Published in final edited form as:
Nurs Res. 2005 ; 54(4): 229–234.

An Analytic Strategy for Modeling Multiple-Item Responses:

A Breast Cancer Symptom Example

Ardith Z. Doorenbos, PhD, RN [Postdoctoral Fellow]

College of Nursing, Michigan State University, East Lansing.

Natalya Verbitsky [Doctoral Candidate]

Department of Statistics, The University of Michigan, Ann Arbor.

Barbara Given, PhD, RN, FAAN [University Distinguished Professor]

College of Nursing, Michigan State University, East Lansing

Charles W. Given, PhD [Professor]

Department of Family Practice, College of Human Medicine, Michigan State University, East Lansing

Abstract

Background—Item Response Theory (IRT) is increasingly applied in health research to combine information from multiple-item responses. IRT posits that a person's susceptibility to a symptom is driven by the interaction of the characteristics of the symptom and person. This article describes the statistical background of incorporating IRT into a multilevel framework and extends this approach to longitudinal health outcomes, where the self-report method is used to construct a multi-item scale.

Methods—A secondary analysis of data from 2 descriptive longitudinal studies is performed. The data include 21 symptoms reported across time by 350 women with breast cancer. A 3-level hierarchical linear model (HLM) was used for the analysis. Level 1 models the item responses, consisting of symptom presence or absence. Level 2 models the trajectory of each individual, representing change over time of the IRT-created latent variable symptom experience. Level 3 explains that trajectory using person-specific characteristics such as age and location of care. The purpose of the analysis is to examine if older and younger women with breast cancer differ in their symptom experience trajectory after controlling for location of care.

Results—Fatigue and pain were the most prevalent symptoms. The symptom experience of women with breast cancer was found to improve over time. Neither age nor location of care was significantly associated with the symptom experience trajectory.

Discussion—Embedding IRT into an HLM framework produces several benefits. The example provided demonstrates benefits through the creation of a latent symptom experience variable that can be used either as an outcome or as a covariate in another model, examining the latent symptom experience trajectory and its relationship with covariates at the individual level, and managing symptom nonresponse.

Keywords

cancer symptoms; hierarchical linear model; Item Response Theory; women with breast cancer

In studying symptoms of disease, exposure to risk, behavior, beliefs, and attitudes, nursing researchers frequently have to combine a number of item responses. In such studies, participants may be repeatedly assessed over time; nested within social settings, such as hospitals, nursing homes, or communities; or both. With social settings in particular, a wide variability of gathered data results from a sometimes unknown multitude of sources. Missing item-level data are often unavoidable as well. Some examples of questions that arise in such cases, requiring robust analytical methodologies are: How does the symptom experience of women with breast cancer change over the course of the chemotherapy treatment? Does age affect engagement in risky behaviors that contribute to AIDS/HIV? Do individual beliefs and cultural attitudes influence the acceptance of differing end-of-life care paradigms?

Item Response Theory

Item Response Theory (IRT) was developed in the 1980s in educational research to address some of the issues of measurement practices in scoring tests (McDonald, 1999; van der Linden & Hambleton, 1997). The IRT models postulate that characteristics of a test item, such as its difficulty, interact with an individual's ability or trait to determine the probability of a correct response to that item (Cheong & Raudenbush, 2000; Lord, 1980). The simplest IRT model, the Rasch model, has only one parameter per item, namely difficulty. The Rasch model makes the assumption that each item is equally discriminating. When this assumption is true, the resulting scale has a clear interpretation—that difficult items will be answered correctly less frequently than easy items. Besides an item-difficulty scale, the IRT can also provide estimates of latent abilities that can be studied as either explanatory or outcome variables in other models. Of particular significance to this class of analysis, IRT reduces the skewness that commonly arises in composite measures, such as the sum or proportion. This framework has recently been applied to health outcomes (Fortinsky, Garcia, Sheehan, Madigan, & Tullai-McGuinness, 2003; Hays, Morales, & Reise, 2000).

