

# Building PROMIS item banks: librarians as co-investigators

Mary Klem · Ester Saghafi · Rebecca Abromitis ·  
Angela Stover · Mary Amanda Dew · Paul Pilkonis

Accepted: 2 June 2009 / Published online: 23 June 2009  
© Springer Science+Business Media B.V. 2009

## Abstract

**Purpose** There is growing interest in the use of item response theory (IRT) for creation of measures of health-related quality of life (HRQOL). A first step in IRT modeling is development of item banks. Our aim is to describe the value of including librarians and to describe processes used by librarians, in the creation of such banks.

**Method** Working collaboratively with PROMIS researchers at the University of Pittsburgh, a team of librarians designed and implemented comprehensive literature searches in a selected set of information resources, for the purpose of identifying existing measures of patient-reported emotional distress.

**Results** A step-by-step search protocol developed by librarians produced a set of 525 keywords and controlled vocabulary terms for use in search statements in 3 bibliographic databases. These searches produced 6,169 literature citations, allowing investigators to add 444 measurement scales to their item banks.

**Conclusion** Inclusion of librarians on the Pittsburgh PROMIS research team allowed investigators to create large initial item banks, increasing the likelihood that the banks would attain high measurement precision during subsequent psychometric analyses. In addition, a comprehensive literature search protocol was developed that can now serve as a guide for other investigators in the creation of IRT item banks.

**Keywords** Databases as topic · Outcome assessment (health care) · Librarians · Interdisciplinary communication

## Introduction

Conventional measures of patient-reported outcomes, including health-related quality of life (HRQOL), have been criticized as cumbersome in nature, burdensome for patients to complete, or lacking a standardized scoring metric [1, 2]. As a result, there has been increasing interest in the application of item response theory (IRT) modeling to health outcomes research, and it seems likely that IRT will be a driving force behind the development of a new generation of HRQOL measures [3].

The patient-reported outcome measurement information system (PROMIS) is an NIH roadmap initiative, with the goal of using IRT to develop item banks, static short forms, and computerized adaptive tests (CATs) for measurement of patient-reported health outcomes among individuals with a variety of health conditions [4]. The network of seven primary PROMIS research sites focused initially on five key health domains for cross-site collaboration and item bank development: emotional distress, social functioning, physical functioning, pain, and fatigue. Within the network, researchers at the Pittsburgh PROMIS site were responsible for the development, testing, and validation of the emotional distress banks for depression, anxiety, anger, and substance abuse. Pittsburgh PROMIS also developed item banks for sleep-wake function.

IRT modeling begins with the collection of existing questionnaires for the purpose of creating a preliminary bank of items [2, 5, 6]. Failure to create a large initial collection of items can lead, during subsequent psychometric

---

M. Klem (✉) · E. Saghafi · R. Abromitis · A. Stover ·  
M. A. Dew · P. Pilkonis  
Pittsburgh, PA, USA  
e-mail: klem@pitt.edu

analyses, to a limited effectiveness range of measurement and a loss of measurement precision [7]. For these reasons, Fries et al. [8] recommend that the search for existing questionnaires be considered analogous to the comprehensive literature searching done for a meta-analysis, i.e., the purpose of the search should be to identify all existing instruments used to assess the domain of interest.

Such comprehensive literature searching requires a depth and breadth of search strategies unfamiliar to many researchers and clinicians [9, 10]. The Cochrane Collaboration, an organization devoted to the production of comprehensive literature reviews on health-related topics, recommends that comprehensive literature searches utilize multiple bibliographic databases and complex search strings consisting of controlled vocabulary, free-text terms, extensive lists of synonyms, and truncation [11]. Evidence suggests, however, that many researchers fail to follow these guidelines [12–15]. Mokkink et al. [16] reviewed the methodological quality of 148 systematic reviews of health status instruments and found that authors searched relatively few databases, used minimal numbers of terms to describe the concepts of interest, made limited use of synonyms, and imposed seemingly arbitrary time period limits. Golder et al. [17] found similar limitations in the search methods of systematic reviews of the adverse effects of drugs, commenting that the methods were of “variable quality” (p. 443). While Mokkink et al. did not address whether or not librarians or other information professionals had assisted with the systematic reviews, Golder et al. did note that systematic reviews with librarians or other information professionals as co-authors appeared more likely to have used complex search strategies and to have adequately described those strategies in the final paper.

