Review

Use of O*NET as a Job Exposure Matrix: A Literature Review

Manuel Cifuentes, MD, ScD,[1]* Jon Boyer, ScD,[2] David A. Lombardi, PhD,[3] and Laura Punnett, ScD[4]

Background *O*NET is a publicly available online database that describes occupational features across US job titles and that has been used to estimate workplace physical and psychosocial exposures and organizational characteristics. The aim of this review is to describe and evaluate the use of O*NET as a job exposure matrix.*
Methods *A review of the peer-reviewed published and gray literature was conducted. Twenty-eight studies were found that used O*NET to estimate work exposures related to health or safety outcomes. Each was systematically evaluated across eight main features.*
Results *Many health outcomes have been studied with O*NET estimates of job exposures. Some studies did not use conceptual definitions of exposure; few studies estimated convergent validity, most used predictive validity. Multilevel analysis was underutilized.*
Conclusion *O*NET is worthy of exploration by the occupational health community, although its scientific value is still undetermined. More studies could eventually provide evidence of convergent validity. O*NET has the potential to allow examination of occupational risks that might have otherwise been ignored due to missing data or resource constraints on field data collection of job exposure information.* Am. J. Ind. Med. 53:898–914, 2010. *© 2010 Wiley-Liss, Inc.*

KEY WORDS: O*NET; occupational information network; occupational exposure; job exposure matrix; ergonomics

INTRODUCTION

The collection of exposure data for specific chemical, physical, or biological agents in the work environment is often a time-consuming and resource-intensive process. The range of exposure assessment methods includes workers' self-report, visual observation, direct measurement, expert opinion, and imputation from job titles or from external sources. Each method has advantages and disadvantages with regard to precision, reliability, resource requirements, accessibility to the workforce and/or the workplace, potential to influence exposures of interest or to interfere with worker activities and productivity, and ability to represent working conditions during the time period etiologically relevant to the health effect of concern [Smith et al., 1991; Harris, 1993; Kilbom, 1994; van der

Additional Supporting Information may be found in the online version of this article.

[1]Center for Disability Research at Liberty Mutual Research Institute for Safety, Hopkinton, and Work Environment Department at University of Massachusetts Lowell, Lowell, Massachusetts
[2]Center for Physical Ergonomics at Liberty Mutual Research Institute for Safety, Hopkinton, and Work Environment Department at University of Massachusetts Lowell, Lowell, Massachusetts
[3]Center for Injury Epidemiology at Liberty Mutual Research Institute for Safety, Hopkinton, Massachusetts
[4]Work Environment Department at University of Massachusetts Lowell, Lowell, Massachusetts
Contract grant sponsor: National Institute for Occupational Safety and Health (NIOSH); Contract grant number: R01-OH07381.
*Correspondence to: Dr. Manuel Cifuentes, Research Scientist, Center for Disability Research, Liberty Mutual Research Institute for Safety, 71 Frankland Road, Hopkinton, MA 01748. E-mail: manuel.cifuentes@libertymutual.com

Accepted 2 March 2010
DOI 10.1002/ajim.20846. Published online 21 April 2010 in Wiley Online Library (wileyonlinelibrary.com)

Beek and Frings-Dresen, 1998; Burdorf and van der Beek, 1999; Checkoway et al., 2004; Punnett and Wegman, 2004].

Imputation of work exposure estimates from external data sources to study subjects via standard occupation codes in a specific study population is appealing because it is usually less expensive than direct measurement [Reed et al., 1989; Gomez et al., 1994; Muntaner et al., 1998; Dosemeci et al., 1999; Le Moual et al., 2000; Green et al., 2005] although the validity of such external data sources has been questioned [Kauppinen, 1996]. Imputation of work exposure is not frequently used due to lack of sources containing reliable and valid estimates regarding working conditions and exposures, concern for misclassification at the individual level, and some technical difficulties with the level of the analyses. Another important limitation in the use of job exposure matrices is the lack of information about within-job exposure variation. If the exposure of interest is not equally distributed within a job group, (e.g., varying by ethnic group or gender [Murray, 2003]) the utilization of a job exposure matrix may provide non-reliable results. Thus, despite the potential appeal of using a job exposure matrix to impute exposure to groups of workers, there are also important challenges. One is the accurate translation of workplace-specific job titles to standardized industry-wide job titles and then to the job coding system used in the database containing the imputed data. Another is in evaluating whether the information contained in the job exposure matrix properly matches the estimated exposure. In technical terms, the potential for misclassification of job titles and accuracy of exposure estimates are challenges to the reliability of job exposure matrices.

One past source of information on job requirements was the US Dictionary of Occupational Titles (DOT), which was designed primarily for vocational and rehabilitation counselors but used occasionally by researchers as a source of work exposure information [Felson et al., 1991]. The DOT was first published in 1938 and focused on blue collar industrial jobs. It was periodically updated until its fourth version, which contained descriptions of more than 12,000 jobs, making it cumbersome for most research applications [Hadden et al., 2004].

However, the shift towards a service and information-based economy decreased the relative representativeness of the DOT [Mariani, 1999]. The DOT was replaced in 1998 by the US Occupational Resource Network (O*NET) [Hadden et al., 2004] after a transition period that included the development of a standardized classification of occupations (Standard Occupational Classification (SOC)). O*NET is an online database that is being developed under the sponsorship of the US Department of Labor/Employment and Training Administration (http://www.onetcenter.org). Additional information related to the features of O*NET are provided in Supplementary Appendix 1. The history and development of O*NET, as well as its structure and organization, have been well described elsewhere [Mariani, 1999; Hadden et al., 2004; Crouter et al., 2006; Alterman et al., 2008], and a full description of the chronological changes in O*NET can be accessed in the ‘‘O*NET Database Update Summary’’ file at http://www.onetcenter.org/database.html.

As a publicly available source of workplace information, the O*NET database could provide estimates of average exposure for studies that have collected job titles for the study subjects but lack specific information about their working conditions. O*NET has been used for this purpose and this will likely continue [Zimmerman et al., 2004] because it is easily accessible, free of cost, and offers a wide variety of information about many job titles. A detailed review of studies that have used O*NET as a source of occupational exposure provides an opportunity to illustrate the challenges of using O*NET and suggest new research approaches to overcome some of those challenges. Thus, the main aims of this paper were to review the existing published and unpublished literature to describe and evaluate how O*NET has been used as a source of job exposure; to illustrate the difficulties that researchers have faced; and to suggest possible strategies for overcoming those challenges. Secondarily, we aimed to evaluate the different types of validation of O*NET data conducted to date.

METHODS

We conducted a systematic search of the peer-reviewed and gray literature to identify research projects that have used O*NET information to investigate health and occupational exposures. Published articles were sought through PubMed using ‘‘O*NET’’ and ‘‘Occupational Information Network’’ as search terms. Ten articles were retrieved [Zimmerman et al., 2004; Cifuentes et al., 2007, 2008a Grandey et al., 2007; Meyer et al., 2007; d’Errico et al., 2007; Alterman et al., 2008; Bell et al., 2008; Forstmeier and Maercker, 2008; Ward et al., 2008], from which reference lists were reviewed to identify other relevant studies [Glomb et al., 2004; Hadden et al., 2004; Shaw and Gupta, 2004; Liu et al., 2005]. Additional published [Zhang and Snizek, 2003; Pransky et al., 2005a,b; Crouter et al., 2006; Lis et al., 2007; Verma et al., 2007; Benjamin et al., 2008; Dierdorff and Ellington, 2008; Harrold et al., 2008]. Unpublished materials [Huzyak, 2008; Spreng, 2008; Boyer et al., in press; Gardner et al., submitted] were found through the O*NET webpage (http://www.onetcenter.org/dl_files/omb2008/AppendixE.pdf), web search engines (e.g., Google), and personal communication with researchers known to have conducted research using O*NET. Studies using O*NET data as a source of exposure, for a health-related purpose or for a health-related outcome were included in the current review. Those using exposure data classified by the Dictionary of Occupational Titles (DOT, the precursor to O*NET) were excluded. The search for O*NET related research using PubMed and web search engines was repeated over time until the study team was satisfied that all available relevant studies were included.

