

Accepted manuscript version (after peer review) of the following article:

Sabariego, C., Al-Kudwah, H.M., & Cieza, A. (2015). Content Comparison of Patient Reported Outcome Instruments used to Measure Burnout. *European Journal of Psychological Assessment*, 31, 2, 91–99. <https://doi.org/10.1027/1015-5759/a000210>

© 2015 Hogrefe Publishing

This version of the article may not completely replicate the final version published in the journal. It is not the version of record

The accepted manuscript is subject to the Creative Commons licence CC-BY-NC.

Title

Content Comparison of Patient Reported Outcome Instruments used to Measure Burnout

Authors: Carla Sabariego¹, Hasnaa Murid Al-Kudwah¹, Alarcos Cieza^{1,2,3}

¹ Chair for Public Health and Health Services Research
Unit for Biopsychosocial Health
Department of Medical Informatics, Biometry and Epidemiology – IBE
Ludwig Maximilians University, Munich

² Faculty of Health and Social Sciences
University of Southampton,
Southampton, UK

³ Swiss Paraplegic Research
Nottwil, Switzerland

Corresponding author:

Dr. Carla Sabariego
Chair for Public Health and Health Services Research
Department of Medical Informatics, Biometry and Epidemiology – IBE
Ludwig-Maximilians-University
Marchioninstr. 17
81377 Munich, Germany
Phone: +49 89 2180 78224
Fax: +49 89 2180 78278
e-Mail: Carla.Sabariego@med.lmu.de

Summary

Objective: To compare the content of patient-reported outcome (PRO) burnout measures and to examine the degree of overlap between them and psychosocial difficulties experienced by persons with burnout.

Methods: The most frequent used PRO and qualitative studies were identified in two systematic literature reviews. Psychosocial difficulties identified in both sources were standardised and a qualitative content comparison was performed.

Results: Seven PROs and seven qualitative studies were selected. Energy and drive, emotional functions and work were key themes of both sources. Disparities were observed for problems in areas such as sleep, attention or family relationships, which were reported in several qualitative studies, but are seldom addressed in PROs.

Conclusion: Several areas important to persons with burnout, such as family relationships, are seldom measured by currently-used PROs. From a biopsychosocial perspective, these instruments cannot therefore be considered comprehensive enough to capture the whole experience of burnout and should be improved.#

Keywords: professional burnout; patient outcome assessment; psychometrics; qualitative research; International Classification of Functioning, Disability and Health (ICF)

Introduction

Burnout syndrome is increasingly gaining importance. This is because it dramatically affects the work force, reducing not only the efficiency and productivity of those affected, but also threatening co-workers through disruptive behaviours or increased workload and, finally, destabilizing the workplace itself (Schaufeli, Leiter, & Maslach, 2009). In high income societies, it is generally recognized that burnout causes substantial costs; it has been associated with absenteeism and high turnover rates (McCray, Cronholm, Bogner, Gallo, & Neill, 2008; Parker & Kulik, 1995).

Burnout is highly prevalent, although prevalence rates differ significantly among professions. A systematic review including 19 studies among interns, mainly conducted in the USA, reports prevalence rates ranging from 17.6% to 82% (Prins et al., 2007). Another review indicates that palliative-care physicians have lower burnout levels than other specializations, such as oncology, where prevalence rates range from 8% to 51% (Pereira, Fonseca, & Carvalho, 2011; Trufelli et al., 2008). The following prevalence rates were reported in the Netherlands: 11% among occupational physicians, 9% among psychiatrists, 9% among teachers, 8% among general practitioners, 7% among social workers, 1% among police officers, 2% among hospice workers and 3% among correctional officers (Schabracq, Winnubst, & Cooper, 2002).

Despite its popularity and burden, a universally acknowledged definition of burnout is still lacking (Kaschka, Korczak, & Broich, 2011). Even though there is no comprehensive definition, the one proposed by Maslach is the most widely used and defines burnout as a multidimensional syndrome comprising three dimensions: emotional exhaustion, depersonalization and reduced personal accomplishment (Maslach, Schaufeli, & Leiter, 2001). Exhaustion refers to "*feelings of being overextended and depleted of one's emotional and physical resources*", depersonalization to "*a negative, callous, or excessively detached response to various aspects of the job*" and reduced personal accomplishment to "*feelings of incompetence and a lack of achievement and productivity at work*" (Maslach et al., 2001).#

The popularity of burnout is reflected by an increasing number of publications. A systematic review of articles published between 2004 and 2009 identified 852 publications, the majority addressing medical topics with a focus on treatment of burnout (over 96%) (Kaschka et al., 2011). Due to the lack of consensus on the conceptualization of burnout, the open question we address in the present work is how burnout was measured in these studies. Several psychometrically sound instruments have been developed so far, such as the Maslach Burnout Inventory (MBI) (Maslach, Jackson, & Leiter, 1996), the Shirom Melamed Burnout Questionnaire (SMBQ) (Lundgren-Nilsson et al., 2012), and the Oldenburg Burnout Inventory (OBI) (Demerouti et al., 2003). As in many other fields, researchers and health professionals may get overwhelmed when deciding which outcome instrument among the many ones available to use.