Failure to conduct a comprehensive literature search may lead to the inadvertent exclusion of relevant studies [15]. For example, Mokkink et al. [16], noting a systematic review based on 17 primary studies of a disability questionnaire, were able to identify an additional 23 relevant primary studies, simply by altering the search terms used, utilizing truncation, and placing no limits on the timeframe

of the search. Omitting relevant studies can, in turn, create concerns about possible bias of a study’s outcome or conclusions [8, 18].

Successful completion of comprehensive literature searches requires specialized skills and knowledge [9]. Librarians, as information professionals, possess extensive knowledge of information resources, including content, coverage, structure, and search vocabulary. As such, librarians can play “...a significant role in the expert retrieval and evaluation of information in support of knowledge and evidence-based clinical, scientific, and administrative decision-making” [9, p. 1]. In this article, we describe in detail the process used by a team of health sciences librarians to develop and complete comprehensive literature searches for the Pittsburgh PROMIS group. In doing so, we demonstrate the value of including librarians as members of a multidisciplinary research team and provide HRQOL researchers with a proven protocol for conducting the thorough literature searches that are an integral part of the IRT modeling process [8].

## Methods

### Identification and searching of information sources

Comprehensive harvesting of search vocabularies and exhaustive searching of the literature for the Pittsburgh PROMIS project required identification and searching of diverse electronic and print information sources from several related fields. Twelve information sources were initially identified by the librarian team as likely to be either good sources for harvesting vocabulary for subsequent searches of the measurement literature or to contain information about measurement instruments. The librarian team then prioritized these twelve sources based on reliability, authoritativeness, relevance, and potential for yielding information on measurement instruments in the emotional distress and sleep-wake function domains of interest to the Pittsburgh PROMIS investigators (Table 1).

**Table 1** Prioritized information sources

Priority I	Priority II	Priority III
MEDLINE	Mental measurements yearbook	Test critiques
PsycINFO	CINAHL (cumulative index to nursing and allied health literature)	Directory of unpublished experimental mental measurements
HaPI (health and psychosocial instruments)	Social work abstracts	Selected measurement textbooks and handbooks
	Ageline	
	ERIC (education resources information center)	
	Test link	

In addition to all twelve sources, National Library of Medicine (NLM) and Library of Congress (LC) classifications were also consulted for vocabulary harvesting.

Creating search vocabularies for selected information sources

The next step in the literature search process was to create a list of vocabulary terms that could be used for searching in priority databases. The purpose of each literature search was to identify citations of studies describing the development or use of self-report measures of the identified domains. Thus, vocabulary lists were constructed to capture three distinct aspects of the investigators’ search requests:

- the constructs of emotional distress (e.g., depression, anxiety) and sleep-wake function
- development or use of a measurement tool
- use of self- or patient-report (as opposed to clinician-administered) instruments

Vocabulary terms to be used in searching the literature for the emotional distress and sleep-wake constructs were identified through the MEDLINE thesaurus online, Medical Subject Headings (MeSH), and the American Psychological Association’s (APA) online Thesaurus of Psychological Terms. These thesauri contain standardized languages known as “controlled vocabulary”. Controlled vocabulary

provides a consistent way to describe and retrieve information that may use different terminology to describe the same concepts (e.g., in PsycINFO, the controlled vocabulary term “anxiety” will retrieve the records of articles in which authors have discussed anxiety, anxiousness, worry, or angst). Search vocabulary terms that reflected the concept of measurement (including the concept of self-report) were identified for each database by searching their thesauri, indexes, and scope notes. Additional terms were identified by hand searching the NLM and LC classification schemes, as well as texts by Bellack and Hersen [19] and Knapp [20].

Vocabulary terms compiled for emotional distress and sleep-wake function, measurement development or use, and self/patient-report were initially entered into separate Excel spreadsheets for each topic. The worksheets were then uploaded into a Microsoft Access database. Each table within the database contained a column for potential search vocabulary for a single domain, and an additional column for a code representing the source from which the vocabulary was harvested. An excerpt from one such table is reproduced in Fig. 1. To allow investigators to view search terms and vote on their relevance to the domain of interest, investigators could quickly scroll through the table and vote for the exclusion or retention of any item by clicking on a “retain” box next to each term. Investigators were

**Fig. 1** Database table containing harvested search terms and investigator votes to retain or discard possible search terms for sleep-wake function

