Development and validation of prognostic models to predict risk of prolonged sickness absence in workers with musculoskeletal disorders: statistical analysis plan

Corresponding author: Tarjei Rysstad, tary@oslomet.no

Authors: Tarjei Rysstad[1], Margreth Grotle[1, 2], Adrian C. Traeger[3, 4], Lene Aasdahl[5, 6], Kate M. Dunn[8], Gwenllian Wynne-Jones[8], Ørjan Vigdal[1], Are Hugo Pripp[1], Fiona Aanesen[5], Egil Fors[6, 7], Steven J. Linton[9], Anne Therese Tveter[1, 10] (final order is not determined).

Statistical advisor: Are Hugo Pripp[1]

Affiliations: [1]Department of Rehabilitation Science and Health Technology, Faculty of Health Sciences, Oslo Metropolitan University, Norway. [2]Research- and Communication Unit for Musculoskeletal Health, Oslo University Hospital, Norway. [3]Institute for Musculoskeletal Health, The University of Sydney and Sydney Local Health District, Sydney, Australia. [4]School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia. [5]National Institute of Occupational Health (STAMI), Norway. [6]Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. [7]Unicare Helsefort Rehabilitation Centre, Rissa, Norway. [8]School of Medicine, Keele University, Keele, UK. [9]Department of Law, Psychology, and Social Work, Örebro University, Örebro, Sweden. [10]Center for treatment of Rheumatic and Musculoskeletal Diseases (REMEDY), Diakonhjemmet Hospital, Oslo, Norway.

Word count: 4692

DEFINITION OF TERMS

Term	Source	Definition
Calendar days	NAOB	Every day on the calendar, including weekends and public holidays.
Disability pension	NAV	Regular income to individuals who are unable to work due to a permanent disability. To be eligible for disability pension, the illness and/or injury must be the main cause of reduced work and earning capacity.
Employment	ILO/WHO	An agreement to produce goods or services for a specific period in time for compensation by a salary, a wage or in kind. Different types of employment exist, among which is self-employment.
Part-time employment	ILO/WHO	When the hours of work are less than the 'normal' hours of work of comparable full-time employment.
Prolonged sickness absence	[1–5]	Sick leave for >90 continuous days or >180 continuous days
Sick leave/sickness absence/work absence	EULAR	Time off from work that workers can use to stay home to address their health and safety needs without losing pay.
Work assessment allowance	NAV	Financial support provided by the government for people who are unable to work due to illness or injury, but are expected to return to work in the near future.
Work disability	ILO	When an individual is unable to perform work-related tasks due to physical or mental impairments or disability.
Workdays	NAOB	Any of those days of a week on which work is done, officially Monday to Friday, a 5-day workweek.

Abbreviations: EULAR, European Alliance of Associations for Rheumatology; ILO, International Labour Organisation; NAOB, Norwegian Academy Dictionary; NAV, Norwegian Labour and Welfare Service; WHO, World Health Organization.

BACKGROUND

Long-term sickness absence due to musculoskeletal disorders (MSDs) is associated with substantial societal costs, constituting a major concern for the welfare state in western societies [6]. In Norway, the sickness absence rate is among the highest in Northern Europe and MSDs accounted for 35-39% of the total sickness absence [7]. Although it is only a small proportion of individuals that experience long-term sick leave, this group is responsible for the majority of the total sick leave cost [6]. Early identification of workers on sick leave at risk of prolonged work absence is therefore important to guide stratified prevention strategies to improve individual outcomes and reduce societal costs [8]. This targeting can be done by using prognostic models [9].

There have been some previous attempts to develop prognostic regression models for predicting work absence. However, most of these models were either developed for individuals without an employment contract or for workers not already on sick leave, aiming at primary prevention strategies [4,10–15]. Of the few prognostic models developed for secondary prevention, the models have either demonstrated insufficient performance, lack external validation [14,16], have been developed in a homogenous sample with a specific diagnosis [4], or in a specific work sector [17]. Thus, a crucial need exists for accurate models to predict risk of long-term, or prolonged, sickness absence used within the earlier stage of an MSD-related sick leave.

To address this, we aim to develop and externally validate prognostic models to predict an individual's risk of prolonged sickness absence due to MSDs in a social insurance setting in Norway. However, the definition of prolonged or long-term work absence outcomes can vary greatly, with most previous prognosis studies using anything from over 14 days to over 365 days [2,10,18,19]. In this study, we have defined three binary outcomes for prolonged sickness absence: (1) being on sick leave for more than 90 consecutive days, (2) more than 180 consecutive days, and (3) any new episode or increase in work assessment allowance (WAA) and/or disability pension (DP) during a 1-year follow-up [1–5]. Therefore, we need to develop three separate prognostic models for each outcome.

Objectives

Our study objectives are to:

1. Develop and internally validate three prognostic models for prolonged sickness absence (>90 consecutive days, >180 consecutive days, and granted WAA/DP) used within the first 4-12 weeks of sick leave due to an MSD in a social insurance setting.
2. Externally validate these models and assess their clinical usefulness in a new sample separate from the sample used in the development stage.

METHODS

Our study design will be informed by the PROGRESS framework [9] and will be reported according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement for prediction studies [20,21].

Data sources

This study involves analysis of data from three data sources: a cohort study and two randomised controlled trials (RCTs). All three studies collected 12 months of sick leave data from participants. We will use all arms of the RCTs (intervention and control groups). The study protocols and details of these studies have been published elsewhere [22–24], and detail on study populations is briefly described below and in Table 1.

Development sample

To develop the models, we will combine data from a prospective cohort study [22] and an RCT (MI-NAV trial) [25]. The prospective cohort study was carried out from November 2018 to February 2020, involving workers on sick leave due to MSDs who were invited to participate through the electronic communication website of the Norwegian Labour and Welfare Service (NAV). The inclusion criteria were workers aged 18 to 67 years on sick leave for a minimum of 4 weeks with a diagnosis within the musculoskeletal (L) chapter of ICPC-2 (International Classification of Primary Care, Second edition) [26]. Exclusion criteria were insufficient Norwegian or English

language skills to participate in the study. This study recruited 549 participants with median age of 50.1 (interquartile range [IQR], 41.9-56.9) years, who had been on sick leave at a median of 37.8 (IQR, 26.1-80.3) workdays the year prior to the baseline assessment.

