NIPH
Norwegian Institute of Public Health

REPORT

2018

SYSTEMATIC MAPPING REVIEW:

Effects of partial sick leave versus full-time sick leave on sickness absence and work participation

Published by	Folkehelseinstituttet Område for helsetjenester
Norwegian title	Effekt av gradert sykmelding vs. full sykmelding på sykefravær og arbeidstilknytning: en systematisk kartleggingsoversikt
English title	Effects of partial sick leave versus full-time sick leave on sickness absence and work participation: a systematic mapping review
Responsible	Camilla Stoltenberg, director
Authors	Jose F. Meneses-Echavez, *project leader*, *researcher* Nikita, Baiju, *researcher* Rigmor C. Berg, *department director*
ISBN	978-82-8082-917-7
Type of report	Systematic mapping review
No. of pages	56 (77 inklusiv vedlegg)
Comissioner	NAV
Subject heading (MeSH)	Sick leave, return to work, disability evaluation
Citation	Meneses-Echavez JF, Baiju N, Berg RC. Effects of partial sick leave versus full-time sick leave on sickness absence and work participation: a systematic mapping review. Report –2018. Oslo: Folkehelseinstituttet, 2018.

Content

CONTENT 3

HOVEDBUDSKAP 5

SAMMENDRAG 6

KEY MESSAGES 9

EXECUTIVE SUMMARY (ENGLISH) 10

PREFACE 13

ABBREVIATIONS 14

BACKGROUND 15

METHODS 21
What is a systematic mapping review? 21
Selection criteria (identifying the research question) 22
Literature search (identifying relevant studies) 23
Study selection 24
Data extraction (charting the data) 24
Quality appraisal of the included studies 24
Collating and summarizing the results 25

RESULTS 26
Search results 26
Description of included studies 27
Quality appraisal of the included studies 35
Summary of main findings from the included studies 35

DISCUSSION 45
Main findings 45
Generalizability and strength of findings 46
Comparison with other reviews 48
Strengths and weaknesses 48
Implications for practice 49

CONCLUSION 50

REFERENCES 51

APPENDIX **57**
Appendix 1. Glossary 57
Appendix 2. Search strategies 58
Appendix 3. Excluded studies 61
Appendix 4. Adjusted analyses and covariates in the registry-based studies 62
Appendix 5. Definition of the statistical analyses in the registry-based studies 73
Appendix 6. Quality appraisal of the randomized controlled trial 75
Appendix 7. Quality appraisal of the registry-based studies 76

Hovedbudskap

Høyt sykefravær blant arbeidstakere er en bekymring i mange land. Gradert sykmelding er en kombinasjon av arbeid og sykepenger som benyttes når arbeidstakeren er delvis arbeidsufør, slik at den ansatte kan være fraværende fra jobb en del av tiden og jobbe en del av tiden. Folkehelseinstituttet fikk i oppdrag av NAV å utføre en kartlegging av den empiriske forskningen om effekten av gradert sykmelding versus full sykmelding.

Metode

Vi utførte en systematisk kartleggingsoversikt. I januar 2018 gjorde vi et omfattende litteratursøk, inkludert søk i store databaser, referanselister, grå litteratur, og vi kontaktet arbeidslivsorganisasjoner og departement. To forskere vurderte uavhengig av hverandre alle identifiserte referanser og den metodisk kvaliteten til de inkluderte studiene. Vi hentet ut data fra de inkluderte studiene og utførte beskrivende analyser. Syntese av individuelle studieresultater inngår ikke i en systematisk kartleggingsoversikt.

Resultater

Vi inkluderte en randomisert kontrollert studie og 12 registerbaserte studier. De 13 studiene inkluderte ca. 2,74 millioner sykmeldte. Studiene hadde følgende kjennetegn:

- Elleve av studiene var fra nordiske land, inkludert fire fra Norge.
- Alle studiene hadde enten moderat eller høy metodisk kvalitet.
- Den randomiserte kontrollerte studien inkluderte finske ansatte (n=62) som var sykmeldt på grunn av muskel- og skjelettplager, mens de registerbaserte studiene hovedsakelig inkluderte ansatte med muskel- og skjelettplager eller psykiske lidelser.
- Det var 15 utfall, hvorav hovedutfallene var arbeidsdeltakelse, varighet av sykefravær, arbeidsførhet og sosiale stønader.

Funnene tyder på at gradert sykmelding er forbundet med flere positive utfall, slik som høyere arbeidsdeltakelse og kortere sykefravær, men sikre konklusjoner om effektene av gradert sykmelding er begrenset på grunn av det store flertallet av observasjonsstudier på dette temaet.

Tittel:
Effekt av gradert sykmelding vs. full sykmelding på sykefravær og arbeidstilknytning: en systematisk kartleggingsoversikt

---

Publikasjonstype:
Systematisk Kartleggingsoversikt

En systematisk karleggingsoversikt kartlegger og kategoriserer eksisterende forskning på et tema og identifiserer forskningshull som kan lede til videre forskning.

Svarer ikke på alt:
- Ingen syntese av resultater
- Ingen vurdering av evidensen

Hvem står bak denne publikasjonen?
Folkehelseinstituttet har gjennomført oppdraget etter forespørsel fra NAV

Når ble litteratursøket utført?
Søk etter studier ble avsluttet januar 2018.

Sammendrag

Bakgrunn

Gitt det høye sykefravær blant arbeidstakere er det i mange land en prioritet å øke arbeidsdeltakelsen blant personer i arbeidsfør alder. Gradert sykmelding er en kombinasjon av arbeid og sykepenger som benyttes når arbeidstakeren er delvis arbeidsufør, slik at den ansatte kan være fraværende fra jobb en del av tiden og jobbe en del av tiden. Det forstås som et gradert fravær, som gjør at folk med redusert arbeidskraft kan jobbe deltid og fortsatt beholde tilknytningen til arbeidsmarkedet. Gradert sykmelding varierer mellom 20 % og 99 %, og kan muliggjøre raskere retur til arbeid. I mange land, spesielt de nordiske landene, er dette tiltaket sett som det første alternativet for å takle økende sykefravær og forhindre ekskludering fra arbeidsmarkedet.

Forskning på gradert sykmelding har hovedsakelig vist lovende resultater, spesielt i de nordiske landene. En nylig uført evaluering av ordningen med gradert sykmelding i Norge identifiserte imidlertid barrierer når det gjelder bruken av ordningen, slik som omfattende byråkrati, vanskelig å forstå- og tilgjengelig tilgang til informasjon samt mangel på kvalitetskontroll. Selv om lovende resultater er publisert har forskningen om effekten av gradert sykmelding blitt kritisert av metodiske grunner, som for eksempel risiko for utvalgsskjevhet og svakheter ved måling av arbeidsdeltakelse. Per i dag fins det ingen systematiske analyser av kunnskapsgrunnlaget for effektene av gradert sykmelding.

Metode

Vi gjennomførte en systematisk kartleggingsoversikt i henhold til internasjonale standarder. En systematisk kartleggingsoversikt (også kjent som *systematic scoping review*) er en kunnskapsoppsummering som kartlegger og beskriver eksisterende litteratur- eller forskningsgrunnlag på et bestemt tema. Oppsummeringen inkluderer ingen systematisk syntese av individuelle studieresultater (f.eks. ingen metaanalyse). Vi inkluderte studier uansett språk som evaluerte effektene av gradert- versus full sykmelding på sykefravær og arbeidsdeltakelse. Relevante studiedesign var prospektive kontrollerte studier samt registerbaserte studier (RB). Prosjektteamet (forskerne) og oppdragsgiveren (NAV) diskuterte og ble enige om prosjektplanen.

Vi søkte i åtte store litteraturdatabaser (fra begynnelsen inntil 2018), kontaktet eksperter, og søkte websidene til arbeidslivsorganisasjoner og relevante departement, referanselistene til alle inkluderte studier og litteraturoversikter på temaet. To forskere vurderte uavhengig av hverandre de identifiserte referansene og datauthentingen ble

dobbeltsjekket. Vi gjennomførte uavhengig kvalitetsvurdering av de inkluderte studiene med bruk av validerte sjekklister. Vi grupperte data hentet fra de inkluderte studiene i henhold til deres hovedkarakteristika, utførte beskrivende analyser og presenterte resultatene i tekst og tabeller.

Resultat

Tretten studier, publisert mellom 2010-2017, møtte inklusjonskriteriene. Alle studiene målte effekter av gradert sykmelding sammenlignet med full sykmelding blant voksne sykmeldte. Vi inkluderte en finsk randomisert kontrollert studie (RCT) (n=62 som var sykmeldt på grunn av muskel- og skjelettplager) og 12 RBer (n=2,742,497 som var sykmeldt på grunn av hovedsakelig muskel- og skjelettplager eller psykiske lidelser). Deltakerne i de tolv RBene var fra Norge, Danmark, Finland, Sverige og Tyskland.

Det var ulike grader av gradert sykmelding i studiene. I den finske RCTen ble 50 % gradert sykmelding gitt til 70 % av de sykmeldte, mens 30 % av de sykmeldte arbeidet kortere timer 3-4 dager i uken. I RBene var 50 % den hyppigst brukte graderingen. RBene analyserte landsdekkende registerdata på sykefravær (både gradert- og full sykmelding) mellom 2001 og 2014. Med hensyn til studienes metodiske kvalitet hadde den inkluderte RCTen moderat kvalitet; det hadde også 11 av de 12 RBene mens en RB hadde høy metodisk kvalitet. Det er viktig å fremheve at registerbaserte studier er dårlig egnet til å påvise kausale sammenhenger.

Sammendrag av hovedfunn fra de inkluderte studiene

Sykefravær: Den finske RCTen og tre RBer rapporterte positive resultater av gradert sykmelding, sammenlignet med full sykmelding, på sykefravær ved ett års oppfølging.

Arbeidsdeltakelse
Return-to-work (RTW): Den finske RCTen fant at gradert sykmelding forbedret arbeidsdeltakelsen sammenlignet med full sykmelding ved ett års oppfølging. Alle de ni RBene som målte arbeidsdeltakelse, bortsett fra en norsk RB, viste bedre resultater blant de med gradert sykmelding sammenlignet med full sykmelding.

Arbeidsledighet: Tre RBer – fra Norge, Tyskland og Finland – rapporterte lavere arbeidsledighet blant personer som hadde vært på gradert sykmelding sammenlignet med personer som hadde vært fulltidssykmeldt.

Grad av uførhet og attføring
Gjentakende sykefravær: Den finske RCTen fant ingen signifikante forskjeller mellom gruppen som hadde gradert sykmelding og de som hadde full sykmelding når det gjelder gjentakende sykefravær.

Funksjonsnivå: Ingen forskjeller mellom gradert- og full sykmelding ble rapportert i den finske RCTen ved ett års oppfølging og i én RB fra Norge, mens to andre RBer (fra Norge og Finland) fant at gradert sykmelding var forbundet med bedringer i funksjonsnivå.

Produktivitetstap: Kun den finske RCTen rapporterte på utfallet produktivitetstap. Data fra denne RCTen viste at det ikke var signifikant forskjell på produktivitetstap mellom gradert- og full sykmelding ved ett års oppfølging.

Uførepensjon: En norsk RB viste at gradert sykmelding var forbundet med en høyere grad av uførepensjon sammenlignet med full sykmelding. De to tyske RBene rapporterte en redusert risiko for å motta uførepensjon hos ansatte på gradert sykmelding. En finsk RB fant at gradert sykmelding var forbundet med lavere risiko for full uførepensjon sammenlignet med full sykmelding, mens motsatt sammenheng ble funnet for risiko for delvis uførepensjon.

Sosiale stønader: Fire RBer fant at gradert sykmelding var forbundet med en lavere grad av sosiale stønader sammenlignet med full sykmelding.

Helserelaterte utfall

Bare to av de inkluderte studiene rapporterte på helserelaterte utfall. Den finske RCTen fant ingen forskjeller mellom gradert og full sykmelding på smerteintensitet, men viste positive resultater for gradert sykmelding på både selvrapportert generell helse og helserelatert livskvalitet. En tysk RB viste at personer med gradert sykmelding fungerte bedre fysisk og følelsesmessig enn de som hadde full sykmelding. Den finske RCTen fant ingen forskjeller mellom gradert og full sykmelding når det gjaldt de sykmeldtes depresjonssymptomer, mens den tyske RBen viste at gradert sykmelding var forbundet med forbedringer med hensyn til depresjon- og angstsymptomer og arbeidsevne, sammenlignet med full sykmelding.

Konklusjon

Forskningsgrunnlaget for effekten av gradert sykmelding sammenlignet med full sykmelding består av én RCT og 12 RBer, med totalt ca 2,74 millioner personer som er sykmeldt på grunn av hovedsakelig muskel- og skjelettplager eller mentale lidelser. De inkluderte studiene viste sammenfallende mønster i resultat i favør av gradert sykmelding. Både RCTen og RBene indikerte at gradert sykmelding er forbundet med kortere sykmelding og høyere arbeidsdeltakelse. Resultatene fra den finske RCTen indikerte at ansatte med muskel- og skjelettsykdommer rapporterte bedre generell helse og livskvalitet med gradert sykmelding. Denne RCTen fant at gradert sykmelding ikke har noen effekt på gjentakende sykefravær, produktivitetstap eller smerte. Positive sammenhenger mellom gradert sykmelding og forbedring av ansattes funksjonshemming og depressive symptomer ble støttet av RBene, men ikke av RCTen. Resultater fra RBene antyder at bruk av gradert sykmelding er forbundet med lavere sannsynlighet for å motta både uførepensjon og sosiale stønader, samt bedre skåre på fysisk- og følelsesmessig fungering, angst og arbeidsevne.

Sikre konklusjoner om effektene av gradert sykmelding er begrenset på grunn av det store flertallet av RBer på dette temaet. Observasjonelle studier som RBer gir en betydelig risiko for systematiske skjevheter i resultatene som gjør det vanskelig å gi sikre svar på spørsmål om årsak og virkning. Flere RCTer av høy metodisk kvalitet er nødvendig for å kunne trekke klare konklusjoner.

Key messages

In many countries, the high sickness absence rate in working age people is a concern. Partial sick leave (PTSL) is a return-to-work strategy that enables employees to be absent from work part of the time and remain working for a proportion of the time. The Norwegian Labor and Welfare Administration (NAV) commissioned the Norwegian Institute of Public Health to map all evidence on the effects of PTSL versus full-time sick leave (FTSL) on sickness absence and work participation.

Methods

We conducted a systematic mapping review. In January 2018, we conducted an extensive literature search, including searches in major databases, reference lists, grey literature, and we contacted labor agencies and international ministries. Two independent reviewers screened all retrieved records and appraised the included studies. We extracted data from the included studies and performed descriptive analyses. Synthesis of individual study results is not part of systematic mapping reviews.

Results

We included one small randomized controlled trial and 12 registry-based studies. The 13 studies included about 2.74 million employees on sick leave. The studies exhibited the following characteristics:

- Eleven of the studies were from Nordic countries, including four from Norway.
- All studies had either moderate or high methodological quality.
- The randomized controlled trial included Finnish employees (n=62) who were sick-listed due to musculoskeletal disorders, while the registry-based studies mostly included employees with either musculoskeletal- or mental disorders.
- There were 15 outcomes, of which the most frequently reported outcomes were work participation, sickness absence duration, disability, and social welfare benefits.

The findings indicated that PTSL may be associated with several favorable outcomes such as shorter sickness absence and higher work participation. However, firm conclusions about the effects of PTSL cannot be drawn due to the overwhelming majority of observational studies in this body of evidence.

Title:
Effects of partial sick leave versus full-time sick leave on sickness absence and work participation: a systematic mapping review

---

Type of publication:
Systematic mapping review
A systematic mapping review maps out and categorizes existing research on a topic, identifying research gaps that can guide future research.

Doesn't answer everything:
No synthesis of the results
No recommendations are made

Publisher:
Norwegian Institute of Public Health

Updated:
Last search for studies:
January 2018.

Executive summary (English)

Background

In many countries, high sickness absence rates mean that the need to increase work participation of working age people is a critical priority. Partial sick leave (PTSL), also called graded sick leave, is a return-to-work strategy that enables employees to be absent from work part of the time and remain working for a proportion of the time. It is understood as a graded leave, which allows people with reduced workability to work part time and still keep the link to the labor market. PTSL varies between 20% up to 99%, and might facilitate a progressive return to work. In many countries, especially the Nordic countries, this return-to-work strategy has been considered the first option to tackle increasing sickness absence rates and to prevent labor market exclusion.

Research on PTSL has shown largely promising results, especially in the Nordic countries. A recent evaluation of the PTSL arrangement in Norway, however, highlighted barriers to its use, such as bureaucratic red tape, lack of easy to understand and access information, and quality control. Although promising results have been published, research on the effects of PTSL has been criticized on methodological grounds, such as risk of selection bias and weaknesses in measuring work participation. To date, no systematic analysis of the evidence base on the effects of PTSL has been undertaken.

Objective

The Norwegian Labor and Welfare Administration (NAV) commissioned the Norwegian Institute of Public Health to map all evidence on the effects of partial sick leave (PLSL) versus full-time sick leave (FTSL) on sickness absence and work participation.

Method

We conducted a systematic mapping review according to international standards. A systematic mapping review (also known as systematic scoping review) is a review that maps and describes the existing literature or evidence base on a particular topic. There is no systematic synthesis of individual study results (e.g., no meta-analysis). We included studies in any language evaluating the effects of PTSL versus FTSL on sickness absence and work participation. Eligible study designs were prospective controlled studies as well as registry-based studies (RBs). The project team (reviewers) and commissioner (NAV) discussed and agreed on the research protocol.

We searched eight major databases, from inception to 2018, contacted experts, and hand searched websites of labor organizations and ministries, the bibliographies of all included studies, and literature reviews in the field. Two independent reviewers

screened the retrieved references and data extraction was double-checked. We conducted independent quality appraisal of the included studies by using validated checklists. We grouped data extracted from the included studies according to their chief characteristics, performed descriptive analyses, and presented the results in text and tables.