Item	Source	Invest 01	Invest 02	Invest 03	Invest 04
airway obstruction	CINAHL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
airway obstruction	Medline	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
ALERT	SWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Alert	MMY	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
alertness	SWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
alertness	MMY	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
ALERT-WAKEFULNESS	SWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
apnea	CINAHL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
apnea	MMY	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
AWAKE	SWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
AWAKEN	SWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
AWAKENED	SWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
AWAKENING	SWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
AWAKENINGS	SWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
AWAKENS	SWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Bedtime	LC	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
biological clocks	CINAHL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
biological clocks	Medline	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Biological rhythms (use for 1967-72)	PSY	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Bruxism	PSY	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
cataplexy	Medline	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
chronobiology disorders	CINAHL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
chronobiology disorders	Medline	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
chronotherapy	CINAHL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
chronotherapy	Medline	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
CIRCADIAN	SWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
circadian preference	PSY	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
circadian rhythm	CINAHL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
circadian rhythm	Medline	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
circadian rhythms	AgeLine	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

also able to suggest additional search terms by making entries into a supplemental table. Librarians tabulated the investigators' votes and used this input to help in final selection of search terms. In some cases, terms rejected by investigators were retained by librarians, based upon their knowledge of how terms are used in a particular database.

#### Determining best search strategies

Based on their knowledge of database content and coverage, the librarian team determined MEDLINE, PsycINFO, and Health and Psychosocial Instruments (HaPI) to be most relevant to the Pittsburgh PROMIS project and assigned one each to a librarian on the team for the purpose of conducting the initial literature searches. Although vocabulary terms were harvested from other sources (Table 1), investigators ultimately decided that literature searches in the above three databases retrieved sufficient measures to yield an optimum number of test items.

Initial literature searches were run in designated databases using complete lists of vocabulary terms generated through the process outlined above. Search terms reflecting each domain of interest (e.g., depression) were combined with measurement terms (e.g., validity, psychometric) and self-report terms (e.g., self-inventory, personal monitoring). This search strategy narrowed the retrieval to the assessment and instrument literature for each domain of interest. Results of the initial searches were reviewed, and, when necessary, librarians edited or otherwise tailored the search strategies to further refine search results. The purpose of any such editing was to maximize retrieval of relevant citations while minimizing retrieval of irrelevant citations. The strategies used by librarians to edit searches are briefly described below.

**MEDLINE** In MEDLINE, to represent emotional domains, most search strategies were constructed using controlled vocabulary (MeSH) and employing the EXPLODE function as needed: this function automatically retrieves all narrower terms associated hierarchically with a broader term. The concepts of measurement and self-report were primarily represented by keywords previously harvested by the librarian team. Given the fact that very large retrievals are possible in MEDLINE searches, in most cases the precision of search strategies was increased by application of a *focus* limit to terms representing the emotional distress and sleep-wake function domains. This limits a search to include only those citations in which the vocabulary term is considered the major point of the entire article.

**PsycINFO** The PsycINFO search utilized a combination of controlled vocabulary and keywords to search the domains and the concepts of measurement and self-report. Several search enhancements were used to improve the

initial search results. Use of the limit *Test and Measures* helped to assure that assessment tools were discussed within the articles and also allowed investigators to quickly identify the *names* of the tests discussed. PsycINFO's classification codes were applied as a second type of search enhancement. Use of the codes cast the broadest net to capture any and all categories of testing in the search formula. Measurement search terms were not *focused*. The use of *focus* could have potentially eliminated journal articles in which multiple measurement instruments were discussed.

**HaPI (Health and Psychosocial Instruments)** HaPI is a specialized database designed to provide users with information on measurement instruments, and as such, search strategies for the database did not require terminology reflecting the concept of measurement. However, because HaPI records are indexed using controlled vocabulary from both MeSH and the PsycINFO Thesaurus, maximum precision in retrieval was achieved by using search statements containing controlled vocabulary from both sources. In addition, the citations of instruments already known to investigators (e.g., Center for Epidemiologic Studies Depression Scale; Radloff, [21]) were excluded from final search results.