The present investigation will provide the information needed to consciously select instruments to measure burnout. Our main objective was to perform a content comparison of the currently most relevant burnout patient-reported outcome (PRO) measures. Additionally, we examined the degree of overlap among their contents and symptoms and psychosocial difficulties reported by persons with burnout, using the “International Classification of Functioning, Disability and Health” (ICF) (WHO, 2001) as a reference framework. In the ICF, functioning encompasses the components “body functions and structures” and “activities and participation” and is determined by interactions between health conditions and contextual factors (environmental and personal factors). Using the ICF, we defined psychosocial difficulties (PSD) as impairments of body functions, activity limitations and participation restrictions in such domains as work, family life and leisure activities experienced by individuals with burnout in interaction with environmental and personal factors, such as treatment received, work environment, attitudes of others or personal beliefs.

Materials and Methods

Two systematic reviews were performed to identify the most relevant PROs and obtain an in-depth understanding of experienced symptoms and PSD. In both reviews the last decade (2002 to 2012) was set as a time frame because of the increasing number of studies targeting burnout in this period. Due to the lack of a standardised definition, broad search strategies were developed to make sure that all relevant studies were identified. The results of both reviews were analysed qualitatively and compared to examine the degree of overlap between the content of PROs and experienced PSD. The specific steps are presented below.

PROs

The electronic databases MEDLINE, and PsycINFO were searched for studies published in English between 2002 and 2012 using validated PROs to measure burnout. Using the vocabulary thesaurus of each database, search terms (terms with a slash are MESH terms, the vocabulary thesaurus used for indexing articles for PubMed) were customized by combining the keywords 'Professional burnout', and 'Occupational Stress' with the following key words: case-control studies/ or retrospective studies/, controlled clinical trials as topic/ or randomized controlled trials as topic/, cohort studies/ or longitudinal studies/ or retrospective studies/, cross-sectional studies/, longitudinal studies/ or follow-up studies/ or prospective studies/. Studies reporting the use of PROs to measure burnout were included. The study population was not limited by gender, age or profession. Reviews, case reports, psychometric studies, and books were excluded. The eligibility check of retrieved references comprised two steps. First, an abstract check was performed to

identify studies fitting the inclusion criteria. If the name of the PRO was already literally mentioned in the abstract, this information was extracted. In a second step, full-text articles of included studies not reporting the name of the PRO literally in the abstract were retrieved and checked. The PRO name was extracted from all included studies.

An index was calculated to select the most relevant PROs for content examination. The index was obtained by dividing the number of retrieved studies using the PRO by the number of years since first use (2012 minus year of first use). Instruments that a) yielded an index higher than one or b) were recently developed (from 2008 on) were eventually included. The later criterion should avoid omitting recently developed instruments.

Qualitative studies on the experience of burnout

A systematic review was conducted to identify qualitative studies reporting symptoms and PSD experienced by persons with burnout. The electronic databases MEDLINE and PsycINFO were searched for studies published between 2002 and 2012 by using the same burnout keywords of the PRO review. The following terms were added to the search: Qualitative research/ Interview, Psychological/ Interviews as topic/ or focus groups/. Qualitative and psychometric studies reporting focus groups or interviews targeting symptoms and PSD were included. Reviews, case reports and books, as well as studies where burnout was not a study aim, were excluded.

The eligibility check comprised two steps. First, an abstract check was performed. Second, the full text of included studies were retrieved and rechecked. Signs, symptoms and PSD, such as irritability, exhaustion and interactions with others, were extracted. Data-collection method, population and setting were extracted as well.

In both reviews retrieved abstracts were checked by a single, trained reviewer (Ms. Al-Kudwah). Twenty percent of the abstracts were then double checked by a second reviewer who was blinded to the decision made by the first one. Kappa statistics and non-parametric, bootstrapped confidence intervals (CI) were calculated using SAS to estimate the level of agreement between reviewers. Disagreements were resolved by consensus.

Overlap between PROs and qualitative studies

Qualitative data obtained from different sources can only be compared if a common and sound reference framework is used. The ICF is a universally accepted framework that provides a shared language (WHO,

2001). In the ICF, categories are the units of the classification; they are organized hierarchically and denoted by specific alphanumeric codes. Categories are arranged in a stem/branch/leaf scheme within each of the four major components (Body Functions, Body Structures, Activity and Participation, Environmental Factors). Each component consists of chapters (first level categories), and each chapter consists of second-, third- and fourth-level categories.