TABLE I. Studies Using O*NET as a Source of Occupational Exposure (Listed in Chronological Order of Publication)

References	Title	Sample features	Occupational coding process	Outcome	Exposure concept	O*NET variable operational definition	Analysis	Results	Type(s) of validation achieved	O*NET version
Zhang and Snizek [2003]	Occupation, job characteristics, and the use of alcohol and other drugs	Complex sampling design. 7,477 full time workers from the National Household Survey on Drug Abuse[a]	No mention to coding process	Current use of alcohol; current heavy use of alcohol; current use of drugs; prior year marijuana; prior year cocaine	Primary exposure. Workload; work independence; job variety; financial compensation; job security; working conditions; job autonomy; skills utilization	Single O*NET items except for the last two concepts, which were constructed as the mean of three and two items respectively. Scale 1–5	SUDAAN software for weighted multivariate logistic regression modeling. No mention to multilevel analysis	Steady employment and job security are the most important predictors of drug abuse. Higher job autonomy is associated with higher consumption of cocaine	Face Predictive	1998; first version
Glomb et al. [2004]	Emotional labor demands and compensating wage differential	Occupation level data from the Occupational Employment Statistics Survey for the outcome and the Current Population Survey for the control variables merged with O*NET and DOT for the exposure	Used the standard occupational classification system. It could merge 560 occupations. Average of job characteristics scores was computed if more than one O*NET job represented an occupation	Occupation level wages	Primary exposure. Jobs demands in general. Emotional labor; cognitive demands; physical demands	Through factorial analysis identified a factor containing 6 items with conceptual overlap with emotional labor demands. Then selected two very specific items and combined them with four other selected items and run a second confirmatory factor analysis. Specificity was prioritized	Job level linear regression weighted by the number of individuals in the Occupational Employment Statis tics Survey sample size	Higher emotional demands asso ciated with lower wages except in group with high cog nitive demands	Face (some items) factorial (some items) predictive	2001; ver sion 3.1
Hadden et al. [2004]	Descriptive dimensions of US occupations with data from the O*NET	900 occupation included in O*NET 4.0. Only job title level data	No coding needed	None/creation of scales	Not clear. Dimensions of US occupations (?)	Level and importance were rescaled between 0 and 1, summarized by multiplying and log transforming them. Three factor analyses using O*NET structure. The first analyzed sub-domains separately; in the second, factors from the first analyses were analyzed; in the third, all 227 variables were analyzed at once	Factorial analysis	Four scales explained 62.4% of variance. Substantive complexity (36.6%), people versus things (13.2%); physical demands (8.3%); and bureaucracy (4.3%)	Factorial	2002; ver sion 4
Shaw and Gupta [2004]	Job complexity, performance, and well-being: when does supplies-values fit matter?	Ad hoc convenience sample of 272 full and part time students employees	Minimal mention "Job titles reported by the participants were matched to. . .O*NET database"	Well-being (somatic complaints) moderated by performance	Primary exposure as effect modifier. Job complexity	Average of 31 complexity items used by Glomb in 2004	Multivariate polinomial regression. No multilevel	In low job performance direct association between job complexity and somatic complaints	Face predictive	??

Zimmerman et al. [2004]	Tinker, tailor, soldier, patient: work attributes and depression disparities among young adults	Complex sampling design. 3,753 men and 3,525 women from the National Longitudinal Survey of youths (NLSY)	Clearly explained. Two steps. First, NLSY census code of occupation cross walked to DOT and, second, DOT codes cross walked to O*NET	Center for Epidemiological Studies Depression (CES-D Scale)	Primary exposure. Attributes compatible with the concepts studied in the literature as linked to depression	Eight domains: recognition, opposition, security, machine's pace, sociability, moral, physically uncomfortable, wage premium. Required Cronbach's alpha >0.85	Negative binomial model followed by logistic regression. Neither multilevel nor weights mentioned	Many predictors were stratified by gender, race. Higher job status decreases risk of depression. In young men. High opposition score increases it. Physically uncomfortable or dangerous job increase risk for young women	Face Predictive	2001; ver sion 3.1
Liu et al. [2005]	The relation of job control with job strains: a comparison of multiple data sources	Collected in 1988 to be matched with the DOT. Convenience sample (it does not refer to random sampling) 232 employees of the University of South Florida representing 68 jobs	From DOT to O*NET by two judges obtaining agreement in 52 jobs. 18 jobs required discussion to be matched. 2 jobs could not be matched	Multiple outcomes. Frustration, anxiety, job satisfaction, turnover intention, total symptoms, doctor visits	Primary exposure. Job autonomy	Job autonomy in the domain work needs is compared with expert-rater and self-reported job autonomy ins it predictive validity of self reported "strains"	Pearson correlations and multiple linear regression analyses. No multilevel	More O*NET autonomy associated with less doctor visits and less absences. No with psychological strains	Face predictive	1998; first version
McElroy et al. [2005]	Breast cancer risk associated with electromagnetic field exposure from computer work ascertained from occupational history data	6,422 cases identified from statewide cancer registries; 7,673 controls from population lists. Matched with O*NET	No mention to coding process	Breast cancer	Primary exposure. Cumulative occupa tional exposure to electromagnetic field	Level of work with computers (from O*NET) multiplied by hours of exposure (from survey)	Not clear. No mention to multilevel	No association to small assoc in one state	Face predictive	?
Pransky et al. [2005a]	Early retirement due to occupational injury: who is at risk	Based on injuries reported to the New Hampshire Department of Labor. Additional information obtained by mail	No mention to coding process	Intent to retire early	Secondary exposure. Overall job physical demands. Job physi cal activity level	1 – 5 original item scale	Logistic regression. No mention to multilevel modeling	No association	Face predictive	1998; first version
Pransky et al. [2005b]	Outcomes in work-related injuries: a comparison of older and younger workers	Based on injuries reported to the New Hampshire Department of Labor. Additional information obtained by mail	No mention to coding process	Poor health outcomes after a work-related injury comparing younger with older workers	Secondary exposure. Overall job physical demands. Job physi cal activity level	1 – 5 original item scale	Multiple linear and logistic regression. No mention to multilevel modeling	No association	Face predictive	1998; first version

(*Continued*)

TABLE I. (*Continued*)

References	Title	Sample features	Occupational coding process	Outcome	Exposure concept	O*NET variable operational definition	Analysis	Results	Type(s) of validation achieved	O*NET version
Crouter et al. [2006]	The O*NET jobs classification system: a primer for family researchers	Family Life Project. ∼1,300 families	Do not explain the coding process and reports that 603 of 613 biological mothers could have their job coded in O*NET as 534 of 552 partners also	None/creation of scales	Occupational self-direction; physical hazards; physical activity; care work; automation/ repetition	Factorial analysis after selecting O*NET items	Factorial. Comparison of exposure means across maternal work shifts	Five factors: self-direction, hazardous condition, physical activity, care work, and automation/ repetition. Mother's shift was significantly related to hazardous condition, physical activity, and automation/ repetition	Face factorial discriminant	2004; ver sion 6.0
Benjamin et al. [2008]	Factors associated with retirement-related job lock in older workers with recent occupational history	Based on injuries reported to the New Hampshire Department of Labor. Additional information obtained by mail	No mention to job coding	Job lock	Secondary exposure. Overall job physical demands. Job physical activity level	1–5 original item scale	Logistic regression. No mention to multilevel modeling	No association	Predictive	1998; first version
Cifuentes et al. [2007]	Inter-method agreement between O*NET and survey measures of psy chosocial exposure among healthcare industry employees	Ad hoc. Health care workers	Barely mentioned. "each digit SOC code was matched to its respective occupation under the O*NET. . ."	Agreement between O*NET data and ad hoc survey data	Primary exposure. Psychosocial exposure: decision latitude, psychological demand, job strain, supervisor support, coworker support, rewards, effort reward ratio, emotional labor	Not described	Multilevel linear regression, two way mixed total agreement, and job level regression and correlation	Good level of agreement on healthcare specific jobs	Face convergent	2005; ver sion 9.0
d'Errico et al. [2007]	Hospital injury rates in relation to socioeconomic status and working conditions	Ad hoc. Health care worker individual level data was collapsed to job level data	Facility job titles were coded to Standard Occupational Classifi cation system by expert coders and cross walked to O*NET	Occupational injury rate by job title	Primary exposure. Psychosocial and physical exposure	Items averaged by scales (1–100)	Job level regression analysis	Exposure predicts injuries (decision latitude; supervisor support; force exertion; temperature extreme)	Face predictive	2002; ver sion 4