Results

Thirteen studies, published between 2010-2017, met our inclusion criteria. All evaluated the effects of PTSL compared to FTSL in sick-listed employees. We included one Finnish randomized controlled trial (RCT) (n=62 sick-listed employees due to musculoskeletal disorders), and 12 RBs (n=2,742,497 sick-listed employees due to mainly musculoskeletal- or mental disorders). The participants in the 12 RBs were from Norway, Denmark, Finland, Sweden, and Germany.

Different grades of PTSL were observed across the studies. In the Finnish RCT, a 50% PTSL was given to 70% of all sick-listed employees, whereas 30% of employees worked shorter hours on 3–4 days per week. In the RBs, the most commonly used PTSL was 50%. The RBs analyzed nationwide registry data on sickness absences (both PTSL and FTSL) granted between 2001 and 2014, which indicated a 13-years' timeline. Regarding the studies' methodological quality, the included RCT had moderate quality, as did 11 of the 12 RBs while one RB had high methodological quality. It is important to note that RBs do not enable researchers to establish causal relationships between an intervention or exposure and outcomes.

Summary of main findings from the included studies

Sickness absence: The Finnish RCT and three RBs reported positive results of PTSL compared to FTSL on sickness absence at one-year follow-up.

Work participation:

Return-to-work: The Finnish RCT found that PTSL improved work participation compared to FTSL at the end of the one-year follow-up period. All of the nine RBs that measured work participation, except one Norwegian RB, showed favorable associations in employees on PTSL compared to FTSL.

Unemployment: Three RBs – from Norway, Germany and Finland – reported favorable effects of PTSL compared to FTSL on employees' unemployment.

Degree of disability and rehabilitation benefits:

Recurrence of sick leave for any cause: The Finnish RCT found no significant differences between PTSL and FTSL in the recurrence of sick leave for any cause.

Disability: No differences between PTSL and FTSL were reported by the Finnish RCT at one-year follow-up and in one Norwegian RB, whereas two other RBs (from Norway and Finland) found that PTSL was associated with improvements on employees' disability.

Productivity loss: Only the Finnish RCT reported on productivity loss. Data from this RCT showed there was no significant difference on productivity loss between PTSL and FTSL up to one-year follow-up.

Disability pension: One Norwegian RB reported that PTSL was associated with a higher rate of receiving disability pension compared to FTSL, whereas the two German RBs reported a decreased risk of receiving disability pension in employees on PTSL. A Finnish RB found that PTSL was associated with a lower risk of full disability pension compared to FTSL, whereas the opposite association was found for the risk of partial disability pension.

Allowance of social welfare benefits: Four RBs observed that PTSL was associated with a lower allowance of social welfare benefits compared to FTSL.

Health-related outcomes

Only two of the included studies reported on health-related outcomes. The Finnish RCT found no differences between PTSL and FTSL on pain intensity, but showed positive results for PTSL on both self-rated general health and health-related quality of life. A German RB showed that PTSL was associated with better physical and emotional functioning in sick-listed employees. The Finnish RCT found no difference between PTSL and FTSL on sick-listed employees' depressive symptoms, while the German RB showed that PTSL was associated with improvements on both depression and anxiety symptoms, and working ability, when compared to FTSL.

Conclusion

The evidence on the effects of PTSL compared to FTSL consists of one small RCT and 12 RBs, with a total of about 2.74 million study participants with mostly musculoskeletal- or mental disorders. The findings indicate PTSL may be associated with several favorable outcomes, such as higher work participation, but due to the overwhelming majority of observational studies in this body of evidence, firm conclusions about the effects of PTSL cannot be drawn. Both study designs suggested PTSL may be associated with shorter sickness absence duration and higher work participation. The Finnish RCT reported that employees with PTSL experienced better general health and quality of life compared to those on FTSL. However, it did not find statistical differences between PTSL and FTSL on sick leave recurrence, employees' productivity loss, and pain. The RBs indicated a lower probability for people on PTSL of receiving both disability pension and allowance benefits, disability, as well as better scores on physical- and emotional functioning, anxiety, depression, and working ability.

Firm conclusions about the effects of PTSL are constrained due to the overwhelming majority of RBs in this body of evidence. Observational designs have considerable risk of systematic differences in the results, which make it difficult to answer questions about cause and effect. Further high quality RCTs are necessary in order to draw firm conclusions.

Preface

The Norwegian Labor and Welfare Administration (in Norwegian: Arbeids- og velferdsetaten, NAV) commissioned the Norwegian Institute of Public Health (NIPH) to map out all evidence on the effects of partial sick leave compared to full-time sick leave on sickness absence and work participation. The commission is part of the ongoing framework agreement between NIPH and NAV.

This systematic mapping review can help to inform and support NAV, other authorities as well as other research organizations in evidence-informed deliberations about the use of partial sick leave for sick-listed employees.

The Division for health services within the Norwegian Institute of Public Health follows a standard approach in conducting systematic mapping reviews, which is described in the NIPH handbook *Slik oppsummerer vi forskning*. We may use standard formulations when we describe the methods, results and discussion of the findings.

Contributors to the project:

- Project coordinator: Jose Meneses-Echavez, *researcher*, NIPH
- Other contributors: Rigmor C Berg, *Department director*; Nikita Baiju, *researcher*, and research librarian Elisabet Hafstad, all from NIPH.

Declared conflicts of interest:
All authors filled out a form to document potential conflicts of interest. No conflicts of interest were declared.

The authors would like to thank Elisabet Hafstad, research librarian at NIPH, for help with developing the search strategy and for running the searches in the major databases, as well as Heid Nøkleby and Doris Tove Kristoffersen, both senior researchers with the NIPH, for their assistance with checking data accuracy and guidance on statistical analyses in the primary studies. We are grateful for expert peer review by professors Arne Mastekaasa, University of Oslo, and Pia Svedberg, Karolinska Institute.

NIPH is responsible for the contents of the review presented in this report. The commissioner of the review and peer reviewers bear no responsibility.

Kåre Birger Hagen	Rigmor C Berg	Jose F. Meneses-Echavez
Research director	*Department director*	*Project leader*

Abbreviations

CBA, Controlled before-after study
CI, Confidence interval
Cochrane EPOC group, The Effective Practice and Organisation of Care Group
FRW, Faster return-to-work
FTSL, Full-time sick leave
GNI, Gross national income
GP, General practitioner
HR, Hazard ratio
IA-avtalen, Inkluderende arbeidsliv avtalen
ITS, Interrupted time series
MD, Mental disorder
MSD, Musculoskeletal disorder
NAV, the Norwegian Labor and Welfare Administration
NIPH, Norwegian Institute of Public Health
NRCT, Non-randomized controlled study / Quasi-randomized study
OECD, Organisation for Economic Co-operation and Development
OR, Odds ratio
PICO, Population, Intervention, Comparison, Outcome
PRISMA-ScR, Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
PTSL, Part-time sick leave
RB, Registry-based study
RCT, Randomized controlled trial
RR, Risk ratio
RTW, Return-to-work
SINTEF, Stiftelsen for industriell og teknisk forskning
SLS, Semiparametric Least Squares analysis

Background

Sickness absence and the need to increase work participation in working age people are critical priorities in developed countries (1). Not only does a high sickness absence rate have substantial cost implications for employers and social security systems, sickness absence affects workers' well-being, self-esteem and livelihood (2). Long-term sickness absence rates are high in many OECD countries (2). Norway exhibits a comparatively high rate of sickness absence (3), which has remained almost unchanged for the past five years (4;5). The most recent data published by Statistics Norway (in Norwegian*: Statistisk sentralbyrå*) indicate the sickness absence rate in the country was 6.5% adjusted for seasonal and influenza variations (self- and doctor certified) in the 3rd quarter of 2017. This rate is considerably higher in females (8.3%) than in males (4.9%) (6).

The Nordic countries have similar social security systems, and in general terms, they are relatively similar in their data on sickness absence. However, comparisons across these countries are constrained due to differences in the registration methods in the national registers. For example, Sweden does not register sickness absences below 15 days (7). Overall, however, data from the Nordic Social Statistical Committee 2015 indicate that Norway and Sweden have high rates of long-term sickness absence, whereas short-term sickness absence (i.e. sickness absence of less than 8 days) is high in Denmark and low in Norway (7). Further, the report states "women have more sickness absence than men. Older employees have more long-term sickness absence than younger employees. Younger employees have more short-term sickness absence than older employees" (7).

To gather the most up-to-date data on sickness absence in the five Nordic countries, we consulted national ministries and relevant organizations (table 1). Briefly, Norway exhibits the highest rate of sickness absence in the region. Norway and Denmark pay the highest percentage of salary to their sick-listed employees (percentage of sickness benefit compared to salary), with the lowest payment being observed in Finland. The highest number of days away from work before a medical certificate must be presented is seen in Sweden and Finland; whereas Denmark and Iceland show the highest number of days in which the employer is responsible for paying sickness benefits.

Table 1. Data on sickness absence across the Nordic countries

Sick leave data	**Norway**	**Denmark**[1]	**Finland**[3]	**Iceland**[4]	**Sweden**[5]
Sickness absence	6.5% in the 3rd quarter of 2017	3.7% in 2014-2015	Ca 5% in 2015	1-2% in 2006	2.9% in the 3rd quarter of 2017
Number of days away from work before a medical certificate must be presented	3 days	2 days	0-7 days	3 days	7 days
For how long can the employee be off sick before losing sickness-benefits?	260 days (52 weeks)	22 weeks within 9 months. Benefits can be paid for a longer period if the recipient fulfils the prolongation conditions[1]	300 days (excluding Sundays) over a 2-year period (for the same illness). 120 days for partial sickness (excluding Sundays) over a 2-year period[2].	52 weeks over a 2-year period. For each month worked, an employee is entitled to wages for two days of sick leave. The general rule is that the sick leave rights of employees increase the longer they work for the same employee.	From 1 February 2016, the time limit was abolished in the sickness benefit system. Hence, there is no longer a maximum time for how long an employee can be off sick.
Percentage of sickness benefit compared to salary	100%	90-100% Sickness cash benefit (sygedagpenge) calculated based on the hourly wage of the employee, with a maximum of DKK 4,245 (€571) per week or DKK 114.73 (€15.43) per hour (37 hours per week), and on the number of hours of work[2]	Ca 70%	A set figure irrespective of salary. Today, government pays IKR 1746 per day, plus IKR 480 per each child supported by the person receiving the benefits (the amount is decided by ministry regulation).	Ca 80%
Number of days before benefit starts being paid	None	None	None	None	1
Number of days the employer is responsible for paying sickness benefit	16 days	30 days[2]	10 working days (including Saturdays)	1 month	14 days

1. Danish data provided by the Danish Agency for Labor Market and Recruitment (STAR) via email communication (12/02/18).
2. Data from MISSOC, the "Mutual Information System on Social Protection".
3. Finnish data confirmed by Eira Viikari-Juntura and colleagues via email communication (07/03/18).
4. Icelandic data were partially confirmed by the Icelandic Health Insurance via email communication (15/03/18).
5. Swedish data confirmed by the Ministry of Health and Social Affairs via email communication (12/02/18).

In Norway, employees are entitled to sickness benefits from day one if they have been in paid work for the last four weeks before the sickness case starts (8). The employer is

responsible for paying the first 16 days of leave, and thereafter NAV assumes the payments. The employer designs a follow-up plan within the first 4 weeks, and a meeting with the sick-listed employee, NAV, and the general practitioner (GP) is arranged to discuss solutions for return to work, ensure dialogue and update the follow-up plan. A similar meeting is also organized at 26 weeks of sick leave. After one year on sick leave, the employee may be granted a transition benefit for up to four years (8;9).

Although the approaches to reduce sickness absence in the Nordic countries vary due to structural differences and variations in sickness policies, the countries are relatively comparable with regard to their policies related to it and measures to reduce sickness absence, as compared to many other countries, using initiatives that are based on close follow-up of the sick-listed person. However, a lack of evidence regarding the effects of those measures has been reported in the literature (10;11). Some of the Nordic measures to reduce sickness absence and disability pension include the following: return to work program (Denmark), charting of the possibilities of the long-term sick listed employees' return to work (Finland), rehabilitation chain (Sweden), and the use of partial sick leave (possible in all Nordic countries) (7).

Partial sick leave (PTSL), also called graded sick leave, has been considered the first option to tackle increasing sickness absence rates and to prevent labor market exclusion, especially in the Nordic countries (3). According to a recent report (7), PTSL is understood as a graded leave, which allows people with reduced workability to work part time and keep the link to the labor market. That is, it enables employees to return to modified duties at the workplace, and might facilitate a progressive return to work (7;12). An underlying premise is the expectation that working, and staying connected to the workplace, in itself will contribute to reduced sick leave duration (13). PTSL varies between 20% up to 99%, independent of the proportion of employment (8). In Norway, use of PTSL gained traction after the 2004 restructuring of the national sick leave regulations. In most cases, whenever an employed person asked for sick leave, the GP should consider PTSL the default option (13). A similar situation developed in Sweden after advice by the national government in the early 2000s to increase use of PTSL. In 2016, PTSL represented around 34% of all sick leave cases among women and 26% of the cases among men (14).

In Norway, the aim to reduce rates of sick leave through improved inclusion mechanisms such as PTSL is anchored in the collective agreement on inclusive working conditions (In Norwegian: Inkluderende arbeidsliv, IA-avtalen). IA-avtalen is a collaborative agreement between the Norwegian government and social partners. The agreement, first signed in 2001, has been renewed several times, most recently in 2014 for the period 2014–2018 (15). Importantly, the 2010 agreement stated the following three main objectives: to reduce sick leave so that sickness absence will not exceed 5.6%, to include more people with reduced functional abilities into working life, and to get people to retire later (16). Participation in the agreement is voluntary for the companies that may or may not choose to sign the agreement and become a so-called 'IA-company' (7). The IA-agreement includes the following five measures (7;16):

1. Increased use of PTSL

2. Changes in the role of the sick leave certifier (e.g., training, feedback on own practices and professional guidance in the work surrounding authorized sick leave)
3. Enterprises and businesses can be members of the IA-agreement
4. The Faster Return-to-Work (FRW) scheme
5. Changes in the follow-up schemes of sick-listed employees

With regard to point four, the Faster Return-to-Work (FRW) scheme (In Norwegian: *Raskere tilbake*) is a collection of measures that intended to prevent unnecessary long-term sick leave (7). The scheme implies offers on individual follow-up, clarification and work-oriented rehabilitation; offer for treatment at specialist health services (purchase of health services); and legislative amendments to ensure closer follow-up of sick people. Thus, the FRW scheme is based on the intention that sick-listed employees get faster clarification, follow-up and work-oriented rehabilitation through NAV. The Directorate of Health and the Labor and Welfare Directorate are responsible for the implementation of the scheme (17). To test the idea that an assessment of functional abilities could strengthen the patient's resources, which in turn could facilitate and encourage an early return to the workplace, a Norwegian cluster randomized controlled trial was conducted (18). It evaluated the effects of teaching GPs about structured functional assessments to change their sick-listing practice, especially in prescribing more PTSL. The study results showed that the intervention GPs prescribed PTSL more often (odds ratio [OR] 1.3, $p<0.05$) than the control GPs. But no differences between groups were seen in sickness absence duration.

An evaluation of the FRW scheme, commissioned by the Directorate of Health, was published in 2010 by Malterud, Mæland, and Ursin (19). This evaluation highlighted the different challenges that GPs and other participants face when both the population and employers have inadequate general knowledge about the sickness benefit scheme. Some of the measures within the FRW scheme were seen as relevant but unstable. Appropriate dialogues were perceived as facilitators of strong cooperation and motivation among all concerned parties. Conversely, the numerous bureaucratic procedures related to sick leave follow-ups were perceived as barriers for the adoption of the FRW scheme, as it could interfere with people's good intentions. Finally, all participants highlighted the need to achieve a broader understanding among Norwegian employers and employees of the correct use of existing welfare systems regarding illness and other absenteeism needs (19). Another evaluation of the FRW scheme published in 2012 by Aavik and colleagues described that waiting time for treatment was eight days shorter on average among sick-listed employees who were enrolled in the scheme compared to peers who were on ordinary list (20).

In line with gaining insight on the measures included within the IA-agreement, in 2005, Olsen and colleagues described that the promotion of PTSL was associated with a drop in the rates of sickness absence by two percentage points among IA-companies compared to those who did not adhere to the agreement (21). However, the authors stated that this small difference was due to a normal fluctuation and concluded that the IA-agreement had not lived up to expectations so far (21). A subsequent evaluation conducted by SINTEF in 2009 (In Norwegian: *Stiftelsen for industriell og teknisk forskning*)

showed improvements on working life in the country after the IA-agreement, and concluded that the IA-companies fulfilled all their obligations within the agreement, and exhibited better cooperation (e.g., better assistance from working life centers and more focus on close follow-up of persons on sick leave). Nevertheless, the evaluation found no effects on sickness absence among IA-companies, and several methodological difficulties were discussed (22).

Continuing its interests in reducing sickness absence in the country, the Norwegian Ministry of Labor arranged a meeting with scientists and experts in April 2013 in order to discuss and review the evidence regarding the different measures the IA-agreement implied (16). Attendees highlighted the lack of empirical research on the IA-agreement measures, and encouraged further research in the area. After reviewing data from different registries and studies derived from them, experts observed that the sickness absence rates decreased in the periods when use of PTSL increased. The decrease was partially attributed to an increased preference among GPs towards graded sick leave (16). In addition, employees on long-term sick leave listed with GPs who often used PTSL exhibited shorter sickness absence and a higher probability of remaining employed two years later. Some explanatory mechanisms discussed by the experts were health benefits, reduced risk of expulsion from work and an effect of PTSL on employee's discipline and attitude to work (16). Recent registry-based analyses published by NAV observed that use of PTSL doubled in the period 2002-2016. Half of the sickness absence cases granted in 2016 were graded sick leave, with half of them graded at 50%. However, PTSL of both higher and lower levels are more commonly used over time. A quarter of people on PTSL move onto full time sick leave (FTSL), and the risk of moving onto FTSL is highest in the first few weeks of the sickness absence period. Moreover, the researchers found that while the duration of sickness absence has remained fairly stable over time, the use of PTSL occurs earlier than before in the course of the sickness absence, with most of the cases graded already from the first day of absence (13).

