Statistical analysis plan (SAP) for:

Modifiable prognostic factors of high costs related to healthcare utilization and productivity loss among people on sick leave due to musculoskeletal disorders (working title)

Project	The MI-NAV project
NCT number	NCT04196634
Document date	25.08.2022

1. Introduction

Scope

This document is a supplement to the MI-NAV protocol (ClinicalTrials.gov Identifier: NCT04196634) and comprises a statistical analysis plan (SAP) for the article "Modifiable prognostic factors of high cost related to healthcare utilization and productivity loss among people on sick leave due to musculoskeletal disorders". The current SAP has been written after data collection was finished. However, the SAP will be uploaded to ClinicalTrials.gov before we enter the study database for the subsequent analyses.

2. Administrative information

Version of SAP

1.0

Study sponsor

This study is part of the MI-NAV project, a large-scale project funded by the Research Council of Norway, through the program "Sickness absence, work, and health" (280431/GE).

Contributors to SAP

Name	Title	Role
Margreth Grotle[1,2]	Professor	Principal Investigator
Rikke Munk Killingmo[1]	PhD-student	Main author of SAP
Mauritz W. van Tulder[3]	Professor	Contributor to SAP
Anne Therese Tveter[4,1]	PhD	Contributor to SAP
Alexander Tingulstad[1]	PhD-student	Contributor to SAP
Tarjei Rysstad[1]	PhD-student	Contributor to SAP

[1]Oslo Metropolitan University, [2]Oslo University Hospital, [3]Vrije University Amsterdam, [4]Diakonhjemmet Hospital

Signatures

I hereby declare that I have reviewed and approved the statistical analysis plan for this study

Rikke Munk Killingmo
Main author of the SAP

Margreth Grotle
Principal Investigator

3. Study aim

The primary aim of this study is to identify modifiable prognostic factors of high costs related to healthcare utilization and productivity loss among people on sick leave due to musculoskeletal disorders. A secondary aim is to identify modifiable prognostic factors of high costs related to separately 1) healthcare utilization, and 2) productivity loss.

4. Method

Design and setting

A prospective observational cohort study with one year of follow-up was conducted within the Norwegian Labour and Welfare Administration (NAV). This cohort study [1] is part of a large-scale project; the MI-NAV project (ClinicalTrials.gov Identifier: NCT04196634).

Study population and recruitment

Eligible participants were people, aged 18 to 67 years, on sick leave (50-100% sick leave rate, ≥ 4 weeks) due to musculoskeletal disorders (diagnosis within the musculoskeletal (L) chapter of the International Classification of Primary Care, 2nd edition (ICPC-2)[2]). Exclusion criteria were insufficient Norwegian or English language skills to participate in the study. Participants were invited through a link on the participants' individual profile page on the NAV website between November 2018 and February 2019. Participants entering the link were asked to consent digitally.

Data collection, outcome, modifiable prognostic factors, and covariates

At baseline, all participants responded to an electronic questionnaire including demographic variables and a set of patient-reported measures. Data on healthcare utilization were collected from public records including the Norwegian Patient Registry (NPR) and the Municipal Patient and User Registry (KPR). Data on productivity loss were collected from public records (NAV), containing dates and grading of absenteeism as well as the related diagnostic code. Data on healthcare utilization and productivity loss were collected in the period from baseline to 3 months retrospectively, and in the one-year follow-up period. All information is stored and will be analysed securely through the Service for sensitive data (TSD) at the University of Oslo, Norway.

Outcomes

The primary outcome of this study is costs related to healthcare utilization and productivity loss aggregated for one year of follow-up and dichotomized as high or low. Having high costs is defined as patients with costs in the top 25th percentile [3, 4]. As noted above, healthcare utilization was collected from public records (NPR, KPR) and included: consultation with healthcare professionals (type and frequency), consultations in emergency room, outpatient activity, number of days of hospitalization and/or institutionalisation, and surgery. Healthcare utilization during the one year of follow-up will be described as shown in Table 2. The total cost of healthcare utilization will be estimated based on reimbursement rates collected from NPR and KPR. Productivity loss was collected from public records (NAV) and included productivity loss related to absenteeism. The total costs of productivity loss will be estimated based on the number of days with absenteeism adjusted for the percentage of absenteeism and multiplied by the average wage rate (from official statistics) in Norway by sex and age groups.

