistical power than in previous MOA studies. A further advantage of the current study is the explicit test of convergent/discriminant validity, which has not been done previously, to our knowledge.

A limitation of the current study is the lack of data on personal or phone interviews using a 'live' interviewer. It is for these MOA that lack of equivalence has been documented most consistently, and the frameworks of social desirability and interviewer style provide good theoretical explanations as to why MOA effects could happen. Also, while our implementation of IVR provided results that were equivalent to the PC MOA, not all IVR implementations may do so. Finally, since respondents in study 1 answered the questionnaires at home, we cannot rule out that they answered the questionnaires in another order than instructed or with a longer interval between the two forms. However, such behavior would introduce random noise and lead to lower parallel forms reliability, which was not observed.

In conclusion, our results provide strong support for the equivalence of score levels, precision, and convergent/discriminant validity for the evaluated MOA: PC, PQ, IVR and PDA. Further, participants showed a strong (PC vs. IVR) to moderate (PC vs. PQ) preference for a screen interface.

Acknowledgments

The Patient Reported Outcomes Measurement Information System (PROMIS) is a National Institutes of Health (NIH) Roadmap initiative to develop a computerized system measuring patient-reported outcomes in respondents with a wide range of chronic diseases and demographic characteristics. PROMIS was funded by cooperative agreements to a Statistical Coordinating Center (Northwestern University PI: David Cella, PhD, U01AR52177) and six Primary Research Sites (Duke University, PI: Kevin Weinfurt, PhD, U01AR52186; University of North Carolina, PI: Darren DeWalt, MD, MPH, U01AR52181; University of Pittsburgh, PI: Paul A. Pilkonis, PhD, U01AR52155; Stanford University, PI: James Fries, MD, U01AR52158; Stony Brook University, PI: Arthur Stone, PhD, U01AR52170; and University of Washington, PI: Dagmar Amtmann, PhD, U01AR52171). NIH Science Officers on this project are Deborah Ader, Ph.D., Susan Czajkowski, PhD, Lawrence Fine, MD, DrPH, Louis Quatrano, PhD, Bryce Reeve, PhD, William Riley, PhD, and Susana Serrate-Sztejn, PhD. This manuscript was reviewed by the PROMIS Publications Subcommittee prior to external peer review. The authors would like to thank two anonymous PROMIS reviewers for comments on a previous version of this manuscript. See the web site at www.nihpromis.org for additional information on the PROMIS cooperative group.

Reference List

1. Dillman DA, Phelps G, Tortora R, et al. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*. 2009; 38:1–18.
2. Dillman, D. Mail and Internet Surveys: The tailored Design Method - 2007 Update with New Internet, Visual, and Mixed-Mode Guide. New York, NY: J. Wiley; 2007.
3. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health*. 2008; 11:322–333. [PubMed: 18380645]
4. Raat H, Mangunkusumo RT, Landgraf JM, Kloek G, Brug J. Feasibility, reliability, and validity of adolescent health status measurement by the Child Health Questionnaire Child Form (CHQ-CF): internet administration compared with the standard paper version. *Qual Life Res*. 2007; 16:675–685. [PubMed: 17286197]
5. Yu SC. Comparison of Internet-based and paper-based questionnaires in Taiwan using multisample invariance approach. *CyberPsychology & Behavior*. 2007; 10:501–507. [PubMed: 17711357]
6. Agel J, Rockwood T, Mundt JC, Greist JH, Swiontkowski M. Comparison of interactive voice response and written self-administered patient surveys for clinical research. *Orthopedics*. 2001; 24:1155–1157. [PubMed: 11770093]
7. Dunn JA, Arakawa R, Greist JH, Clayton AH. Assessing the onset of antidepressant-induced sexual dysfunction using interactive voice response technology. *J Clin Psychiatry*. 2007; 68:525–532. [PubMed: 17474807]

8. Rush AJ, Bernstein IH, Trivedi MH, et al. An evaluation of the quick inventory of depressive symptomatology and the hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. *Biol Psychiatry*. 2006; 59:493–501. [PubMed: 16199008]
9. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care*. 2007; 45:S3–S11. [PubMed: 17443116]
10. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. 2007; 45:S22–S31. [PubMed: 17443115]
11. Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware JE Jr. Difference in Method of Administration Did Not Significantly Impact Item Response: An IRT-based Analysis from the Patient-Reported Outcomes Measurement Information System (PROMIS) Initiative. *Qual Life Res*. 2014; 23:217–27. [PubMed: 23877585]
12. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual Life Res*. 2009; 18:873–880. [PubMed: 19543809]
13. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Family Medicine*. 2004; 36:588–594. [PubMed: 15343421]
14. Buxton J, White M, Osoba D. Patients' experiences using a computerized program with a touch-sensitive video monitor for the assessment of health-related quality of life. *Qual Life Res*. 1998; 7:513–519. [PubMed: 9737141]
15. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*. 2010; 63:1179–1194. [PubMed: 20685078]
16. Cohen, J. *Statistical power for the behavioral sciences*. Hillsdale NJ: Erlbaum; 1988.
17. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 1988; 75:800–802.
18. Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value Health*. 2009; 12:419–429. [PubMed: 19900250]