There have been debates on the effects of the increased emphasis on PTSL. Empirical research on PTSL has shown largely promising results, especially in the Nordic countries. For example, data from a Finnish nationwide registry-based study showed beneficial effects of PTSL compared to FTSL on return to work and work participation (23). PTSL has also been associated with increased work retention and decreased use of full disability pension in long-term assessments in Finland (24). Similar findings were reported in a trial among Finnish workers with musculoskeletal disorders (MSDs) (25). In Norway, Kann and colleagues (26), found a decline in the proportion of individuals on sick leave when the rate of PTSL increased, as well as shorter sickness absence duration. Data from another Norwegian registry-based study (27) indicated that sick-listed employees who were on PTSL when they completed a work-related rehabilitation program were more likely to return to work compared to those who were on FTSL when they left the rehabilitation clinic. In Sweden, PTSL was found to be associated with an earlier return to work in people with mental disorders (MDs) after 60 days of FTSL (28), whilst no effect of PTSL on early return to work was observed among people with MDs in another study conducted in Denmark (29). Grasdal (30), who reviewed a hand-

ful of empirical studies published prior to 2016, concluded that overall, the results indicated that PTSL "contributes to reduction in sickness absence. Specifically, grading seems to contribute to reducing the sick leave period, but there is considerable uncertainty about the size of this effect" (p.114).

Although notable promising results have been published, research on the effects of PTSL has been criticized for having weak external validity, and methodological flaws, such as selection bias, high use of self-reported data, and weaknesses in measuring work participation (1;31). To date, no systematic analysis of the evidence base on the effects of PTSL compared to FTSL has been undertaken. Therefore, this systematic mapping review aimed to map all quantitative evidence on the effects of PTSL versus FTSL on sickness absence and work participation.

Methods

We conducted a systematic mapping review to answer the question: what evidence exists and what does it say about the effects of PTSL versus FTSL on sickness absence and work participation? The project team (reviewers) and commissioner (NAV) discussed and agreed on the research protocol, which is available upon request.

What is a systematic mapping review?

Systematic mapping reviews (also known as systematic scoping reviews) are reviews that map and describe the existing literature or evidence base on a particular topic (32). Such literature reviews take stock of the research available in a particular field. This type of review produces a useful end product in its own right, describing the empirical research that has been undertaken within a particular field of study, but also provides an overview of a research area, highlighting where empirical research is located and where there are gaps. It does not include a synthesis of individual study results (32;33). In a typology of reviews, Grant and Booth (34) explain that such reviews "map out and categorize the existing research on a particular topic, identifying research gaps from which to commission further reviews and/or primary studies."

The present mapping review was methodologically guided by a framework proposed by Arksey and O'Malley (32), as well as Levac and colleagues' (35) recommendations on clarifying and enhancing each stage of the review. Thus, the methodological steps were:

1. Identifying the research question
2. Identifying relevant studies
3. Selecting studies
4. Charting the data
5. Collating, summarizing and reporting the results
6. Optional consultation.

As seen from this outline of the methodological steps of systematic mapping reviews, quality appraisal is not a defined step within such reviews (32;33). However, we decided to appraise the methodological quality of the included studies upon agreement with the commissioner (NAV). Quality appraisal was performed as part of step 4, charting the data. This systematic mapping review is reported in accordance with the PRISMA-ScR reporting guideline (36).

Selection criteria (identifying the research question)

The selection criteria were discussed and agreed with the commissioner ahead of the literature search. We included quantitative research addressing the effects of PTSL versus FTSL on sickness absence and work participation. The main study inclusion criterion was a substantial emphasis on the effect of PTSL versus FTSL as the subject matter. The selection criteria were guided by the following PICO (population, intervention, comparison, and outcome) elements:

Population: Part-time or full-time adult employees (16-69 years old). We excluded studies of people who were described as self-employed.

Intervention: Partial sick leave (PTSL). Following the 2015 report by the Nordic Social Statistical Committee (7), we defined PTSL as graded leave, that allows people with reduced workability to work part time. PTSL varies between 20% up to 99% (7;12). We present any differences in the conceptualization of PTSL across the included studies in the results.

Comparison: Full-time sick leave (FTSL), i.e., no physical presence at the work place.

Outcomes:

Sickness absence (extent, duration, and similar): man-days lost due to own sickness as a percentage of contractual man-days (6). One man-day corresponds to the length in time of one working day for a person in a full-time position (100% position).

Work participation (extent, being fired, and similar) measured as position proportion (values might range between 0 to 1) (6).

Degree of disability and rehabilitation benefits: whether the person in question is on full- or PTSL, and to what degree, is indicated by the degree of disability. The degree of disability is between 20% and 100%, where 100% means full disability. If a sickness absence case consists of more than one medical certificate, researchers might report the average of the degrees of disability for the medical certificates in question (6).

Health-related outcomes: Disease severity and disability.

Study design: Due to the fact that this commission was tied to an effectiveness question, we aimed to include randomized controlled trials (RCTs) as well as non-randomized studies with a control condition. We included non-randomized studies because we anticipated that few, if any, RCTs had been conducted in this field.

Registry-based studies (RBs) (also known as panel data analysis) were included in this review after rounds of consultation with the commissioner. RBs involve the statistical analysis of data sets from registries containing multiple observations over time of a sampling unit (37). RBs can be conducted by pooling time-series observations across a variety of cross-sectional units, including individuals, countries, or companies (37;38).

However, these studies do not enable researchers to establish causal relationships among an intervention or exposure and outcomes (38).

In sum, we considered the following study designs for inclusion:

- Randomized controlled trials (RCT)
- Non-randomized controlled studies / Quasi-randomized studies (NRCT)
- Controlled Before-After studies (CBA)
- Interrupted Time Series (ITS) with at least three measurement points before and after the intervention
- Registry-based studies (RBs)

We followed the definitions of non-randomized studies proposed by the Effective Practice and Organisation of Care Cochrane Group (39) (see Glossary in Appendix 1).

Publication date: Studies published between 1990 and 2018.

Language: We included all languages as long as there was an abstract in English. Any studies meeting the inclusion criteria and published in languages not mastered by the review team (English, Spanish, Norwegian, Swedish, Danish, German) would have been translated with Google translate or by a colleague at the NIPH.

Context: Studies conducted in high-income economies (GNI per capita of $12,236 or more) as defined by the World Bank (Link: https://data-helpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups).

Literature search (identifying relevant studies)

After extensive dialogue with the commissioner to agree on the research question and the selection criteria, a research librarian (Elisabet Hafstad) planned and executed systematic searches in the following databases (from inception to January 2018):

- Cochrane Library: CENTRAL
- Embase
- MEDLINE
- PsycINFO
- PubMed
- Sociological Abstracts & Social Sciences Abstracts
- SveMed+
- Web of Science

The search strategy was adapted for each database. The final search strategy is provided in Appendix 2.

Searching other sources
To identify additional studies, we hand searched the bibliographies of all included studies, as well as any literature reviews and seminal reports about PTSL. We searched the

website of the Norwegian Institute of Public Health, Idunn (Nordic Journals online), the Norwegian and Nordic index to periodical articles (Norart), OpenGrey, Google, and Google Scholar and screened the first 200 hits. Two reviewers (JM, RB) also hand searched on the Nordic Labour Journal (http://www.nordiclabourjournal.org/), the Campbell Library and the following websites:

Nordic organizations for labor and work environment

- The Danish National Research Centre for the Working environment
- The Danish Agency for Labor Market and Recruitment
- Finnish Institute of Occupational Health
- The Norwegian Labor and Welfare Administration (NAV)
- The Norwegian Ministry of Labor and Social Affairs
- Försäkringskassan (Sweden)
- The Swedish Ministry of Health and Social Affairs

Ongoing and recently completed clinical trials

- World Health Organization International Clinical Trials Registry Platform (http://www.who.int/trialsearch/)
- National Institute of Health clinical trials database (http://clinicaltrials.gov)

Study selection

All records retrieved through the literature searches were independently screened for eligibility against the selection criteria by two researchers (JM and NB) by using a pre-designed screening form. We first screened titles and abstracts and then proceeded to full-text screening of relevant records to decide final inclusion or exclusion. Inclusion was decided by consensus and any discrepancies were solved by discussion. If necessary, we would have involved a third researcher (RB) to solve discrepancies.

Data extraction (charting the data)

We designed a data abstraction form to gather relevant information from each study, including characteristics of study participants, settings, context, percentage of sick leave evaluated in the study, comparisons, study designs, methods, statistical analyses and covariates, and results. One reviewer (JM) extracted all data from the included studies and a second reviewer (NB) checked the information for accuracy and completeness. HN and DTK assisted with checking data accuracy. Disagreements were solved by discussion, consensus, and participation of RB.

Quality appraisal of the included studies

Two review authors (JM, NB/RB) appraised the methodological quality of each included study independently. We resolved disagreements by consensus. If necessary, we would have involved another researcher.

RCTs were appraised by using the Cochrane tool for assessment of risk of bias of RCTs (40). Thus, we assessed the following criteria:

- Random sequence generation (selection bias).
- Allocation concealment (selection bias).
- Blinding (performance bias and detection bias), blinding of participants and personnel assessed separately from blinding of outcome assessment.
- Incomplete outcome data (attrition bias).
- Selective reporting (reporting bias).
- Other bias.

We critically appraised the RBs by using the checklist for cohort studies described in the NIPH handbook '*Slik oppsummerer vi forskning*' (41). This 10-items checklist evaluates known sources of bias, such as selection bias, incomplete or lack of reporting of outcome assessment, dropouts, confounding factors, and blinding of outcome assessment. We used this tool because it is the most suitable appraisal tool we could identify for RBs. We search extensively and asked methodological experts, and their recommendation was to use the checklist for cohort studies. It has been used by us to appraise RBs in previous systematic reviews. While this checklist has limitations, to the best of our knowledge, there is no unique checklist for appraising RBs.

In the event that non-randomized controlled studies, including CBAs and ITSs, had been included, we would have appraised the methodological quality of such studies with the tool suggested by the Cochrane EPOC Group (42). This tool includes additional items (relative to the Cochrane risk of bias tool for RCTs) to assess the risk of selection bias and subsequent confounding. The additional items are "were baseline outcome measurements similar?" and "were baseline characteristics similar?" (42).

Collating and summarizing the results

As described above, mapping reviews provide an overview- and *description* of existing research. Data synthesis is limited, relative to full systematic reviews: A systematic mapping review does not include a synthesis, such as meta-analysis, of individual study results. In accordance with the aim- and methodological scope of systematic mapping reviews, we analysed the data descriptively, with frequencies and percentages, and presented results in text, tables, and figures. We grouped studies into categories according to how they were seen to relate to each other, following a data driven approach. For clarity, we presented information separately for RCTs and RBs as well as research from Nordic countries. We note that RBs do not enable researchers to establish causal relationships among an intervention or exposure and outcomes.

Results

Search results

The electronic searches in the major databases yielded 676 references, and additional searches in grey literature sources added 30 references. A total of 300 duplicates were removed. We excluded 380 out of the 406 references screened at title/abstract level, and we read the remaining 26 references in full-text. Thirteen studies met our inclusion criteria. Three out of the 13 included studies were identified after consulting labor agencies and international ministries (15;43;44). Figure 1 depicts the flow diagram for the selection of the studies.

No relevant ongoing studies were identified by searching in the trial registries. The interventions under evaluation in the registered protocols about return-to-work trials included behavioral interventions, self-management, psychotherapy, motivational interviewing, and other integrated active return-to-work programs.

Excluded studies

Most of the 13 excluded studies read in full-text either did not evaluate the effects of PTSL or were not empirical research. A Norwegian RCT that evaluated active sick leave did not meet the inclusion criteria as the authors excluded employees on partial sick leave (45). This trial found that active return-to-work improved neither the number of days on sick leave nor the proportion of patients returning to work in workers with low back pain from 65 Norwegian municipalities (45). Additionally, two Norwegian RBs were excluded because of the lack of a comparison group who were on FTSL. These two studies evaluated solely data from sick-listed employees who received PTSL (26;46). We provide the main reasons for exclusion of the 13 references excluded after full-text consideration in Appendix 3.

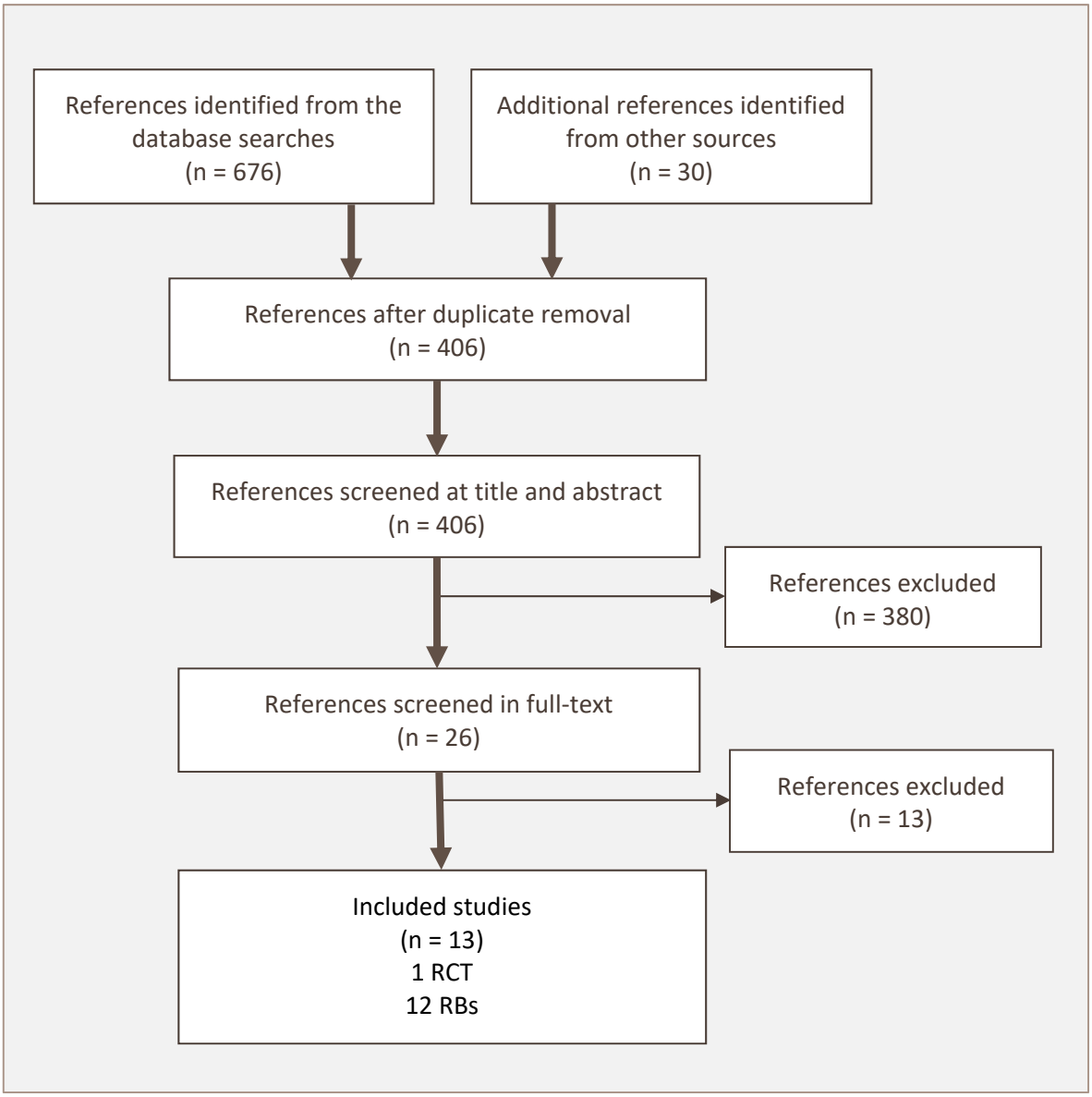

Figure 1. Flow diagram of the selection of studies

Description of included studies

The evidence presented in this systematic mapping review consists of 13 studies. We included one RCT from Finland, reported in two publications (25;47), while the remaining 12 studies were RBs (1;15;17;24;28;43;44;48-52). These studies used observational methods to explore the interactions between PTSL and different variables, such as return-to-work (RTW), in sets of panel data from registries.

Research aim

Overall, all 13 included studies examined the effects of PTSL compared to FTSL for sick-listed employees (table 2).

Randomized controlled trial

Researchers of the Finnish Institute of Occupational Health conducted the only RCT included in this mapping review. The trial, which followed a prospective parallel design and was reported in two different publications, evaluated the effects of early PTSL on RTW and sickness absence (25), and on health-related outcomes (47) among workers with MSDs.

Registry-based studies

All the 12 included RBs evaluated the effects of PTSL compared to FTSL for sick-listed employees. We note that two German studies evaluated employees who completed a rehabilitation program, which is mandatory in the country (17;50). Two other studies, from Finland, addressed the transition to disability pension (24) and the introduction of new legislation of PTSL (1). Lastly, one RB (24) has a companion paper with analyses of the same data set, which we applied when relevant (23).

Setting

In general, it can be stated that most of the evidence base on the effects of PTSL comes from Nordic countries, as 11 studies (85%) were done in such settings (1;15;24;25;28;43;44;47-49;51). The two remaining RBs were conducted in Germany (17;50). See table 2.