#### Additional search conventions

Finally, the following search conventions were used across all databases:

- Drug classes were used to locate studies relating to the emotional domains of interest (e.g., "antidepressants" identified studies of depression); specific trade or generic drug names were not included.
- Names of actual tests or known measures were not included as search terms.
- Language limits (e.g., *English language-only*) were not applied. Although initially requested by investigators, librarians determined that use of this limit would have minimal impact on search results.
- The limit *abstract only* was not used, allowing inclusion of all pertinent citations regardless of the presence or absence of an abstract.
- While no age limits are available in HaPI, the limit *all adult* was applied in MEDLINE searches, and the limit *adulthood* was applied in PsycINFO searches.

## Results

#### Results of vocabulary harvesting and database searches

The extensive lists of vocabulary terms from each of the five domains, combined with the measurement terms and the self-report terms, were used to create the search

strategies for locating the scales. Initially, 3,123 terms were harvested from the twelve print/electronic resources identified by the librarians. This list was further refined through review by librarians and investigators (see Creating Search Vocabularies for Selected Information Sources). From this list, a total of 525 terms were used to search three primary databases (MEDLINE, PsycINFO and HaPI). The combined search results from the three databases, across five domains, yielded 6,169 citations (1,204 for depression, 1,738 citations for anxiety, 1,415 for substance use, 1,277 for anger, 535 for sleep-wake function).

#### Review of initial searches

The results of the first literature search (measures of anxiety) were reviewed jointly by PROMIS investigators and the collaborating librarians. Reassured that the results were satisfactory, subsequent searches for the remaining domains were tailored by librarians without further review by investigators.

The initial search results for each of the domains were reviewed and coded by Pittsburgh PROMIS research staff. Those citations not meeting inclusion criteria, including duplicate citations, citations of instruments already known to investigators, and irrelevant citations, were removed from the sets of results. For each identified instrument that met inclusion criteria, staff located the original article describing the instrument's development and psychometric properties and performed a cited reference search for that article using the ISI Web of Knowledge database. The cited reference search provided an estimate of the number of times the original article had been cited by authors of other scientific publications, and thus was viewed as a rough measure of the instrument's "acceptability" or use within the scientific community. Investigators used results of the cited reference searches, as well as expert content knowledge, to select a final set of 444 potential measures (Fig. 2).

Individual items from the measures were then placed into initial item pools for each of the domains. Investigators organized the items into "bins" or categories reflecting conceptual characteristics of the larger domains. For example, items in the depression pool were placed into the categories of mood, cognition, behavior, somatic complaints, or suicidality. If necessary, subcategories were also created. Investigators then reviewed all items, with the aim of reducing redundancy among each domain's items, as well as removing vague or confusing items, or those with insufficient face validity [22]. Review of items continued until each pool was reduced to between 100 and 200 items.

The reduced item pools then underwent qualitative item review utilizing such techniques as focus groups and cognitive interviews with psychiatric outpatients. Initial pilot testing of the items using computerized administration was

