Study Protocol

Screening for the risk on long-term sickness absence

Kaat Goorts,[1] Sofie Vandenbroeck,[1,2] Dorina Rusu,[3,4] Marc Du Bois,[1] Lode Godderis[1,2]

[1]University of Leuven, Centre for Environment and Health, Leuven; [2]IDEWE, External Service for Prevention and Protection at Work, Leuven; [3]Occupational and Environmental Medicine, Department of Public Health Sciences, University of Liège; [4]SPMT-ARISTA, External Service for Prevention and Protection at Work, Leuven, Belgium

Significance for public health

In this research, we tested a generic instrument to screen for long-term sickness absence, regardless the cause of the sickness absence or the political context. Both biomedical factors and psychosocial factors (such as the patients' own prediction) are questioned in the prediction model, which is thus adapted to the modern view on sick leave. A screening method to detect high risk of long-term sickness absence among the large group of sick employees might help to use resources (e.g. money, services) in a more efficient way. Physicians will be able to focus on patients with a high risk on long-term sick leave, and the return to work process of employees at high risk can start much earlier. The instrument will, next to the physicians' prognosis, offer support in prioritizing patients' files. Hence, a lot of patients will resume their work spontaneously. Patients who need support will experience shorter follow-up periods, and better quality of care. In addition, the relationship between predicting factors of the biopsychosocial model and long-term sickness absence will be examined.

Abstract

Background: Long-term sickness absence is a growing concern in Belgium and other European countries. Since 2017, Belgian physicians of the sickness funding organisations are required to assess the re-integration possibilities within the first two months of sickness absence. Given the shortage of physicians in the assessment of work disability and the growing number of people in sickness absence, there is a need for a triage tool, allowing to assign return-to work support to patients having a high-risk profile not to resume work.

Methods/design: The current study comprises a comprehensive validation process of a screening tool that supports Belgian physicians in guiding people back to work. The study consists of a theoretical construct validation (face validity and content validity), and an empirical construct validation (concurrence validity, factorial validity, predictive validity, hypothesis testing validity and known- group validity).

Expected impact of the study for Public Health: The screening instrument assessing the risk for long-term sickness absence is a tool developed to support physicians who work for sickness funds and for occupational health and safety organisations. Both professionals play an important role in the return to work process and the prevention of long-term sickness absence. The screening tool aims at making a distinction between people who will resume their work independently and people who will need support to do so. Generation of this prediction model will help physicians to focus effort and resources in the high-risk group. Results may also help understand the relationship between the biopsychosocial model and long-term sick-leave.

Introduction

Long-term sickness absence − defined as absence from work longer than six months due to sickness − is increasing in the 27 European member states and Norway.[1] The Organization for Economic Co-operation and Development (OECD) estimates the costs of disability and sickness to be 2.5 times higher than those of unemployment.[1] Sickness absence is not only costly for the organization and for society, but also for the individual. Previous studies have shown that long-term sickness absence is a prognostic marker of even more future absence from work, early retirement due to ill health and mortality. Long-term sickness absence can also be associated with future unemployment, financial difficulties, psychological problems, and social exclusion.[2]

Therefore, several countries have taken initiatives to decrease long-term sickness absence, and lower the costs related to disability and sickness. However, despite the growing interest in decreasing sickness absence across Europe and many countries that implement promising projects, it remains challenging to dedicate the resources to the people that need it most, i.e. sick listed employees with a high risk of not returning to work. Hence, most sick listed employees are returning to work spontaneously within six weeks and do not need additional support.[3] Existing instruments do not make a clear distinction between employees who will spontaneously return to work and employees who have a high risk of long-term sickness absence. Early intervention is key, because it is widely acknowledged that the longer an employee is off work, the smaller his chances of ever returning to work.[4] Above this, most countries have to deal with great shortage of physicians who judge disability for sickness fund organisations and insurance companies.[5]

Currently, many initiatives are taken worldwide. In Belgium, new legislation urges physicians of sickness fund organisations and insurance companies to make a first estimation about re-integration possibilities about every patient before two months of sickness absence. In addition, a close cooperation with the occupational health physician is advised in this legislation. A screening method to detect high risk of long-term sickness absence among the large group of sick employees might thus be a useful contribution to support both physicians to meet the requirements of the new legislation. As such, resources (*e.g.* money, services) can be provided in a more efficient way, and the return to work process of employees at high risk can start much earlier.