Table 2. Country and research aim of the included studies (n=13)

Study, year	Country	Research aim
Andrén 2012 (48)	Sweden	To examine the benefits of being on PTSL compared to FTSL in individuals with musculoskeletal disorders.
Andrén 2014 (28)	Sweden	To analyze the impact of PTSL on the probability of returning to work with full recovery of lost work capacity within 1 year for employees with mental disorders.
Bethge 2016 (17)	Germany	To determine the effects of PTSL on disability pension and regular employment in a random sample of rehabilitation patients who finished a rehabilitation program between 2002 and 2009.
Grødem 2015 (15)	Norway	To study employees who completed their period of sickness benefits (2-3 years after completion) but are still unable to fully return to work.
Høgelund 2010 (52)	Denmark	To examine the effect of a national graded return-to-work program on the probability of sick-listed workers returning to regular working hours.
Kausto 2012 (24)	Finland	To estimate the effects of PTSL on the transition to disability pension applying propensity score methods.
Kausto 2014 (1)	Finland	To examine the effects of the new legislation on PTSL on work participation of employees with long-term sickness absence.
Lie 2014 (43)	Norway	To evaluate the effects of PTSL vs FTSL on sickness absence.
Markussen 2012 (49)	Norway	To examine whether PTSL can reduce absenteeism and subsequent social insurance dependency, and promote self-sufficiency.
Nossen 2013 (44)	Norway	To explore the role of different definitions of PTSL compared to FTSL on sickness absence duration.

Shiri 2013 (47) Viikari-Juntura 2012 (25) RCT	Finland	To determine the health-related effects of early PTSL among employees with musculoskeletal disorders (47). To evaluate the effects of early PTSL on return to work and sickness absence among patients with musculoskeletal disorders (25).
Streibelt 2017 (50)	Germany	To determine the effect of PTSL in addition to a multimodal rehabilitation program on long-term work participation in people with chronic mental disorders.
Viikari-Juntura 2017 (51)	Finland	To assess the effectiveness of the use of PTSL at the early stage of work disability (first 12 weeks) due to mental disorder or musculoskeletal disease on sustained return to work (RTW) and overall work participation.

Type of publication and publication year

Most of the studies included in this systematic mapping review were published in peer-reviewed journals (11 studies, 85%). Two Norwegian RBs were published as organizational reports (15;43). The Finnish RCT was published in 2012-2013 (25;47). The RBs were published between 2010 and 2017, with most published around 2014.

Participants

Below, we present the characteristics of the participants in the RCT and RBs separately.

Randomized controlled trial

The Finnish trial (25;47) included 62 employees on sick leave due to MSDs. They were recruited from six occupational health units of medium- and large-size private or public enterprises. They had a permanent or long-term contract, working full-time (37-38 hours per week), and had not been on sick leave due to their MSDs for >2 weeks during the preceding month and >30 days during the preceding 3 months. Most of the participants worked in the healthcare sector or retail trade, and a minority from call-centres or meat-processing industry. Around half of the participants had higher vocational school but none of them had completed university studies (table 3).

Table 3. Characteristics of participants, intervention and control group in the Finnish randomized controlled trial

Randomized controlled trial (RCT)	**Population**	**Intervention: Part-time sick leave, PTSL N (%)**	**Comparison: Full-time sick leave, FTSL N (%)**
Viikari-Juntura 2012 (25), Shiri 2013 (47) Finland	N=62 employees Age: mean age 44 (standard deviation 10) Sex: 97% female Diagnosis: Musculoskeletal disorders (e.g., neck or shoulder region, back or upper or lower extremities) Ethnicity: not reported	N=31 (50%) 70% received 50% PTSL 30% worked shorter hours on 3–4 days a week	N=31 (50%)

Registry-based studies

The participants in the twelve registry studies were from Norway (four studies), Denmark (one study), Finland (three studies), Sweden (two studies), and Germany (two studies). In total, 2,741,563 participants were analyzed in these RBs (range 627 – 1,400,094). In general terms, the vast majority of participants across the RBs shared common characteristics, such as having a regular job contract, being female (up to 78%) and around 45 years old (range 18-64). Grødem and colleagues included Norwegian employees older than 60 (15), while Lie (43) and Nossen and Brage (44) did not provide information on participants' age, gender or diagnosis. Only the two Swedish studies provided data on participants' ethnicity (28;48). See table 4.

National registries used for analysis

All twelve RBs used data from national registries. The four studies conducted in Norway all utilized data from the Norwegian Labor and Welfare Administration (NAV), as follows:

- Markussen analyzed data from all sickness absences granted from 2001 through 2005 (49).
- Nossen and Brage analyzed data from all spells granted in 2011 (44).
- Lie (43) analyzed data from spells granted to 10% of all people reported in NAV's registry during 2002-2010.
- Grødem (15) analyzed data from employees who terminated the sickness benefits period during the first half of 2011 after one year of sick leave.

This suggests some overlap in the cases based on the NAV registry, between Lie (43) and Markussen and colleagues (49) and between Nossen and Brage (44) and Grødem and colleagues (15). We provide further descriptions in the section "partial sick leave in the included studies" below. There were no indications of overlap between data sources in the remaining studies. Andrén and collaborators used the same data set from the 2002 sample of the Swedish Social Insurance Agency containing data of sick-listed employees due to MSDs (48) and mental disorders (MDs) (28). Høgelund and colleagues used data from the national register of payments of sickness benefits in Denmark, supplemented with survey information (52). In the two German studies, Bethge and colleagues (17) studied employees who continued on sick leave after they completed a rehabilitation program (January-June 2007), while Streibelt and colleagues (50) in 2012 recruited employees who had completed a rehabilitation program and were eligible for PTSL. Both studies used data from the German Pension Insurance Agency.

Table 4. Characteristics of participants, intervention and control groups in registry-based studies (n=12)

Registry-based studies	**Population**	**Part-time sick leave, PTSL N (%)**	**Full-time sick leave, FTSL N (%)**
Andrén 2012 (48)	N=1170; Sweden Age: 20-64 years (50% >46 years) Diagnosis: MSDs Gender: around 60% female Ethnicity: 85% born in Sweden	N=140 (12%)	N=1030 (88%)

Andrén 2014 (28)	N=627; Sweden Age: 20-64 years (50% >46 years) Diagnosis: MDs Gender: around 60% female Ethnicity: 85% born in Sweden	N=79 (13%)	N=548 (87%)
Bethge 2016 (17)	N=3,750; Germany Age: average 45 years (range 18-60) Diagnosis: Around 63% MSDs. Other diagnoses: cardiac, oncological, psychosomatic Gender: around 50% female Ethnicity: not reported	N=1875 (50%)	N=1875 (50%)
Grødem 2015 (15)	N=17,077; Norway Age: >60 years (born after 1951) Diagnosis: MSDs (45.5%), MDs (25.4%) Gender: 58% female Ethnicity: not reported	N=5294 (31%) >50% PTSL (8.7%) 50% PTSL (14.6%) <50% PTSL (7.7%)	N=11783 (69%)
Høgelund 2010 (52)	N=934; Denmark Age=43.8 years Diagnosis: ca 16% MDs Gender: ca 61% female Ethnicity: not reported	N= 265 (28%)	N=669 (72%)
Kausto 2012 (24)	N=26,259; Finland Full-time workers with long-term sickness absence due to MSDs, MDs, trauma or tumors. Age: 48 years (SD 11) Diagnoses: MSDs (46%), MDs (27%), traumas (18%), tumours (9%) Gender: 54% female Ethnicity: not reported	N=1012 (4%)	N=25247 (96%)
Kausto 2014 (1)	N=58,091; Finland Age: average 46 years (SD 10) Diagnoses: MDs (38% in PTSL, 25% in FTSL), MSDs (35.9% in PTSL, 36.3% in FTSL), tumours (6.4% in PTSL, 5.4% in FTSL), traumas (7.8% in PTSL, 14.8% in FTSL), other (11.7% in PTSL, 18.4% in FTSL) Gender: 71% female Ethnicity: not reported	N=1685 (3%)	N=56406 (97%)
Lie 2014 (43)	N=890,726; Norway (10% of all people reported in NAV's registry 2002-2010) Age, diagnosis, gender, ethnicity: not reported	N=62,699 (7%) 0.8% >50% PTSL, 4.2% 50% PTSL, 2% <50% PTSL	N=828,027 (93%)
Markussen 2012 (49)	N=339,251; Norway Age: average 44.1 years in PTSL vs 42.4 years in FTSL Diagnoses: MDs (23.6% in PTSL, 22.3% in FTSL), MSDs (46.4% in PTSL, 47.2% in FTSL) Gender: 67.% female in PTSL vs 53% in FTSL Ethnicity: not reported	N=77,655 (23%)	N=261,596 (77%)
Nossen 2013 (44)	N=1,400,094; Norway Age, diagnoses, gender, and ethnicity: not reported	N=250,617 (17.9%)	N=1,149,477 (82.1%)

Streibelt 2017 (50)	N=762; Germany Age: average 47.8 years Diagnosis: MDs (65% affective disorders) Gender: 78% female Ethnicity: not reported	N=381 (50%)	N=381 (50%)
Viikari-Juntura 2017 (51)	N=3756; Finland Age: 20-64 (41% 20-44 years, 37% 45-54 years, 22% 55-64 years) Diagnoses: MDs and MSDs Gender: 77.5% females Ethnicity: not reported	N=1878 (50%)	N=1878 (50%)

MSD= Musculoskeletal disorder, MD= Mental disorder, SD= standard deviation

Intervention (partial sick leave = PTSL)

In the section below, we present the characteristics of the intervention, PTSL, in the RCT and RBs separately.

Randomized controlled trial

In the Finnish RCT (25;47), the GPs gave the patient a fit note, indicating the duration of partial work disability, whether certain physical loads should be reduced, and whether any additional work modifications were deemed necessary. A 50% PTSL was given to 70% of all sick-listed employees, whereas due to difficulties in arranging a half work day, 30% of employees worked shorter hours on 3–4 days a week. Some task modifications were also implemented if necessary.

Registry-based studies

In the 12 RBs, all data on the use of PTSL were taken from the national registries in the respective countries (see above *National registries used for analysis*). The most commonly used PTSL was 50%, which was granted in around 70% of the graded work absences for most of the time. The study from Denmark describes PTSL as a workplace intervention program whereby "sick-listed workers return to their pre-sick leave job on temporarily reduced working hours" (52). The RBs analyzed data about sickness absences (both PTSL and FTSL) granted between 2001 and 2014, which gives a timeline of 13 years. Figure 2 illustrates the dispersion of the years for data analysis across the RBs (i.e., the timeline from which the data were taken in the registry). The two Swedish RBs, both with Andrén as first author, used the same years from the registry, but included workers with different disorders, MSDs and MDs, respectively (28;48).

It is important to highlight that all RBs, except for two Norwegian studies published by Grødem and colleagues (15) and Nossen and Brage (44), explored the influence of different covariates on the relationship between PTSL and the outcome measures. Most of the RBs conducted propensity-score matching analyses for balancing the samples on the probability of being assigned to PTSL (six studies). The most common covariates used across the RBs were gender/sex and age (five studies), followed by type of occupation and diagnosis (four studies), and geographic area, income and data on the physician who granted the sick leave (three studies). Other covariates included the sick leave duration, previous sick leaves, level of education, etc. Details on the adjusted analyses and the corresponding covariates across RBs can be found in Appendix 4, and definitions of the statistical analyses are presented in Appendix 5.

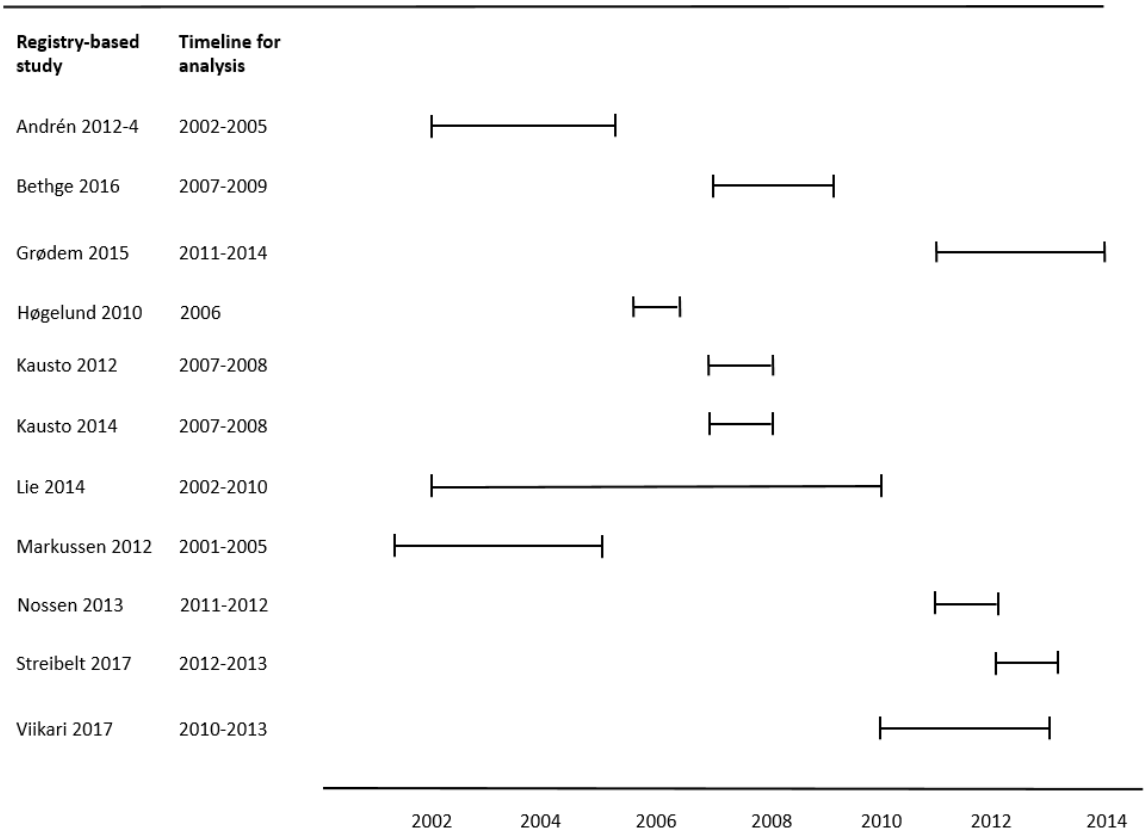

Figure 2. Dispersion of the years for data analysis across the registry-based studies (n=12)

Comparison

Both the Finnish RCT and the 12 RBs used FTSL as comparator.

Outcome measures

In total, 15 outcomes were reported in the 13 included studies. There were four main types of outcomes: sickness absence, work participation, degree of disability, and health-related outcomes. Work participation (RTW) was the most common outcome, reported in ten studies (i.e., in the RCT and nine RBs), followed by sickness absence duration, disability, disability pension and allowance of social benefits (each measured by four studies). Because 12 of the 13 included studies were RBs, registry data were the most common source of measurement. Table 5 shows the different outcomes, numbers of studies and tools.

Table 5. Outcome measures and tools in the included studies

Outcome measure	Number of studies	Tools
Sickness absence duration	1 RCT (25) 3 RBs (44;49;50)	Registry data
Work participation		
Return-to-work (RTW)	1 RCT (25)	Registry data

	9 RBs (1;15;28;43;48-52)	
Unemployment	3 RBs (49-51)	Registry data
Degree of disability		
Recurrence of sick leave for any cause	1 RCT (25)	Registry data
Disability	1 RCT (47) 3 RBs (15;43;51)	Registry data Shiri et al. 2013 (47) used the following tools: Oswestry Disability Index to assess the disability level due to back pain; the Neck Disability Index to assess cervical spine-related disabilities; the QuickDASH to assess the disabilities of the arm, shoulder, and hand; and the Comprehensive Osteoarthritis Test (COAT) to assess the symptoms of the hip or knee.
Productivity loss	1 RCT (47)	Two questions recommended by Brouwer et al., "The subjects were asked to consider the latest full or partial working day and compare it to their normal workday when answering the questions: (i) assess the amount of work you were able to perform, and (ii) assess the quality of your work. For both questions, the scale ranged from 0–10 (0=very poor to 10=regular quantity or quality). In case the reported value was <10 for either question, the respondent was requested to indicate whether the reason was musculoskeletal, other health-related, or another problem. To estimate productivity loss, we used a formula "[1 - (quality/10) . (quantity/10)] × 100%" modified from Hoeijenbos et al."
Disability pension	4 RBs (15;17;24;50)	Registry data
Allowance of social welfare benefits	4 RBs (15;17;43;49)	Registry data
Health-related outcomes		
Pain (intensity and interference with work)	1 RCT (47)	Pain intensity on a scale from 0–10 (0=no pain to 10=the worst possible pain) and pain interference with work and sleep during the last 7 days (from 0=no interference at all to 10=the worst possible interference)
Self-rated general health	1 RCT (47)	The respondents were asked to assess their self-rated health with a numerical scale from 0–10 (0=worst possible health state to 10=best possible health state).
Health-related quality of life	1 RCT (47)	EQ-5D.
Physical and emotional functioning	1 RBs (50)	The 36-item Short-Form Health Survey of the Medical Outcomes Study.
Depression	1 RCT (47) 1 RB (53)	Shiri et al. (54) evaluated depression using two validated questions: (i) During the past month have you often been bothered by feeling down, depressed, or hopeless? and (ii) During the past month have you often been bothered by little interest or pleasure in doing things? Depression was defined as a "yes" response to both questions. Streibelt et al. (50) used the 2-item PHQ-4 scale.
Anxiety	1 RB (53)	The 2-item PHQ-4 scale.
Working ability	1 RB (50)	Streibelt et al. (50) used two items from the Work Ability Index (WAI) for assessing the self-rated current physical and mental work ability.

Quality appraisal of the included studies

While not a required step in a systematic mapping review, we assessed the methodological quality of the 13 included studies. We used different checklists for the RCT and the 12 RBs (see methods).