Hierarchical Linear Models

Statistical models that account for nesting of data (e.g., hierarchical linear models [HLMs]) have been growing in popularity (Raudenbush & Bryk, 2002). There has been an increase in the use of HLM in nursing research, especially in research examining patient and organizational outcomes (Cho, 2003; Cho, Ketefian, Barkauskas, & Smith, 2003; Whitman, Davidson, Sereika, & Rudy, 2001). In the past, when confronted with data on individuals nested in organizations, a researcher had to decide whether to perform the analysis at the individual level, thus ignoring the nested structure of the data, or whether to aggregate the variables to the higher level, thus ignoring individual variation within the organizations. In using an HLM analysis, the researcher no longer has to decide at which level to perform the analysis. This avoids problems of misestimating standard errors and of incorrect statistical inference.

There are several benefits of incorporating the IRT into an HLM framework for nursing researchers: (a) It includes the ability to examine multiple dimensions of abilities, traits, or symptoms; (b) it can separate the variation between social settings, such as hospitals, nursing homes, or communities, from the variation between individuals who are nested within these settings; (c) it provides a way to examine the measurement error in the assessment of social settings where individuals are used as informants about their social setting; (d) it allows the researcher to examine the relationship between explanatory variables at various levels (e.g., individual or setting) and the ability or trait; (e) it provides a framework for incorporating repeated observations of item responses to examine changes in the latent ability over time; and (f) the combined framework also provides a way to manage item nonresponse (Raudenbush & Bryk, 2002). These items exemplify the benefits of embedding an IRT model into an HLM framework as a tool for studying symptoms and other self-reported health behavior. A more detailed theoretical discussion regarding incorporating the IRT into an HLM framework can

be found in Raudenbush, Johnson, and Sampson (2003), and Johnson and Rauden-bush (in press).

Having described the statistical background of incorporating IRT into an HLM framework, the purpose of this article is to illustrate this methodology using an example of the symptom experience for women with breast cancer. This demonstrates the methodology by extending the approach to longitudinal data with health outcomes, where the self-report method is used to construct a multi-item scale. The aim of the analysis is to examine if older and younger women differ in their symptom experience trajectory after controlling for location of care.

An Example: Three-Level HLM Model Incorporating a Symptom IRT

An important aspect of symptom research is how symptom experience varies over time according to the characteristics of the individual and setting. For example, the symptom experience may change differently over time for each woman with breast cancer. Age may influence the relationship, as older women may tend to report fewer symptoms and thus have better symptom experience than younger women at diagnosis and start of chemotherapy. However, younger women may tend to return to the prediagnosis symptom experience faster than older women. Additionally, medical care can affect the symptom experience trajectory. Women receiving care at urban hospitals may have a greater accessibility to medical treatments and thus experience fewer symptoms overall than those at rural hospitals. Incorporating IRT into an HLM framework allows us to examine these and other similar questions.

This example describes a longitudinal Rasch model, which incorporates repeated measures on 21 symptoms at four time points over a 1-year period. Following HLM terminology, we have symptoms at Level 1 nested in repeated measures at Level 2 that are, in turn, nested in individuals at Level 3. To keep the model simple, only three covariates are included in the model: two individual characteristics (age and location of care) and time since diagnosis. In this analysis, the Rasch model orders the responses to a set of items (symptom presence or absence) according to a symptom's characteristic of prevalence in lieu of the traditional "item difficulty." The analog to the typical IRT latent ability is then a latent symptom experience.

Data and Participants

This example involves a secondary analysis of data from two descriptive longitudinal studies conducted from 1990 to 1998. There were 242 women from urban hospitals in the first study; 108 women from rural hospitals participated in the second study. These 350 women were newly diagnosed with breast cancer and undergoing chemotherapy. The participants were followed for 1 year and completed telephone interviews on four occasions. At each interview, the presence of 21 symptoms was recorded along with other characteristics.