In this context, the present study is planned as a part research project with the objective to develop and validate a new, more generic questionnaire that is applicable in all contexts to predict long-term sickness absence. The construction of the instrument, more specifically selecting and combining scales, has already been executed in a previous study. In the current study, we will

perform a complete validation of the questionnaire, testing both the theoretical and empirical construct validity, according to a validation model.[6]

Informed consents are available for insight and were approved by the ethical committee responsible for the project.

Materials and Methods

Aims and objectives

The main objective of this study is to evaluate the validity of a new screening instrument, aiming to prevent long-term sickness absence.

Patients will complete the questionnaire after six weeks of sick leave in order to make a first assessment about their risk for long-term sick leave. The goal is to screen patients for their risk on long-term sick leave. This way, the physician can focus on the high-risk population and thanks to the content of the questionnaire, the physician will possess more information about the opportunities and threats in the work resumption process. The instrument is available through an online platform, allowing patients to complete it at home. Afterwards, results are immediately shared with the patient, and the responsible physician of the sickness benefits organisation. After both the validation and implementation process, the patient will also be able to share the results with his general practitioner and his occupational health physician (optional).

The following questions will be addressed:
- Is the screening tool valid to predict long-term sickness absence according to patients who experience(d) long-term sick leave. Are there predicting factors that should be added or removed from the questionnaire?
- Is the screening tool a valid measure for long-term sickness absence when compared to the occupational health physicians' estimation of the probability of work resumption?
- What is the internal validity of the different scales of the questionnaire? Are the scales valid in different populations (occupational and insurance medicine)?
- Is the screening tool a valid predictor of long-term sickness absence using the actual time until work-resumption as a reference?
- What does the screening tool predict about the time until work-resumption for cancer patients? Does this converge with findings from other studies about cancer patients?
- Is the screening tool able to confirm hypothesis about sickness absence based on scientific evidence (*e.g.* higher risk for certain professions)?

Outcome measures

The primary outcome is time until work-resumption. This outcome will be identified through different kinds of measures: a subjective estimation of the occupational health physician, an objective registration of the actual time until resumption, the patients' estimation of time until work resumption, hypothesis about return to work in cancer patients and hypothesis based on the Belgian statistics about long-term sick leave.

Data collection

In total, three data samples will be collected to perform the validation.

The first data sample will be collected by inviting patients to five focus groups on different locations in Belgium. The patients will be invited using the network of the researchers and using the network of both Dutch and French patient organisations. Both patient organisations will spread the invitation to their connected organisations and to the patients in their network.

The second data sample will be collected by inviting occupational health physicians to participate in the study. The invite will be sent by the researchers themselves and by CoPrev, the umbrella organisation of all external health and safety services in Belgium. The recruited occupational health physicians on their turn will recruit patients for the study who are in sick leave and who come for a consult concerning return to work. After a period of three weeks, the physicians sent all questionnaires back to the researchers by regular mail. All questionnaires will be scanned using special scanning software.

The third data sample will be collected sending the questionnaires via all sickness fund organisations to all patients who are in sick leave for six weeks. The patients will receive an invite (either by regular mail or by email) in which they are asked to use the electronic available platform to fill in the questionnaire. A helpdesk is foreseen. After the data collection, the work resumption data of all participants will be collected via the sickness fund organisations. A random code is used for every patients file to reassure that the researchers never know patients' identities. The sickness fund organisations sent all work-resumption dates to the researchers who will put them together in the database. Two inclusion periods are foreseen (6 and 6 weeks).