The MI-NAV trial was a multi-arm randomised controlled trial investigating the effectiveness of adding motivational interviewing or stratified vocational care intervention to usual care in a sample of workers on sick leave due to MSDs. The inclusion criteria in this trial were similar to the prospective cohort study: participants were eligible if they were aged 18-67 years, employed full or part-time, and on sick leave due to MSDs (L-chapter of ICPC-2) for at least 50% of their contracted work hours for more than 7 weeks. Exclusion criteria included serious somatic or mental health disorders, pregnant women, those who were unemployed, freelancers, self-employed workers and people lacking sufficient Norwegian or English language skills. This trial consisted of 514 participants, median age 48.8 (IQR, 40.8-55.3) years, who had been on sick leave at a median of 36.8 (IQR, 29.3-50.0) workdays the year prior to the baseline assessment. The MI-NAV trial showed a 7-day non-statistically significant reduction in sickness absence over 6 months for the intervention groups compared to usual care.

Table 1. Characteristics of the datasets that will be used in this study.

Study	N	Study type	County, Country	Diagnosis coding	Follow-up time	Inclusion criteria	Linkage to sickness absence registry data	Intended use of database
Prospective cohort study [22]	549	Prospective cohort	All counties, Norway	ICPC-2	12 months	Sick leave due to MSD for at least 4 weeks. Aged 18-67 years.	Yes	Development
MI-NAV trial [25,27]	514	Three-arm parallel RCT	Vestfold, Norway	ICPC-2	12 months	Sick leave due to MSD for at least 7 weeks (>50 % of contracted working hours). Aged 18-67 years.	Yes	Development
MI study [24]	865	Three-arm parallel RCT	Trøndelag, Norway	ICPC-2	12 months	Sick leave due to any diagnoses for at least 8 weeks (>50 % of contracted working hours). Aged 18-60 years.	Yes	External validation

Abbreviations: ICPC-2, International Classification of Primary Care 2nd version; MI, motivational interviewing; MSD, musculoskeletal disorders; RCT, randomised controlled trial.

Validation sample

We will externally validate the models using data from all arms of a second RCT (MI study) [24]. This study evaluated the effect of motivational interviewing on return to work in a social security setting using a parallel group design. Participants aged 18 to 60 who had been on sick leave for more than 8 weeks were recruited through the NAV system. Individuals who were not employed or pregnant were excluded. This trial comprised 865 participants, median age 45.0 (IQR, 37.0-52.0) years, who had been on sick leave at a median of 52.1 (IQR, 40.4-65.6) workdays the previous year. The results from this trial have not been published yet.

The inclusion criteria for all studies are pragmatic and unrestricted, indicating that they are likely to be representative of the target population. We plan to use data from all arms (intervention and control) of the RCTs.

Source population

Our prognostic models will be focused on information collected between the initial 4-12 weeks of a sick leave spell due to MSDs. As such, participants of each dataset will be eligible if they at baseline were:

- On sick leave due to a musculoskeletal diagnosis within the musculoskeletal (L) chapter of ICPC-2 [26].
- On sick leave for at least 50% of their contracted work hours for 4-12 consecutive weeks.

Individuals on sick leave for more than 12 weeks at baseline were excluded.

Setting

All data sources are from a social insurance setting in Norway.

Sickness absence data

Sickness absence days will be measured using registry data from the National Social Security System Registry where all individuals receiving any form of benefits in Norway

are registered by their social security number. We will extract sick leave data for each participant from 12 months before and 12 months after baseline.

In Norway, after 12 months of sick leave, it is possible to apply for long-term medical benefits: work assessment allowance, and disability pension. Both provide approximately 66% of the income and can be granted for 20-100% of ordinary working hours. Disability pension can be combined with part-time work. Therefore, a person receiving disability pension can also be on part-time sick leave.

In this study, we will calculate days of medical benefits according to a 7-day week for every month during follow-up. Days on sick leave will be adjusted for part-time sick leave, work assessment allowance, and disability pension. Any new episode or increase in disability pension during follow-up will be counted as sick leave. Furthermore, we will adjust days on sick leave based on the participant's contracted working hours. For example, a person who works part-time at 50% and is 100% on sick leave will have 3.5 calendar days of sick leave during a week.

Start-point (time of prediction)

The start-point, the time of intended prediction, will be within the initial 4-12 weeks of sick leave due to an MSD.

We chose this time window (within 4-12 weeks) based on two aspects. Firstly, the majority of workers on sick leave with MSDs will return to work within the first 4 weeks [5], and we therefore believe that it is not beneficial to screen all individuals within their first weeks on sick leave. Secondly, it is important not to wait too long with intervention strategies, as the longer people are away from work the less likely they are to return [28,29]. This time window is in line with recommendations for when work disability prevention strategies should be initiated [30–32].

End-points (outcome definitions)

The end-points (outcomes) to be predicted are three binary end-points that characterise prolonged sickness absence during the follow-up of 12 months:

1. Over 90 days of work absence during the first sick leave spell, defined as the number of calendar days during a continuous episode of full-time or part-time sick leave. Coded as 0 = ’≤90 days’, 1 = ’>90 days’.
2. Over 180 days of work absence during the first sick leave spell, defined as the number of calendar days during a continuous episode of full-time or part-time sick leave. Coded as 0 = ’≤180 days’, 1= ‘>180 days’.
3. Any new episode or increase in work assessment allowance (WAA) and/or disability pension (DP) during follow-up. Coded as 0 = ’no WAA/DP’, 1 = ’new episode/increase in WAA/DP’.

For end-point 1 and 2, the end of the first sick leave spell will be censored at ‘full sustainable return to work’ or end of follow-up. We defined ‘full sustainable return to work’ as having returned to work for 4 weeks or more without benefits [24,33].

We considered a large number of end-points to capture the concept of prolonged sickness absence since the definition of prolonged or long-term sickness absence varies in different studies and no consensus exists. End-points predicted by existing prognostic models were also considered. We regarded a cut-off of 90- and 180-days to be clinically meaningful. This agrees with a recent report on sick leave in Nordic countries [5] and previous studies [2,4,34,35]. Furthermore, these end-points are in line with a recent recommendation for measuring work participation, stating that absenteeism should include sick leave duration in days [36]. However, since no consensus exists, we plan to conduct a sensitivity analysis for different end-points (see “Sensitivity analysis” section).

Identification of candidate prognostic factors

When developing a prognostic model, selection of candidate prognostic factors should be guided by evidence from the literature and consultations with content experts [37]. The included studies collected candidate predictors based on existing literature and expert knowledge using sociodemographic and clinical variables. The predictor data from the cohort and the MI-NAV study are nearly identical, but differ slightly from the MI study. For model development, we identified those prognostic factors that are available in all three datasets. Since the aim of our prognostic models is to predict

prolonged sickness absence with as good accuracy as possible, we included both modifiable and non-modifiable prognostic factors.