Secondary outcomes of this study are costs related to separately 1) healthcare utilization aggregated for one year of follow up and dichotomized as high and low, and 2) productivity loss aggregated for one year of follow up and dichotomized as high and low.

Modifiable prognostic factors

Potential modifiable prognostic factors are factors expected to have the potential to be modified or improved by appropriate care or treatment, and therefore classified as modifiable. Potential modifiable prognostic factors of high costs related to healthcare utilization and productivity loss are based on previous scientific literature on patients with back pain primarily [3-11], as well as people with musculoskeletal disorders [12-15] and non-diagnosis-specific studies [16-23], and includes the following self-reported variables measured at baseline:

- Pain severity [3-7, 12-15] measured by the numeric rating scale (NRS) [24] from the Keele STarT MSK tool (STarT MSK) [25]
- Disability [3-8, 11, 13-15] measured by a single item (Q4) from the Musculoskeletal Health Questionnaire (MSK-HQ) [26]
- Health-related quality of life [7, 12] measured by the EuroQol 5 dimensions (EQ-5D-5L) [27]
- Emotional well-being [3, 4, 8-10, 14] measured by a single item (Q11) from the MSK-HQ [26]
- Sleep quality [21, 22] measured by a single item (Q4) from the Örebro Musculoskeletal Pain Screening Questionnaire Short Form (ÖMPSQ-SF) [28]
- Health literacy [23] measured by a single item (Q12) from the MSK-HQ [26]
- Future disease expectation measured by a single item (Q6) from the STarT MSK [25]
- Return to work expectancy measured by a single item (Q8) from ÖMPSQ-SF [28]
- Work satisfaction measured by a single item (numeric rating scale, 0=not satisfied, 10=satisfied)

Covariates

Prognostic factor research may vary depending on context (time, place, healthcare setting) and characteristics of the study population. We therefore plan to adjust for potential covariates when evaluating the modifiable prognostic factors. Potential covariates are based on previous scientific literature (as described above), and include the following self-reported variables measured at baseline:

- Sex [5-7, 11, 15-19]
- Age [5, 7, 11, 13, 15, 16, 18, 19]
- Education level [5, 8, 17, 20] measured as the highest education completed, and categorised into low vs. high (university level)
- Pain duration [3, 15] measured by a single item (Q1) from the ÖMPSQ-SF [28]

In addition, the following public records variables will be included as covariates:

- Absenteeism-related diagnosis type at baseline collected from the NAV registry and categorized into “upper/lower limb conditions”, “back and neck conditions”, “joint/inflammatory conditions”, “injury or trauma” or “other MSK conditions”
- Total costs related to healthcare utilization during a period of 3 months prior to inclusion. Healthcare utilization prior to inclusion was collected from public records (NPR, KPR) as described above. Total costs of healthcare utilization will be estimated as described above.
- Total costs related to productivity loss during a period of 3 months prior to inclusion [12, 14]. Productivity loss prior to inclusion was collected from public records (NAV) as described above. Total costs of productivity loss will be estimated as described above.

Other variables

Included participants will also be described with respect to the following baseline characteristics: mother tongue, days of absenteeism one year prior to inclusion, and healthcare utilization prior to inclusion.

Sample size

This study contains secondary analyses embedded in the MI-NAV project. Details on sample size calculation are provided in the MI-NAV protocol [1]. To determine statistical power of this study we used number of events per parameter (EPP) [46-50] and the rule-of-thumb of "10 events per parameter included" [51-54]. With a fixed sample size of 549 participants included in the MI-NAV project, we anticipate 137 participants to be in the top 25th percentile of costs and categorised as having high costs (yes/no) (events). An EPP of 10 will allow a maximum of 13 parameters to be included in the final multivariable prediction model.

Statistical analyses

General analysis considerations

All analyses described in this SAP are considered a priori in that they have been defined in the protocol and/or in this SAP. All post hoc analyses will be identified as such in the article if relevant. All analyses will be carried out using SPSS or Stata and controlled by a senior researcher/statistician. We consider our study as explanatory. Thus, no correction for multiple testing will be performed and p-values < 0.05 will be considered statistically significant. All statistical tests will be two-sided. All confidence intervals will be reported as 95%. Preliminary analyses assessing the influence of missing data and assumptions of normality for continuous variables will be conducted. The assumption of normal distribution will be investigated using histograms and QQ-plots. Normally distributed data will be presented with means and standard deviations (SDs), skewed data with medians and interquartile range (IQR). Categorical data will be reported as counts and percentages.