Table 1
Analyses of score differences. Comparison method of administration is Personal Computer

Study 1						
	Physical Function		Fatigue		Depression	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
PQ	-0.39	(-1.04 to 0.25)	0.16	(-0.61 to 0.92)	0.16	(-0.56 to 1.20)
IVR	-0.62	(-1.25 to 0.01)	0.28	(-0.46 to 1.02)	-0.14	(-1.00 to 0.72)
Last place in form	0.16	(-0.42 to 0.73)	-1.01	(-1.69 to -0.33)	-1.94	(-2.57 to -1.31)
Form B	-0.74	(-1.20 to -0.28)	0.96	(0.42 to 1.51)	-2.12	(-2.91 to -1.33)
Study 2						
	Physical Function		Fatigue		Depression	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
PDA	-0.56	(-1.38 to 0.27)	0.20	(-0.89 to 1.29)	-0.40	(-1.66 to 0.85)
Last place in form	-0.80	(-1.63 to 0.02)	-1.51	(-2.60 to -0.42)	-4.68	(-5.93 to -3.43)
Form B	-1.73	(-2.45 to -1.02)	-0.28	(-1.23 to 0.66)	-3.41	(-4.5 to -2.33)

PQ: Paper questionnaire, IVR: Interactive Voice Response, PDA: Personal Digital Assistant

Table 2

Reliability estimates

Alternate forms reliability										Cronbach's alpha					
Physical Function			Fatigue			Depression			Physical Function		Fatigue		Depression		
Study 1	ICC	95% CI	ICC	95% CI	ICC	95% CI	Form A	Form B	Form A	Form B	Form A	Form B	Form A	Form B	
PQ	0.91	0.86 to 0.94	0.92	0.90 to 0.94	0.91	0.89 to 0.93	0.92	0.95	0.96	0.97	0.96	0.96	0.96	0.95	
IVR	0.92	0.90 to 0.94	0.89	0.86 to 0.91	0.87	0.83 to 0.90	0.92	0.94	0.96	0.96	0.96	0.96	0.94	0.93	
PC	0.94	0.91 to 0.96	0.94	0.92 to 0.95	0.91	0.89 to 0.93	0.93	0.92	0.96	0.96	0.96	0.96	0.95	0.94	
Study 2															
PDA	0.93	0.90 to 0.95	0.91	0.87 to 0.94	0.85	0.80 to 0.88	0.93	0.93	0.97	0.97	0.94	0.94	0.94	0.94	
PC							0.92	0.92	0.96	0.96	0.92	0.92	0.92	0.94	

PQ: Paper questionnaire, IVR: Interactive Voice Response, PDA: Personal Digital Assistant, PC: Personal Computer

Table 3
Association with exogenous variables and interactions between exogenous variables and method of administration

Label	Physical Function			Fatigue			Depression		
	Effect*	P main effect	P interaction with MOA	Effect*	P main effect	P interaction with MOA	Effect*	P main effect	P interaction with MOA
Overall health status	5.27	<.0001	0.73	-5.24	<.0001	0.26	-3.20	<.0001	0.06
Extent able to carry out everyday physical activities	5.52	<.0001	0.81	-4.56	<.0001	0.07	-2.80	<.0001	0.01
Past 7 days - Rate fatigue on average	-4.85	<.0001	0.20	8.78	<.0001	0.17	6.30	<.0001	0.02
Past 7 days - Frequency been bothered by emotional problems	-1.46	<.0001	0.42	3.78	<.0001	0.30	7.09	<.0001	0.02
Past 28 days - Number of days had to reduce activity for more than half day	-0.44	<.0001	0.77	0.61	<.0001	0.57	0.46	<.0001	0.27
Past 28 days - Number of days taken off for health reason	-0.33	<.0001	0.57	0.41	<.0001	0.90	0.35	<.0001	0.35
Diagnosed - Arthritis or rheumatism	-5.68	<.0001	0.97	3.15	<.0001	0.87	-1.13	0.04	0.13
Diagnosed - Chronic lung disease (COPD) or chronic bronchitis or emphysema	-5.00	<.0001	0.59	0.54	0.27	0.04	-3.75	<.0001	0.53
Diagnosed - Depression	-0.73	0.08	0.18	5.08	<.0001	0.84	10.83	<.0001	0.87

* Parameters estimated in a model without interaction terms.