The data collected throughout the validation process will be saved either on the servers of KULeuven (qualitative analysis and patients sample 1), or on the servers of the sickness funds organisations (patients sample 2) (Figure 1). From both sample 1 and 2 all directly identifiable personal data (name, social security number, address etc.) is replaced by a unique ID-number before data are sent to the Research group. Researchers are never able to iden-
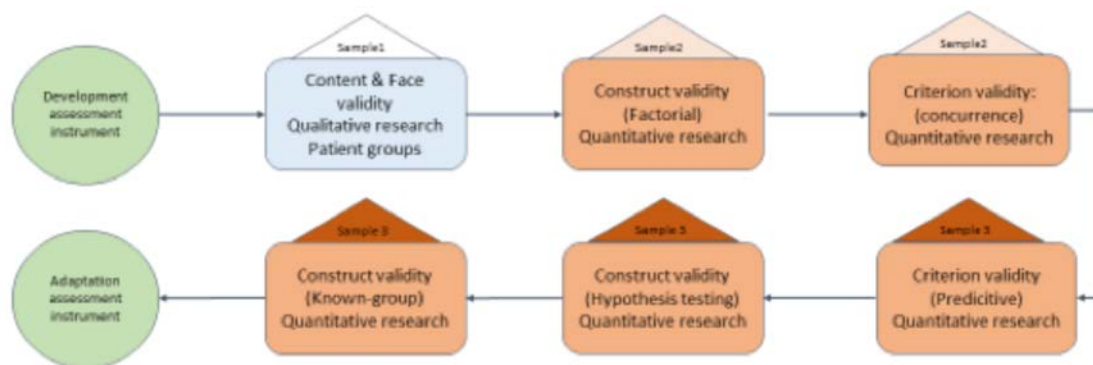

Figure 1. Research process flow.

tify patients thorough the research process. In the qualitative research project, the researchers know the identity of the patients. An informed consent assures the patients that no personal data will be spread in any case. The personal data of the patients (name and e-mail address) are stored at secured KULeuven servers.

Study design

One of the most common tasks often encountered in social science research is ascertaining the validity and reliability of a measurement tool.[7] Researchers always wish to know if the measurement tool actually measures the intended research concept or construct (is it valid? or true measures?) or if the measurement tools used to quantify the variables provides stable or consistent responses (is it reliable? or repeatable?).

As simple as this may seems, it is often omitted or just mentioned passively in the research proposal or report.[7] This has been adduced to the dearth of skills and knowledge of validity and reliability test analysis among social and health science researchers. In this research the psychometric quality is therefore studied very extensively.

The study design is presented in Figure 1 was inspired by the validation framework of Bolarinwa and colleagues.[6] Figure 2 depicts the various types of validity that should be tested to support the psychometric evidence of the screening tool.[6]

As we aim to validate our screening tool, we will conduct various studies. All validation studies are marked in rectangles in Figure 1. The triangle on top of it shows which patient sample was used to conduct the test (1, 2 or 3). Both the theoretical and the empirical construct will be validated as described by Bolarinwa and will be performed simultaneously.[6]

Development of the screening instrument

The development process of the screening tool for long-term

sickness absence was based on two approaches: a literature review followed by a snowball sampling technique to address the following research questions: i) Which are the predictors for long-term sickness absence? ii) Which questionnaires measure these predictors for long-term sickness absence in a valid and reliable manner?

To answer the first research question and to determine the in- and exclusion criteria and search string a PICO strategy (population, intervention, comparison and outcome) was applied: Population: active population (18-65y), Long-term sickness absence (invalidity, incapacity), Different conditions who are at the cause of the sickness absence; musculoskeletal disorders, mental disorders, heart conditions and colon cancer, Predictors, Outcome: return to work/not (part-time, fulltime, with/without support). The search inquiries were run in the scientific databases Medline and Scopus.

Exclusion criteria were: children, adolescents, elderly, persons with an intellectual disability, other specific disorders, descriptions of interventions or rehabilitation programs, an evaluation of the efficiency of interventions or rehabilitation programs, an evaluation of a treatment.

For the first research question, the search retrieved 998 publications in Medline and 247 in Scopus. After removal of duplicates, a total of 1010 publications were retained. After reading the title, abstract and full text, 53 publications were selected. In total, 21 predictors were found in literature, which were assigned to four categories: stressful life events, work-related factors, person related factors and functioning factors.