Randomized controlled trial

The Finnish RCT exhibited moderate methodological quality (25;47). There were some concerns about both performance and detection bias due to the lack of blinding of both participants and outcome assessor. In addition, the trial reported some outcomes that were not pre-specified in the protocol. We found no major concerns for selection or attrition bias. See Appendix 6 for a complete description of the quality appraisal of the RCT.

Registry-based studies

We used a checklist from the NIPH handbook for systematic reviews (41) to assess the overall quality of the RBs, which resulted in a categorization of studies into low, moderate or high methodological quality, as follows:

- High quality: low risk of bias in ≥ 8 items.
- Moderate quality: low risk of bias in 5-7 items.
- Low quality: low risk of bias in ≤ 4 items.

Eleven out of the twelve included RBs exhibited moderate methodological quality (1;15;17;24;28;43;44;48;49;51;52). One showed high methodological quality (50). All studies included large numbers of individuals, were judged as representative of their population, and measured both exposure and outcomes equally and reliably. The PTSL and FTSL groups were comparable on important background factors in five studies (1;15;28;50;51). Follow-up time was judged as adequate in all studies. See Appendix 7 for a complete description of the methodological appraisal of the RBs.

Summary of main findings from the included studies

We provide a brief summary of the main findings reported by the study authors for the outcomes included in this review, i.e., sickness absence, work participation, degree of disability and rehabilitation benefits, and health-related outcomes. Further data can be retrieved in the full-text publications. A summary table, indicating direction of results across the studies, is provided at the end of the chapter (table 6).

Sickness absence duration

The Finnish RCT of moderate methodological quality, and three RBs, of moderate and high methodological quality, reported on sickness absence. The Finnish trial showed a lower proportion of sickness absence days in employees with PTSL than in peers with FTSL throughout the 12-month follow-up period (20% lower on average) (25).

Three RBs measured sickness absence duration. The Norwegian RB published by Markussen and colleagues (49) found that PTSL predicted reductions in sickness absence duration by more than 60 days after controlling for patient/job covariates and physician characteristics, compared to FTSL. This reduction increased up to 74 days when spells exceeding 12 weeks were analyzed. Another Norwegian study, by Nossen and Brage (44), observed that the use of PTSL in spells lasting at least 15 days and that were on FTSL during the first 14 days led to a 21 days shorter sickness absence duration compared to the use of FTSL. Larger reductions in favor of PTSL were seen after 8 weeks in spells lasting longer than 2 weeks and graded during the first 2 weeks (166 days in PTSL vs 199 days in FTSL; mean difference= -33 days). This difference in favor of PTSL increased up to 39 days at 12 weeks. Conversely, crude analyses revealed shorter sickness absences among those assigned to FTSL compared to those in PTSL (115 days in PTSL vs 23 days in FTSL; mean difference= 92 days). This difference was reduced to 13 days when data for spells graded after 2 weeks were analyzed[1]. The analysis of spells graded at <80% showed a difference of 79 days in favor of FTSL, whereas smaller differences were seen in spells lasting at least 2 weeks. However, only spells lasting longer than 12 weeks had a slightly shorter duration among those assigned to PTSL compared to FTSL (212 days in PTSL vs 216 days in FTSL; mean difference= -4 days).

In a German RB, Streibelt and colleagues (50) reported that the mean sick leave duration was 7.1 weeks in the PTSL group compared to 13.4 weeks in the FTSL group at 15 months follow-up ($p<0.001$). No statistically significant difference between PTSL and FTSL was found in employees who reported a positive subjective prognosis of RTW at baseline (−1.5 weeks, $p= 0.338$). However, employees assigned to FTSL with a negative RTW subjective prognosis experienced 12 weeks more than peers assigned to PTSL (-11.7 weeks, $p < 0.001$). Thus, the RCT and three RBs that reported on sickness absence found positive results of PTSL on this outcome.

Work participation

Return-to-work (RTW)

The Finnish RCT and nine RBs reported on RTW. All of these studies were assessed to have moderate methodological quality, except for one RB from Germany that had high methodological quality. In the Finnish trial (25), participants in the PTSL group returned to work earlier throughout the one-year follow-up period. Time to sustained RTW ≥2 weeks was similar in both groups (median time: 9 days in both groups), while it tended to be shorter at >4 weeks in the PTSL group (median 12 vs 20 days, $p=0.10$) at the end of the follow-up period. Age-adjusted Hazard Ratio (HR)[2] for RTW for ≥4 weeks was 1.60 (95% CI 0.98–2.63), which means that participants who received PTSL were

[1] This analysis excluded spells from pregnant women, people over the age of 66, persons with 60 days sick leave or more started in 2010, spells with a median duration of <15 days (serious diagnoses), and spells graded at 90-99%.

[2] Hazard ratio (HR) is a measure of an effect of an intervention on an outcome of interest over time. Hazard ratio is reported most commonly in time-to-event analysis or survival analysis (i.e. when we are interested in knowing how long it takes for a particular event/outcome to occur).

more than one and a half times as likely to RTW than those on FTSL. Controlling for previous sickness absence during the preceding 30 days decreased the HR by 11% and controlling for body mass index decreased it by 3%. Overall HR for RTW, controlling for age, pain interference with sleep, and previous sickness absence, was 1.76 (95% CI 1.21–2.56).

Three of the Norwegian RBs reported on work participation. Grødem and colleagues (15) reported that employees on PTSL stayed connected to their jobs more than their peers on FTSL at two years after termination of the sickness benefits (38% in FTSL vs 84% in <50% PTSL at 2.5 years after termination). Work participation among employees who received PTSL at 50% - largest group - was 72% at the same time. Lie (43) found that employees who started their spell with PTSL were 63% less likely to RTW than employees with FTSL were (HR=0.37, 95% CI 0.36-0.38; p<0.001), after adjusting for covariates (i.e., age, sex, diagnosis, time in job, and physician data). Employees assigned to FTSL earlier than 8 weeks returned to work more often than peers assigned to PTSL. Conversely, a slightly higher probability of RTW in favor of PTSL was seen after 24 weeks (adjusted HRs ranged between 1.13 at 24 weeks up to 1.25 at 40 weeks). In addition, subgroup analyses suggested no clear difference among employees with recurrent cases (i.e., second and third sick leaves), but there may be a difference in favor of PTSL for the longest sickness benefit cases. Finally, and using data from the same registry, Markussen and colleagues (49) found that PTSL reduced the expected number of workdays (full-time equivalent) lost during the spell by more than 90 days compared with FTSL throughout the observation period between 2001-2005, after controlling for patient/job covariates and physician characteristics.

In Denmark, Høgelund and colleagues (52) evaluated the effect of PTSL on the probability of sick-listed workers returning to regular working hours. With a multivariate mixed-proportional-hazard-rate model (which controlled for >10 variables, including gender, age, education, type of occupation, visits to general practitioner, enrollment in vocational rehabilitation), the results showed that for each week of being on sick leave, the sick-listed workers on PTSL had a 51% higher probability of returning to regular working hours than did sick-listed workers who were on FTSL (coefficient 0.414, p=0.021). In Sweden, Andrén and Svensson (48) found that employees who were sick-listed for MSDs had a 0.25 higher likelihood[3] of full recovery if assigned to PTSL rather than FTSL (330 days or less). Being male, married and younger than 56 years were associated with full recovery. Having been on sick leave in the previous year was negatively associated with full recovery. After adjusting for occupational type, results indicated being assigned to PTSL was associated with a higher probability of full recovery in the lengths of time analyzed. Regression coefficients varied from 1.50 for spells lasting equal to or less than 30 days to 1.20 for spells lasting equal to or less than 330 days. In sick-listed employees for mental disorders, Andrén (28) concluded that PTSL was as-

[3] Average treatment effect (ATE) expresses the average difference between the probability that the individual will fully recover after PTSL and the probability that the individual will fully recover after FTSL.

sociated with a greater likelihood of full recovery compared to FTSL (average treatment effect 0.015)[2] when it is assigned in the beginning of the spell. Andrén also observed a strong positive effect of PTSL (average treatment effect 0.387)[2], and statistically significant, when assigned after 60 days of FTSL at the end of the 330 days observation period. Distributional analysis of the effect parameters showed that in a group of randomly selected employees on sick leave for more than two weeks due to a MD, 17.8% of them would fully recover their lost work capacity if assigned to PTSL in the beginning of the spell, but would not have fully recovered their lost work capacity without the PTSL treatment. However, 16.3% of them would not fully recover if assigned to PTSL.

The other three RBs that reported on work participation were from Finland (two studies) and Germany. In Finland, Kausto and colleagues (1) observed reductions in the level of work participation for both the PTSL and the FTSL groups during the one-year follow-up, the absolute reduction being larger in the FTSL group (−26.5%) as compared with the PTSL group (−21.2%), which means a difference of 5.3% (95% CI 3.1% to 7.5%). This difference increased up to 9.8% (95% CI 5.9% to 13.7%) in the propensity score matched subsample (i.e., the conditional probability of being assigned to PTSL given observed covariates). Subgroup analyses showed that in all age categories, work participation declined more in the FTSL group than in the PTSL group. The difference in the decline was significant in age categories 45–54 and 55–65. There was no effect in those aged 35–44. In subgroup analyses, a statistically significantly larger effect in favor of PTSL was found for people with mental disorders as compared with the other diagnostic categories (difference 12.8%, 95% CI 9.0% to 16.5%).

Another Finnish study published in 2017 by Viikari-Juntura and colleagues (51) found an absolute risk difference of 8.0% and a relative risk difference of 10.9% in favor of PTSL on sustained RTW. In addition, the authors observed that the mean overall time spent at work was 77.4%; it was 10.5% higher in the PTSL group compared to the FTSL group during the 2-year follow-up. Subgroup analyses showed that the difference was larger among men than women and for people with MDs compared to MSDs. In Germany, the RB of high methodological quality, by Streibelt (50), found that 88.4% of the PTSL group had returned to work at 15 months follow-up compared to only 72.6% of the controls (Relative Risk [RR]=1.22, 95% CI 1.13–1.31). The relative risk of returning to work was greater in the PTSL group compared to FTSL. The greatest effect of PTSL on RTW was observed among employees who did not believe that they would go back to work after rehabilitation (74% in PTSL vs 49% in FTSL).

Unemployment

Three RBs reported on unemployment. In Norway, Markussen and colleagues (49) observed that PTSL predicted a rise in the employment probability two years after the start of the sick leave by around 16% (Semiparametric Least Squares (SLS) estimate 0.16, standard error 0.04), after controlling for patient/job covariates and physician characteristics. This probability increased up to 20% when spells exceeding 12 weeks were analyzed. In Germany, Streibelt (50) found that sick-listed employees assigned to PTSL had a 60% lower risk of unemployment compared to peers who were assigned to

FTSL (RR 0.41, 95% CI 0.26 to 0.65). Finally, The Finnish RB published by Viikari-Juntura and colleagues (51) observed that sick-listed employees who received PTSL spent less time unemployed during the 2-year follow-up compared to those who received FTSL. The difference in proportions was about 1.8-fold (3.2% in FTSL vs 1.8% in PTSL). This difference was larger among men than women and in workers in manufacturing compared to other industries.

In summary, one RCT from Finland and all of the nine RBs that measured RTW, except the Norwegian RB by Lie (43), indicated positive effects on RTW for employees on PTSL compared to FTSL (15;49-51). Favorable effects of PTSL compared to FTSL on unemployment were reported by all of the three RBs that measured this outcome (49-51).

Degree of disability and rehabilitation benefits

There were five types of outcomes with regard to degree of disability and rehabilitation benefits: recurrence of sick leave for any cause, disability, productivity loss, disability pension, and allowance of social welfare benefits. We report the study results for each of these five outcomes separately below.

Recurrence of sick leave for any cause

The Finnish RCT (moderate methodological quality) (25) indicated that time to first recurrent sick leave was similar in the PTSL and FTSL groups. However, the number of recurrent sick leaves per person year after the initial sickness absence period was about 20% lower in the PTSL group at one-year follow-up.

Disability

The Finnish RCT and three RBs reported on disability. All had moderate methodological quality. No differences between PTSL and FTSL regarding disability were reported by the Finnish RCT at one-year follow-up (47).

The three RBs reporting on disability were from Norway (two studies) and Finland. Using Norwegian data, Lie (43) found no statistically significant differences between PTSL and FTSL on employees' disability after controlling for covariates (i.e., age, sex, diagnosis, time in job, and physician data). Grødem and colleagues (15) observed the largest recovery among employees who had <50% PTSL, with 70% of them being employed without receiving benefits at two-years follow-up. People with 50% or higher PTSL did worse, with less than 40% employed without receiving benefits. The rate was 25% in the FTSL group, which also exhibited the highest percentage of full disability benefits.

Finnish researchers analysed data on both full- and partial disability retirement (51). The proportion of full disability retirement was almost threefold in the FTSL group compared to the PTSL group (6.9% versus 2.4%) after two years follow-up. Subgroup analyses showed that the positive effects of PTSL were similar in both genders, higher in the oldest age group compared with the younger age groups, slightly higher in people with MDs than MSDs, and remarkably high among people with technical and scientific work (i.e., greater effect meaning that fewer people went on full disability retirement). However, the proportion of partial disability retirement in the PTSL group was

4.5-fold compared to the FTSL group (7.9% versus 1.8%); the overall absolute risk difference was -6.1% (95% CI -7.1 to -4.9) (negative value indicating increase in risk). Higher risks were seen among women, the oldest employees, and people with MSDs compared with MDs, and among people in the public sector and healthcare and social work. These results remained even after adjusting for residual imbalance in baseline covariates (e.g., age, major region, employment sector, socioeconomic status, and annual gross income).

In sum, one RCT from Finland (47) and one RB from Norway (43) found no differences between PTSL and FTSL on disability, whereas two Nordic RBs suggested positive effects of PTSL on employees' disability (15;51), except with regard to partial disability retirement in the Finnish RB (51).

Productivity loss

Only the Finnish RCT reported on productivity loss. Data from this RCT of moderate methodological quality found there were no significant difference on productivity loss between PTSL and FTSL up to one-year follow-up (regression coefficient -0.6; 95%CI -9.1 to 7.9; p-value= 0.88). However, this effect became favorable for PTSL after adjusting for body mass index, follow-up time, time since beginning of symptoms (number of elapsed days) and the baseline measure but did not reach statistical significance (regression coefficient 2.3; 95%CI -4.8 to 9.5; p-value= 0.52) (47).

Disability pension

Four RBs, from Norway, Finland, and Germany (two studies), reported on disability pension. Three of the RBs had moderate methodological quality and one had high methodological quality. In Norway, Grødem and colleagues (15) reported that the highest rate of receiving disability pension was observed in employees who were on 50% PTSL (12% at the termination of sickness benefits, and 19% two years later), whereas the rate varied from 7% up to 13% at the end of the observation period (January 2014) among employees who were on FTSL. The lowest likelihood of receiving a disability pension was seen in the group of <50% PTSL (6% two years after sickness benefit). Partial disability pension was the most common solution among employees who received PTSL.

Finnish researchers (24) observed that employees on PTSL had a 70% lower crude risk of full disability pension and a threefold lower crude risk for partial disability pension compared with those on FTSL at one-year follow-up. Similar estimates were found after adjusting for covariates (i.e., age, gender, diagnostic group, occupational group, gross income, insurance district, length of sick leave before treatment, and length of sick-leave in connection with treatment) and propensity score matching analysis. The risk of full disability pension was around 0.5-fold in the PTSL group compared with the FTSL group, in both genders. Furthermore, the use of PTSL was associated with statistically significant reductions in both absolute and relative risks of full disability pension by 6% and 41%, respectively. On the other hand, the absolute and relative risks of partial disability pension increased among those who had been on PTSL compared with those who had been on FTSL by 8% and 159%, respectively. Subgroup analyses showed that this effect was stronger in women than in men and among people with

MDs than in MSDs. No associations were observed between PTSL and the transition to any disability pension (partial and full disability pension combined).

Lastly, two German RBs examined disability pension. Bethge and colleagues (17) reported that assigning PTSL to sick-listed employees reduced their risk of receiving a disability pension by 40% (HR=0.62, 95% CI 0.49–0.80) at one-year follow-up. Streibelt and colleagues (50) reported that PTSL reduced the risk of receiving a disability pension by 60% (RR=0.40, 95% CI 0.23-0.70) compared to FTSL at 15 months follow-up. This study had high methodological quality.

In sum, across the four RBs, the results for disability pension were mixed. The Norwegian RB suggested a higher rate of receiving disability pension among employees on PTSL compared to FTSL (15), whereas the two German RBs reported a decreased risk of receiving a disability pension in employees on PTSL (17;50). The Finnish study (24) found that PTSL reduced the risk of full disability pension compared to FTSL, whereas the opposite result was found for partial disability pension.

Allowance of social welfare benefits

Four RBs, from Norway (three studies) and Germany, reported on allowance of social welfare benefits. All four of these RBs had moderate methodological quality.

In a Norwegian setting, Markussen and colleagues (49) found that PTSL was associated with fewer social security claims (regression coefficient adjusted for patient/job covariates and physician characteristics: -79.7 (standard error 11.8)) in the 2-year period following just after the end of the absence spell. This reduction increased up to 99 days receiving benefits when spells exceeding 12 weeks were analyzed. Lie (43) found that the probability of receiving social benefits was lower among those on PTSL than FTSL, although the differences were small and non-significant (HR ranged from 0.95 at 4 weeks to 0.80 at 28 weeks) after controlling for covariates (i.e., age, sex, diagnosis, time in job, and doctor's data). The sequential analysis did not reveal differences between second and third sick leaves. Grødem and colleagues (15) reported that all groups, but not <50% PTSL, showed reductions in the allowance of social benefits. The highest allowance of benefits was seen in the FTSL group (72% of the employees received social benefits in the period after the termination of the sickness benefits period, and the rate decreased to 46% two years later), followed by employees who received PTSL at >50% (69% at termination, and 40% two years after), and 50% PTSL (59% at termination, and 30% two years after). The allowance of social benefits remained stable at 10% in the <50% PTSL group. The total allowance of income-based benefits was higher among workers with FTSL and those with >50% PTSL.