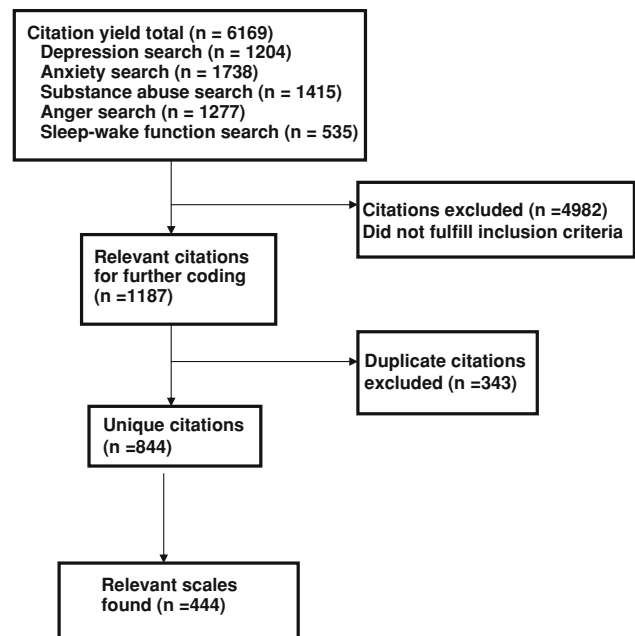


Fig. 2 Literature search results

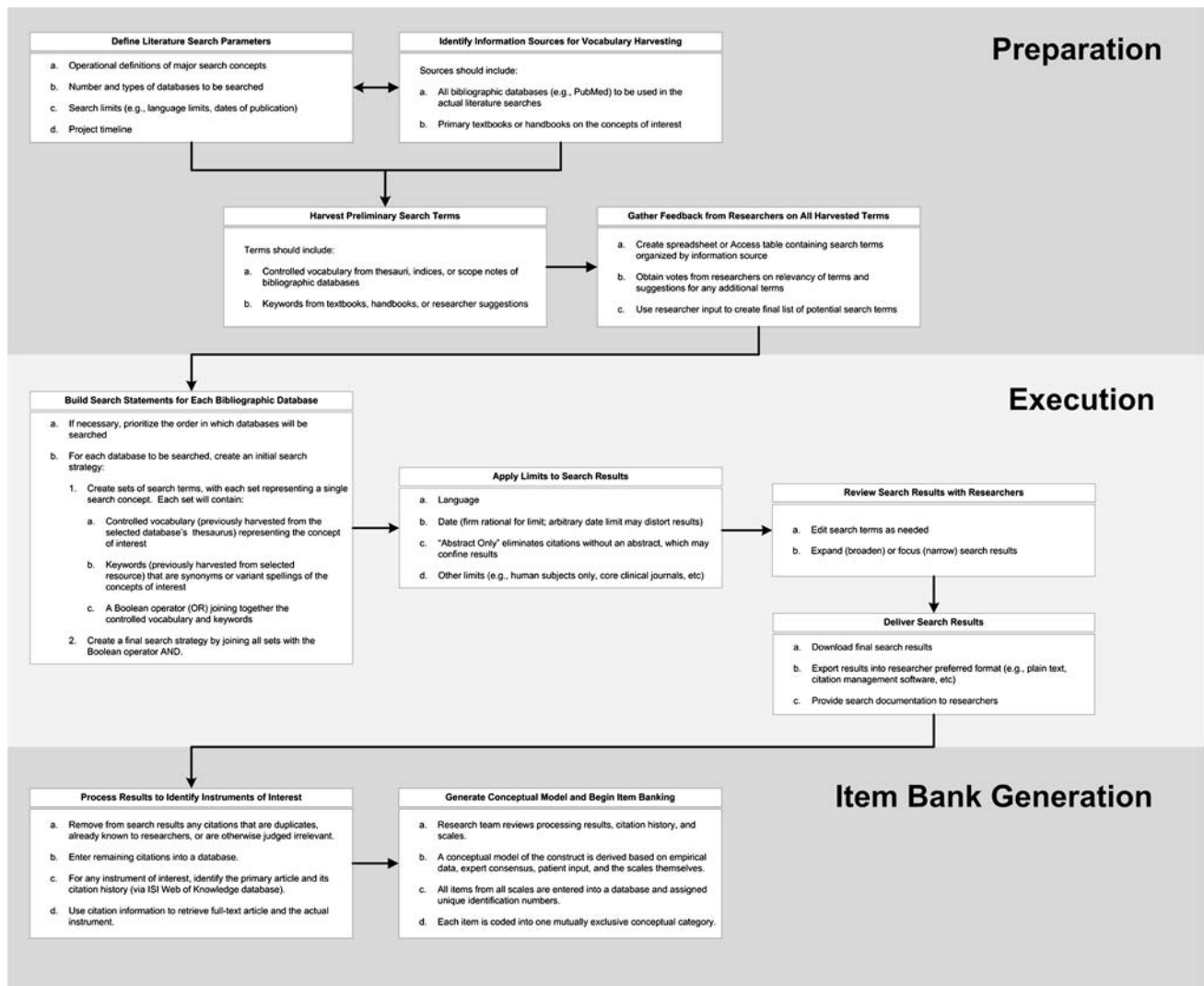
also conducted. The purpose of these steps was to gather information from both patients and content experts about the importance, clarity, and usefulness of items in describing the domains of interest [8]. Using this information, inadequate items were removed, remaining items were refined, and additional items created. The revised item pools were administered to large general population and clinical samples and calibrated using models based on IRT. Using these IRT calibrations, CAT algorithms have been generated. For more information about PROMIS item banks, see [www.nihpromis.org](http://www.nihpromis.org).

#### Creation of a literature search protocol

Finally, the PROMIS librarians created a literature search protocol (Fig. 3) that outlines the steps required to prepare for, execute, and use the results of, comprehensive literature searches. This protocol is currently being used to guide the development of additional literature searches for the Pittsburgh PROMIS group.