To address the second research question, a snowball sampling technique was used to identify existing questionnaires. Snowball sampling (also known as chain-referral sampling) is a non-probability (non-random) sampling method used when sample characteristics are rare and difficult to find. This sampling method involves primary data sources nominating another potential pri-
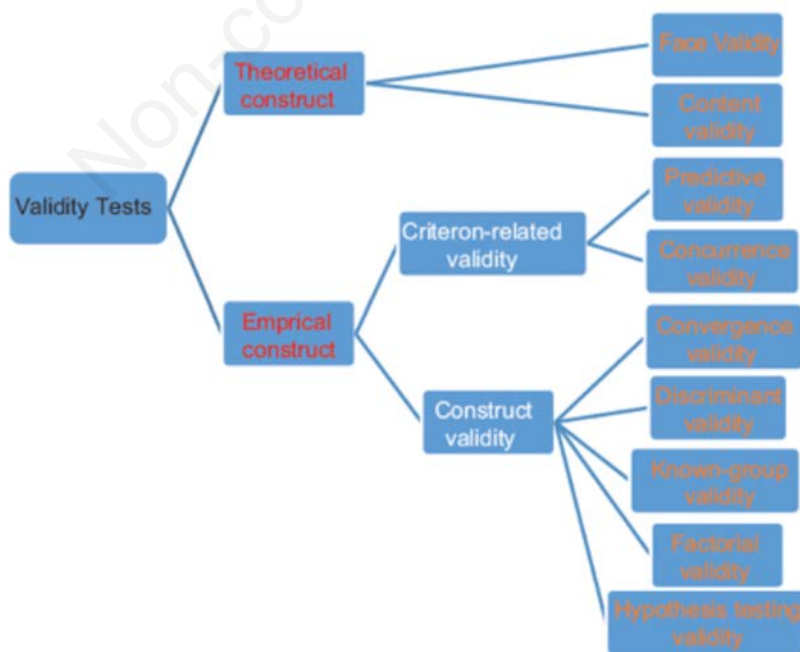

Figure 2. Graphical representation of the subtypes of various forms of validity tests.

mary data source to be used in this research. In other words, snow-ball sampling method is based on referrals from initial subjects to generate additional subjects.[8] We applied the snowball sampling method, because we wanted to include grey literature sources about (non-published) best practices in our study.

There are different kinds of snowball sampling. In this research, exponential non-discriminative snowball sampling was conducted. The first subject recruited to the sample group provides multiple referrals. Each new referral is explored until primary data from sufficient amount of samples are collected.[8]

The inclusion criteria were: a questionnaire about at least one factor that predicts long-term sickness absence (out of the 21 factors we found in literature). The exclusion criteria were: covering more than one disease and if the questionnaire was a long-form then the short form was taken.

The snowball sampling started at two points: six questionnaires from the literature review of phase 1 were retained and three questionnaires from the grey literature. The exponential non-discriminative snowball sampling was conducted by an examination of all sources of the articles concerning each questionnaire. The selected sources from the articles are entered in PubMed and a search for 'similar articles' was performed to screen the literature in a systematic way.

Finally, the results of both search strategies were compiled.

Participants, recruitment, inclusion and exclusion

The participant's recruitment depends on the phase of the research process. The initial screening instrument was based on literature and existing questionnaires, which did not require patient participation. As mentioned earlier, three main patient samples will be collected to conduct five out of six validation tests. The sixth validation test will be performed using data from a cooperative research project. We describe the recruitment process of all three samples in the next section.

In view of the qualitative evaluation (content and face validity), patients will be recruited to participate in focus groups via patient organisations, and social network sites. Eligible patients are at working age (18-65y) and have experienced at least six weeks of sickness absence. For the second and third part of the validation process, (factorial and concurrence validity) patients were selected via occupational health physicians. Several occupational health physicians were asked to include every patient whose consult concerned return-to-work support, during an inclusion period of six weeks. The inclusion criteria for the patients were: 1) being on sick leave, 2) made an appointment with an occupational physician to discuss re-integration, and 3) French or Dutch speaking. Exclusion criteria were: 1) Not receiving sickness benefits, 2) having an appointment with the occupational physician for any other reason than discussing re-integration, 3) not speaking French or Dutch.