Inclusion criteria for the primary studies required that women with breast cancer be at least 21 years of age; cognitively intact; and able to speak, read, and write English. Women under the care of a psychologist or psychiatrist, or with a diagnosed emotional or psychological disorder, were excluded. Nurse recruiters approached women who met the inclusion criteria, explained the studies, and obtained written consent. At mutually convenient times, the participants were interviewed by telephone; they also completed self-administered questionnaires. The ages of the participants ranged from 28 to 98, with a mean of 67.72 years ($SD = 11.36$).

Measures

Symptoms were assessed using the self-report Physical Symptom Experience tool (Given et al., 1993). Participants responded regarding the presence of 21 symptoms commonly experienced by individuals with cancer, indicating whether they experienced the symptom (1) or not (0).

Time was coded in days since diagnosis. Demographic information included age and location of care. Location of care was coded rural = 1 if a rural hospital and rural = 0 if an urban hospital. To render the intercept of the regression line meaningful, age was grand-mean centered.

Analysis

Level 1 Model

The Level 1 model is a standard one-parameter item response or Rasch model, with random effects. In applying the Rasch model, item difficulty was used as symptom prevalence. Let $Y_{ijk} = 1$ if the symptom i was present at time j for person k and 0 otherwise. The probability of a symptom being present, $\Pr(Y_{ijk} = 1)$, is denoted by μ_{ijk} . At this level, there are 20 dummy variables, D_{mijk} , representing 20 of the 21 symptoms measured. So the Level 1 equation is

$$\log\left(\frac{\mu_{ijk}}{1 - \mu_{ijk}}\right) = \pi_{0jk} + \sum_{m=1}^{20} \pi_{mjk} D_{mijk} \quad (1)$$

π_{mjk} is interpreted as the prevalence of the symptom m at time j for person k , compared with the reference symptom (i.e., the symptom for which a dummy variable was not included in the model). This model creates an interval scale for the symptoms, where large values of π_{mjk} indicate more prevalent symptoms while low values indicate less frequent symptoms. By IRT convention, the prevalence for the reference symptom (fatigue) is fixed at 0. This generates the IRT ordering of symptoms by prevalence.

The IRT latent variable describing aggregated symptoms is symptom experience, or π_{0jk} , which indicates the overall symptom experience at time j for person k . π_{0jk} becomes an outcome at Level 2, where the symptom experience trajectory is examined. Larger values of π_{0jk} indicate a higher relative prevalence of symptoms, while smaller values indicate a lower relative prevalence of symptoms.

Level 2 Model

The Level 2 model accounts for variation in symptom experience over time for each woman with breast cancer. Equation 2 models parameters from the Level 1 model, π_{0jk} and π_{mjk} . To conform to the Rasch methodology, we fixed the prevalence of each symptom (π_{mjk}) across time (Level 2) and individuals (Level 3) in the model. This constraint reflects the belief that, given a symptom experience, random samples of women with breast cancer will experience a symptom with the same prevalence. Otherwise, the symptom may be regarded as biased against a subset of women with breast cancer. At Level 2, the symptom experience (π_{0jk}) is described as a function of time.

$$\begin{aligned} \pi_{0jk} &= \beta_{00k} + \beta_{01k} \times \text{time}_{jk} + u_{0jk} \\ \pi_{mjk} &= \beta_{m0k}, \text{ for } m = 1, \dots, 20 \end{aligned} \quad (2)$$

β_{00k} and β_{01k} represent the initial symptom experience, and the linear daily rate of change in symptom experience for individual k , respectively. The random effects, u_{0jk} , are the deviations at time j of individual k 's symptom experience from the predicted. β_{m0k} represents the prevalence of the symptom m , compared with the reference symptom for individual k .