We identified a set of prognostic factors for the model development based on an iterative approach:

1. **We identified candidate predictors of sickness absence from available literature**: The first author (TR) reviewed all available best-evidence synthesis of systematic reviews [28], systematic reviews [38–43], prognostic model studies [10–15,17,18,35,44–51], and expert opinions from Delphi-studies [52,53]. Additional focused searches on PubMed (https://pubmed.ncbi.nlm.nih.gov) for relevant prognosis studies were also conducted to identify additional evidence for candidate prognostic factors.
2. **Clinical grounds through discussion in a multidisciplinary group**: The candidate prognostic factors from the literature review were shared with the research group, and through discussion we selected a set of prognostic factors to be used for model development, restricted to data availability and sample size calculation (see “Sample size” section).

The prognostic factors that were identified and that will be included in the model are summarised in Table 2. Because we will use data from randomised controlled trials, we must consider the extraneous trial effects which can limit the generalisability of the model validation [54]. Thus, we will model for treatment effect by following the guidance provided by Pajouheshnia et al. [54]. We will code the intervention predictor as a binary variable, to avoid overcomplicating the model.

Table 2. Summary of selected prognostic factors for model development with definition, variable type, unit/categories, measurement method, number of predictor parameters, and evidence.

Prognostic factor	Definition	Variable type	Unit/categories	Measurement method	No. of predictor parameters	Evidence
Expectation of RTW	Expectation of probability of returning to work within 3 months	Continuous	0-10, 10 = best	Self-reported using item from ÖMPSQ-SF [55]	1	[14,17,28,40,52,56–63]
Pain intensity	Pain intensity last week	Continuous	0-10, 10 = worst	Self-reported	1	[17,38,43,52,53,61–65]

Depression/ anxiety	Self-reported depression or anxiety	Categorical	Recategorized from 5 to 3 categories	Self-reported using item from EQ5D [66]	3	[10,28,43,67]
Education level	Completed education	Binary	0 'Lower education: primary/secondary/vocational school' 1 'Higher education: College/University. Recategorized from 4 to 2 categories	Self-reported	1	[10,13,14,18,28,44,45,52,53,59]
Age	Age	Continuous	Years	Self-reported	1	[4,10,14,17,18,28,38–40,43–46,52,53,56,58–63]
Previous sick leave	Days of sick leave previous year	Continuous	Days	Register data	1	[4,10,17,18,28,45,46,52,53,56]
Disability pension status	Receiving work assessment allowance or disability pension at starting-point	Binary	0 'No' 1 'Yes'	Register data	1	[4,10,17,18,28,45,46,52,53,56]
General health	Self-perceived general health	Continuous	0-100	Self-reported using VAS from EQ5D [66]	1	[10,18,40,44,45,52,53,61,62]
Fear-avoidance	Fear that movement or activity will worsen the injury	Binary	0 'No' 1 'Yes'	Self-reported using item from Keele STarT MSK tool [68,69]	1	[39,40,43,53]
Workability (job performance)	Self-reported current workability compared to at its best	Continuous	0-10, 10=best	Self-reported using item from Work Ability Index [35]	1	[18,35,44,46,52,53]
Intervention	Participant in a trial arm with an effective intervention	Binary	0 'No' 1 'Yes'		1	[70]
					Total 13	

Abbreviations: EQ5D, Euroqol questionnaire 5-dimensions; ÖMPSQ-SF, Örebro Musculoskeletal Pain Screening Questionnaire short form; RTW, return to work; SA, sickness absence; VAS, visual analogue scale.

Exploratory prognostic factors

The review and discussion process also resulted in a list of exploratory prognostic factors, that are less well-established in the literature but are considered to be important factors for sickness absence. These prognostic factors will be investigated to see if the model performance improves by adding one or more of these exploratory factors to the model. The scope of this exploratory analysis depends on the sample size calculation and the number of predictor parameters of the full model (see 'Sample size' section). Therefore, this will be investigated outside the planned principal analysis.

Sample size

The flow of study participants for both the development sample and the validation sample is presented in Figure 1.

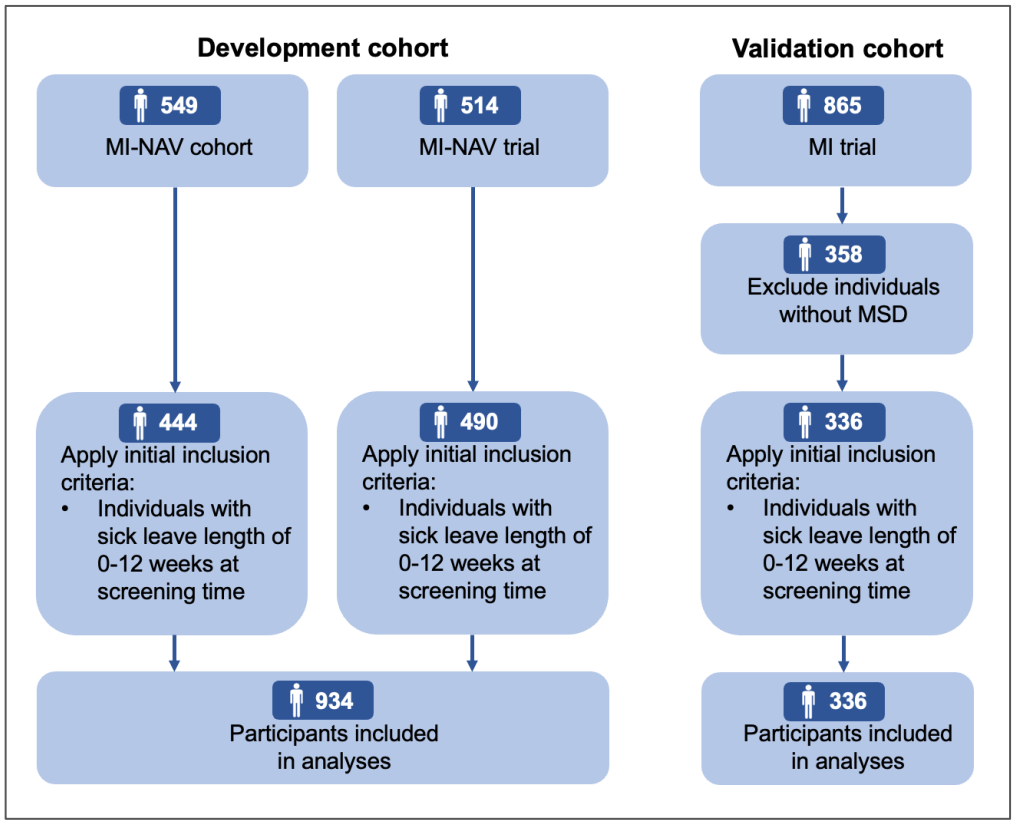

Figure 1. Flow of participants through study.