#### Discussion

Although existing models for research collaboration frequently include expertise from multiple subdisciplines (e.g., biostatistics), librarians have not traditionally been included as members of health sciences research teams. However, our study shows that a team of librarians was able to create a literature search protocol that allowed Pittsburgh PROMIS investigators to retrieve the self-report



**Fig. 3** Identifying items for inclusion in item banks: a step-by-step process

instruments required for development of IRT item banks. Thus, librarians added a “quality component” [23, p. 90] to this complex research project.

To complete exhaustive searches for the Pittsburgh PROMIS group, librarians worked first with investigators to clarify and focus their information needs. Using their understanding of those needs and their knowledge of pertinent resources, librarians then identified information sources likely to yield measurements of interest and completed systematic and comprehensive searches of those sources.

The librarian team meticulously documented all aspects of the search process, including information sources used, search term harvesting, search strategy construction, and management of search results, and then used this documentation to create a comprehensive literature search protocol (Fig. 3). These actions benefited the Pittsburgh PROMIS group in several ways. First, Pittsburgh PROMIS

investigators can easily show that searches for measures were comprehensive and unbiased—important requirements in development of IRT-based CAT applications [4]. Secondly, such detailed record-keeping enhances the reproducibility of the project [25]—the Pittsburgh group will be able to easily replicate their searches in the future and identify any newly published instruments measuring their domains of interest.

Other HRQOL researchers may also benefit from the work conducted by the PROMIS librarians. As noted earlier, there is increasing interest in the use of IRT as a methodology for creating QOL measures [3], and an important step in the use of IRT is the development of item banks [2, 5, 6]. While the PROMIS literature search protocol was developed by a research group interested in identifying self-report instruments measuring emotional distress, the information retrieval principles and strategies outlined in the protocol

would be effective across a variety of search topics. Thus, the PROMIS protocol can serve as a guide to all QOL researchers involved in item banking activities. The protocol may also be relevant to researchers interested in or involved in the writing of systematic reviews.

Librarians' involvement with the Pittsburgh PROMIS project also benefited research team members in some unanticipated ways. PROMIS staff were trained by a librarian to conduct cited reference searches in the ISI Web of Knowledge database, allowing them to gather data on instrument acceptability within the scientific community. Over the course of the project, librarians conducted several additional literature searches to assist in investigator discussions of psychometric issues (e.g., choice of response sets) and were able to notify investigators about new information sources pertinent to the goals of the PROMIS project (e.g., the university's purchase of new citation management software). Librarians benefited from the involvement as well: their subject area knowledge expanded as they discussed the emotional domains and conceptual issues of interest to Pittsburgh PROMIS investigators and as they engaged in the iterative process of developing initial vocabulary lists. Finally, the librarians also gained a better appreciation of the time and work commitments required for participation in a large federally funded, multi-center research project.

**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 (Evanston Northwestern Healthcare, 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, Laura Lee Johnson, PhD, Louis Quatrano, PhD, Bryce Reeve, PhD, William Riley, PhD, Susana Serrate-Sztejn, MD, and James Witter, MD, PhD. See the web site at [www.nihpromis.org](http://www.nihpromis.org) for additional information on the PROMIS cooperative group. We would like to acknowledge the contributions of all Pittsburgh PROMIS co-investigators and staff.