Grandey et al. [2007]	Verbal abuse from outsiders versus insiders: comparing frequency, impact on emotional exhaustion, and the role of emotional labor	Two samples, one from the National Survey of Workplace and Safety (complex survey design) and a convenience sample. Both are examined separately. Only the first one was merged with O*NET	Well explained. Two independent coders; meetings with third author; search for agreement; minimum kappa value required	Emotional exhaustion	Primary exposure. Emotional labor demands; emotional expectations	Interactions that require to deal with external customers; communicating with people outside the organization while representing the organization; performing for people or dealing directly with the public. Statistical reasons to build the emotional expectations scale	Weighted (to the US workforce) linear regression analyses with Taylor series linearization. No multilevel	Emotional labor predicts more abuse from outsiders and does not modify the association between verbal abuse and emotional exhaustion	Face factorial (for emotional expectation) predictive	2005; version 9
Lis et al. [2007]	Association between sitting and occupational LBP	Review	No mention to coding	Low back pain	Complement to primary exposure. O*NET used to estimate amount of sitting required by occupation when selected studies reported LBP by occupation	Unclear	Unclear	Unclear	Predictive	2003; version 5.1
Meyer et al. [2007]	Job control, substantive complexity, and risk for low birth weight and preterm delivery: an analysis from a state birth registry	Connecticut birth registry in the Connecticut Department of Public Health. Year 2000. Matched with O*NET	Well described. Coding performed by using algorithm developed by NIOSH and National Center for Health Statistics and then cross walked to O*NET	Low birth weight Preterm delivery	Primary exposure. Substantive complexity of work (as analogous to job control obtained from 1970 census), people versus things, and physical demands	Factorial analysis. Included 10 variables with highest loading in each factor. Substantive complexity of work (analogous to job control) and physical demands. Grouped in tertiles. Based on Hadden and Muntaner	Multiple logistic regression. No multilevel	O*NET physical demands assoc with LBW. A gradient between LBW and substantive complexity	Predictive	2004; version 6.0
Verma et al. [2007]	Occupational physical demands and same-level falls resulting in fracture in female workers: an analysis of workers' compensation claim	Ad hoc sample	Minimal mention: "Reported occupations were assigned O*NET codes. . . by a person blinded to injury outcomes"	Occupational related fracture	Primary exposure. Occupational physical demands	Levels of general physical activity and standing and sitting duration	Log-binomial regression. No multilevel	Only in over 50-year-old female workers, moderate work related physical activity decreases risk of fracture	Face predictive	2004; version 6

(*Continued*)

TABLE I. (*Continued*)

References	Title	Sample features	Occupational coding process	Outcome	Exposure concept	O*NET variable operational definition	Analysis	Results	Type(s) of validation achieved	O*NET version
Alterman et al. [2008]	Examining associations between job characteristics and health: Linking Data From the Occupational Information Network (O*NET) to two US National Health Surveys	Complex sampling design. National Health and Nutrition Examination Survey III and National Health Interview Survey from years 2001 to 2003	Minimal mention "The O*NET 98 was linked to NHANES III. . . and NHIS. . . by using occupational titles provided by NCHS."	Cardiovascular disease Health risk behaviors (current smoking, heavy drinking, over weight) and diseases (hypertension, angina, coronary heart disease, depressive ymptoms)	Primary exposure. Gaining knowledge and information processing; interpersonal relationship, assisting, guiding work of others; physical activities, repairing and maintaining equipment; hazardous work exposure; dealing with people, diversity of tasks; competitive work context, importance of being precise; psychosocial work environment; work ing with others; worker and management relations	Principal factor analysis within each O*NET sub-domain; mean of factor based scores for each individual; each O*NET factor was dichotomized in a binary variable at the median	Multiple logistic regressions. SUDAAN 8.0 for complex survey design of health surveys. No mention of multilevel analysis	Significant as expected. "Seven of the nine factors were significantly associated with health risk behaviors in both surveys, but only one of the nine factors (gaining knowledge and information processing) was significant for a health outcome (depression) in both surveys"	Factorial predictive	1998; first version
Bell et al. [2008]	Maternal work and birth outcome disparities	Complex sampling design. Combination of US Bureau of Labor Statistics' National Longitudinal Survey of Youths and O*NET 5.1	Well explained. Two steps. First, NLSY census code of occupation cross walked to DOT and, second, DOT codes cross walked to	Birth weight; preterm birth; fetal growth restriction	Primary exposure. Job demands, control, status, autonomy, and physical exertion	Factor analysis over selected items. Three indices obtained. Status and recognition, physical demands, and exposure to conflict	Multiple linear regression and logistic regression. Robust standards errors to account for mother/births clustering. No mention of job level clustering. No mention of complex sampling design	Significant as expected work related physical demands lower birth weight and higher risk of preterm; low job status and recognition risk for fetal growth restriction	Face factorial predictive	2003; ver sion 5.1
Cifuentes et al. [2005a]	Job strain predicts survey response in healthcare industry workers	Ad hoc. Health care workers	Fair description. Facility job titles were coded using SOC system and then double-checked and matched to O*NET	Survey response	Primary exposure. Job strain and effort reward imbalance	Karasek and Siegrist similes	Multilevel linear regression	Job strain predicts survey response	Face convergent predictive	2005; ver sion 9.0

Dierdorff and Ellington [2008]	It is the nature of the work: examining behavior based sources of work-family conflict across occupations	General Social Survey (GSS) from the National Opinion Research Center (University of Chicago) that uses full-probability sampling of households merged with O*NET	Well described. Sequential process. Exclusion of jobs with no O*NET match	Work-family conflict from GSS	Primary exposure. Behavior-based antecedents of work-family conflict	Confirmatory factor analysis after careful selection of items. Interdependence, responsibility for others, interpersonal conflict	Multilevel linear regression	Interdependence and responsibility for others predict work-family conflict	Face factorial predictive	?
Forstmeier and Maercker [2008]	Motivational reserve: lifetime motivational abilities contribute to cognitive and emotional health in old age	147 60- to 99-year-old healthy adults living in Zurich, Switzerland	Well described. Independent two coders; meetings with first author; inter-rater agreement measured	Current cognitive status, psychological wellbeing	Motivational and cognitive abilities	Motivational abilities: O.NET Item 4.A.2.b.6, and Item 4.A.1.b.3, Cognitive abilities Item 1.A.1.g.1, Item 1.A.1.b.3, Item 2.A.2.d, and Item 2.B.1.a	Hierarchical multiple regression analyses and multiple logistic regression. No multilevel	O*NET motivational abilities, not cognitive abilities, predicted the outcomes	Face convergent predictive	June 2007. version 12
Harrold et al. [2008]	Understanding the role of sex differences in work injuries: implications for primary care practice	Based on injuries reported to the New Hampshire Department of Labor. Additional information obtained by mail	No mention to coding	Return to work; injury related lost time; re-injury; current work status; future work concerns	Secondary exposure. Job physical activity level	1 – 5 original item scale	Linear and logistic regressions. No multilevel	Men have more physical demands. Physical demands are not associated with work injuries	Face predictive	?
Huzyak [2008]	Work and domestic violence: examining spillover among women	Complex sampling design. National Survey of Households and Families Wave I. Over-sampled for some groups	No mention to coding	Women domestic violence	Primary exposure. Occupation violence	Item: deal with the physical aggression of violent individuals	Logistic regression. No multilevel. No weights	No association between occupation violence and domestic violence	Face predictive	2006; version 11
Spreng [2008]	Occupation attributes relate to origin and extent of atrophy in frontotemporal lobar degeneration	Multicenter medical records review of 588 patients	Coding explained but no description of the process	Origin and extent of atrophy in frontotemporal lobe	Primary exposure. Occupation attributes	Factorial analysis captured 5 factors: verbal, physical, mechanical, mathematical, and visuospatial	Multivariate methods. Partial consideration of multilevel analysis	Attributes relates to origin and extent of atrophy	Factorial	2006; version 11.0
Ward et al. [2008]	Occupational physical activities and long term functional and radiographical outcomes in patients with ankilosing spondilyitis	397 patients participating in the prospective study of outcomes in AS	No mention to coding	Functional limitations (BATH AS functional index) and radiographic damage (BATH AS radiology index for the spine). As second measure, the disability index of the health assessment questionnaire modified for spondyl-arthropaties	Primary exposure. Occupational physical activities	13 physical activities that involve the axial skeleton and 13 work context descriptors. In quartiles	Multivariable ANOVA and ordinal logistic regression. No multilevel	More functional limitations in jobs requiring dynamic flexibility and more image damage in jobs requiring more dynamic flexibility, extent flexibility, and whole body vibration	Face predictive	2006; version 11