For the other parts of the validation (predictive and hypothesis-testing) a third sample of patients will be recruited. All sickness fund organisations will send the questionnaire to every patient who reaches the 7th week of sick leave. The recruitment period will cover 6 weeks for the first period and 4 weeks for the second period. Working in two periods is designated because practical issues with the new online platform can be solved in the meantime. For example; if people are asking the same questions about where to fill in their individual code, then we can make this clearer in the next inclusion period. Only employees are included, because civil servants and self-employed people have other sick leave systems, which are handled by different authorities in Belgium. The factorial validity will be retested in this population sample and the results will be compared with the population sample in occupational medicine.

Validation of the screening instrument

The validation of the screening instrument encompasses six components: a theoretical construct validation, a criterion validation (concurrence), a construct validation (factorial), a criterion (predictive) validity test, a construct (hypothesis) validity test, and a construct (known- group) validity test. The validation process is a mixed method approach, since both quantitative and qualitative research methods are applied. We will discuss all parts of the validation process below.

Theoretical construct validation

Theoretical construct validity refers to the degree to which the instrument fully assesses or measures the construct of interest.[6] The development of a content valid instrument is typically achieved by a rational analysis of the instrument by raters (experts) familiar with the construct of interest or experts on the research subject.[6] In fact, expert consultation is considered by some to be the *sin qua non* of content validity (Messick, 1995).[9] According to Vogt *et al*., members of the population may also provide useful input at the item-development stage, as they can review items for their ease of understanding and relevance to and representativeness of the construct. Although the term expert has typically been used to refer to researchers who are knowledgeable in the specific topic area, members of the population under study may also be considered *experts* in some cases.[10]

To test for theoretical construct validity of the screening instrument in this study, patients will be invited to participate in focus groups. In total five focus group conversations will take place in two of the national languages, 3 in Dutch, and two in French. Consensus techniques, such as the nominal group technique (NGT), are commonly used with health professionals for developing clinical practice and setting priorities. The aim of consensus methods is to determine the degree to which experts agree about a particular issue. The use of such methods is valuable where unanimity of opinion does not exist because of lack of scientific evidence or the complexity of the issue.[11]

Nominal group technique will be applied to pose the following questions:
- What are the major bottlenecks and support needs that you (have) experience(d) in order to resume work? Patients will note five issues or opportunities. Afterwards we will categorise all bottlenecks or opportunities (work-related, person-related, …). Next, patients are given some time to read the questionnaire instrument that has been developed by the research team.
- Are there other factors that they want to add now? Are all the factors that they named in the first phase retrievable in the questionnaire?
- Finally, every patient may make a top five of the most important factors, that certainly should be questioned, and write it down on a piece of paper (individually).

The NGT uses multiple rounds of ranking to achieve consensus and is conducted through face-to-face meetings to facilitate discussion and consensus-forming. The systematic and democratic process equalizes decision making among participants and allows individual opinions to be informed by the community of experts.[11]

To analyse the results of the focus groups, all recordings will be transcribed. Then, using NVIVO 11 software, all factors named by patients will be compared to the factors that are questioned in the questionnaire. Potential additional factors will be discussed in the research group. The transcripts will be evaluated by at least two people of the research group, as well as all focus groups will be guided by at least two people to prevent for interpretation (measurement) bias.[12] Additionally, the scores that the patients gave to the most important factors will be summed for each outcome

selected across all participants. A percentage of the maximum possible score will then be calculated for each outcome [(score from Round 3 / 15 × number of participants) × 100]. The process of analysis is described in Sanderson *et al.*[11]

The other qualitative data (why are people selecting certain factors), will be analysed using grounded theory paradigm. Grounded theory is a general methodology with systematic guidelines for gathering and analysing data to generate middle-range theory. The analytic process consists of coding data; developing, checking, and integrating theoretical categories; and writing analytic narratives throughout inquiry. A grounded theory of a studied topic starts with concrete data and ends with rendering them in an explanatory theory.[13] In this case, we will analyse the specific motivation of people choosing certain factors as *the most important,* or *priorities*. These motivations are hidden in the unstructured data of the transcripts of the Focus groups. Coding these data, we will try to formulate a theory about why people are prioritizing the factors in a certain way.