Description of study flow

The flow of participants through the study will be reported with a flow chart according to the REMARK guidelines [29]. Reasons for dropout will be provided where known. Differences between responders and non-responders will be evaluated.

Missing data

We anticipate few missing values within this study. Information on the primary and secondary outcome will be obtained from public records (NRP, KPR, NAV) where all individuals receiving any form of benefits are registered by their social security number. Furthermore, we anticipate no or minimal missing data for the majority of the potential modifiable prognostic factors and the covariates, as a requirement to answer all questions on key questionnaires was implemented in the electronic baseline questionnaire. Nevertheless, missing value pattern will be visually explored and handled by multiple imputation if relevant (if >5% data is missing).

Participant characteristics

Baseline characteristics of included participants will be presented as shown in Table 1.

Healthcare utilization, productivity loss and cost estimation

Type and frequency of use of different healthcare resources will be calculated for the one-year follow-up period. Healthcare utilization will be presented as shown in Table 2. Costs related to healthcare utilization will be estimated based reimbursement rates collected from NPR and KPR. Days of absenteeism will be calculated for the one-year follow-up period and adjusted for percentage of absenteeism. Cost related to productivity loss will be estimated based number of days with 100% absenteeism and national average wage rates (from official statistics) in Norway by sex and age groups. All costs will be presented in euros (€) 2020 and estimated with both mean and median values with 95%CI, using bias-corrected and accelerated (BCa) bootstrapping as presented in table 3. The BCa will be conducted with a bootstrap sample size of 1000. Cost data are commonly skewed,

thus both mean and median values will be presented to support the result interpretation. Values in Norwegian kroner (NOK) were recalculated to euros using the exchange rate from February 2020 (1€=NOK 10).

Identification analysis

Univariable and multivariable binary logistic regression models will be used to investigate associations (crude and adjusted for selected covariates) between each predefined modifiable prognostic factor and costs related to 1) healthcare utilization and productivity loss, 2) healthcare utilization, and 3) productivity loss. The cost score will be entered into the model as a dependent dichotomous variable (high cost defined as patients with cost in the top 25th percentile, yes/no). Non-linear relationships in the modelling process will be explored using cubic splines or multivariable fraction polynomials, as these are recommended approaches for modelling continuous prognostic factors in prognosis research [30]. The results will be presented as crude and adjusted odds ratios (OR) with 95% confidence intervals (CI) as shown in table 4.

Sensitivity analysis

To assess credibility of the total cost calculation related to the primary analyses, the calculation will be conducted without outliers. Outliers will be identified with simple scatterplots by visual inspection and defined as patients with remarkably high total costs. If multiple imputation on missing data is conducted, the univariable and multivariable logistic regression analyses related to the primary analyses will be performed on complete case data to test credibility of the imputation procedure.

6. Ethics approval

This study is a part of the MI-NAV project [1]. The MI-NAV project (ClinicalTrials.gov Identifier: NCT04196634) has been classified as a quality assessment study by the Norwegian Regional Committee for Medical Research Ethics (reference no. 2018/1326/REK sør-øst A) and approved by the Norwegian Social Science Data Service (reference no. 861249) in 2018.

Table 1. Participants characteristics and clinical status at baseline

	All participants (n=549)	Missing, n (%)
Female, n (%)		
Age in years		
Education at university level, n (%)		
Mother tongue Norwegian, n (%)		
Days of absenteeism last year*		
Diagnosis (ICPC-2)**, n (%)		
Upper limb conditions		
Lower limb conditions		
Neck conditions		
Back conditions		
Joint or inflammatory conditions		
Injurie or trauma		
Other MSK conditions		
Pain severity average last week (NRS, 0-10)		
Pain duration, n (%)		
< 3 months		
3-6 months		
> 6 months		
Disability (MSK-HQ, Q4)		
Health-related QOL (EQ-5D-5L, -0.59-1)		
Emotional well-being (MSK-HQ, Q11)		
Sleep quality (ÖMPSQ-SF, Q4)		
Health literacy (MSK-HQ, Q12)		
Future disease expectation (STarT MSK, Q6)		
Return to work expectancy (ÖMPSQ-SF, Q8)		
Work satisfaction (1-10)		
*Healthcare utilization prior to inclusion****		
Primary care consultation last 3 months, n (%)		
General practitioner		
Physiotherapist		
Chiropractor		
Emergency room		
Secondary care consultation last 3 months, n (%)		
Outpatient activity		
Hospitalization (non-surgery)		
Rehabilitation stay		
Surgery		