Development

The sample size is fixed at 934 participants in the development sample. To adequately power the regression analysis and potential exploratory analysis, we will follow recent sample size recommendations by Riley et al. [71,72] using the *pmsampsize* module in STATA 16.1 (StataCorp LLC, Texas, USA). This prediction sample size calculator is also available online: https://riskcalc.org/pmsamplesize/.

In order to include all prognostic factors, corresponding to Table 1 and including 2 parameters for each continuous predictor to account for potential non-linear trends, 19 parameters (*P*) are needed. We will target an expected shrinkage factor (*S*) of 0.9, to

reflect small optimism in predictor effect estimates, as recommended by Riley et al. [72]. We assume that the models will have a modest Nagelkerke's R^2 of 25% which corresponds to a Cox-Snell R^2 of 0.185.

- For the first end-point (>90 days), the event rate is 0.403 (416 events of 934), which gives a minimum required sample size (*n*) of 826 (with 368 events), resulting in an event per predictor parameter (EPP) of 19.35.
- For the second end-point (>180 days), the event rate is 0.172 (161 events of 934), which results in a minimum required sample size (*n*) of 826 (with 143 events) and an EPP of 7.48.
- For the third end-point (WAA/DP), the event rate is 0.108 (101 events of 934), which gives a minimum required sample size (*n*) of 826 (with 90 events) and an EPP of 4.7.

Our fixed sample size exceeds these minimum requirements. Therefore, we expect our study to be adequately powered which will lead to minimal shrinkage and reliable results.

External validation

Our validation cohort has a fixed sample size of 336. To ensure unbiased and precise performance measures (e.g., calibration and discrimination), Collins et al. [73] recommend that external validation samples have a minimum of 100 events. We will adhere to this guideline when interpreting the results from the external validation.

Missing data

We plan to perform a complete case analysis if <5% of the participants are excluded from the regression analyses due to missing data [37]. If >5% are missing, we will use multiple imputation by chained equations. Multiple imputation was chosen as an adequate method, as it avoids removing people from the study and is a suitable technique to deal with missing prognostic factor data [74]. We will create imputed datasets using the *mim* module in STATA. To reduce bias and increase the precision of the imputation, we will include auxiliary variables in the imputation model. Auxiliary variables are variables within the data that are not included in the analysis but are

highly correlated to the prognostic variables of interest [75]. To ensure reliable imputations across the imputed datasets, we will compare observed and imputed values by visual inspection (tables and plots). Using registry data, we do not expect missing data on the primary outcome (sickness absence days). Nevertheless, multiple imputation of the outcome will not be performed.

Statistical analysis methods

Model development

Considering the sample sizes of our data, we will develop the prognostic models based on the 'full model approach' (Figure 2). That is, using the pre-selected prognostic factors based on literature review and clinical grounds without any being excluded at a later stage [37]. This approach has several advantages compared to other selection methods (such as stepwise methods), as it avoids the risk of removing clinically important prognostic factors, reduces estimation bias and overfitting, and avoids the selection process to be overly data-driven [37,76,77].

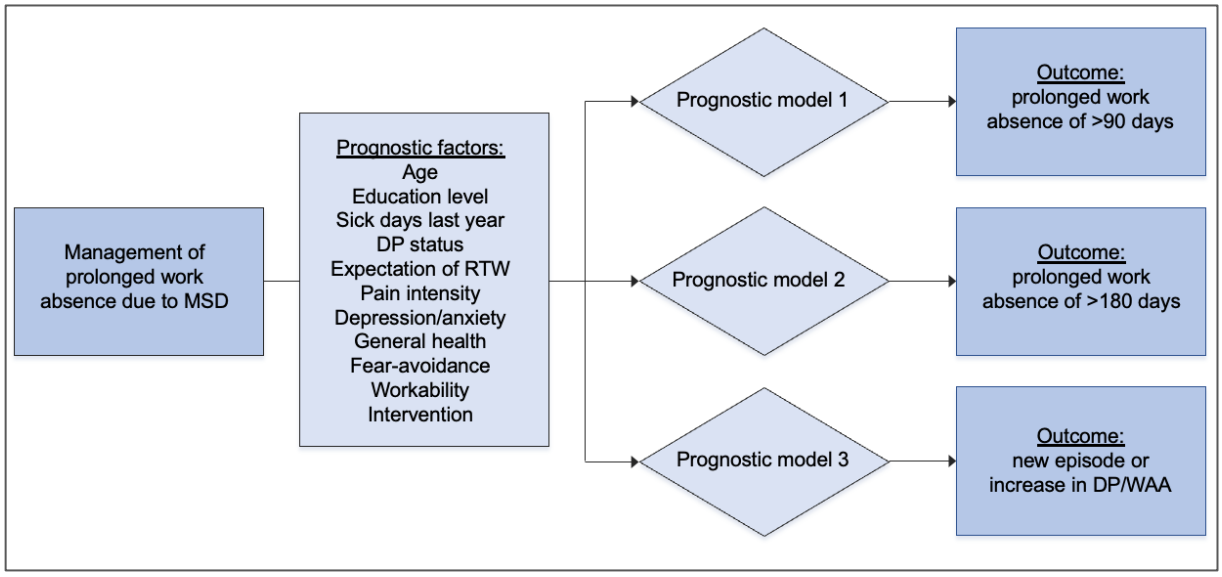

Figure 2. Flow diagram of model development. Abbreviations: DP, disability pension; MSD, musculoskeletal disorder; RTW, return to work; WAA, work assessment allowance.

We will use multivariable logistic regression, since this method is recommended when the outcome is binary, and the independent variables are dichotomous, continuous, categorical or a combination [20,37]. Continuous variables will be kept continuous and not categorised to avoid losing prognostic information [78]. Nonlinear transformations will be checked using multivariable fractional polynomials using STATA package *mfp* (or *mfmpi* for multiple imputation data) [79] or other methods for nonlinear modelling (e.g., splines or segmented regression).

Model performance

The predictive performance of the apparent model in the development sample will be assessed using measures of overall performance, calibration, and discrimination. Overall performance will be assessed using Nagelkerke's pseudo R^2. Calibration will be examined by plotting (calibration plot) the observed proportions of events against the predicted risks for groups defined by ranges of individual predicted risks. To quantify the performance of calibration, the calibration will be described by calibration-in-the-large (intercept) and calibration slope, with perfect calibration characterised by an intercept of 0 and a slope of 1 [37]. We will also assess the ratio of Expected to Observed cases (E/O). To explore the model's ability to discriminate those at higher risk of having an event from those at lower risk, *c*-statistic will be used. The *c*-statistic is equivalent to the area under the curve (AUC) statistic from receiver operating characteristic (ROC) curves. An index measure of 1 indicates perfect discrimination, whilst 0.5 demonstrates that the discrimination of the model is no better than by chance alone [37]. A *c*-statistic of 0.60–0.69 represents poor discrimination, 0.70–0.79 fair, 0.80–0.89 excellent, and ≥0.90 outstanding discrimination [80].