Construct (factorial) validity

Factorial validity is an empirical extension of content validity. This is because it validates the contents of the construct employing the statistical model called factor analysis. It is usually employed when the construct of interest is in many dimensions, which form different domains of a general attribute. In the analysis of factorial validity, the several items put up to measure a particular dimension within a construct of interest is supposed to be highly related to one another than those measuring other dimensions. Construct validity will be evaluated through a series of confirmatory factor analyses.[6] In this study, we will test a hypothesized measurement model for each category of factors (*i.e.* work-related factors, functioning, stressful life-events and person-related factors). Model fit of these measurement models will be evaluated using the following fit indices: the Comparative Fit Index (CFI), the Non-Normed Fit Index (NNFI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Residual (SRMR). The χ2 difference test will be applied to compare the hypothesized measurement model with alternative, nested models. Analyses will be performed with SPSS 24 and AMOS 24. The analysis will be performed for two samples, and the results will be compared (for sample 2 from occupational medicine and sample 3 in insurance medicine).

Criterion (concurrence) validity

Criterion validity assesses the newly developed questionnaire against a highly rated existing standard. When the criterion exists at the same time as the measure, we talk about concurrent validity.[6] Path analysis will be applied to investigate the concurrence validity. We will test multiple models including interrelationships between the explanatory variables from the patients' questionnaire and occupational health physicians' estimation of patients' return to work. In this case, the estimation of the occupational health physician serves as *gold standard* for actual return to work, since this criterion exists at the same time as the questionnaire measurement. Standardized regression weights will be considered significant at the 0.05 level. Analyses will be performed with SPSS 24 and AMOS 24.

Criterion (Predictive) validity

Predictive validity assesses the ability of the questionnaire

(instrument) to forecast future events, behaviour, attitudes or outcomes. This is assessed using correlation coefficient(6). To test for predictive validity of the screening tool, the time until work resumption for every patient will be measured. Survival analysis will be applied to discover possible relationships between the measured variables and the time until work resumption. Survival analysis suits the research question the best, because at the end of the study period, not all employees will yet have resumed labour, and some maybe never will. Analysis will be performed using software R and SPSS 24.

Construct (Hypothesis testing) validity

Hypothesis testing validity is the evidence that a research hypothesis about the relationship between the measured concept (variable) or other concepts (variables), derived from a theory, is supported.[6] In this study, we will test different hypothesis using the same dataset as for the predictive validity test. Statistics from the Belgian organisation for sickness and disability benefits are giving us the opportunity to test our prediction model for various hypothesis from which we know they should be (not) true according to the provided statistics. Following hypotheses will be tested: i) Psychological problems and musculoskeletal disorders are the main causes of occupational disability; ii) Workers are more often absent due to illness than white-collar workers; iii) Women are more often absent due to illness than men; iv) Absence due to sickness increases with age. If the screening tool concludes comparable ideas about these hypotheses stated in statistical reviews from available data, we could conclude that the hypothesis testing validity is met.

Construct (Known group) validity

In known-group validity, a group with already established attribute of the outcome of construct is compared to a group in whom the attribute is not yet established. Since the attribute of the two groups of respondents is known, it is expected that the measured construct will be higher in the group with related attribute but lower in the group with unrelated attribute.[6] The known group validity test will take place using data from another research project that examines pain related disability in breast cancer patients. We will use the completed questionnaires from the cancer patients, from whom it is known that they generally have a high risk on long-term sickness absence given their diagnosis. Other information (such as participation in return-to work programs) about the known group will be collected in cooperation with the other research team. The work resumption prediction of the questionnaire will be evaluated, the same way as we did for the predictive validity. The same variables will be measured.