(*Continued*)

TABLE I. (*Continued*)

References	Title	Sample features	Occupational coding process	Outcome	Exposure concept	O*NET variable operational definition	Analysis	Results	Type(s) of validation achieved	O*NET version
Boyer et al. [in press]	Ergonomic and Socioeconomic Risk Factors for Hospital Workers' Compensation Injury Claims	Ad hoc. 1,483 healthcare employees	Explained by referring to previously published studies	All injuries, strains and sprains, and back body part claims	Primary exposure. Proxy ergonomic physical and psychosocial exposure variables	O*NET was part of a job exposure matrix for healthcare workers	Poisson regression models. Multilevel not needed because there was no clustering by occupations	Physical and organizational factors predict injury claim risk	Face Predictive	2005; version 9.0
Gardner et al. [submitted]	Reliability of job-title based physical work exposures for the upper extremity: comparison to self-reported and observed exposure estimates	Subsample of the predictors of carpal tunnel syndrome study	No mention to coding	Agreement between self-report, direct observation and O*NET	Primary exposure. Upper extremity physical work exposure	Specific items selected and averaged handling strength dexterity physical repetition	Two-way random effects model to estimate total agreement (intraclass correlation)	Good agreement	Face Convergent	2007; version 12

[a]The National Household Survey on Drug Abuse changed its name and it is currently known as the National Survey on Drug Use and Health.

Each study was reviewed to extract eight features regarding the use of O*NET as a source of job exposure data (Table I). ‘‘Sample features’’ refers to data sources used in addition to O*NET and specific features of each sample (e.g., the use of complex sampling design) that could influence the way the data were analyzed or validated. ‘‘Coding process’’ presents special situations that relate to the challenge of coding job titles in the O*NET coding system. ‘‘Outcome’’ refers to the dependent health variable under study. ‘‘Exposure concept’’ represents, where available, the idea or concept defining the specific independent variable(s). ‘‘O*NET variable operational definition’’ describes, in as much detail as possible, how O*NET items were used to construct the exposure measures. ‘‘Analysis’’ focuses on the statistical methods and any specialized software utilized. ‘‘Results’’ summarizes the main findings of the study in relation to the O*NET variables. The column ‘‘Types of validation’’ refers to whether the results provide evidence of face, factorial, convergent, and/or predictive validity. The last column, ‘‘O*NET version,’’ indicates which of the 13 possible O*NET versions was used and the year it was available to the public.

RESULTS

General Characteristics of Studies That Have Used O*NET

A total of 28 studies were identified that used any version of O*NET to estimate work exposures in relation to health or safety outcomes. The first study we found that explicitly used O*NET data was published in 2003 [Zhang and Snizek, 2003]. Five other studies were reported in 2004, four in 2005, only one in 2006, seven in 2007 and eight in 2009. Two studies were graduate theses, accessible only online [Huzyak, 2008; Spreng, 2008]. Two other manuscripts were under review for publication [Gardner et al., submitted]. One of which has now been accepted [Boyer et al., in press].

Different O*NET versions have been used by study investigators (Table I), which makes it difficult to compare results across studies directly. In general, the version used was the most recent O*NET update available, although there were exceptions. For example, one of the more recent papers used the first version of O*NET [Alterman et al., 2008]. In some cases, researchers did not identify which O*NET version they used [Shaw and Gupta, 2004; McElroy et al., 2005; Dierdorff and Ellington, 2008; Harrold et al., 2008].

A total of 15 studies used O*NET to impute job exposure information into already existing study databases. The databases that have been merged with O*NET can be seen in the ‘‘Sample Features’’ column of Table I. As an example, Liu used a sample that was originally matched with the DOT in 1988 but was later matched with O*NET in 2005 [Liu et al., 2005]. Eleven other studies collected data with the explicit purpose of merging them with O*NET or they added O*NET to more recently collected data. One study [Forstmeier and Maercker, 2008], for example, used a sample of elders from Switzerland, exemplifying the wide range of potential O*NET uses. All of the other studies used US population samples.

Fifteen studies reported associations between the O*NET occupational exposure and a specific health outcome. The health outcomes of interest covered a broad spectrum: cardiovascular diseases, health risk behaviors, obstetric outcomes, mental health, health-related work activities and wages, occupational injury, and others (Table I, fourth column).

Several studies utilized the O*NET data for health services research or similar purposes [Glomb et al., 2004; Pransky et al., 2005a; Benjamin et al., 2008; Cifuentes et al., 2008a]. For example, Cifuentes et al. [2008a]used O*NET to obtain job exposure for survey non-respondents. In some studies there was no defined outcome [Hadden et al., 2004; Crouter et al., 2006; Cifuentes et al., 2007; Gardner et al., submitted]; the objective of these studies was to evaluate agreement of the O*NET exposure indicators with other measures or to describe the construction of exposure scales.

Non-statistically significant associations were found in three studies in which O*NET was not the main source of information on exposure (physical demands) [Pransky et al., 2005a,b; Benjamin et al., 2008]. In two unpublished papers, the primary exposure variables from O*NET were not associated with the health outcomes. One was as a doctoral dissertation in which the likelihood of occupational violence was not related to domestic violence in women [Huzyak, 2008], and the other was a conference proceeding examining occupational exposure to electromagnetic fields and breast cancer [McElroy et al., 2005].

The Process of Coding Occupations

Occupational coding is the first step in using O*NET data. Researchers need to link the jobs held by study subjects with the portion of the O*NET database they are interested in, by assigning an O*NET code to each job title. Each O*NET job code has eight digits which can be obtained by typing a job title of interest (as a keyword) directly into the search window of the O*NET webpage (http://online.onetcenter.org/find/). The search engine will produce a list of occupations and associated O*NET codes that match the inputted job title. Each potential match has a relevance score, which is defined in the O*NET website (http://www.onetcenter.org/questions/20.html).

Another way to obtain O*NET job codes is to use any of the cross walk tools available at http://online.onetcenter.org/crosswalk/. These tools can be used to assign O*NET job codes if the researcher already has the job titles pre-coded in any of the SOC, DOT, Military Occupational Classification (MOC), Registered Apprenticeship Partners Information

Data System (RAPIDS), or Classification of Instructional Programs (CIP) systems. The National Cross Walk Service Center has made files available that contain several O*NET cross walks, which have been used in some studies [Zimmerman et al., 2004; d'Errico et al., 2007; Bell et al., 2008; Cifuentes et al., 2008a]. While these may be useful to match a large initial number of jobs, our experiences and those of others [Glomb et al., 2004; Liu et al., 2005; d'Errico et al., 2007] indicate that the process must be completed through complex and intensive manual case-by-case matching [Meyer et al., 2007].