To decrease the possibility of overfitting the model, the amount of optimism in the model will be assessed and corrected by internal validation techniques, using bootstrap re-sampling to estimate the degree of optimism due to overfitting [81]. To adjust for this overfitting, we will use the *bsvalidation* module in STATA or similar commands in STATA or R software. The bootstrap procedure will include all modelling steps for the assessment of the model's performance. The model building will be repeated for 500 bootstrap samples [82,83]. Initially, the difference in bootstrap performance and test performance will be averaged over the 500 samples. This will

give a single estimation of the optimism of the model. Finally, this estimate will be subtracted from the estimates of the apparent model performance.

After the optimism estimate has been subtracted from the model's performance, a new calibration plot will be performed to provide a shrinkage factor. This factor will be applied to all predictor effects in the apparent model to shrink for overfitting. A re-estimation of the intercept of the model will then be performed [21,37]. After these adjustments, the final apparent model is complete.

External validation

We will externally validate the models to determine the validity of predictions for workers on sick leave outside the development sample. We will use data from the MI study, which included participants from a different geographical region than the development sample. Hence, this validation is called external geographic validation. External geographic validation has been found to have higher generalisability than internal validation and is essential before a prognostic model can be applied to practice [37,84].

To assess the extent of transportability or reproducibility, we will estimate the relatedness between the development and validation samples according to Debray et al. [85]. This will be performed using two approaches. In the first approach, we will examine relatedness of the samples by estimating a binary logistic regression model (membership model), to predict the probability that a participant belongs to the development sample ("1") compared with the validation sample ("0"). The independent variables in this model will include the primary outcome (long-term sickness absence), age, sex, education level, and all the predictors from the prognostic model. Thereby, we will assess the discriminative ability of this membership model by quantifying the *c*-statistic. Low values (close to 0.5) indicate that the samples are largely the same and that the case-mix is indistinguishable between the datasets. In the second approach, we will compare the mean and standard deviation of the linear predictor of the prognostic model in the development and validation sample. A large difference in the linear predictor indicates a heterogeneity of the case-mix between the two samples.

Model performance will be tested in the external validation sample using the predictive performance statistics calculated as detailed above in the model performance section. It will be estimated using the model coefficients from the developed models to calculate the individual risk scores [86]. In addition to these performance measures, we will also assess the clinical usefulness of the prognostic models by calculating decision-curve analysis for the model. This will evaluate the ability to make better decisions with a prognostic model than without, by using a decision-curve analysis that estimates the net benefit and compares this to decision-making strategies (treat all/treat none) [82,87].

Sensitivity analysis

We will perform several sensitivity analyses to assess the robustness of the findings based on our primary analyses. The results of the sensitivity analyses will either be published in the main paper or as an additional paper.

1. *Candidate predictors.* To ensure that important predictors are not missed during the development, we plan to conduct analyses by adding other important candidate predictors that are only present in the development sample and not in the validation sample. This is to determine if some of these candidate predictors improve the accuracy of the models.
2. *End-point definitions.* Since no consensus on the definition of long-term sickness absence exists, we will perform sensitivity analysis to explore the differences in model performance between varying end-points, such as >120 days and >270 days.
3. *Sick leave days adjustment*. As graded sick leave only applies in some countries (e.g., Norway and Sweden), we will also conduct sensitivity analyses where we do not adjust sick leave days for sick leave grading or percentage of contracted work hours.
4. *Full-time absenteeism.* Those who are on full-time sick leave have the highest risk of long-term disability benefits [88]. Therefore, we will undertake sensitivity analyses to evaluate the models' predictive ability to predict the end-points (>90 days and >180 days) when we only include full-time absenteeism in the calculation of these end-points.

5. *Predicting a second dialogue meeting*. Within week 26 of the sick leave process, the Norwegian Labour and Welfare Service (NAV) must summon the employee and the employer to a second dialogue meeting. The purpose of this meeting is for the parties to review the situation and create a follow-up plan for the employee's return to work. Dialogue meetings are legally mandated, and all parties are required to participate. Today, it is the NAV caseworker's assessments that form the basis for the decision on whether to hold a second dialogue meeting or not. A recent report from NAV [89] has highlighted that caseworkers use a lot of resources to assess if the second dialogue meeting should be held or not, and that they find this assessment difficult as the information base is scarce. Moreover, NAV must already by week 17 decide if a second dialogue meeting will be necessary, that is, whether the sick employee will be fit to work again within week 26 or not. Therefore, we will also explore if our model can be used to predict the risk of still being on sick leave at week 26.
6. *Inclusion criteria.* Since the duration of sick leave is an important prognosis factor of prolonged sickness absence [28], a sensitivity analysis will be performed to evaluate the predictive performance of the prognostic models after the removal of participants who had been on sick leave for more than 8 weeks at the start-point.
7. *Imputation of missing data.* Depending on how we handle missing data (multiple imputation or complete cases) as described in 'Missing data'-section, we will perform a sensitivity analysis to compare the proposed methods.
8. *External validation.* We also want to evaluate the model's predictive ability in the entire validation sample, that is, to include participants with all types of sick leave causes (e.g., mental disorders).

DISCUSSION

This study will provide externally validated prognostic models for risk of prolonged sickness absence in workers on sick leave due to MSDs. This will be explored using three different outcomes of prolonged work absenteeism over a 1-year follow-up [36,90]. The models will be developed using predictors that are selected on the grounds of the best available evidence and clinical acceptability, which ensures that

the model development is based on existing evidence of importance and validity. This approach ('full-model approach') will also minimise the potential overfitting of the new models [37,76]. To our knowledge, this study will also be the first to assess the clinical usefulness of a prognostic model for predicting prolonged sickness absence. If the models show good predictive performance, we plan to create an online risk calculator that provides risk individual estimates, which automates the complex modelling methods using an interactive graphical interface. Essentially, before we develop a web-based risk calculator, we will consider issues regarding privacy and data storage, as well as transparent reporting of model equations and possible model changes over time [91].