Level 3 Model

At Level 3, the symptom experience trajectory is explained using person-specific characteristics, such as age and location of care.

$$\begin{aligned}
\beta_{00k} &= \gamma_{00} + \gamma_{001} \times (\text{age}_k - 67.72) \\
&\quad + \gamma_{002} \times \text{rural}_k + \nu_{00k} \\
\beta_{01k} &= \gamma_{010} + \gamma_{011} \times (\text{age}_k - 67.72) \\
&\quad + \gamma_{012} \times \text{rural}_k + \nu_{01k} \\
\beta_{m0k} &= \gamma_{m00}, \text{ for } m=1, \dots, 20
\end{aligned} \tag{3}$$

In this equation, γ_{000} is the expected initial symptom experience for women with breast cancer who are receiving care in urban hospitals; γ_{001} is the expected difference in the initial symptom experience between two women who differ by 1 year in age; γ_{002} is the expected difference in the average initial symptom experience between the rural and urban locations of care; γ_{010} is the expected average daily rate of change in symptom experience for women receiving care at the urban hospitals; γ_{011} is the difference in the expected daily rate of change in symptom experience between two women who differ by 1 year in age; γ_{012} is the difference in the expected daily rate of change in symptom experience between the rural and urban studies; and γ_{m00} is the prevalence of the symptom m compared with the reference symptom.

By combining the Levels 1, 2, and 3 models, the hierarchical generalized linear model can be estimated. The combined model tests how the log-odds of experiencing a symptom vary with time- and person-specific characteristics, such as age and location of care. This three-level hierarchical model can be viewed as an item–response model embedded within a hierarchical structure, in which repeated measures are nested within women with breast cancer.

Missing data were addressed by using complete case analyses (Little & Rubin, 2002); that is, if at least one symptom was recorded as present or absent during an interview, then this interview information was used in the analysis. The number of symptoms recorded (present or absent) during interviews ranged from 10 to 21. Some women had fewer than four interviews, resulting in 1,184 interviews (rather than $350 \times 4 = 1,400$) included at Level 2. Since every woman had at least one interview, and all women had both age and location of care recorded, 350 individuals were included in the analysis at Level 3.

We present two models, an unconditional and a conditional, estimated using HLM 6.20 (Raudenbush, Bryk, Cheong, & Congdon, 2004). The unconditional model has no covariates at Level 2 or 3, which yields a readily interpretable ordering of symptoms as well as the unadjusted symptom experience estimates for each person at each occasion that they recorded presence or absence of at least one symptom. We examined the symptom experience over time by testing linear and quadratic trajectories at Level 2. To test the associations between individual variables and symptom experience, these variables were incorporated into the multivariate model at Level 3. In the final model, Level 1 remains the same as in the unconditional model (see Equation 1), but now, entered into the model are the time-level variable (days since diagnosis) at Level 2, and the individual-level variable (age and location of care) at Level 3, as previously shown in Equations 2 and 3.

Model Results

Unconditional Model

Fatigue, the most common symptom in the raw data, was used as the reference symptom (Table 1). The results of the unconditional model yield a readily interpretable ordering of symptoms. Figure 1 shows the symptoms organized by their prevalence (γ_{m00}); the more prevalent symptoms appear at the top (high values), while the less prevalent appear at the bottom (low values). Symptoms appearing close together in Figure 1 have similar symptom prevalence. Construct validity for this scale was confirmed by the fact that pain and fatigue occurred with greatest frequency, which has been well-established in the cancer literature (Given, Given,

Azzouz, Kozachik, & Stommel, 2001; Mock, 2003; Patrick et al., 2003). The lowest frequency symptom during chemotherapy treatment for breast cancer was dehydration (Table 1).

This model produced an unadjusted symptom experience estimate for each individual and each occasion when the presence of at least one symptom was recorded (Figure 2). These symptom experience estimates are approximately normally distributed and may be used as either a covariate or an outcome in other models.

Conditional Model

Since our focus was to determine the trajectory of symptom experience as well as its association with important individual variables, partial output for Levels 2 and 3 is reported in Table 2. Both the linear and quadratic trajectories were tested; however, the quadratic term did not significantly improve the model fit. Therefore, a linear trajectory was used in the final model. Time was significantly, negatively associated with symptom experience ($\gamma_{010} = -.002, p < .001$). So, as women with breast cancer moved through the year, on average their symptom experience improved. Since we hypothesized that the symptom experience trajectory may differ according to women's age and location of care, age and rural were used to explain initial symptom experience and change in symptom experience. Neither age nor location of care was found to be statistically significantly associated with change of symptom experience over time ($p = .085$ and $p = .819$, respectively), nor were they found to be statistically significantly associated with symptom experience ($p = .173$ and $p = .150$, respectively; Table 2).