The German study, by Bethge and colleagues (17), reported that employees with PTSL received fewer welfare benefits due to sickness absence and unemployment up to the end of the study period than those in FTSL. The accumulated time of receiving sickness benefits was reduced by 52 days (95% CI 40–64 days), short-term unemployment benefits by 58 days (95% CI 49–67 days), and long-term unemployment benefits by 15 days (95% CI 10–20 days) at one-year follow-up.

Thus, all four RBs (15;17;43;49) observed a lower allowance of social welfare benefits in PTSL compared to FTSL.

Health-related outcomes

Only two of the included studies reported on health-related outcomes. This was the Finnish RCT (25;47) and a German RB (50). The RCT had moderate methodological quality and the RB high methodological quality. The Finnish RCT included people with MSDs while the German study included people with MDs, primarily affective disorders. Below, we report the results for the six health-related outcomes separately. These were: pain, self-rated general health and health-related quality of life, physical and emotional functioning, depression, anxiety, and working ability.

Pain (intensity and interference with work)

Results from the Finnish RCT (47) showed reductions in both pain intensity and interference with work in all groups during the first 8 weeks and stabilized thereafter. No differences between the PTSL and FTSL groups were observed during a 12 weeks follow-up period after adjusting for body mass index, follow-up time, time since beginning of symptoms (number of elapsed days) and the baseline measure. Thus, pain intensity (≤3 months) showed a regression coefficient of -0.4 (95%CI -1.3 to 0.4; p=0.31); pain interference with work (≤3 months) -0.7 (95%CI -1.6 to 0.3; p=0.15); pain interference with sleep (≤3 months) -0.12 (95%CI -0.9 to 0.7; p=0.77), and pain at 1 year -0.2 (95%CI -0.7 to 0.4; p=0.48).

Self-rated general health and health-related quality of life

The Finnish trial (47) found that employees who received PTSL self-reported better general health than those in the FTSL group (regression coefficient 0.5, 95%CI -0.0 to 1.0; p=0.07), and higher health-related quality of life at one-year follow-up (regression coefficient -0.5, 95%CI -0.9 to -0.01; p=0.02). These analyses were adjusted for body mass index, follow-up time, time since beginning of symptoms (number of elapsed days) and the baseline measure of the outcome.

Physical and emotional functioning

Streibelt and colleagues' (50) results in Germany showed that employees who received PTSL had a higher physical (regression coefficient +7.9, p=0.01) and emotional (regression coefficient +6.8, p=0.025) role function compared to those in the FTSL group at 15 months follow-up.

Depression

No differences were observed in the Finnish RCT between the PTSL and FTSL groups on sick-listed employees' depression symptoms (47). Conversely, in the German RB (50), people in the PTSL group did better than people in the FTSL group (regression coefficient −0.6, 95%CI −1.1 to −0.1; p-value= 0.03) at 15 months follow-up.

Anxiety

Data from the German RB (50) showed that people in the PTSL group improved with respect to anxiety symptoms (measured with the same tool as depression), compared

to FTSL (regression coefficient −0.6, 95%CI −1.1 to −0.1; p-value= 0.03) at 15 months follow-up.

Working ability

The German RB (50) found that people in the PTSL group did better than people in the FTSL group with respect to working ability (regression coefficient 0.1, 95%CI -0.1 to 0.3; p=0.05) at 15 months follow-up, but the difference was not statistically significant.

Summary of results across the included studies

Table 6 summarizes the direction of results across the 13 included studies (15 outcomes). With regard to causal effects, the RCT presents the strongest study design and is therefore highlighted. However, given the study's small sample size (n=62) firm conclusions about the effects of PTSL cannot be drawn.

Table 6. Summary of direction of results across the included studies (n=13)

Outcome measure ↓	Andrén 2012 [48]	Andrén 2014 [28]	Bethge 2016 [17]	Grødem 2015 [15]	Høgelund 2010 [52]	Kausto 2012 [24]	Kausto 2014 [1]	Lie 2014 [43]	Markussen 2012 [49]	Nossen 2013 [44]	**Shiri 2013 [47] * Viikari-Juntura 2012 [25] ***	Streibelt 2017 [50] #	Viikari-Juntura 2017 [51]
Sickness absence									+	+	+	+	
Work participation													
Return-to-work	+	+		+	+		+	-	+		+	+	+
Unemployment									+			+	+
Degree of disability and rehabilitation benefits													
Recurrence of sick leave for any cause											=		
Disability				+				=			=		+
Productivity loss											=		
Disability pension			+	-		+						+	
Allowance of social welfare benefits			+	+				+	+				
Health-related outcomes													
Pain (intensity and interference with work)											=		
Self-rated general health											+		
Health-related quality of life											+		
Physical and emotional functioning												+	
Depression											=	+	
Anxiety												+	
Working ability												+	

Legend: + favors the intervention (partial sick leave, PTSL); - favors the control group (full-time sick leave, FTSL); = no difference between PTSL and FTSL; * Randomized-controlled trial (RCT); # study of high methodological quality, remaining studies had moderate methodological quality. Empty cell means the study did not examine the outcome.

Discussion

Main findings

To date, 12 RBs (from the Nordic countries and Germany) and one RCT from Finland have investigated the effects of PTSL compared to FTSL in more than 2.74 million sick-listed employees. All studies had moderate methodological quality, except one RBs, which had high methodological quality. This indicates consistent high internal validity in the research methods used, but it is important to stress that RBs have limited capacity to detect causal effects. The Finnish RCT studied employees who were sick-listed due to MSDs, while the RBs included employees with primarily either MSDs or MDs.

While firm conclusions cannot be drawn, the RCT and the RBs suggested PTSL may be associated with shorter sickness absence and higher work participation. The Finnish RCT reported that employees with PTSL experienced better general health and quality of life compared to those on FTSL. However, it did not find statistical differences between PTSL and FTSL on sick leave recurrence, employees' productivity loss, and pain. The RBs indicated a lower probability for people on PTSL of receiving both disability pension and allowance benefits, disability, as well as better scores on physical- and emotional functioning, anxiety, depression, and working ability.

The results regarding sickness absence and work participation are supported by a Norwegian RB (26), which used the same NAV dataset as Lie (43) and partially Markussen and colleagues (49), both included in this review. Kann and colleagues (26) demonstrated that an increase in the rate of use of PTSL in a municipality of one percentage point (e.g. from 13% to 14%) was associated with a reduction in the sickness absence rate of 1.79%, and shorter sickness absence duration. That is, when the proportion of days of sick leave that were graded increased by one percentage point, the rate of sickness absence decreased by about two percent. Thus, the researchers concluded that greater use of PTSL can in subsequent months lead to a reduction in the sickness absence rate, duration of sickness absence, and number of individuals on sick leave. A similar analysis with data from 2000 to 2011 reached the same conclusions, although the associations in this analysis were weaker (46).

Our systematic mapping review identified some gaps with regard to the different outcomes measured across studies. Work participation was the most common outcome reported, measured in ten studies, followed by sickness absence duration, disability, disability pension and allowance of social welfare benefits, with four studies each. In contrast, the recurrence of sick leave was only measured in one study (the RCT). There was

an important lack of knowledge regarding the effect of PTSL for health-related outcomes, as only two studies addressed this issue (table 6). Further, it should be noted that the overwhelming majority of the study populations were adults suffering from musculoskeletal- or mental disorders. None of the studies specifically included people on sick leave for other reasons. It is possible that effects of PTSL are related to diagnosis. For example, Høgelund and colleagues (29), who combined survey and register data on about 850 Danish workers, found that PTSL had no effect on the duration until returning to regular working hours for employees with MDs, but significantly reduced the duration until returning to regular working hours for employees with other disorders.

Generalizability and strength of findings

The OECD has reported that the insufficient labor force participation among people with health issues and disability, their low income, and the high costs of sickness and disability benefit schemes represent a serious problem for governments, and the use of PTSL might be a suitable measure to counteract them (2). An important point in favor of the generalizability of the favorable findings from the studies included in this systematic mapping review is the predominance of Nordic studies. Eleven of the 13 studies were from a Nordic country, including four from Norway. Further, a considerable group of employees in the included studies were sick-listed due to MSDs and MDs. The former represents the main occupational diseases suffered by European workers, according to the European Observatory of Working Life (55). In Norway, MSDs are among the most common reasons for consultation to GPs and emergency primary health care, and represent the majority of day treatments in the national health system (56). Moreover, with regard to MDs, data from the Swedish Social Security Agency show that MDs accounted for around half of all sick leave cases among women and 40% of all cases among men in December 2016. In 2016, PTSL represented around 30% of all sick leave cases and was especially high among people with MDs (14).

All RBs used data from structured nationwide registries. The Norwegian studies used data from NAV, which might facilitate the formulation of public policies and further research of the use of PTSL for controlling sickness absence. We note that the two German studies concerned a total of about 4500 employees – and therefore contributes less to the overall findings – who completed a rehabilitation program, which is mandatory in Germany, but not standard practice in Norway where people generally start PTSL early in the course of the sickness absence. In Germany, PTSL is granted when the sick-listed employee has completed the rehabilitation program but is still unable to perform full duties, and it is approved by the rehabilitation physician and the social worker with consent from the employer, the patient, the general practitioner and the occupational physician (57-59). As part of the scheme developed by the rehabilitation physician, the sick-listed employee starts working for at least two hours/day and gradually increases the working time (58;59).

Strength of evidence

The two types of study designs and research approaches that represent the body of evidence on the effects of PTSL versus FTSL merit some discussion in terms of the possibility to draw causal inferences from their findings. The Finnish RCT represents the strongest study design to answer the research question that guided this systematic mapping review, as it is widely accepted that well-conducted prospective, experimental studies have the greatest capacity to detect causal effects due to their possibility to randomly assign individuals to different exposures, and therefore ensure that groups are similar (60). Essential support for this statement may be based on the counterfactual approach for evaluating causal inference in epidemiology, which argues that a high comparability between exposed and unexposed individuals is needed to estimate any causal effect, as it is not possible to observe an individual's outcome both at the time when he is exposed and at the same time when he is not (61-63). Nevertheless, random assignment is not sufficient to ensure high validity of experimental studies, as also these may have limitations that can lead to biased estimates of causal effects (e.g., low adherence to the intervention, high attrition, and outcome measurement errors) (40-42;64;65). Additionally, ethical constraints for conducting experimental research are a common concern that turns the focus to observational data.

In spite of the preference for experimental studies when addressing questions about effect, and as noticed in our findings, the body of evidence about the benefits/drawbacks of PTSL compared to FTSL is mostly represented by RBs that use observational data derived from nationwide registries. While a broad discussion about causal inference goes beyond the scope of this systematic mapping review, we offer a few notes about the capacity to draw causal inferences from observational data. Observational studies estimate differences on a certain outcome "X" among individuals with different levels of an exposure "Y" (63;66), and numerous methodological approaches are described in the literature to estimate causal effects based on their data (62;63;66). Some of the major constraints to detect causal inferences in observational data are tied to the susceptibility of selection bias due to a non-random distribution of the exposure, the possibility of confounding, reverse causation, and the option to see outcome data before the allocation of the individuals to the groups of study (62;67). Outcome data analysis is not preceded by the allocation of the individuals to different exposure levels, thus an equal distribution of covariates between exposed and unexposed individuals may not be ensured (66;67). The coexistence of these factors leads to ambiguous causal inferences. Therefore, the association measures obtained from observational data might not equal any causal effect; these association measures might not reflect causation or serve as valid estimates of the true causal effect (66). Relative to the RBs included here, it is possible that the people on PTSL and FTSL systematically differed on measured and non-measured factors, including work capacity and motivation.

Drawbacks with respect to detecting causal inferences notwithstanding, an opportunity of large-scale observational data, as the case of the RBs with 2.74 million participants included in this mapping review, is the ability to test confounders, as long as they are clearly defined and measured appropriately (63;67). It is important to notice that all but two RBs from Norway (15) (44) tested the interaction of important confounders

such as gender, age and previous sick leaves on the benefits/drawbacks of receiving PTSL, and most used propensity-score matching analyses for balancing the samples. Which covariates to include is a matter of debate, and complicated by the fact that some researcher state that people on PTSL, relative to those on FTSL, are sicker and therefore have longer and more illnesses (68), and others that they have better health and greater work capacity (44). Interaction analyses might demonstrate that the associations between PTSL with favorable outcomes in sick-listed employees emerge from an interaction of multiple factors, potentially accumulative, and not merely from a unique component-cause relationship. Further details on the conceptual mechanistic models of causation from observational data are described elsewhere, e.g. by Martin in 2014 (66).

Comparison with other reviews

We identified neither systematic reviews nor mapping reviews addressing the effects of PTSL versus FTSL on sickness absence and work participation. A literature review of the use, effects and feasibility of PTSL in Nordic countries was published by Kausto and other Finnish colleagues in 2008 (69). Similar to our study, the review authors conducted a comprehensive systematic search of the literature in major databases and contacted social insurance institutions in Nordic countries. However, they did not appraise the quality of the included studies. The authors found few studies on the effects of PTSL and stated more research and more rigorous studies are needed to determine whether PTSL is feasible and beneficial in keeping those with reduced work ability in work life (69). Mirroring our conclusions, also the literature review by Grasdal (30) cautiously concluded that PTSL seems to contribute to reducing the sick leave period. Lastly, a more recent systematic review, from 2016 (70), found there was a lack of evidence on early interventions compared to usual care for return to work. The review included three RCTs, including Viikari-Juntura and colleagues (25), which also we included here, concluding that there was some evidence of benefits of intervening during the first two weeks of the sickness absence among people with MSDs.

Strengths and weaknesses

The main strengths of this systematic mapping review are the close collaboration between the commissioner, NAV, and the research team, in planning the scope of the mapping review, detailed in a project plan, as well as the comprehensive and extensive searches of the literature, the contact with national directorates to retrieve additional data, and the adherence to international methodological standards for systematic mapping reviews. That three of the 13 included studies were identified through our grey literature searches illustrates the benefit of applying an inclusive and broad search strategy for a topic like graded sick leave. However, we provided only a narrative summary on the main findings from the included studies as pooled data analysis is not part of mapping reviews. We presented data on the adjusted analyses across the RBs in order to facilitate a better understanding of the findings, even when a synthesis goes beyond the scope of this report. An important strength of including RBs is that those registers have no loss of follow-up, providing a representative estimate of the national data.

Lastly, we encourage the development of a checklist for methodological quality assessment of registry-based studies that attends to the statistical approach used.

With the existing evidence base, we cannot draw firm conclusions regarding the effects of PTSL versus FTSL on sickness absence and work participation. The recommendation of further RCTs to strengthen the potential to draw causal inferences about the effects of PTSL on sickness absence must be considered in light of the ethical constraints, governmental regulations on the intervention, and data sharing difficulties to obtain highly representative samples (71-73). Although unlikely, given the existing body of evidence, a well-conducted systematic review might assist in drawing firmer conclusions.

Implications for practice

Findings from the included studies suggest PTSL may be associated with several favorable outcomes for sick-listed employees. Such benefits might also reach employers, companies, policy makers, social security systems, and other stakeholders. Acceptability of PTSL also appears to be high. Kausto and colleagues' literature review (69) concluded that acceptance of PTSL was good in Sweden, Norway, Denmark, and Finland. A Swedish survey found that more than nine of ten people on long-term sick leave asserted that PTSL was beneficial for them. They expressed positive attitudes towards PTSL and the researchers concluded that there was a potential for an increased use of the measure among people on long-term sick leave (74).

In general terms, the included studies shared similar research aims and methodological structures, which might facilitate the commission of further research and the subsequent progress in the study of this subject matter, especially in western European countries. Finally, our findings must be interpreted in line with the current policies for sickness absence for each country.

Conclusion

The evidence on the effects of PTSL compared to FTSL consists of one small RCT and 12 RBs, with a total of about 2.74 million study participants with mostly musculoskeletal- or mental disorders. The results indicate PTSL may be associated with several favorable outcomes, such as shorter sickness absence and higher work participation, although firm conclusions about the effects of PTSL cannot be drawn.

The RCT results, derived from a sample of 62 people, indicated employees with MSDs had shorter sick leave duration and higher work participation, and experienced better general health and quality of life with PTSL. However, the trial failed to detect significant differences between PTSL and FTSL on leave recurrence, employees' productivity loss, pain, disability, and depressive symptoms. Results from the RBs suggested use of PTSL may be associated with shorter sickness absence, lower probability of disability, receiving disability pension and allowance benefits, as well as better scores on physical- and emotional functioning, anxiety, depression, and working ability.

Firm conclusions about the effects of PTSL are constrained due to the overwhelming majority of RBs in this body of evidence. Observational designs are not well suited to answer questions about effect, as causation cannot be confirmed. To draw firm conclusions, well-conducted RCTs are necessary.