REFERENCES

1 Westerlund H, Ferrie J, Hagberg J, *et al.* Workplace expansion, long-term sickness absence, and hospital admission. *Lancet* 2004;**363**:1193–7. doi:10.1016/S0140-6736(04)15949-7

2 Holm J, Frumento P, Almondo G, *et al.* A prognostic model for predicting the duration of 20, 049 sickness absence spells due to shoulder lesions in a population-based cohort in Sweden. *PLoS One* 2023;**54**:1–13. doi:10.1371/journal.pone.0280048

3 Gémes K, Frumento P, Almondo G, *et al.* A prediction model for duration of sickness absence due to stress-related disorders. *J Affect Disord* 2019;**250**:9–15. doi:10.1016/j.jad.2019.01.045

4 Ropponen A, Gémes K, Frumento P, *et al.* Predicting the duration of sickness absence spells due to back pain: A population-based study from Sweden. *Occup Environ Med* 2020;**77**:115–21. doi:10.1136/oemed-2019-106129

5 Thorsen SV, Friborg C, Lundstrøm B, *et al.* The sickness absence in the Nordic countries. Copenhagen: : Nordic Social Statistical Committee 2015.

6 Henderson M, Glozier N, Holland Elliott K, *et al.* Long term sickness absence. *BMJ* 2005;**330**:802–3. doi:10.1136/bmj.330.7495.802

7 Eurostat. Eurostat. http://epp.eurostat.ec.europa.eu/

8 Linton SJ, Nicholas M, Shaw W. Why wait to address high-risk cases of acute low back pain? A comparison of stepped, stratified, and matched care. *Pain* 2018;**00**:1. doi:10.1097/j.pain.0000000000001308

9 Steyerberg EW, Moons KGM, van der Windt DA, *et al.* Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med* 2013;**10**:e1001381. doi:10.1371/journal.pmed.1001381

10 Airaksinen J, Jokela M, Virtanen M, *et al.* Prediction of long-term absence due to sickness in employees: Development and validation of a multifactorial risk

score in two cohort studies. *Scand J Work Environ Health* 2018;**44**:274–82. doi:10.5271/sjweh.3713
11 Roelen CAM, Heymans MW, Twisk JWR, *et al.* Updating and prospective validation of a prognostic model for high sickness absence. *Int Arch Occup Environ Health* 2015;**88**:113–22. doi:10.1007/s00420-014-0942-9
12 Roelen CAM, Stapelfeldt CM, Heymans MW, *et al.* Cross-National Validation of Prognostic Models Predicting Sickness Absence and the Added Value of Work Environment Variables. *J Occup Rehabil* 2015;**25**:279–87. doi:10.1007/s10926-014-9536-3
13 Bosman LC, Roelen CAM, Twisk JWR, *et al.* Development of Prediction Models for Sick Leave Due to Musculoskeletal Disorders. *J Occup Rehabil* 2019;**29**:617–24. doi:10.1007/s10926-018-09825-y
14 Louwerse I, van Rijssen HJ, Huysmans MA, *et al.* Predicting Long-Term Sickness Absence and Identifying Subgroups Among Individuals Without an Employment Contract. *J Occup Rehabil* 2020;**30**:371–80. doi:10.1007/s10926-020-09874-2
15 Bosman LC, Dijkstra L, Joling CI, *et al.* Prediction models to identify workers at risk of sick leave due to low-back pain in the Dutch construction industry. *Scand J Work Environ Health* 2018;**44**:156–62.
16 Steenstra IA, Busse JW, Tolusso D, *et al.* Predicting Time on Prolonged Benefits for Injured Workers with Acute Back Pain. *J Occup Rehabil* 2015;**25**:267–78. doi:10.1007/s10926-014-9534-5
17 Roelen CA, Speklé EM, Lissenberg-Witte BI, *et al.* Predicting long-term sickness absence among retail workers after four days of sick-listing. *Scand J Work Environ Health* 2022;**44**:67–74. doi:10.5271/sjweh.4041
18 Roelen C, Thorsen S, Heymans M, *et al.* Development and validation of a prediction model for long-term sickness absence based on occupational health survey variables. *Disabil Rehabil* 2018;**40**:168–75. doi:10.1080/09638288.2016.1247471
19 Schouten LS, Joling CI, van der Gulden JWJ, *et al.* Screening manual and office workers for risk of long-term sickness absence: cut-off points for the Work Ability Index. *Scand J Work Environ Health* 2015;**41**:36–42. doi:10.5271/sjweh.3465
20 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med* 2015;**13**:1–10. doi:10.1186/s12916-014-0241-z
21 Moons KGM, Altman DG, Reitsma JB, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015;**162**:W1–73. doi:10.7326/M14-0698
22 Tveter AT, Øiestad BE, Rysstad TL, *et al.* Risk assessment for prolonged sickness absence due to musculoskeletal disorders: Protocol for a prospective cohort study. *BMC Musculoskelet Disord* 2020;**21**:1–8. doi:10.1186/s12891-020-03354-7
23 Øiestad BE, Aanesen F, Løchting I, *et al.* Study protocol for a randomized controlled trial of the effectiveness of adding motivational interviewing or stratified vocational advice intervention to usual case management on return to work for people with musculoskeletal disorders. The MI-NAV study. *BMC Musculoskelet Disord* 2020;**21**:1–9. doi:10.1186/s12891-020-03475-z