The content of the PROs and the symptoms and PSD extracted from the qualitative studies were linked to the ICF using standardized linking rules (Cieza et al., 2005). The linkage procedure started with the identification of meaningful concepts from the PRO items and the extracted data from the qualitative studies. Two reviewers translated the concepts into corresponding ICF categories. Cieza gives an example using the sentence "Pain doesn't prevent me from walking any distance", which contains two different concepts: "pain" and "walking any distance". These concepts are linked to the ICF categories b280 "Sensation of pain" and d450 "Walking" (Cieza et al., 2005). Concepts that could not be linked to the ICF were documented as "other factors". These concepts were grouped into thematically-related, overarching categories. For example, feeling worthless, feeling insecurity and feeling of failure were grouped as "feelings about oneself". The linking procedure was conducted by two linkers working independently. When linkers disagreed, the opinion of a third linker was considered to reach a consensus. No kappa statistics were calculated because nearly all concepts were checked by both reviewers.

Commonalities and divergences between concepts identified in the PROs and symptoms and PSD identified in the qualitative studies were summarized in a table using the ICF categories as the reference.

Results

PROs

The electronic searches yielded 894 studies. After the eligibility check, 383 studies were included. Studies were frequently excluded for not having burnout as a study aim (83.94%).

Nineteen PROs were identified (data not shown). The MBI was by far the most frequently used PRO. The following instruments were selected for content examination: MBI - Human Services Survey (MBI-HSS), index: 24.10, (Maslach et al., 1996); MBI - General Survey (MBI-GS), index: 6.70, (Maslach et al., 1996); SMBQ, index: 2.89, (Lundgren-Nilsson, Jonsdottir, Pallant, & Ahlborg, 2012); Athlete Burnout Questionnaire (ABQ), index: 1.57; (Raedeke & Smith, 2001); OBI, index: 1.57, (Demerouti, Bakker, Vardakou, & Kantas, 2003), Copenhagen Burnout Inventory (CBI), index: 1.50, (Yeh, Cheng, Chen, Hu, & Kristensen, 2007); Professional Quality of Life Scale (ProQOL R-V), index: 0.50, (Stamm, 2010). While the first six PROs were included for having an index higher than 1, the ProQOL-V was included because it was developed in 2009.

Two further instruments should have been included as well: the Nursing Burnout Scale (Garrosa, Moreno-Jimenez, Liang, & Gonzalez, 2008), index: 0.75, developed in 2000, and the Occupational Burnout Scale, index: 0.50, developed in 2010. However, full versions of these scales were not available. The psychometric properties of included PROs are presented in Table 1.

Qualitative studies on the experience of burnout

The electronic searches yielded 409 references. Seven studies (Billeter-Koponen & Freden, 2005; Fischer, Kumar, & Hatcher, 2007; Gillespie & Melby, 2003; Gustafsson, Norberg, & Strandberg, 2008; Pouloupoulos, 2010; Rehnsfeldt & Arman, 2008; Severinsson, 2003) were included (Table 2) after the eligibility check. Burnout was not a study aim in 65.40% of the excluded studies. Our broad search strategy and the low number of qualitative studies on burnout are probably the reasons for keeping less than 2% of identified references.

The kappa coefficient [95% CI] of the 20% double abstract check was 0.77 [0.67; 0.85] for the PROs and 0.75 [0.46; 0.95] for the qualitative studies.

Overlap between PROs and qualitative studies

Altogether 222 of the 269 meaningful concepts identified in the PROs and 161 of the 243 meaningful concepts identified in the qualitative studies were linked to the ICF.

In the PROs, 41.89% of the concepts were linked to “Body Functions”, 56.76% to “Activity and Participation”, and three to “Environmental Factors”. In the qualitative studies, 44.72% of the concepts were linked to “Body Functions”, 36.65% to “Activity and Participation”, and 18.63% to “Environmental Factors”.

Considering “Body Functions”, energy and drive (e.g. energy level, motivation) and emotional functions (e.g. fear, anxiety) are the key themes both in the PROs and qualitative studies (Table 3). Additionally, the MBI-GS, SMBQ, OBI, and ProQOL-V address personality functions (e.g. optimism, confidence), also reported in several qualitative studies. Disparities emerged in problems with sleep, attention, and pain, reported in several qualitative studies, but seldom addressed in PRO.

While commonalities across PROs were common in “Body Functions”, no clear overlap was seen in “Activity and Participation”, with the exception of work, which was addressed in all PROs and almost all qualitative studies (Table 3). The MBI-HSS addressed the highest number of “Activity and Participation” domains, whereas the SMBQ addressed the fewest. Problems in undertaking multiple tasks and handling stress were reported in six qualitative studies, but were only covered by three PROs. Likewise, difficulties in looking after

one's health (e.g. maintaining a balanced diet), and recreation and leisure (e.g. hobbies, socializing) were quite often reported in qualitative studies, but were rarely assessed in the PROs. The negative impact of burnout on diverse interpersonal relationships was reported in qualitative studies. PROs focussed, however, on "formal relationships" with clients and recipients.