References

1. Kausto J, Viikari-Juntura E, Virta L, Gould R, Koskinen A, Solovieva S. Effectiveness of new legislation on partial sickness benefit on work participation: a quasi-experiment in Finland. BMJ Open 2014;4.
2. OECD. Sickness, Disability and Work. Keeping on track in the economic downturn. High-Level Forum, 14-15 May 2009. Stockholm: 2009.
3. Grødem S. Partial sick leave in Norway: high hopes for social inclusion. European Commission; 2015. ESPN - Flash report. 2015/58.
4. OECD. Health status: absence from work due to illness[cited 09 January 2018]. Available from: http://stats.oecd.org/index.aspx?queryid=30123#
5. Sundell T. Utviklingen i sykefraværet, 3. kvartal 2017. Oslo: NAV; 2017.
6. Statistics Norway. Sickness absenceOslo[updated 19 December 2017; cited]. Available from: http://www.ssb.no/en/sykefratot
7. NOMESCO-NOSOSCO. Sickness Absence in the Nordic Countries. Copenhagen: Nordic Council of Ministers; 2015.
8. NAV. Sickness benefitsOslo[updated 06.01.2016; cited 15.05.2018]. Available from: https://www.nav.no/en/Home/Benefits+and+services/Sickness+benefits
9. Husabo E, Monstad K, Holmås TH, Oyeflaten I, Werner EL, Maeland S. Protocol for the effect evaluation of independent medical evaluation after six months sick leave: a randomized controlled trial of independent medical evaluation versus treatment as usual in Norway. BMC Public Health 2017;17(1):573.
10. Försäkringskassen. Nordiska strategier för att begränsa sjukfrånvaro. Socialförsäkringsrapport. 2008. 1.
11. Ministerråd N. Nordiske initiativer til nedbringelse af sygefravær. Norden - Nordisk ministerråd. 2005.
12. Øyeflaten I, Lie S, Ihlebæk C, Eriksen H. Prognostic Factors for Return to Work, Sickness Benefits, and Transitions Between These States: A 4-year Follow-up After Work-Related Rehabilitation. J Occup Rehabil 2014;24(2):199-212.
13. Nossen JP, Lysø N. Gradert sykmelding – økt gradering, stabil varighet. Arbeid og velferd 2018;2:101-25.
14. Försäkringskassan (Swedish Social Insurance Agency). Social Insurance in Figures 2017. Stockholm: 2017.
15. Grødem A, Orupabo J, Pedersen A. Gradert Sykemelding: oppfølging og trygdemottak etter sykepengeperiodens utløp. Oslo, Norway: Institutt for Samfunnsforskning; 2015. 5.
16. Mykletun A, Brinchmann B. Effekter av tiltak under IA-avtalen. Rapport fra forskermøte på oppdrag fra Arbeidsdepartementet. Oslo: Nasjonalt Folkehelseinstittut; 2013.
17. Bethge M. Effects of graded return-to-work: a propensity-score-matched analysis. Scand J Work Environ Health 2016;42(4):273-9.
18. Osteras N, Gulbrandsen P, Kann IC, Brage S. Structured functional assessments in general practice increased the use of part-time sick leave: a cluster randomised controlled trial. Scand J Public Health 2010;38(2):192-9.

19. Malterud K, Mæland S, Ursin H. Hvor trykker skoen? Sykefraværsoppfølging i Norge 2009 Raskere tilbake. Bergen: UniHelse; 2010. Available from: http://uni.no/nb/uni-helse/stress-helse-og-rehabilitering/sykefravrsoppflging-i-norge-2009-raskere-tilbake/
20. Aakvik A, Holmås T, Kjerstad E. Hospital Capacity, Waiting Times and Sick Leave Duration – An Empirical Analysis of a Norwegian Health Policy Reform. Working papers in Economics. No.10/12. Universitetet i Bergen; 2012.
21. Olsen T, Svendal A, Amundsen I. More inclusive workplaces: fact or fiction? The case of Norway. International review of psychiatry (Abingdon, England) 2005;17(5):375-84.
22. Ose S, Bjerkan A, Johnsen A, Paulsen B, Mo T, Saksvik P, et al. Evaluering av IA-avtalen (2001–2009). SINTEF; 2009. Available from: https://evalueringsportalen.no/evaluering/evaluering-av-ia-avtalen-%25282001-2009%2529
23. Kausto J, Virta L, Luukkonen R, et al. Associations between partial sickness benefit and disability pensions: initial findings of a Finnish nationwide register study. BMC Public Health 2010;10(361).
24. Kausto J, Solovieva S, Virta L, et al. Partial sick leave associated with disability pension: propensity score approach in a register-based cohort study. BMJ Open 2012;2(e001752).
25. Viikari-Juntura E, Kausto J, Shiri R, Kaila-Kangas L, Takala E, Karppinen J, et al. Return to work after early part-time sick leave due to musculoskeletal disorders: a randomized controlled trial. Scandinavian journal of work, environment & health [Internet]. 2012; 38(2):[134-43 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/664/CN-00971664/frame.html
26. Kann I, Brage S, Kolstad A, Nossen J, Thone O. Har gradert sykmelding effekt på sykefraværet? Arbeid og velferd 2012;2:60-70.
27. Oyeflaten I, Lie SA, Ihlebaek CM, Eriksen HR. Multiple transitions in sick leave, disability benefits, and return to work. - A 4-year follow-up of patients participating in a work-related rehabilitation program. BMC Public Health 2012;12:748.
28. Andren D. Does part-time sick leave help individuals with mental disorders recover lost work capacity? J Occup Rehabil 2014;24:344-60.
29. Hogelund J, Holm A, Eplov L. The effect of part-time sick leave for employees with mental disorders. J Ment Health Policy Econ 2012;15:157-0.
30. Grasdal AL. De helserelaterte trygdeytelsene. Betydningen av økonomiske insentiver og samspill mellom trygdeordninger. Tidsskrift for velferdsforskning 2016;19(2):102–24.
31. Viikari-Juntura E, Kausto J, Shiri R, et al. Return to work after early part-time sick leave due to musculoskeletal disorders: a randomized controlled trial. Scand J Work Environ Health 2012;38:134-3.
32. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005;8(1):19-32.
33. Armstrong R, Hall R, Doyle J, Waters E. Cochrane update. 'Scoping the scope' of a Cochrane review Journal of Public Health 2011;33(1):147-50.
34. Grant M, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Information and Libraries Journal 2009;26(2):91–108.
35. Levac D, Colquhoun H, O'Brien K. Scoping studies: advancing the methodology. Implement Sci 2010;20(5):69.
36. Tricco AC, Lillie E, Zarin W, et al. Prisma extension for scoping reviews (prisma-scr): Checklist and explanation. Annals of Internal Medicine 2018.
37. Lavrakas P. Encyclopedia of Survey Research Methods. 2008.

38. Frees E. Longitudinal and Panel Data: Analysis and Applications in the Social Sciences: Cambridge University Press; 2004.
39. Effective Practice and Organisation of Care (EPOC). What study designs should be included in an EPOC review? EPOC Resources for review authors. Oslo: 2016.
40. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011. Available from: www.handbook.cochrane.org.
41. Nasjonalt kunnskapssenter for helsetjenesten. Slik oppsummerer vi forskning. Håndbok for Nasjonalt kunnskapssenter for helsetjenesten. 4. reviderte utg. Oslo: Nasjonalt kunnskapssenter for helsetjenesten; 2015.
42. Cochrane Effective Practice and Organisation of Care (EPOC). Suggested risk of bias criteria for EPOC reviews. EPOC resources for review authors. 2013. Available from: epoc.cochrane.org/epoc-specific-resources-review-authors 2013
43. Lie S. Liten effekt av gradert sykmelding for raskere retur til arbeid: En kausal analyse av registerdata for varighet av sykefravær, ulike trygdeytelser og retur til jobb. Bergen, Norway: Institutt for Klinisk Odontologi, University of Bergen; 2014.
44. Nossen J, Brage S. Effekt av gradert sykemelding -betydningen av metodevalg Arbeid og velferd 2013;3:68-81.
45. Scheel I, Hagen K, Herrin J, Carling C, Oxman A. Blind faith? The effects of promoting active sick leave for back pain patients: a cluster-randomized controlled trial. Spine (Phila Pa 1976) 2002;27(23):2734-40.
46. Ose S, Kaspersen S, Reve SH, Mandal R, Jensberg H, Lippestad J. Sykefravær– gradering og tilrettelegging. Trondheim, Norway: SINTEF Teknologi og samfunn; 2012. 6OH18730.
47. Shiri R, Kausto J, Martimo KP, Kaila-Kangas L, Takala EP, Viikari-Juntura E. Health-related effects of early part-time sick leave due to musculoskeletal disorders: a randomized controlled trial. Scandinavian Journal of Work, Environment & Health 2013;39(1):37-45.
48. Andren D, Svensson M. Part-Time Sick Leave as a Treatment Method for Individuals with Musculoskeletal Disorders. Journal of Occupational Rehabilitation 2012;22(3):418-26.
49. Markussen S, Mykletun A, Røed K. The case for presenteeism—Evidence from Norway's sickness insurance program. Journal of Public Economics 2012;96:959–72.
50. Streibelt M, Burger W, Nieuwenhuijsen K, Bethge M. Effectiveness of Graded Return to Work After Multimodal Rehabilitation in Patients with Mental Disorders: A Propensity Score Analysis. J Occup Rehabil 2017.
51. Viikari-Juntura E, Virta LJ, Kausto J, Autti-Ramo I, Martimo KP, Laaksonen M, et al. Legislative change enabling use of early part-time sick leave enhanced return to work and work participation in Finland. Scandinavian Journal of Work, Environment and Health 2017;43(5):447-56.
52. Hogelund J, Holm A, McIntosh J. Does graded return-to-work improve sick-listed workers' chance of returning to regular working hours? J Health Econ 2010;29(1):158-69.
53. Godoy A. Profiting from presenteeism? Effects of an enforced activation policy on firm profits. Labour Economics 2016;43:122-8.
54. Annerblom ML, Sjöström S. Partiell sjukskrivning - bra rehabilitering för kvinnor? Svensk Rehabilitering 2001;2:54-6.
55. Edwards P, Greasley K. Absence from work. Dublin, Ireland: European Foundation for the Improvement of Living and Working Conditions; 2010. Available from: https://www.eurofound.europa.eu/observatories/eurwork/comparative-information/absence-from-work

56. Statistics Norway. Patients Statistics: 7 000 000 treatments in Norwegian hospitals Oslo[cited 25 June 2018]. Available from: https://www.ssb.no/en/helse/artikler-og-publikasjoner/7-000-000-treatments-in-norwegian-hospitals
57. Voelter-Mahlknecht S, Stratil JM, Kaluscha R, Krischak G, Rieger MA. Experiences, attitudes and possibilities for improvement concerning the cooperation between occupational physicians, rehabilitation physicians and general practitioners in Germany from the perspectives of the medical groups and rehabilitation patients – a protocol for a qualitative study. BMJ Open 2017;7(4).
58. Deutsche Rentenversicherung (DRV). [Rehabilitation report 2015]. Berlin: DRV; 2015.
59. Bundesministerium für Arbeit und Soziales [Federal Ministry for work and social affairs] (BMAS). [Statement of the scientific expert group RehaFutur on the future of vocational rehabilitation in Germany. Research report F393],. Bonn: BMAS; 2009.
60. West SG, Duan N, Pequegnat W, Gaist P, Des Jarlais DC, Holtgrave D, et al. Alternatives to the Randomized Controlled Trial. American Journal of Public Health 2008;98(8):1359-66.
61. Rubin DB. Estimating the causal effects of smoking. Statistics in medicine 2001;20(9-10):1395-414.
62. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Statistics in medicine 2007;26(1):20-36.
63. Hernan MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health 2004;58(4):265-71.
64. Hemming K, Eldridge S, Forbes G, Weijer C, Taljaard M. How to design efficient cluster randomised trials. BMJ (Clinical research ed) 2017;358.
65. Vader JP. Randomised controlled trials: A User's guide. BMJ (Clinical research ed) 1998;317(7167):1258.
66. Martin W. Making valid causal inferences from observational data. Preventive veterinary medicine 2014;113(3):281-97.
67. McGue M, Osler M, Christensen K. Causal Inference and Observational Research: The Utility of Twins. Perspectives on psychological science : a journal of the Association for Psychological Science 2010;5(5):546-56.
68. Eklund M vGH, Marklund S.,. Deltidssjukskrivning - individ, arbetsplats och hälsa. In: Hogstedt C, Bjurvald M, Marklund S, Palmer E, Theorell T, editors. Den höga sjukfrånvaron - sanning och konsekvens. Statens Folkhälsoinstitut; 2004.
69. Kausto J, Miranda H, Martimo KP, Viikari-Juntura E. Partial sick leave--review of its use, effects and feasibility in the Nordic countries. Scand J Work Environ Health 2008;34(4):239-49.
70. Vargas-Prada S, Demou E, Lalloo D, Avila-Palencia I, Sanati KA, Sampere M, et al. Effectiveness of very early workplace interventions to reduce sickness absence: a systematic review of the literature and meta-analysis. Scand J Work Environ Health 2016;42(4):261-72.
71. Skivington K, McCartney G, Thomson H, Bond L. Challenges in evaluating Welfare to Work policy interventions: would an RCT design have been the answer to all our problems? BMC Public Health 2010;10:254.
72. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Social science & medicine (1982) 2018;210:2-21.
73. Burdorf A, van der Beek AJ. To RCT or not to RCT: evidence on effectiveness of return-to-work interventions. Scandinavian Journal of Work, Environment & Health 2016;(4):257-9.
74. Sieurin L, Josephson M, Vingard E. Positive and negative consequences of sick leave for the individual, with special focus on part-time sick leave. Scand J Public Health 2009;37(1):50-6.

75. OECD. Glossary of statistical terms [cited 13.11.2018]. Available from: https://stats.oecd.org/glossary/detail.asp?ID=630
76. OECD. Glossary of statistical terms [cited 13.11.2018]. Available from: https://stats.oecd.org/glossary/detail.asp?ID=2480
77. Beemster T, Velzen J, Bennekom C, Frings-Dresen M, Reneman M. Cost-effectiveness of 40-hour versus 100-hour vocational rehabilitation on work participation for workers on sick leave due to subacute or chronic musculoskeletal pain: study protocol for a randomized controlled trial. Trials [Internet]. 2015; 16:[317 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/626/CN-01170626/frame.html
78. Floderus B, Göransson S, Alexandersson K, Aronsson G. Self-estimated life situation in patients on long-term sick leave. Journal of Rehabilitation Medicine 2005;37(5):291-9.
79. Kausto J, Solovieva S, Virta LJ, Viikari-Juntura E. Partial sick leave and subsequent work participation: A register based study of working-aged Finnish women and men. European Journal of Epidemiology 2013;1):S62-S3.
80. Maeland JG. Å gradere eller ikke - det er sykemeldingspörsmålet. Utposten 2011;40(2):7-9.
81. Olaya-Contreras P, Styf J. Illness behavior in patients on long-term sick leave due to chronic musculoskeletal pain. Acta Orthopaedica 2009;80(3):380-5.
82. Scheel IB, Hagen KB, Herrin J, Oxman AD. A call for action - A randomized controlled trial of two strategies to implement active sick leave for patients with low back pain. Spine 2002;27(6):561-6.
83. Steenstra I, Anema J, Bongers P, Vet H, Knol D, Mechelen W. The effectiveness of graded activity for low back pain in occupational healthcare. Occupational and environmental medicine [Internet]. 2006; 63(11):[718-25 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/584/CN-00566584/frame.html
84. Vermeulen SJ, Anema JR, Schellart AJ, van Mechelen W, van der Beek AJ. Cost-effectiveness of a participatory return-to-work intervention for temporary agency workers and unemployed workers sick-listed due to musculoskeletal disorders: design of a randomised controlled trial. BMC Musculoskeletal Disorders 2010;11:60.
85. Jones A. Applied econometrics for health economists. Oxford: Radcliffe Publishing; 2007.
86. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health 2012;66(12):1182-6.
87. Ding YY. Risk adjustment: towards achieving meaningful comparison of health outcomes in the real world. Annals of the Academy of Medicine, Singapore 2009;38(6):552-7.
88. Rencher AC, Schaalje GB. Linear models in statistics. Second edition. New Jersey: John Wiley & Sons; 2008.
89. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22.
90. Hernán MA. The Hazards of Hazard Ratios. Epidemiology (Cambridge, Mass) 2010;21(1):13-5.
91. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. Third Edition. New Jersey: John Wiley & Sons; 2013.
92. Long J, Freese J. Regression Models for Categorical and Limited Dependent Variables Using Stata. College Station. Texas: Stata Press; 2006.
93. Mendes D, Alves C, Batel-Marques F. Number needed to treat (NNT) in clinical literature: an appraisal. BMC Medicine 2017;15(1):112.

94. Judkins DR, Porter KE. Robustness of ordinary least squares in randomized clinical trials. Statistics in medicine 2016;35(11):1763-73.
95. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research 2011;46(3):399-424.
96. Marra G, Radice R. Estimation of a semiparametric recursive bivariate probit model in the presence of endogeneity. Canadian Journal of Statistics 2011;39(2):259-79.
97. Xia Y. Semiparametric Regression Models. In: Lovric M, editor. International Encyclopedia of Statistical Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 1296-. Available from: https://doi.org/10.1007/978-3-642-04898-2_508

Appendix

Appendix 1. Glossary

Concept	Definition
Controlled before-and-after study (39)	A study in which observations are made before and after the implementation of an intervention, both in a group that receives the intervention and in a control group that does not.
Disability (75)	"Disability reflects any limitation or lack of ability that a person experiences in performing an activity in the manner or within the range considered normal for a person, in other words, a limitation in learning, speaking, walking or some other activity (individual dimension)."
Interrupted time series study (39)	A study that uses observations at multiple time points before and after an intervention (the 'interruption'). The design attempts to detect whether the intervention has had an effect significantly greater than any underlying trend over time.
Non-randomized controlled trial (39)	An experimental study in which people are allocated to different interventions using methods that are not random.
Parallel randomized-controlled trials (65)	In these studies (also called parallel trials or RCTs with parallel group design), each group of participants is exposed to only one of the study interventions. For instance, if a group of investigators uses a parallel design to evaluate the effects of a new analgesic compared with those of a placebo in patients with migraine, they would give the new analgesic to one group of patients and placebo to a different group of patients.
Social benefits (76)	"Social benefits are current transfers received by households intended to provide for the needs that arise from certain events or circumstances, for example, sickness, unemployment, retirement, housing, education or family circumstances."