For example, the healthcare sector includes many DOT coded job titles that can be matched with a single O*NET code (such as "registered nurse"). Depending on the study goals, it might be necessary to distinguish between nurses working with patients and those working in predominantly management or administrative tasks. In this case, information beyond the DOT code (such as alternative workforce administrative databases, phone calls, interviews with key industry informants, or web searches into workforce descriptions) is needed to identify the "best" O*NET code. Ultimately, some jobs cannot be matched to O*NET and would likely be excluded [Dierdorff and Ellington, 2008].

Some research projects have access only to original self-reported job titles (e.g., "part-time assistant"), which may not provide enough specific information for coding. (This is a common problem in occupational epidemiology, whether or not O*NET is utilized.) Additional information about the specific industry or even the department in which the worker performs the job could help narrow the search by job family or career cluster. Therefore, researchers planning to assign O*NET codes to their own database should be prepared for the possibility that additional job description information may be needed to minimize the possibility of coding errors that could result in exposure misclassification. On the other hand, studies with a relatively small number of well-defined and documented jobs will not present great difficulties.

Using a single job coder may seem resource-efficient but an additional qualified person can help to improve coding reliability and also to resolve uncertainties through consensus [Liu et al., 2005; Grandey et al., 2007; Forstmeier and Maercker, 2008]. It is realistic to assume that some misclassification of job titles may result in some exposure misclassification. The level of exposure misclassification will be conditional on the proximity of the mismatching job titles and might not necessarily be severe [Mannetje and Kromhout, 2003]. This is mainly due to the fact that jobs grouped under the same families tend to be much more similar than those from other job families. The reduction of misclassification is always an imperative and the proper utilization of additional information is usually a requirement to decrease it as much as possible. However, when the optimal O*NET match has to be selected between two similar O*NET job codes, the resulting exposure misclassification should not be extreme.

Computational Aspects

Several studies have used single O*NET items directly from the database and/or combined them into additive measures of exposure [McElroy et al., 2005; Pransky et al., 2005a,b; Verma et al., 2007; Benjamin et al., 2008; Harrold et al., 2008; Huzyak, 2008; Ward et al., 2008]. Although some of the original O*NET survey question contained five categorical response options (e.g., never, less than half of the time, about half of the time, more than half of the time, continuously or almost continuously), each answer category is linked to an ordinal numeric value and in the O*NET database those values are reported as arithmetic means for each job title.

In the process of creating O*NET-based variables, ratios, logs, means, medians, standardized scores (*Z* score), and combinations of them have been used. The most common transformation process used among the reviewed studies was to convert the O*NET original item scale into a percentage scale in which the minimum theoretical original value is set to zero and the maximum to 100. Transformation have limitations; for example, if ratios are used, the researcher should be aware that there is a distortion when comparing the values under and over the null ratio value (usually one), in which the values under the ratio have a smaller range than values over the ratio to express differences of magnitude; a ratio ten times higher would be 10 (a distance of nine with the null value ratio) and a ratio ten times lower will be just 0.1 (a distance of 0.9 with the null ratio). This problem could have been present in some studies that used ratios [d'Errico et al., 2007; Cifuentes et al., 2008a; Boyer et al., in press]. *Z* scores are comparable across different scales, but they require a symmetric like-normal original distribution to be plausible.

Distinguishing Conceptual and Operational Exposure Definitions

It is important to make a clear distinction between the conceptual definition of exposure that a researcher considers etiologically important and the operational definition that is formulated by selecting one or more O*NET items as the best available representation of the concept. Some researchers appeared to have operationalized exposure variables without first identifying the explicit concepts of theoretical interest. For example, some relied on factorial analyses to group, weight, and select O*NET items according to their correlations with each other, instead of their relevance to the outcome. Others constructed exposure scales within the framework provided by O*NET dimensions or sub-dimensions [Hadden et al., 2004; Meyer et al., 2007; Alterman et al., 2008] but without stating a clear a priori concept. Similarly, other authors were more concerned about describing certain O*NET job features than categorizing them into occupational exposures according to the traditional conceptual groups of physical, chemical, ergonomic, and psychosocial factors. Using the Hadden study as an example, the factorial analyses were performed using O*NET dimensions and sub-dimensions and the final factorial analysis was performed with all the existing variables at once. In this final analysis, data reduction was elected rather than conceptual grouping of occupational exposure [Hadden et al., 2004].

Other studies had well-defined conceptual definitions of occupational exposure that were used to search for specific O*NET variables. These researchers explored up to hundreds of O*NET items to find those that were most related to exposure concepts of a priori interest [Zhang and Snizek, 2003; Glomb et al., 2004; Shaw and Gupta, 2004; Zimmerman et al., 2004; Cifuentes et al., 2007, 2008a; Grandey et al., 2007; d'Errico et al., 2007; Bell et al., 2008; Ward et al., 2008; Boyer et al., 2009; Gardner et al., submitted]. In some cases, they evaluated the internal consistency of each scale with measures like the Cronbach alpha coefficient [Zimmerman et al., 2004; Cifuentes et al., 2007, 2008a; d'Errico et al., 2007] or factor analysis [Glomb et al., 2004]. The study conducted by Glomb et al. [2004] focused on emotional demands using a combined approach. Initially, the authors used factor analysis in the ''work activities'' O*NET dimension to reduce the number of items and to exclude those that overlapped with other competing concepts. Next, they identified specific and supplemental items in the ''context'' O*NET dimension and added them to the first step items to create a scale of emotional demands.

Among all studies that used O*NET, the most challenging operationalization of exposure concepts to measures occurred where O*NET values were compared to or combined with exposure estimates obtained from other sources. For example, when more specific and precise exposure estimates were needed, the general O*NET values were supplemented with alternative data sources [Boyer et al., in press]. Good examples of exposure supplementation among the reviewed studies were the measurement of highly specific upper extremity exposures [Gardner et al., submitted], trunk postures [Boyer et al., in press], or psychosocial models [Cifuentes et al., 2007; d'Errico et al., 2007; Forstmeier and Maercker, 2008; Boyer et al., in press]. At the same time, this process is challenging because matching or complementing two operational definitions for the same conceptual exposure definition requires both subjective assessment of comparability and quantitative alignment and calibration of the scores.

Many currently used O*NET measures may lack the specificity desired for occupational health research. For example, the Body Positions item ''time bending and twisting the body,'' does not specify the body segments involved or the levels of postural deviation. Similarly, the O*NET Generalized Work Activity item for ''handling and moving objects'' does not specify the weight thresholds, the postures, or the types of objects handled. In both the Body Positions and General Work Activities groups, the exposure dimensions of duration, repetition, and level are mixed together, hampering interpretation of estimates.

Other measurements that were of occupational health interest have been discontinued and are not scheduled to appear in future versions of O*NET. For example, the survey questions used to collect responses necessary to the original O*NET ''values'' scales were discontinued in 2008 [Rounds et al., 2008]. These items are needed to characterize some estimators of the Demand Control and the Effort Reward Imbalance models [Cifuentes et al., 2007, 2008a; d'Errico et al., 2007]. A temporary solution, until it is no longer an option, is to extract data from O*NET version 11, the last one that contained the discontinued items.

These limitations have led some researchers to explore alternative methods to improve convergent and predictive validity for physical ergonomic exposures used in musculoskeletal epidemiology analyses. At least one group of researchers has tried to overcome the limitations in O*NET data by developing a job exposure matrix that summarizes information from multiple sources [Boyer et al., in press]. In this study, when O*NET did not provide information that properly matched the concept, another source of information was used (i.e., survey or work sampling by expert observation). Conversely, when the researchers' own data lacked information about working conditions for certain jobs due to sampling limitations, they used O*NET variables to eliminate missing values. Combining O*NET with supplementary information from other data sources was successfully used in a study of the healthcare sector [Boyer et al., in press]. The utility of this method is mostly for specific exposures (e.g., body postures and manual handling forces with specific frequency and intensity levels) that were not defined in O*NET.