Discussion

The benefits of using IRT are illustrated by the results of the unconditional model. First, IRT created a meaningful metric that reflects the varying prevalence of symptoms in women with breast cancer (Figure 1), while reducing the skewness that commonly arises in composite measures of symptoms (Figures 2 and 3). Second, the analysis provided estimates of the latent symptom experience for each person at each occasion when the presence or absence of at least one symptom was recorded. These symptom experience estimates can be used as explanatory or outcome variables in other models.

Several of the numerous benefits of embedding IRT into an HLM framework were illustrated in the example above. First, this methodology provided a framework for incorporating repeated observations on the presence of symptoms to examine changes in the latent variable symptom experience over time. The results on the conditional model showed that the symptom experience of women with breast cancer improved over time. Second, this methodology allowed us to examine the relationship between individual variables and the latent symptom experience. Contrary to our hypothesis, controlling for the location of care, no statistically significant association of age with the symptom experience trajectory was found.

Moreover, the hierarchical framework provides a way to manage item nonresponse. Presently, two common approaches for combining symptom information use the sum or the proportion of the symptoms present. When using a summary score, a nursing researcher must make an arbitrary decision regarding how to handle item non-response. Using the sum assumes that everyone has the same number of symptoms recorded, but not necessarily present; the symptoms that are not recorded are assumed to be not present. The proportion approach assumes that each symptom contributes the same amount of information, which is again problematic, since some symptoms occur more frequently than others. A researcher can use an IRT model without having to decide what to do with missing data, as long as the data are assumed missing at random, a comparatively mild assumption (Little & Rubin, 2002).

The task of combining information from multiple-item responses arises frequently in studies of health outcomes. In many of these studies, the items are measured over time and nested within individuals, and item-level missing data are often unavoidable. This report demonstrates how embedding a Rasch model into HLM can address these research challenges.

Acknowledgements

The authors thank Professor Stephen W. Raudenbush, PhD, School of Education, University of Michigan, for his review of and comments on this article. The study was funded by the US Army Medical Research and Materiel Command under W81XWH-04-1-0469, PI Ardith Doorenbos, and the National Cancer Institute Grant #R01 CA79280: Family Home Care for Cancer: A Community-Based Model PI Barbara Given.