24 Aasdahl L, Bagøien G, Standal MI, *et al.* Motivational interviewing in long-term sickness absence: study protocol of a randomized controlled trial followed by qualitative and economic studies. *BMC Public Health* 2018;**18**. doi:10.1186/s12889-018-5686-0
25 Aanesen F, Grotle M, Rysstad TL, *et al.* Effectiveness of adding motivational interviewing or a stratified vocational advice intervention to usual case management on return to work for people with musculoskeletal disorders: the MI-NAV randomised controlled trial. *Occup Environ Med* 2022;:1–9.
26 WONCA Classification Committee, Committee WC. International Classification of Primary Care, Second Edition (ICPC-2). *Oxford University Press* 1998.
27 Øiestad BE, Aanesen F, Løchting I, *et al.* Study protocol for a randomized controlled trial of the effectiveness of adding motivational interviewing or stratified vocational advice intervention to usual case management on return to work for people with musculoskeletal disorders. The MI-NAV study. *BMC Musculoskelet Disord* 2020;**21**:1–10. doi:10.1186/s12891-020-03475-z
28 Cancelliere C, Donovan J, Stochkendahl MJ, *et al.* Factors affecting return to work after injury or illness: Best evidence synthesis of systematic reviews. *Chiropr Man Therap* 2016;**24**:1–23. doi:10.1186/s12998-016-0113-z
29 Waddell G, Burton a K. Is Work Good Well-Being ? Good for Your Health. *The Stationery Office, United Kingdom* Published Online First: 2006. doi:978 0 11 703694 9
30 Schultz IZ, Chlebak CM, Law AK. Bridging the gap: evidence-informed early intervention practices for injured workers with nonvisible disabilities. In: *Handbook of Return to Work*. Springer 2016. 223–53.
31 Aasdahl L, Fimland MS. Is there really a “golden hour” for work disability interventions? a narrative review. *Disabil Rehabil* 2019;**0**:1–8. doi:10.1080/09638288.2018.1503735
32 Wynne-Jones G, Cowen J, Jordan JL, *et al.* Absence from work and return to work in people with back pain: A systematic review and meta-analysis. *Occup Environ Med* 2014;**71**:448–58. doi:10.1136/oemed-2013-101571
33 Aasdahl L, Pape K, Vasseljen O, *et al.* Effect of Inpatient Multicomponent Occupational Rehabilitation Versus Less Comprehensive Outpatient Rehabilitation on Sickness Absence in Persons with Musculoskeletal- or Mental Health Disorders: A Randomized Clinical Trial. *J Occup Rehabil* 2018;**28**:170–9. doi:10.1007/s10926-017-9708-z
34 Truchon M, Schmouth MÈ, Côté D, *et al.* Absenteeism screening questionnaire (ASQ): A new tool for predicting long-term absenteeism among workers with low back pain. *J Occup Rehabil* 2012;**22**:27–50. doi:10.1007/s10926-011-9318-0
35 Lundin A, Leijon O, Vaez M, *et al.* Predictive validity of the Work Ability Index and its individual items in the general population. *Scand J Public Health* 2017;**45**:350–6. doi:10.1177/1403494817702759
36 Ravinskaya M, Verbeek JH, Langendam MW, *et al.* Preferred Methods of Measuring Work Participation: An International Survey Among Trialists and Cochrane Systematic Reviewers. *J Occup Rehabil* Published Online First: 2022. doi:10.1007/s10926-022-10031-0
37 Steyerberg EW. *Clinical prediction models : a practical approach to development, validation, and updating*. Cham, Switzerland: : Springer 2019.
38 Steenstra IA, Verbeek JH, Heymans MW, *et al.* Prognostic factors for duration of sick leave in patients sick listed with acute low back pain: a systematic

review of the literature. *Occup Environ Med* 2005;**62**:851–60. doi:10.1136/oem.2004.015842
39 Steenstra IA, Munhall C, Irvin E, *et al.* Systematic Review of Prognostic Factors for Return to Work in Workers with Sub Acute and Chronic Low Back Pain. *J Occup Rehabil* 2017;**27**:369–81. doi:10.1007/s10926-016-9666-x
40 de Wit M, Wind H, Hulshof CTJ, *et al.* Person-related factors associated with work participation in employees with health problems: a systematic review. *Int Arch Occup Environ Health* 2018;**91**:497–512. doi:10.1007/s00420-018-1308-5
41 Rashid M, Kristofferzon ML, Nilsson A, *et al.* Factors associated with return to work among people on work absence due to long-term neck or back pain: A narrative systematic review. BMJ Open. 2017;**7**. doi:10.1136/bmjopen-2016-014939
42 Gragnano A, Negrini A, Miglioretti M, *et al.* Common Psychosocial Factors Predicting Return to Work After Common Mental Disorders, Cardiovascular Diseases, and Cancers: A Review of Reviews Supporting a Cross-Disease Approach. *J Occup Rehabil* 2018;**28**:215–31. doi:10.1007/s10926-017-9714-1
43 Oosterhuis T, Smaardijk VR, Kuijer PPF, *et al.* Systematic review of prognostic factors for work participation in patients with sciatica. *Occup Environ Med* 2019;**76**:772–9. doi:10.1136/oemed-2019-105797
44 van der Burg LRA, van Kuijk SMJ, ter Wee MM, *et al.* Long-term sickness absence in a working population: development and validation of a risk prediction model in a large Dutch prospective cohort. *BMC Public Health* 2020;**20**:699. doi:10.1186/s12889-020-08843-x
45 Nyberg S, Airaksinen J, Pentti J, *et al.* Predicting work disability among people with chronic conditions : a large prospective cohort study. *medrxiv* 2022.
46 Streibelt M, Bethge M. Prospective Cohort Analysis of the Predictive Validity of a Screening Instrument for Severe Restrictions of Work Ability in Patients with Musculoskeletal Disorders. *Am J Phys Med Rehabil* 2015;**94**:617–26. doi:10.1097/PHM.0000000000000220
47 Nicholas MK, Costa DSJ, Linton SJ, *et al.* Predicting Return to Work in a Heterogeneous Sample of Recently Injured Workers Using the Brief ÖMPSQ-SF. *J Occup Rehabil* 2018;**0**. doi:10.1007/s10926-018-9784-8
48 Roelen CA, Bültmann U, van Rhenen W, *et al.* External validation of two prediction models identifying employees at risk of high sickness absence: Cohort study with 1-year follow-up. *BMC Public Health* 2013;**13**. doi:10.1186/1471-2458-13-105
49 Norder G, Roelen CAM, van der Klink JJL, *et al.* External Validation and Update of a Prediction Rule for the Duration of Sickness Absence Due to Common Mental Disorders. *J Occup Rehabil* 2017;**27**:202–9. doi:10.1007/s10926-016-9646-1
50 van Hoffen MFA, Norder G, Twisk JWR, *et al.* External validation of a prediction model and decision tree for sickness absence due to mental disorders. *Int Arch Occup Environ Health* 2020;**93**:1007–12. doi:10.1007/s00420-020-01548-z
51 van Hoffen MFA, Norder G, Twisk JWR, *et al.* Development of Prediction Models for Sickness Absence Due to Mental Disorders in the General Working Population. *J Occup Rehabil* 2020;**30**:308–17. doi:10.1007/s10926-019-09852-3
52 Dekkers-Sánchez PM, Wind H, Sluiter JK, *et al.* What factors are most relevant to the assessment of work ability of employees on long-term sick leave? the