O*NET Exposure Validity

O*NET dates to 1998 and thus is likely to represent the exposure conditions that workers have experienced during the last decade. Exposures could vary due to external circumstances; for example, an economic crisis could change the answer provided by a construction worker on the question, How many hours do you work in a typical week on your current job? However, in relative terms most jobs probably stay within the same relative range of exposure compared to other, qualitatively different jobs [Mannetje and Kromhout, 2003]. Even with the reduction in the number of job titles from previous to the current O*NET version (from more than 1,000 to 809), for most applications job titles can be relatively accurately recoded to one of the job titles in current use. Thus misclassification seems unlikely to present a serious limitation, as long as the jobs under study represent a broad range of working conditions. Further, the goal of O*NET is to be replenished with new data every few years (http://www.onetcenter.org/dataPublication.html). Therefore, it seems reasonable to assume that O*NET will remain up-to-date for the foreseeable future, barring major changes in resources allocated to this effort.

Four different types of validation were identified in the reviewed studies: face, factorial, convergent, and predictive validity. In terms of face validity, most studies found proper items for their conceptual definition of exposure. In some studies there was no a priori conceptual definition of the specific exposure of interest (see Distinguishing Conceptual and Operational Exposure Definitions Section) or face validation had already been performed by another investigator [Hadden et al., 2004; Meyer et al., 2007; Alterman et al., 2008]. As an example, Meyer et al. [2007] supported his study with the construct validation by Muntaner et al. [1993] and followed Hadden et al. [2004] procedures to obtain the exposure. For all of the others, the procedure was simply to match the O*NET item descriptions with an explicit conceptual exposure definition. Important benefits of using face validity are the economy in resources and the utilization, at least implicitly, of a priori exposure concepts, improving the communicability and understandability of the exposure variable(s).

Factorial validity was tested by studies that used a data-based approach to build scales. Most of them were inspired by Cain and Treiman [1981] and later Muntaner et al. [1993], who used factorial analysis of data in the Dictionary of Occupational Titles to group variables within the limits of O*NET's main dimensions. Sometimes this approach was not based on a specific theory or model but on the design of an internally consistent scale named after the items it contained. For example, the utilization of all O*NET variables to perform a factorial analysis seemed to follow that logic [Hadden et al., 2004]. In other cases factorial or other data reduction procedures were used after screening and selecting items with proper face validity [Glomb et al., 2004; Crouter et al., 2006; Cifuentes et al., 2007, 2008a; d'Errico et al., 2007; Bell et al., 2008; Dierdorff and Ellington, 2008; Boyer et al., 2009].

In general, researchers should be cautious when using factor analysis with exposure data. It has been demonstrated that this method (and other shared-variance techniques) may not always be the most appropriate primary technique for building measurement scales and can undermine scale validity when not built on solid a priori models [Delis et al., 2003]. In particular, the results will be highly dependent on which jobs are included in the study and how many subjects are in each job, so the results may have no generalizability from one setting to another. In addition, exposures that coincide because of work organization features do not necessarily have the same etiological relevance to a specific health outcome.

Three studies measured the convergent validity of O*NET based exposure variables [Cifuentes et al., 2007; Forstmeier and Maercker, 2008; Gardner et al., submitted]. Each of the three had an a priori conceptual model of exposures and used O*NET to obtain another measure in line with that theoretical model. The initial process of variable creation went through a face validation stage and eventually moved to the use of statistical tools to improve internal consistency. The convergent validity then was measured using correlation or level of agreement estimates with similar measures of the same theoretical model or conceptual definition. In the last step of this process, a conclusion can be reached about how far or close an O*NET variable is from those other measures of the corresponding exposure.

Finally, most studies in our review estimated the predictive validity of the O*NET scales. Few of these studies had other sources of job exposure measurement to allow parallel analyses and comparisons with O*NET predictive validity [Shaw and Gupta, 2004; Liu et al., 2005]. As noted previously, all but two studies (both unpublished) found associations between the O*NET variables and the health outcome(s) of concern when the O*NET variable was the most important exposure under study. There could be other unpublished negative studies, due to the inherent bias in research toward publishing ''significant'' associations.

Analysis of O*NET Data With Individual Outcomes

O*NET entries are identified by job title, so every exposure variable created from those items refers to a job or group of jobs. When these job-level exposure data are combined with worker-level outcome data, then a multilevel problem may arise because the data are clustered. The health outcomes occur at the individual level, whereas the exposures result from work requirements that are experienced by a group of workers. The practical consequence is that normal regression methods produce underestimates of the correct standard errors, increasing the likelihood of making type I errors (finding significant associations when there are none). Some of the studies reviewed here could have been affected by this situation [Shaw and Gupta, 2004; Liu et al., 2005; Meyer et al., 2007; Verma et al., 2007; Ward et al., 2008]; however, other studies [Glomb et al., 2004; d'Errico et al., 2007] aggregated the individual level outcome data to the job level, making common (''flat'') regression analyses appropriate because the predictor and the outcome were at the job level. However, this choice leads to an increase in the likelihood of type II errors, because the number of records reflects the number of jobs rather than individual persons. While analysis at the job level may sometimes be necessary, when possible the better approach would be to use a multilevel design or an analysis with robust variance methods. Both will provide the proper estimates for the standard errors; the former will additionally include information about the variability of the outcome between and within job titles.

Another set of analytical challenges arises when O*NET data are merged with survey data collected using a complex sampling design. This design gathers a sample by selecting people from strata within clusters and every person has an unequal probability of being selected. Due to the complex sampling design, common regression analysis, which assumes a simple random sample, might have the dual problem of missing the central point estimate and underestimating the standard error of the coefficients. The effect of the complex sampling design can be solved by using the proper analytical tools available in most important statistical software packages (STATA, SAS, SUDAAN, SPSS) [Siller and Tompkins, 2006]. A very useful basic resource for STATA is the guide available at the Stanford University Social Science Data and Software website (http://stanford.edu/group/ssds/cgi-bin/drupal/files/Guides/software_docs_stata_complexsvy.pdf). For SAS, support information is highly detailed (available at http://support.sas.com/rnd/app/da/new/dasurvey.html). SPSS has created the additional routine PWAS Complex Sample. SUDAAN is software designed to analyze complex survey data and its information is available at http://www.rti.org/SUDAAN. The analysis of such complex data requires specialized training, regardless of which of these packages is employed.

A more complex problem arises when the regression analysis has to solve simultaneously the design effect and the multilevel effect. This area of analysis is still under development and there is no widely accepted solution to date. Studies that faced this problem using O*NET data either did not attempt to address it [Zimmerman et al., 2004; Bell et al., 2008; Huzyak, 2008] or arrived at a partial solution by solving the design effect while ignoring the multilevel issue [Zhang and Snizek, 2003; Grandey et al., 2007; Alterman et al., 2008]. Other O*NET related [Li et al., 2006] and non-related [Cifuentes et al., 2008b] studies have tried to address both issues simultaneously using the GLIMMIX procedure in SAS with the empirical option, weights, and random effects. A comparison of alternative solutions can be found elsewhere [Zhang et al., 2000]. Currently SUDAAN announces on its webpage (accessed on October 6, 2009) that it is able to solve both issues simultaneously. We have not found any publication that could be used to exemplify this assertion.

Finally, although not unique to O*NET, statistical problems might arise when multiple hypotheses are tested in the same study (e.g., testing the association of multiple exposures with one health outcome). According to our review, none of the studies even discussed the issue of multiple comparisons, although some did examine many exposure variables [Zhang and Snizek, 2003; Zimmerman et al., 2004; Cifuentes et al., 2007].

Availability of Detailed Methods and Scale Sharing

Some of the papers provided clear and detailed explanations that could allow readers to reproduce the generation of O*NET values of exposure using exactly the same scales [Glomb et al., 2004; Forstmeier and Maercker, 2008]. Most studies did not provide all of the information for an accurate reproduction of the O*NET exposure scales. In light of database evolution with loss of information (see above), the preservation and distribution of extant datasets is important. Some researchers are willing to make available to the research community an updated database with all of the O*NET scales they have created and used [Cifuentes et al., 2007, 2008a; d'Errico et al., 2007; Boyer et al., 2009], plus pertinent documentation. Fluid communication among researchers using O*NET may result in improved approaches to meet the challenges of validity and analysis.