Although environmental factors play a significant role in the burnout syndrome, and were reported in several qualitative studies, only two PROs addressed attitudinal aspects (Table 3).

Regarding concepts not linked to the ICF (Table 3), only the ProQOL-V addresses several concepts. Fulfilling expectations, feelings about oneself, needs and health-related issues were reported in several qualitative studies, but are rarely targeted in PROs. The working environment, frequently reported in the qualitative studies, is not included in any PRO. Instead, PROs focus on value given to work, expressions about burnout, and indifference/commitment towards work.

Discussion

This study shows that currently-used burnout PRO measures mainly address specific aspects (usually symptoms related to mental functions), in particular energy and drive and emotional functions. Psychosocial difficulties with regard to activities persons perform in their everyday lives and areas of participation are not consistently included in the PROs. While PRO measures seem to be appropriate screening tools due to their detailed targeting of symptoms, from a biopsychosocial perspective, as proposed in the ICF, they are not comprehensive enough to address the burden experienced by persons with burnout in their day-to-day lives. The comparison across PROs and the information derived from qualitative studies is relevant to researchers and clinicians selecting PROs to assess the burden experienced by patients or to evaluate the effectiveness of their interventions. In light of this study, they can be aware of the symptoms and psychosocial difficulties they are addressing and the ones they are not considering when selecting a PRO. For example, when evaluating an intervention to enhance the difficulties in handling stress, only instruments addressing these difficulties should be considered. Otherwise, the effectiveness of an intervention may not be confirmed.

The fact that PROs are very heterogeneous with respect to their contents (with the exception of energy and drive, emotional functions, and work) does not come as a surprise, considering how controversial burnout is in medical circles. One question that is frequently discussed in the literature is what are the symptoms and psychosocial difficulties that distinctively and specifically define burnout and whether it is "simply" a form of depression (Brenninkmeyer, Van Yperen, & Buunk, 2001). The heterogeneity might also be due to the fact that some instruments were developed for different settings, which have own peculiarities, such as the ABQ for athletes.

The MBI-HSS was an outstanding instrument regarding frequency of use. It yielded an index of 24 compared to the index of 6 of the second-most-frequently-used instrument, the MBI-GS, which is an adaptation of the MBI-HSS for another setting. This is probably due to the fact that researchers choose the PRO that is well known and has been frequently used in previous studies (Schaufeli, Bakker, Hoogduin, Schaap, & Kladler, 2001). A positive consequence is the standardization of assessment, i.e. comparability across studies and settings. However, the assessment of burnout is narrowed to the three dimensions measured by MBI-HSS. Psychosocial difficulties not covered by the MBI-HSS, such as problems with sleep, attention and family relationships have not been systematically considered.

Environmental and personal factors are differently covered across PROs. While most of them address several personal factors, almost none assessed environmental factors, i.e. facilitating or hindering features of the physical, social and attitudinal world (WHO, 2001). Environmental factors are, however, important acknowledged determinants of burnout (Schaufeli et al., 2009). This was confirmed in our investigation, where all but one qualitative study comprehensively addressed environmental factors. We therefore recommend that investigations include specific instruments to assess the impact of the environment on the lives of persons with burnout. This approach will help generate evidence to develop efficient interventions targeting environmental changes.

Surprisingly, we identified only seven qualitative studies targeting burnout. It remains an open question why this syndrome has not been more comprehensively investigated from a qualitative perspective. One possible reason might be the lack of a tool for diagnosing burnout at the individual level; burnout is not a distinct ICD-10 entity. More systematic research in this direction may bring clarity about diagnostic characteristics that could differentiate burnout from other conditions, such as depression. Consequently, the diagnosis of burnout patients would be improved and the treatment specifically tailored to the needs of these patients.

This study has some limitations. Firstly, not using CINAHL, a database specifically covering nursing literature, is a shortcoming. However, we tried to avoid over-representing nursing literature even more than we already did with our selection of studies. Secondly, even though we intended to compare all available instruments to measure burnout, this was not possible due to restricted human resources and the time-consuming character of the task. We therefore restricted our comparison to the most relevant instruments. Thirdly, the index created to establish a cut-off point for the inclusion of different burnout instruments is somehow arbitrary and might have led to the exclusion of more comprehensive instruments than the ones included. Fourth, despite our efforts, we could not assess two PROs. We might therefore have missed two PROs with a potentially greater overlap with psychosocial difficulties than reported in qualitative studies.

In general, our investigation shows that there is insufficient qualitative research in the field of burnout, the