EQ-5D-5L indicates EuroQol 5 dimensions; ICPC-2, International Classification of Primary Care 2ed edition; MSK-HQ, Musculoskeletal Health Questionnaire; NRS, Numeric Rating Scale; ÖMPSQ-SF, Örebro Musculoskeletal Pain Screening Questionnaire Short Form; STarT MSK, Keele STarT MSK tool. *collected from the Norwegian Labour and Welfare Administration (NAV) registry, measured as calendar days, and adjusted for percentage of absenteeism. **Absenteeism related diagnoses type at baseline, collected from the NAV registry. ***collected from public records; the Norwegian Patient Registry (NPR), the Municipal Patient and User Registry (KPR), and the Norwegian Control and Payment of Health Refunds Database (KUHR).

Table 2. Healthcare utilization throughout one-year of follow-up

	All participants (n=549)	Missing, n (%)
Primary care		
Patients with primary care consultation, n (%)		
General practitioner		
Physiotherapist		
Chiropractor		
Emergency room		
No primary care consultations		
Numbers of consultations, median (IQR)*		
General practitioner		
Physiotherapist		
Chiropractor		
Emergency room		
Secondary care		
Patients with secondary care consultation, n (%)		
Outpatient activity		
Hospitalization (non-surgery)		
Duration of stay in days, median (IQR)		
Rehabilitation stay		
Duration of stay in days, median (IQR)		
Surgery		
No secondary care consultations		

*Numbers of consultations is calculated on basis of patients who have reported primary care consultations.

Table 3. Costs (€) due to healthcare utilization and productivity loss throughout one year of follow-up (n=)

	Mean (95% CI*)	Median (95% CI*)	Patients with zero cost, n (%)
Healthcare utilization			
Primary care			
Secondary care			
Productivity loss			
Absenteeism			
Total			

*Bias-corrected and accelerated bootstrapping (1000 simulations).

Table 4. Binary logistic regression analyses; individual associations between modifiable prognostic factors and high costs

	High costs related to healthcare utilization and productivity loss		High costs related to healthcare utilization		High costs related to productivity loss	
	Crude OR (95% CI)	Adjusted OR* (95% CI)	Crude OR (95% CI)	Adjusted OR* (95% CI)	Crude OR (95% CI)	Adjusted OR* (95% CI)
Pain severity (NRS)						
Disability (MSK-HQ)						
Health-related QOL (EQ-5D-5L)						
Emotional well-being (MSK-HQ)						
Sleep quality (ÖMPSQ-SF)						
Health literacy (MSK-HQ)						
Future disease expectation (STarT MSK)						
Return to work expectancy (ÖMPSQ-SF)						
Work satisfaction						

EQ-5D-5L indicates EuroQol 5 dimensions; ÖMPSQ-SF, Örebro Musculoskeletal Pain Screening Questionnaire Short Form; NRS, Numeric Rating Scale; STarT MSK, Keele STarT MSK tool; MSK-HQ, Musculoskeletal Health Questionnaire. *Adjusted by sex, age, education level, absenteeism related diagnosis type, pain duration, and costs related to healthcare utilization and productivity loss (absenteeism) prior to inclusion.