DISCUSSION

This summary of the O*NET-related literature showed that a wide range of health outcomes has been studied in relation to occupational features and exposures extracted directly or produced from the O*NET database. Five major issues were presented: (1) coding of job titles; (2) challenges in conceptualizing and operationalizing O*NET data; (3) validation of O*NET exposures; (4) analysis of O*NET data; and (5) sharing developed O*NET scales among investigators for resource efficiency and to permit evaluation of reproducibility.

In most of the published papers reviewed, when O*NET data were the primary source of exposure they were generally associated with the health endpoint under study. In other words, despite the apparent weaknesses of O*NET as a source of occupational exposure, the available data have been shown preliminarily to have predictive value for a range of outcomes, from substance abuse to low birth weight to workplace injury. However, more replication of such findings would greatly enhance the strength of this conclusion.

There are two approaches to the use of O*NET measures for health research: (1) to obtain a classification of work-related exposures and (2) to obtain descriptive job characteristics. This is not just a semantic distinction. The job requirement information available in O*NET seems to be readily classified according to the accepted constructs of biological, physical, chemical, and psychosocial hazards that are commonly used in occupational health. Although interesting, using O*NET to describe job features not linked to exposure models (as in the study of Hadden, in which scales were obtained through a computational procedure that did not require any conceptual definition of occupational exposure [Hadden et al., 2004]) would face difficulty with translation into the occupational safety and health field. However, this can be an important area of research for social disciplines interested in the description and impact of job features.

The strategy of using O*NET scales developed in previous studies should proceed with caution. It may be confusing to adopt concepts that are not operationalized in a way that serves the researcher's goals and each conceptual exposure can potentially have more than one O*NET construct. We found at least three different operationalizations of the conceptual exposure to general physical demands in four studies [Glomb et al., 2004; Pransky et al., 2005a,b; Verma et al., 2007]. Psychosocial models can also be formulated in different ways. Further, scales that were created through factorial analysis could change their formulation (e.g., to have different load factors) if the occupations under study were different than the occupations in the study where the scale was first developed. As discussed above, although factorial analysis could be useful for increasing the specificity of the exposure estimates, it is not advisable to rely on it as the only tool for generating predictor variables.

Although O*NET appears promising to date, it would be premature to draw conclusions about the intrinsic scientific value of the occupational exposure information contained in its database. There is a clear need for additional research on convergent validity and for identification of exposure formulations that provide the most consistent predictive validity. For example, what is the difference between the different operationalizations of physical demands [Glomb et al., 2004; Pransky et al., 2005a,b; Verma et al., 2007]. Is there a better way to compute an indicator of job strain or effort reward imbalance? [Cifuentes et al., 2007, 2008a; d'Errico et al., 2007; Boyer et al., 2009]. Convergent validity is one good criterion and predictive validity could offer a complementary view of what the best exposure measures are. The use of predictive validity could become an important validation method but only if findings are replicated across diverse conditions and populations.

The mere presence of a significant association between a variable based on O*NET and a health outcome does not assure the scientific value of such exposure measurement. Without demonstration of convergent validation there could be other reasons for some of the relatively weak associations observed in the studies collected here. Still, in light of the updating of an important portion of the O*NET database every year; it is possible that any demonstrated convergent validation might become outdated. Eventually, the best strategy for maintaining convergent validity across changing O*NET database versions would be to keep performing validation studies of relevant job titles. O*NET data, like evolving working conditions, are not a static target.

Finally, we emphasize the importance of utilizing multilevel analysis [Greenland, 2001]. Traditional flat regression and similar non-multilevel analyses should ideally only be used when there is no clustering of the exposure values by jobs or other structural characteristics. Even very consistent and convincing findings could be misleading where clustering exists and the multilevel structure is ignored. Unfortunately, the solution of examining associations at the job level leads to a decrease in statistical power. In addition, there is no clear solution to the analysis of O*NET data merged with surveys collected using complex sampling design, despite the popularity of this cost-effective approach to examining the impact of work exposures on population health.

CONCLUSION

O*NET is a database worthy of exploration by the occupational health community. It contains information that could be used as indicators of occupational exposures when other measures are not available. Except perhaps for analyses of some psychosocial and physical exposures mainly in the healthcare sector, the scientific validity of O*NET as a job exposure matrix in occupational health research efforts remains unclear. Coding job titles properly, keeping clear conceptual frames with common language, properly testing and improving the validity of the indicators (specifically convergent validity), performing appropriate statistical analyses, and sharing measures of exposure could eventually result in a better understanding of when and how to rely upon these data. Concerted efforts in these areas could yield databases that allow examination of occupational risks in studies that might otherwise have overlooked these due to resource constraints on field exposure assessment. Although the original intention of the O*NET developers was not that it be used for job exposure characterization, the research summarized here suggests that it is reasonable to do so.

ACKNOWLEDGMENTS

We thank Dr. Nils Fallentin, Dr. Leslie MacDonald, Dr. Glenn Pransky, and Dr. SangWoo Tak for their reviews of and comments on earlier versions of the manuscript. Study partly funded by National Institute for Occupational Safety and Health (NIOSH), Contract grant number: R01-OH07381. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of NIOSH.

REFERENCES

Alterman T, Grosch J, Chen X, Chrislip D, Petersen M, Krieg E, Jr., Chung H, Muntaner C. 2008. Examining associations between job characteristics and health: Linking data from the occupational information network (O*NET) to two US National Health Surveys. J Occup Environ Med 50:1401–1413.

Bell JF, Zimmerman FJ, Diehr PK. 2008. Maternal work and birth outcome disparities. Matern Child Health J 12:415–426.

Benjamin KL, Pransky G, Savageau JA. 2008. Factors associated with retirement-related job lock in older workers with recent occupational injury. Disabil Rehabil 30:1976–1983.

Boyer J, Galizzi M, Cifuentes M, d'Errico A, Gore R, Punnett L, Slatin C. 2009. Ergonomic and socioeconomic risk factors for hospital workers' compensation injury claims. Am J Ind Med 52:551–562.

Burdorf A, van der Beek A. 1999. Exposure assessment strategies for work-related risk factors for musculoskeletal disorders. Scand J Work Environ Health 25(Suppl. 4):25–30.

Cain PS, Treiman DJ. 1981. The dictionary of occupational titles as a source of occupational data. Am Sociol Rev 46:253–278.

Checkoway H, Pearce N, Kriebel D. 2004. Research methods in occupational epidemiology. 2nd edition. New York, New York: Oxford University Press.

Cifuentes M, Boyer J, Gore R, d'Errico A, Tessler J, Scollin P, Lerner D, Kriebel D, Punnett L, Slatin C. 2007. Inter-method agreement between O*NET and survey measures of psychosocial exposure among health-care industry employees. Am J Ind Med 50:545–553.

Cifuentes M, Boyer J, Gore R, d'Errico A, Scollin P, Tessler J, Lerner D, Kriebel D, Punnett L, Slatin C. 2008a. Job strain predicts survey response in healthcare industry workers. Am J Ind Med 51:281–289.

Cifuentes M, Sembajwe G, Tak S, Gore R, Kriebel D, Punnett L. 2008b. The association of major depressive episodes with income inequality and the human development index. Soc Sci Med 67:529–539.

Crouter AC, Lanza ST, Piretti A, Goodman WB, Neebee E. 2006. The O*Net jobs classification system: A primer for family researchers. Fam Relat 55:461–472.

Delis DC, Jacobson M, Bondi MW, Hamilton JM, Salmon DP. 2003. The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment. J Int Neuropsychol Soc 9:936–946.

d'Errico A, Punnett L, Cifuentes M, Boyer J, Tessler J, Gore R, Scollin P, Slatin C. 2007. Hospital injury rates in relation to socioeconomic status and working conditions. Occup Environ Med 64:325–333.

Dierdorff EC, Ellington JK. 2008. It's the nature of the work: Examining behavior-based sources of work-family conflict across occupations. J Appl Psychol 93:883–892.