## References

- Reeve, B. B., Hays, R. D., Chang, C. H., & Perfetto, E. M. (2007). Applying item response theory to enhance health outcomes assessment. *Quality of Life Research*, 16(Suppl 1), 1–3. doi:10.1007/s11136-007-9220-6.
- Chang, C. H. (2007). Patient-reported outcomes measurement and management with innovative methodologies and technologies. *Quality of Life Research*, 16(Suppl 1), 157–166. doi:10.1007/s11136-007-9196-2.
- Hays, R. D., & Lipscomb, J. (2007). Next steps for use of item response theory in the assessment of health outcomes. *Quality of Life Research*, 16(Suppl 1), 195–199. doi:10.1007/s11136-007-9175-7.
- Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., et al. (2007). The patient-reported outcomes measurement information system (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. *Medical Care*, 45(5, Suppl 1), S3–S11. doi:10.1097/01.mlr.0000258615.42478.55.
- Bjorner, J. B., Chang, C. H., Thissen, D., & Reeve, B. B. (2007). Developing tailored instruments: Item banking and computerized adaptive assessment. *Quality of Life Research*, 16(Suppl 1), 95–108. doi:10.1007/s11136-007-9168-6.
- Revicki, D. A., & Cella, D. F. (1997). Health status assessment for the twenty-first century: Item response theory, item banking and computer adaptive testing. *Quality of Life Research*, 6(6), 595–600. doi:10.1023/A:1018420418455.
- Cook, K. F., Teal, C. R., Bjorner, J. B., Cella, D., Chang, C. H., Crane, P. K., et al. (2007). IRT health outcomes data analysis project: An overview and summary. *Quality of Life Research*, 16(Suppl 1), 121–132. doi:10.1007/s11136-007-9177-5.
- Fries, J. F., Bruce, B., & Cella, D. (2005). The promise of PROMIS: Using item response theory to improve assessment of patient-reported outcomes. *Clinical and Experimental Rheumatology*, 23(5, Suppl 39), S53–S57.
- Medical Library Association. (2005). Role of expert searching in health sciences libraries. *Journal of the Medical Library Association: JMLA*, 93(1), 42–44.
- Davidoff, F., & Florance, V. (2000). The informationist: A new health profession? *Annals of Internal Medicine*, 132(12), 996–998.
- Lefebvre, C., Manheimer, E., & Glanville, J. (2008). Chapter 6: Searching for studies. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 95–150). England: Wiley.
- Delaney, A., Bagshaw, S. M., Ferland, A., Laupland, K., Manns, B., & Doig, C. (2007). The quality of reports of critical care meta-analyses in the Cochrane database of systematic reviews: An independent appraisal. *Critical Care Medicine*, 35(2), 589–594. doi:10.1097/01.CCM.0000253394.15628.FD.
- Glenny, A., Esposito, M., Coulthard, P., & Worthington, H. (2003). The assessment of systematic reviews in dentistry. *European Journal of Oral Sciences*, 111(2), 85–92. doi:10.1034/j.1600-0722.2003.00013.x.
- Kelly, K., Travers, A., Dorgan, M., Slater, L., & Rowe, B. (2001). Evaluating the quality of systematic reviews in the emergency medicine literature. *Annals of Emergency Medicine*, 38(5), 518–526. doi:10.1067/mem.2001.115881.
- Moher, D., Soeken, K., Sampson, M., Ben-Porat, L., & Berman, B. (2002). Assessing the quality of reports of systematic reviews in pediatric complementary and alternative medicine. *BMC Pediatrics*, 2, 3. doi:10.1186/1471-2431-2-3.
- Mokkink, L. B., Terwee, C. B., Stratford, P. W., Alonso, J., Patrick, D. L., Riphagen, I., et al. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Quality of Life Research*, 18(3), 313–333. doi:10.1007/s11136-009-9451-9.
- Golder, S., Loke, Y., & McIntosh, H. M. (2008). Poor reporting and inadequate searches were apparent in systematic reviews of adverse effects. *Journal of Clinical Epidemiology*, 61(5), 440–448. doi:10.1016/j.jclinepi.2007.06.005.
- Sampson, M., Barrowman, N. J., Moher, D., Klassen, T. P., Pham, B., Platt, R., et al. (2003). Should meta-analysts search

- Embase in addition to Medline? *Journal of Clinical Epidemiology*, 56(10), 943–955. doi:[10.1016/S0895-4356\(03\)00110-0](https://doi.org/10.1016/S0895-4356(03)00110-0).
19. Bellack, A. S., & Hersen, M. (1998). *Behavioral assessment: A practical handbook* (4th ed.). Boston: Allyn and Bacon.
  20. Knapp, S. D. (2000). *The contemporary thesaurus of search terms and synonyms: A guide for natural language computer searching* (2nd ed.). Phoenix, Ariz.: Oryx Press.
  21. Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385–401. doi:[10.1177/014662167700100306](https://doi.org/10.1177/014662167700100306).
  22. Pilkonis, P. A. (2006). Item identification and pooling. Inaugural PROMIS Conference; Gaithersburg, MD.
  23. Ward, D., Meadows, S. E., & Nashelsky, J. E. (2005). The role of expert searching in the family physicians' inquiries network (FPIN). *Journal of the Medical Library Association: JMLA*, 93(1), 88–96.