References

- Cheong YF, Raudenbush SW. Measurement and structural models for children's problem behaviors. *Psychological Methods* 2000;5:477–495. [PubMed: 11194209]
- Cho SH. Using multilevel analysis in patient and organizational outcomes research. *Nursing Research* 2003;52:61–65. [PubMed: 12552178]
- Cho SH, Ketefian S, Barkauskas VH, Smith DG. The effects of nurse staffing on adverse events, morbidity, mortality, and medical costs. *Nursing Research* 2003;52:71–79. [PubMed: 12657982]
- Fortinsky RH, Garcia RI, Sheehan J, Madigan EA, Tullai-McGuinness S. Measuring disability in Medicare home care patients: Application of Rasch modeling to the outcome and assessment information set. *Medical Care* 2003;41:601–615. [PubMed: 12719685]
- Given CW, Given B, Azzouz F, Kozachik S, Stommel M. Predictors of pain and fatigue in the year following diagnosis among elderly cancer patients. *Journal of Pain and Symptom Management* 2001;21:456–466. [PubMed: 11397603]
- Given CW, Stommel M, Given B, Osuch J, Kurtz ME, Kurtz JC. The influence of cancer patients' symptoms and functional status on patients' depression and family care-givers' reaction and depression. *Health Psychology* 1993;12:277–285. [PubMed: 8404801]
- Hays RD, Morales LS, Reise S. Item Response Theory and health outcomes measurement in the 21st century. *Medical Care* 2000;38(Suppl II):28–42.
- Johnson, C.; Raudenbush, SW. A repeated measures, multilevel Rasch model with application to self-reported criminal behavior. In: Bergeman, CS.; Boker, SM., editors. *Methodological issues in aging research*. Erlbaum; Mahwah, NJ: in press
- Little, RJ.; Rubin, DB. *Statistical analysis with missing data*. 2nd ed.. Wiley; Hoboken, NJ: 2002.
- Lord, FM. *Applications of item response theory to practical testing problems*. Erlbaum; Hillsdale, NJ: 1980.
- McDonald, RP. *Test theory: A unified treatment*. Erlbaum; Mahwah, NJ: 1999.
- Mock V. Clinical excellence through evidence-based practice: Fatigue management as a model. *Oncology Nursing Forum* 2003;30:790–796.
- Patrick DL, Ferketich SL, Frame PS, Harris JJ, Hendricks CB, Levin B, et al. National Institutes of Health State of the Science Conference statement: Symptom management in cancer. *Journal of the National Cancer Institute* 2003;95:1110–1117. [PubMed: 12902440]
- Raudenbush, SW.; Bryk, AS. *Hierarchical linear models*. 2nd ed.. Sage; Thousand Oaks, CA: 2002.
- Raudenbush, SW.; Bryk, A.; Cheong, YF.; Congdon, R. *HLM 6: Hierarchical linear and nonlinear modeling*. Scientific Software International; Lincolnwood, IL: 2004.
- Raudenbush SW, Johnson C, Sampson RJ. A multivariate, multilevel Rasch model with application to self-reported criminal behavior. *Sociological Methodology* 2003;33:169–211.
- van der Linden, W.; Hambleton, RK. *Handbook of modern item response theory*. Springer; New York: 1997.
- Whitman GR, Davidson LJ, Sereika SM, Rudy EB. Staffing and pattern of mechanical restraint use across a multiple hospital system. *Nursing Research* 2001;50:356–362. [PubMed: 11725937]

FIGURE 1.
Symptom prevalences (γ_{m00}) according to the unconditional model.

FIGURE 2.

Histogram of symptom experience from the unconditional model, π_{0jk} . Some of the 350 women had fewer than four interviews, resulting in 1,184 total interviews.

FIGURE 3. Histogram of total number of symptoms present. Some of the 350 women had fewer than four interviews, resulting in 1,184 total interviews.

TABLE 1

Raw Data Symptom Frequency (All Observations Combined)

Symptoms	Yes	No	Total	Yes (%)
Fatigue	693	491	1,184	58.53
Pain	462	722	1,184	39.02
Insomnia	426	758	1,184	35.98
Dry mouth	402	781	1,183	33.98
Loss of feeling	326	857	1,183	27.56
Urinary frequency	301	882	1,183	25.44
Weakness	289	894	1,183	24.43
Cough	262	921	1,183	22.15
Constipation	225	958	1,183	19.02
Nausea	208	975	1,183	17.58
Concentration	187	997	1,184	15.79
Poor appetite	182	1000	1,182	15.40
Shortness of breath	169	1015	1,184	14.27
Weight loss	161	1017	1,178	13.67
Coordination problems	103	1081	1,184	8.70
Diarrhea	102	1082	1,184	8.61
Mouth sores	66	1117	1,183	5.58
Difficulty swallowing	57	1124	1,181	4.83
Vomiting	46	1138	1,184	3.89
Fever	45	1138	1,183	3.80
Dehydration	26	1156	1,182	2.20

Note. Some of the 350 women had fewer than four interviews and the presence of fewer than 21 symptoms recorded at each interview resulting in 1,184 interviews and varying totals.

TABLE 2

Partial Output of Estimates for the Conditional Model (Excluding Symptoms Prevalence)

Variable	Coefficient	Standard Error	<i>p</i>
Intercept, γ_{000}	.605	0.106	<.001
Age, γ_{001}	-.010	0.007	.173
Rural, γ_{002}	.268	0.185	.150
Time, γ_{010}	-.002	0.0003	<.001
Time $\times$ Age, γ_{011}	.00005	0.00003	.085
Time $\times$ Rural, γ_{012}	.0002	0.0007	.819