Sample size calculation

Since we are validating the screening tool trough different validation tests, all tests (except for the qualitative research) need the same sample size calculation. According to power-calculations a sample size of 269 participants is sufficient within a population of ±400 000 people (± amount of patients in primary sickness absence in 2013),[14] a confidence level of 90% and a margin error of 5%.[15]

The population that will use our screening tool are the people who are in sick-leave for six weeks. The population size will be a little smaller than 400 000 since primary sickness absence starts at four weeks of sick leave. Sample sizes around 300 will thus be sufficient to conduct our analysis in all validation studies.

Discussion and Conclusions

The development and validation of a screening instrument to screen early in the sickness period for risk on long-term sickness absence is an important study for the Belgian government since it is trying to reduce long-term sickness absence and promote return-to work projects. An early screening can help physicians to focus on high-risk patients who will need their support to resume work. Since the screening tool is very generic, and makes use of existing validated instruments, it might be applicable in various international contexts.

A complete validation of the questionnaire is important to persuade all stakeholders involved (patients, physicians, policy makers, …) of the necessity, and to provide an instrument that is evidence based with proven psychometric value.

The results of this study might help public health systems to prevent long-term sickness absence and to support patients more efficiently in their rehabilitation process. A modern view on predictors for long-term sickness absence is applied in this research. Therefore, not only biomedical factors, but also psychosocial items are questioned in the questionnaire. This means that the results of the research will help us to gain more insight in the relationships between biopsychosocial factors and long-term sickness absence.

Strengths in the study design include a combination approach of different validation methods in different study populations.

References

1. Edwards P, Greasley K. Absence from work In: conditions EFftiolaw, editor. Eurofound 2010.
2. Kausto J, Pentti J, Oksanen T, et al. Length of sickness absence and sustained return-to-work in mental disorders and musculoskeletal diseases: a cohort study of public sector employees. Scand J Work Environ Health 2017;43:358-66.
3. Tompa E, de Oliveira C, Dolinschi R, Irvin E. A systematic review of disability management interventions with economic evaluations. J Occup Rehabil 2008;18:16-26.
4. Wiencke M, Cacace M, Fischer S. Healthy at work interdisciplinary perspectives. Switzerland: Springer international Publishing; 2016.
5. de Wind A, Donceel P, Dekkers S, Godderis L. The role of European physicians in the assessment of work disability: A comparative study. Edorium J Disabil Rehabil 2016;2:78-87.
6. Bolarinwa OA. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. Nig Postgrad Med J 2015;22:195.
7. Kember D, Leung DYP. Establishing the validity and reliability of course evaluation questionnaires. Assess Eval Higher Educ 2008;33:341-53.
8. Dudovskiy J. Snowball Sampling 2016. Available from: http://research-methodology.net/sampling-in-primary-data-collection/snowball-sampling/.
9. Messick S. Standards of validity and the validity of standards in performance asessment. Educat Measure Issues Pract 1995;14:5-8.
10. Vogt DS, King DW, King LA. Focus groups in psychological assessment: enhancing content validity by consulting members of the target population. Psychol Assess 2004;16:231.
11. Sanderson T, Hewlett S, Richards P, et al. Utilizing qualitative data from nominal groups: exploring the influences on treatment outcome prioritization with rheumatoid arthritis patients. J Health Psychol 2012;17:132-42.
12. Sutton J, Austin Z. Qualitative Research: Data Collection, Analysis, and Management. Canad J Hosp Pharm 2015;68:226-31.
13. Charmaz K, Belgrave LL. Grounded theory. The Blackwell encyclopedia of sociology. 2007.
14. RIZIV. Ziekteverzuim in de periode van primaire arbeidsongeschiktheid: analyse en verklarende factoren: 2010 -2015 Brussel: RIZIV, insurance Nifsad; 2016.
15. Rosner B. Estimation of sample size and power for comparing two binomial proportions. Fundamentals of Biostatistics 7th ed Boston: Brooks/Cole. 2011:381-90.