References

1. Tveter, A.T., et al., *Risk assessment for prolonged sickness absence due to musculoskeletal disorders: protocol for a prospective cohort study.* BMC Musculoskelet Disord, 2020. **21**(1): p. 326.
2. Classification Committee WONCA, *ICPC-2: International Classification of Primary Care*. 2nd ed. 1998, Oxford: Oxford University Press.
3. Engel, C.C., M. von Korff, and W.J. Katon, *Back pain in primary care: predictors of high health-care costs.* Pain, 1996. **65**(2-3): p. 197-204.
4. Becker, A., et al., *Low back pain in primary care: costs of care and prediction of future health care utilization.* Spine (Phila Pa 1976), 2010. **35**(18): p. 1714-20.
5. Wenig, C.M., et al., *Costs of back pain in Germany.* Eur J Pain, 2009. **13**(3): p. 280-6.
6. Ferreira, M., et al., *Factors defining care-seeking in low back pain - A meta-analysis of population based surveys.* European Journal of Pain, 2010. **14**(7): p. 747.e1-747.e7.
7. Mutubuki, E.N., et al., *Predictive factors of high societal costs among chronic low back pain patients.* Eur J Pain, 2019.
8. Lim, K.L., P. Jacobs, and S. Klarenbach, *A population-based analysis of healthcare utilization of persons with back disorders: results from the Canadian Community Health Survey 2000-2001.* Spine (Phila Pa 1976), 2006. **31**(2): p. 212-8.
9. Stewart, W.F., et al., *Patterns of health care utilization for low back pain.* J Pain Res, 2015. **8**: p. 523-35.
10. Keeley, P., et al., *Psychosocial predictors of health-related quality of life and health service utilisation in people with chronic low back pain.* Pain, 2008. **135**(1-2): p. 142-50.
11. Steenstra, I.A., 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**(12): p. 851-60.
12. Lentz, T.A., et al., *Factors associated with persistently high-cost health care utilization for musculoskeletal pain.* PLoS One, 2019. **14**(11): p. e0225125.
13. Lentz, T.A., J.M. Beneciuk, and S.Z. George, *Prediction of healthcare utilization following an episode of physical therapy for musculoskeletal pain.* Bmc Health Services Research, 2018. **18**.
14. Budtz, C.R., S. Mose, and D.H. Christiansen, *Socio-demographic, clinical and psychological predictors of healthcare utilization among patients with musculoskeletal disorders: a prospective cohort study.* BMC Health Serv Res, 2020. **20**(1): p. 239.
15. Lotters, F. and A. Burdorf, *Prognostic factors for duration of sickness absence due to musculoskeletal disorders.* Clin J Pain, 2006. **22**(2): p. 212-21.
16. Wammes, J.J.G., et al., *Systematic review of high-cost patients' characteristics and healthcare utilisation.* Bmj Open, 2018. **8**(9).
17. Rosella, L.C., et al., *Predicting High Health Care Resource Utilization in a Single-payer Public Health Care System: Development and Validation of the High Resource User Population Risk Tool.* Med Care, 2018. **56**(10): p. e61-e69.
18. Rattay, P., et al., *[Utilization of outpatient and inpatient health services in Germany: results of the German Health Interview and Examination Survey for Adults (DEGS1)].* Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz, 2013. **56**(5-6): p. 832-44.
19. Chechulin, Y., et al., *Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada).* Healthc Policy, 2014. **9**(3): p. 68-79.
20. Hoebel, J., et al., *Socioeconomic Status and Use of Outpatient Medical Care: The Case of Germany.* PLoS One, 2016. **11**(5): p. e0155982.
21. Ensrud, K.E., et al., *Multidimensional sleep health and subsequent health-care costs and utilization in older women.* Sleep, 2020. **43**(2).
22. Kaufmann, C.N., et al., *Insomnia and health services utilization in middle-aged and older adults: results from the Health and Retirement Study.* J Gerontol A Biol Sci Med Sci, 2013. **68**(12): p. 1512-7.
23. Rasu, R.S., et al., *Health Literacy Impact on National Healthcare Utilization and Expenditure.* Int J Health Policy Manag, 2015. **4**(11): p. 747-55.
24. Von Korff, M., M.P. Jensen, and P. Karoly, *Assessing global pain severity by self-report in clinical and health services research.* Spine (Phila Pa 1976), 2000. **25**(24): p. 3140-51.
25. Hill, J.C., et al., *Does a modified STarT Back Tool predict outcome with a broader group of musculoskeletal patients than back pain? A secondary analysis of cohort data.* BMJ Open, 2016. **6**(10): p. e012445.
26. Hill, J.C., et al., *Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways.* BMJ Open, 2016. **6**(8): p. e012331.
27. Herdman, M., et al., *Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L).* Qual Life Res, 2011. **20**(10): p. 1727-36.
28. Linton, S.J., M. Nicholas, and S. MacDonald, *Development of a short form of the Orebro Musculoskeletal Pain Screening Questionnaire.* Spine (Phila Pa 1976), 2011. **36**(22): p. 1891-5.
29. McShane, L.M., et al., *REporting recommendations for tumour MARKer prognostic studies (REMARK).* Br J Cancer, 2005. **93**(4): p. 387-91.
30. Riley, R.D., et al., *Prognosis Research in Healthcare. Concepts, Methods, and Impact*. First ed. 2019, Oxford: Oxford United Press.