Dosemeci M, Cocco P, Chow WH. 1999. Gender differences in risk of renal cell carcinoma and occupational exposures to chlorinated aliphatic hydrocarbons. Am J Ind Med 36:54–59.

Felson DT, Hannan MT, Naimark A, Berkeley J, Gordon G, Wilson PW, Anderson J. 1991. Occupational physical demands, knee bending, and knee osteoarthritis: Results from the Framingham Study. J Rheumatol 18:1587–1592.

Forstmeier S, Maercker A. 2008. Motivational reserve: Lifetime motivational abilities contribute to cognitive and emotional health in old age. Psychol Aging 23:886–899.

Gardner B, Lombardi D, Dale A, Franzblau A, Evanoff B. Reliability of job-title based physical work exposure for the upper extremityy: Comparison to self-reported and observed exposure estimates. Occup Environ Med (in press).

Glomb TM, Rotundo M, Kammeyer-Mueller JD. 2004. Emotional labor demands and compensating wage differentials. J Appl Psychol 89:700–714.

Gomez MR, Cocco P, Dosemeci M, Stewart PA. 1994. Occupational exposure to chlorinated aliphatic hydrocarbons: Job exposure matrix. Am J Ind Med 26:171–183.

Grandey AA, Kern JH, Frone MR. 2007. Verbal abuse from outsiders versus insiders: Comparing frequency, impact on emotional exhaustion, and the role of emotional labor. J Occup Health Psychol 12:63–79.

Green RS, Gold EB, Samuels SJ, Dosemeci M. 2005. The relation of occupational organic solvent exposure to symptom reporting in a sample of white and Chinese midlife women. J Occup Environ Med 47:410–423.

Greenland S. 2001. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. Int J Epidemiol 30: 1343–1350.

Hadden WC, Kravets N, Muntaner C. 2004. Descriptive dimensions of US occupations with data from O*NET. Soc Sci Res 33:64–78.

Harris RL. 1993. Guideline for collection of industrial hygiene exposure assessment data for epidemiologic use: Chemical manufacturers association. p 1–35.

Harrold LR, Savageau JA, Pransky G, Benjamin K. 2008. Understanding the role of sex differences in work injuries: Implications for primary care practice. Disabil Rehabil 30:36–43.

Huzyak TM. 2008. Work and domestic violence: Examining spillover among women. Available online at: IR http://etd.ohiolink.edu/send-pdf.cgi/Huzyak%20Teresa%20M.pdf?acc_num=akron1217266823. Accessed April 2, 2010. Akron, Ohio: University of Akron.

Kauppinen T. 1996. Exposure assessment—A challenge for occupational epidemiology. Scand J Work Environ Health 22:401–403.

Kilbom A. 1994. Assessment of physical exposure in relation to work-related musculoskeletal disorders—What information can be obtained from systematic observations? Scand J Work Environ Health 20 (Spec No):30–45.

Le Moual N, Bakke P, Orlowski E, Heederik D, Kromhout H, Kennedy SM, Rijcken B, Kauffmann F. 2000. Performance of population specific job exposure matrices (JEMs): European collaborative analyses on occupational risk factors for chronic obstructive pulmonary disease with job exposure matrices (ECOJEM). Occup Environ Med 57:126–132.

Li J, Alterman T, Deddens JA. 2006. Analysis of large hierarchical data with multilevel modeling using PROC GLIMMIX. SUGI 31. San Francisco, California: SAS Users Group International. p 151.

Lis AM, Black KM, Korn H, Nordin M. 2007. Association between sitting and occupational LBP. Eur Spine J 16:283–298.

Liu C, Spector PE, Jex SM. 2005. The relation of job control with job strains: A comparison of multiple data sources. J Occup Organ Psychol 78:325–336.

Mannetje A, Kromhout H. 2003. The use of occupation and industry classifications in general population studies. Int J Epidemiol 32:419–428.

Mariani M. 1999. Replace with a database: O*NET replaces the dictionary of occupational titles. Occup Outlook Q 43:3.

McElroy J, Newcomb P, Trentham-Dietz A, Titus-Ernstoff L, Hampton JM, Egan K. 2005. Breast cancer risk associated with electromagnetic field exposure from computer work ascertained from occupational history data. In: ISSE editor. The Seventeenth Conference of the International Society for Environmental Epidemiology Johannesburg: Epidemiology. p S99–S100.

Meyer JD, Warren N, Reisine S. 2007. Job control, substantive complexity, and risk for low birth weight and preterm delivery: An analysis from a state birth registry. Am J Ind Med 50:664–675.

Muntaner C, Eaton WW, Garrison R. 1993. Dimensions of the psychosocial work environment in a sample of the U.S. metropolitan population. Work & Stress 7:351–363.

Muntaner C, Nieto FJ, Cooper L, Meyer J, Szklo M, Tyroler HA. 1998. Work organization and atherosclerosis: Findings from the ARIC study. Atherosclerosis Risk in Communities. Am J Prev Med 14:9–18.

Murray LR. 2003. Sick and tired of being sick and tired: Scientific evidence, methods, and research implications for racial and ethnic disparities in occupational health. Am J Public Health 93:221–226.

Pransky GS, Benjamin KL, Savageau JA. 2005a. Early retirement due to occupational injury: Who is at risk? Am J Ind Med 47:285–295.

Pransky GS, Benjamin KL, Savageau JA, Currivan D, Fletcher K. 2005b. Outcomes in work-related injuries: A comparison of older and younger workers. Am J Ind Med 47:104–112.

Punnett L, Wegman DH. 2004. Work-related musculoskeletal disorders: The epidemiologic evidence and the debate. J Electromyogr Kinesiol 14:13–23.

Reed DM, LaCroix AZ, Karasek RA, Miller D, MacLean CA. 1989. Occupational strain and the incidence of coronary heart disease. Am J Epidemiol 129:495–502.

Rounds J, Armstrong PI, Liao H-Y, Rivkin D, Lewis P. 2008. Second Generation Occupational Value Profiles for the O*NET System: Summary Washington, DC: The National Center for O*NET Development and U.S. Department of Labor, Employment and Training Administration. Office of Workforce Investment. Division of Workforce System Support. p 5.

Shaw JD, Gupta N. 2004. Job complexity, performance, and well-being: When does supplies-values fit matter? Pers Psychol 57:847–879.

Siller AB, Tompkins L. 2006. The big four: Analyzing complex sample survey data. Using SAS®, SPSS®, STATA®, and SUDAAN®. SUGI31. San Francisco, California: SAS Users Group International.

Smith TJ, Hammond K, Hallock M, Woskie SR. 1991. Exposure assessment for epidemiology: Characteristics of exposure. Appl Occup Environ Hyg 6:441–447.

Spreng RN. 2008. Occupation attributes relate to origin and extent of atrophy in frontotemporal lobar degeneration. Toronto, Ontario, Canada: University of Toronto.

van der Beek AJ, Frings-Dresen MH. 1998. Assessment of mechanical exposure in ergonomic epidemiology. Occup Environ Med 55:291–299.

Verma SK, Sorock GS, Pransky GS, Courtney TK, Smith GS. 2007. Occupational physical demands and same-level falls resulting in fracture in female workers: An analysis of workers' compensation claims. Inj Prev 13:32–36.

Ward MM, Reveille JD, Learch TJ, Davis JC, Jr., Weisman MH. 2008. Occupational physical activities and long-term functional and radiographic outcomes in patients with ankylosing spondylitis. Arthritis Rheum 59:822–832.

Zhang Z, Snizek WE. 2003. Occupation, job characteristics, and the use of alcohol and other drugs. Soc Behav Personal Int J 31:395.

Zhang F, Salvucci S, Cohen M. 2000. Multilevel Linear Regression Analysis of Complex Survey Data. Proceedings of the Section on Survey Research Methods at the American Statistical Association Conference Alexandria, VA. p 197–202.

Zimmerman FJ, Christakis DA, Vander Stoep A. 2004. Tinker, tailor, soldier, patient: Work attributes and depression disparities among young adults. Soc Sci Med 58:1889–1901.