physicians' perspective. *Int Arch Occup Environ Health* 2013;**86**:509–18. doi:10.1007/s00420-012-0783-3
53 de Geus CJC, Huysmans MA, van Rijssen HJ, *et al.* Return to work factors and vocational rehabilitation interventions for long-term, partially disabled workers: a modified Delphi study among vocational rehabilitation professionals. *BMC Public Health* 2022;**22**:1–11. doi:10.1186/s12889-022-13295-6
54 Pajouheshnia R, Groenwold RHH, Peelen LM, *et al.* When and how to use data from randomised trials to develop or validate prognostic models. *The BMJ* 2019;**365**:1–8. doi:10.1136/bmj.l2154
55 Linton SJ, Nicholas M, Macdonald S. Development of a Short Form of the Örebro Musculoskeletal Pain Screening Questionnaire. *Spine (Phila Pa 1976)* 2011;**36**:1891–5. doi:10.1097/BRS.0b013e3181f8f775
56 Dekkers-Sánchez PM, Hoving JL, Sluiter JK, *et al.* Factors associated with long-term sick leave in sick-listed employees: A systematic review. *Occup Environ Med* 2008;**65**:153–7. doi:10.1136/oem.2007.034983
57 Hayden JA, Wilson MN, Riley RD, *et al.* Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor review. *Cochrane Database of Systematic Reviews* 2019;:1–130. doi:10.1002/14651858.CD011284.pub2
58 Nowrouzi-Kia B, Nadesar N, Sun Y, *et al.* Prevalence and predictors of return to work following a spinal cord injury using a work disability prevention approach: A systematic review and meta-analysis. *Trauma (United Kingdom)* 2022;**24**:14–23. doi:10.1177/14604086211033083
59 Street TD, Lacey SJ. A systematic review of studies identifying predictors of poor return to work outcomes following workplace injury. *Work* 2015;**51**:373–81. doi:10.3233/WOR-141980
60 Kuijer W, Groothoff JW, Brouwer S, *et al.* Prediction of sickness absence in patients with chronic low back pain: A systematic review. *J Occup Rehabil* 2006;**16**:439–67. doi:10.1007/s10926-006-9021-8
61 Lötters F, Burdorf A. Prognostic factors for duration of sickness absence due to musculoskeletal disorders. *Clinical Journal of Pain* 2006;**22**:212–21. doi:10.1097/01.ajp.0000154047.30155.72
62 Lotters F, Franche R-LL, Hogg-Johnson S, *et al.* The prognostic value of depressive symptoms, fear-avoidance, and self-efficacy for duration of lost-time benefits in workers with musculoskeletal disorders. *Occup Environ Med* 2006;**63**:794–801. doi:10.1136/oem.2005.020420
63 Nicholas MK, Costa DSJ, Linton SJ, *et al.* Predicting Return to Work in a Heterogeneous Sample of Recently Injured Workers Using the Brief OMPSQ-SF. *J Occup Rehabil* 2019;**29**:295–302. doi:10.1007/s10926-018-9784-8
64 Kent P, Cancelliere C, Boyle E, *et al.* A conceptual framework for prognostic research. *BMC Med Res Methodol* 2020;**20**:1–13.
65 Valentin GHGH, Pilegaard MSMS, Vaegter HBHB, *et al.* Prognostic factors for disability and sick leave in patients with subacute non-malignant pain: A systematic review of cohort studies. *BMJ Open* 2016;**6**. doi:10.1136/bmjopen-2015-007616
66 EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208. doi:10109801
67 Duijts SF, Kant I, Swaen GM, *et al.* A meta-analysis of observational studies identifies predictors of sickness absence. *J Clin Epidemiol* 2007;**60**:1105–15. doi:10.1016/j.jclinepi.2007.04.008

68 Rysstad T, Grotle M, Aasdahl L, *et al.* Stratifying workers on sick leave due to musculoskeletal pain: translation, cross-cultural adaptation and construct validity of the Norwegian Keele STarT MSK tool. *Scand J Pain* 2022;**22**:325–35. doi:10.1515/sjpain-2021-0144

69 Dunn KM, Campbell P, Lewis M, *et al.* Refinement and validation of a tool for stratifying patients with musculoskeletal pain. *European Journal of Pain* 2021;**25**:2081–93. doi:https://doi.org/10.1002/ejp.1821

70 Pajouheshnia R, Groenwold RHHH, Peelen LM, *et al.* When and how to use data from randomised trials to develop or validate prognostic models. *The BMJ* 2019;**365**:1–8. doi:10.1136/bmj.l2154

71 Riley RD, Ensor J, Snell KIE, *et al.* Calculating the sample size required for developing a clinical prediction model. *The BMJ* 2020;**368**:1–12. doi:10.1136/bmj.m441

72 Riley RD, Snell KI, Ensor J, *et al.* Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019;**38**:1276–96. doi:10.1002/sim.7992

73 Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: A resampling study. *Stat Med* 2016;**35**:214–26. doi:10.1002/sim.6787

74 Marshall A, Altman DG, Royston P, *et al.* Comparison of techniques for handling missing covariate data within prognostic modelling studies: A simulation study. *BMC Med Res Methodol* 2010;**10**. doi:10.1186/1471-2288-10-7

75 Hardt J, Herke M, Leonhart R. Auxiliary variables in multiple imputation in regression with missing X: A warning against including too many in small sample research. *BMC Med Res Methodol* 2012;**12**:1–13. doi:10.1186/1471-2288-12-184

76 Steyerberg EW, Eijkemans MJCC, Harrell Jr FE, *et al.* Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 2000;**19**:1059–79. doi:10.1177/0272989X0102100106

77 Harrell FE. *Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis*. Spinger 2015.

78 Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006;**25**:127–41.

79 Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *J R Stat Soc Ser A Stat Soc* 1999;**162**:71–94.

80 Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. John Wiley & Sons 2013.

81 Riley RD, van der Windt D, Croft P, *et al. Prognosis Research in Healthcare : Concepts, Methods, and Impact*. Oxford: : Oxford University Press 2019.

82 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur Heart J* 2014;**35**:1925–31. doi:10.1093/eurheartj/ehu207

83 Steyerberg EW, Harrell Jr FE, Harrell Jr. FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 2016;**69**:245–7. doi:10.1016/j.jclinepi.2015.04.005.Prediction

84 König IR, Malley JD, Weimar C, *et al.* Practical experiences on the necessity of external validation. *Stat Med* 2007;**26**:5499–511.

85 Debray TPA, Vergouwe Y, Koffijberg H, *et al.* A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015;**68**:279–89. doi:10.1016/j.jclinepi.2014.06.018
86 Moons KGMM, Kengne AP, Grobbee DE, *et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;**98**:691–8. doi:10.1136/heartjnl-2011-301247
87 Traeger AC, Hübscher M, McAuley JH. Understanding the usefulness of prognostic models in clinical decision-making. J Physiother. 2017;**63**:121–5. doi:10.1016/j.jphys.2017.01.003
88 Nossen JP, Brage S. Forløpsanalyse av sykefravær: Når blir folk friskmeldt? *Arbeid og velferd* 2016;**3**:75–100.
89 NAV/Oslo Economics. Bruk av dialogmøte 2 i sykefraværsoppfølgingen. Oslo: 2021.
90 Ravinskaya M, Verbeek JH, Langendam M, *et al.* Extensive variability of work participation outcomes measured in randomized controlled trials: a systematic review. *J Clin Epidemiol* 2022;**142**:60–99. doi:10.1016/j.jclinepi.2021.10.013
91 Bonnett LJ, Snell KIE, Collins GS, *et al.* Guide to presenting clinical prediction models for use in clinical settings. *BMJ* 2019;:l737. doi:10.1136/bmj.l737