Appendix 2. Search strategies

Cochrane Library

#1	((active or graded or partial or part-time) near/4 (sick leave or sick-leave or sickness-leave or medical-leave or sickness-absence))	42
#2	((partial* or partly or part-time) near/4 (sicklist* or sick-list*))	6
#3	((partial or graded or graduated or part-time) near/4 (sickness-benefit* or sickness-certificat*))	0
#4	((graded or gradual* or partial* or part-time or progressive or step-by-step or step-wise or stepwise) near/4 (return-to-work or RTW or occupational-reintegration or occupational-re-integration or work-exposure))	32
#5	#1 or #2 or #3 or #4 in Trials	42

OVID-databaser søkt 20180109:
Embase 1974 to 2017 January 08,
MEDLINE® Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE® 1946 to Present,
PsycINFO 1806 to January Week 1 2018

1	((active or graded or partial or part-time) adj4 (sick leave or sickleave or sickness-leave or medical-leave or sickness-absence)).mp.	230
2	((partial* or partly or part-time) adj4 (sicklist* or sick-list*)).mp.	27
3	((partial or graded or graduated or part-time) adj4 (sickness-benefit* or sickness-certificat*)).mp.	27
4	((graded or gradual* or partial* or part-time or progressive or step-by-step or step-wise or stepwise) adj4 (return-to-work or RTW or occupational-reintegration or occupational-re-integration or work-exposure)).mp.	321
5	or/1-4	554
6	remove duplicates from 5	289
7	6 use ppez [MEDLINE]	125
8	6 use oemezd [Embase]	131
9	6 use psyh [PsycINFO]	33

PubMed

1	(active-sick leave[tw] OR graded-return-to-work[tw] OR graded-sick leave[tw] OR graded-work-exposure[tw] OR gradual-return-to-work[tw] OR partially-sick-listed[tw] OR partial-return-to-work[tw] OR partial-sick leave[tw] OR partial-sick-listing[tw] OR partial-sickness-absence[tw] OR partial-sickness-benefit*[tw] OR partly-sick-listed[tw] OR part-time sick-list*[tw] OR part-time-medical-leave[tw] OR part-time-return-to-work[tw] OR part-time-sick leave[tw] OR part-time-sickness-certificat*[tw] OR progressive-return-to-work[tw] OR stepwise-occupational-reintegration[tw] OR stepwise-return-to-work[tw] OR step-wise-return-to-work[tw])	37

Sociological Abstracts + Social Services Abstracts

1	((active or graded or partial or part-time) n/3 (sick leave or sickleave or sickness-leave or medical-leave or sickness-absence)) OR ((partial* or partly or part-time) n/3 (sicklist* or sick-list*)) OR ((partial or graded or graduated or part-time) n/3 (sickness-benefit* or sickness-certificat*)) OR ((graded or gradual* or partial* or part-time or progressive or step-by-step or step-wise or stepwise) n/3 (return-to-work or RTW or occupational-reintegration or occupational-re-integration or work-exposure))	64

SveMed+

1	((active OR graded OR partial OR part-time) AND (sick leave OR sickleave OR sickness-leave OR medical-leave OR sickness-absence))	49
2	((partial* OR partly OR part-time) AND (sicklist* OR sick-list*))	0
3	((partial OR graded OR graduated or part-time) AND (sickness-benefit* OR sickness-certificat*))	0
4	((graded OR gradual* OR partial* OR part-time OR progressive OR step-by-step OR step-wise OR stepwise) AND (return-to-work OR RTW OR occupational-reintegration OR occupational-re-integration OR work-exposure))	0
5	#1 OR #2 OR #3 OR #4	49

Web of Science: SCI-EXPANDED, SSCI

#1	TS=((“active” OR “graded” OR “partial” OR “part-time”) NEAR/3 (“sick leave” OR “sickleave” OR “sickness-leave” OR “medical-leave” OR “sickness-absence*”))	94
#2	TS=((partial* OR “partly” OR “part-time”) NEAR/3 (sicklist* OR “sick-list*”))	13
#3	TS=((“partial” OR “graded” OR “graduated” OR “part-time”) NEAR/3 (“sickness-benefit*” OR “sickness-certificat*”))	9
#4	TS=((“graded” OR gradual* OR partial* OR “part-time” OR “progressive” OR “step-by-step” OR “step-wise” OR “stepwise”) NEAR/3 (“return-to-	100

	work" OR "RTW" OR "occupational-reintegration" OR "occupational-re-integration" OR "work-exposure"))	
#5	#1 OR #2 OR #3 OR #4	195

Appendix 3. Excluded studies

Reference (First author, year)	Main reason for exclusion
Annerblom ML, 2001 (54)	Not a study about effect
Beemster TT, 2015 (77)	Study protocol of a randomized-controlled trial, which evaluates a multimodal intervention combining PTSL with cognitive-behavioral therapy and relaxation
Floderus B, 2005 (78)	Not an effect study
Godøy A, 2016 (53)	Registry-based study focusing on firm profits
Kann IC, 2012 (26)	Registry-based study with no comparison group (examined association between proportion of graded sick leave and outcomes like overall sickness absence)
Kausto J, 2013 (79)	Conference abstract derived from a registry-based study included in this systematic mapping review (1)
Maeland JG, 2011 (80)	Non-systematic literature review
Olaya-Contreras P, 2009 (81)	Registry-based study with no focus on PTSL
Ose SO, 2012 (46)	Registry-based study with no comparison group
Scheel IB, 2002 (45)	RCT with no focus on the effects of PTSL
Scheel IB, 2002 (82)	This is a secondary publication from the RCT listed above (45)
Steenstra IA, 2006 (83)	No focus on PTSL
Vermeulen SJ, 2010 (84)	Study protocol of a randomized-controlled trial, which includes unemployed people

Appendix 4. Adjusted analyses and covariates in the registry-based studies

The following tables present data on the adjusted analysis across RBs, including the covariates used by the authors. The Norwegian RBs published by Grødem and colleagues (15), and Nossen and Brage (44) did not conduct any adjusted analysis and therefore do not appear in this section.

Andrén 2012 (48)			
Outcome: Disability (reported as full-recovery)			
Statistical analysis	Recursive bivariate probit model Dependent variable: full recovery within the time period	Average treatment effect (ATE): mean difference in the probability of recovering	Treatment effects on the treated (TT): mean difference in the probability of recovering on those who were on PTSL
Covariates	Occupational type* (i.e., legislators and managers, professionals, clerks, service and shop sales, craft and related trades, plant/machine operators, elementary occupations)		
Results	Time-period/Coefficient estimates** ≤30 days (1.50) ≤90 days (1.29) ≤150 days (1.25) ≤210 days (1.22) ≤270 days (1.21) ≤330 days (1.20)	Time-period: ATE (Standard Deviation, SD) ≤30 days: 0.52 (0.03) ≤90 days: 0.35 (0.08) ≤150 days: 0.29 (0.09) ≤210 days: 0.27 (0.09) ≤270 days: 0.25 (0.09) ≤330 days: 0.25 (0.09)	Time-period: TT (SD) ≤30 days: -0.18 (0.10) ≤90 days: -0.01 (0.04) ≤150 days: -0.01 (0.05) ≤210 days: -0.01 (0.06) ≤270 days: 0.00 (0.03) ≤330 days: -0.01 (0.06)

* Occupational type is only used in the selection equation of the regression analysis as instrument for the part-time variable.

** All coefficients were statistically significant $p<0.01$

Andrén 2014 (28)		
Outcome: Disability (reported as full-recovery)		
Statistical analysis	Average treatment effect (ATE): mean difference in the probability of recovering	Treatment effects on the treated (TT): mean difference in the probability of recovering on those who were on PTSL
Covariates	Duration of sick leave, geographic area, male, age, occupation, type of physician who granted the sick leave, level of education, employer, marital status, country of birth, previous sick leaves, income	

Results	Sample: ATE (Standard error) Sample 1: 0.015 (0.27) Sample 2: 0.004 (0.25) Sample 3: 0.387* (0.12) Sample 4: -0.027 (0.14) Sample 5: 0.009 (0.23)	Sample: TT (Standard error) Sample 1: -0.126 (0.21) Sample 2: -0.023 (0.41) Sample 3: 0.428* (0.13) Sample 4: 0.004 (0.19) Sample 5: 0.364* (0.03)

Partial sick leave (PTSL)

Sample 1: all employees on sick leave (either part- or full-time) for a mental disorder; sample 2: employees who started their sick leave due to a mental disorder on full-time basis; sample 3: employees who started their long-term sick leave (more than 60 days) due to a mental disorder on full-time basis; sample 4: sample 2 extended to all diagnoses; sample 5: all diagnoses excluding mental disorders.

*p<0.01

Andrén 2014 (28)					
Outcome: Disability (reported as full-recovery)					
Analysis: Distributional parameters for average treatment effect (ATE) and the effect of treatment on the treated (TT) by sample[1]					
	Only mental disorders (MDs)			All diagnoses	All diagnoses excluding MDs
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Distributional parameters for ATE					
PE	0.178	0.177	0.423	0.129	0.160
PI	0.603	0.581	0.340	0.671	0.631
NI	0.056	0.068	0.202	0.043	0.058
NE	0.163	0.174	0.036	0.156	0.151
Distributional parameters for TT					
PE	0.078	0.187	0.465	0.185	0.446
PI	0.693	0.507	0.179	0.556	0.313
NI	0.025	0.095	0.320	0.078	0.160
NE	0.204	0.210	0.037	0.181	0.082

[1] Positive effect (PE) means that PTSL would result in recovery within a given time span, while FTSL would not; positive indifference (PI) means that the individual will recover within a given time span, regardless of the type of sick leave; negative indifference (NI) means that the individual would not recover regardless of the type of sick leave; negative effect (NE) means that PTSL would not result in recovery within a given time span, while FTSL would.

Bethge 2016 (17)					
Outcome: Disability pension					
	HR (95%CI) Unadjusted	NNT Unadjusted	HR and NNT Adjusted 1	HR and NNT Adjusted 2	ARR
Covariates	NA	NA	Age, sex, diagnostic group	Prompt for rehabilitation by health insurance agency and sickness absence duration	NA

Result	0.62 (0.49 to 0.80)	31	There were no significant interactions. Data were not reported by authors.	Prompted by health insurance agency [HR 0.34; 95% CI 0.18 to 0.62] NNT: 13 Not prompted by health insurance agency [HR 0.72; 95% CI 0.55 to 0.95] NNT: 47 Interaction (p=0.027) Sickness absence duration before rehab (<3 months) [HR 0.81; 95% CI 0.51 to 1.30] Sickness absence duration before rehab (≥3 months) [HR 0.57; 95% CI 0.42 to 0.76] Interaction (p=0.207)	3.2%

Absolute risk reduction (ARR); confidence interval (CI); hazard ratio (HR); not applicable (NA); number needed to treat (NNT)

Bethge 2016 (17)	
Outcome: Allowance of social benefits (reported as duration of receiving welfare benefits due to sickness absence and unemployment from 2007-2009)	
	Unadjusted mean difference and 95% CI
Sickness benefits (days)	52 days; 95% CI 40 to 64 days
Short-term unemployment benefits (days)	58 days; 95% CI 49 to 67 days
Long-term unemployment benefits (days)	15 days; 95% CI 10 to 20 days

Confidence interval (CI)

Høgelund 2010 (52)	
Outcome: return-to-work (RTW)	
	Regression coefficient (standard error)
Total sample (N=934)	0.414 (0.179)[a]
Graded return-to-work durations above 8 weeks (N=862)	0.604 (0.170)[b]
Participants who returned to regular working hours before the 16th week of the sick leave	2.770 (0.324)[b]
Participants who returned to regular working hours after the 16th week of the sick leave	0.773 (0.337)[c]

[a] Statistically significant at 10%

[b] Statistically significant at 1%

[c] Statistically significant at 5%

Kausto 2012 (24)		
Outcome: Disability pension		
Analysis 1: ORs (95% CI) for associations between partial sick leave and disability pensions (OR for full-time sick leave =1.0)		
	Full disability pension	Partial disability pension
Full sample		
Crude	0.3 (0.2 to 0.4)	2.9 (2.4 to 3.5)
Adjusted 1	0.6 (0.4 to 0.8)	2.1 (1.6 to 2.7)
Adjusted 2	0.6 (0.4 to 0.7)	2.3 (1.8 to 3.0)
PS matching	0.6 (0.4 to 0.7)	2.8 (2.0 to 4.0)
Men		
Crude	0.4 (0.3 to 0.6)	3.2 (2.1 to 5.0)
Adjusted 1	0.6 (0.4 to 1.0)	2.7 (1.6 to 4.5)
Adjusted 2	0.5 (0.3 to 0.8)	2.5 (1.4 to 4.4)
PS matching	0.5 (0.3 to 0.8)	2.1 (1.1 to 4.3)
Women		
Crude	0.3 (0.2 to 0.4)	2.5 (1.9 to 3.1)
Adjusted 1	0.6 (0.4 to 0.8)	1.9 (1.4 to 2.5)
Adjusted 2	0.5 (0.4 to 0.8)	2.3 (1.7 to 3.1)
PS matching	0.6 (0.4 to 0.8)	3.1 (2.0 to 4.7)
Mental disorders		
Crude	0.3 (0.2 to 0.4)	3.2 (2.3 to 4.5)
Adjusted 1	0.6 (0.4 to 0.8)	3.2 (2.1 to 4.8)
Adjusted 2	0.6 (0.4 to 0.8)	3.3 (2.2 to 5.0)
PS matching	0.5 (0.4 to 0.8)	5.2 (2.6 to 10.3)
Musculoskeletal disorders		
Crude	0.2 (0.1 to 0.3)	2.4 (1.8 to 3.2)
Adjusted 1	0.5 (0.3 to 0.8)	1.7 (1.2 to 2.5)
Adjusted 2	0.5 (0.3 to 0.8)	1.9 (1.3 to 2.7)
PS matching	0.5 (0.3 to 0.8)	2.1 (1.3 to 3.4)
Traumas and tumors		
Crude	0.6 (0.4 to 1.0)	3.7 (2.0 to 6.9)
Adjusted 1	0.7 (0.4 to 1.4)	1.3 (0.6 to 2.9)
Adjusted 2	0.8 (0.4 to 1.5)	1.2 (0.5 to 3.2)
PS matching	0.8 (0.4 to 1.6)	2.0 (0.7 to 5.7)

Confidence interval (CI); odd ratio (OR); propensity score (PS)

Adjusted 1: Multinomial regression adjusted for covariates (i.e., age, gender, diagnostic group, occupational group, gross income, insurance district, length of sick leave before treatment, and length of sick- leave in connection with treatment), reference=no disability pension.

Adjusted 2: Multinomial regression adjusted for: PS and variables with residual imbalance, reference=no disability pension.

PS matching: This test was used to estimate the effects among the subpopulation of those who either had been granted PTSL or were comparable to them on the covariates (matched sample). It was obtained by suing generalized estimating equation to take into account the matching pairs design.

Kausto 2012 (24)						
Outcome: Disability pension						
Analysis 2: Risk reductions (a risk difference between partial and full sick leave groups)* of the use of partial and full disability pension in the matched subsample						
	Full disability pension			Partial disability pension		
	ARR (95% CI)	RRR (95% CI)	NNT	ARR (95% CI)	RRR (95% CI)	NNT
Total	6 (3 to 9)	41 (24 to 55)	16	-8 (-10 to -5)	-159 (-264 to -84)	-14
Men	10 (4 to 16)	47 (20 to 65)	10	-5 (-10 to -0)	-102 (-302 to -2)	-21
Women	4 (15 to 7)	38 (13 to 55)	26	-9 (-12 to -6)	-180 (-315 to -89)	-12
Mental disorders	8 (3 to 13)	42 (17 to 59)	12	-10 (-14 to -7)	-361 (-795 to -138)	-10
Musculoskeletal disorders	6 (2 to 11)	51 (19 to 70)	15	-7 (-12 to -2)	-94 (-202 to -24)	-15
Traumas and tumors	2 (-4 to 8)	17 (-54 to 56)	53	0 (-8 to 2)	-95 (-409 to 25)	-31

Absolute risk reduction (ARR); confidence interval (CI); relative risk reduction (RRR)

* Negative values mean an increase in risk

Kausto 2014 (1)				
Outcome: Work participation				
Analysis 1: comparison of work participation (%) between PTSL and FTSL (GLM repeated measures design)				
	Pre-intervention Mean (95% CI)	Post-intervention Mean (95% CI)	Post-pre difference Mean (95% CI)*	DID Mean (95% CI) DID in the PS-matched subsample Mean (95% CI)
All[1]				
PTSL	86.6 (85.2 to 88.1)	65.4 (63.4 to 67.4)	−21.2 (−23.4 to −19.1)	0.40 (3.1 to 7.5) **
FTSL	79.4 (79.1 to 79.6)	52.9 (52.5 to 53.2)	−26.5 (−26.9 to −26.2)	9.8 (5.9 to 13.7) **
Males[2]				
PTSL	86.6 (84.0 to 89.1)	62.7 (59.0 to 66.5)	−23.9 (−27.9 to −19.9)	6.3 (2.3 to 10.3) **
FTSL	80.3 (80.0 to 80.7)	50.2 (49.7 to 50.7)	−30.1 (−30.7 to −29.6)	12.4 (6.9 to 17.9)**
Females[2]				
PTSL	85.4 (83.7 to 87.0)	66.9 (64.6 to 69.3)	−18.4 (−21.0 to −15.9)	4.9 (2.4 to 7.5) **
FTSL	78.6 (78.2 to 78.9)	55.2 (54.7 to 55.7)	−23.4 (−23.9 to −22.9)	7.2 (3.1 to 11.4)**
16–34 years[1]				
PTSL	89.3 (85.8 to 92.8)	75.5 (70.2 to 80.9)	−13.8 (−19.6 to −8.0)	2.8 (−1.1 to 10.6)
FTSL	84.6 (84.1 to 85.1)	66.1 (65.3 to 66.8)	−16.6 (−20.8 to −12.5)	

				8.5 (0.5 to 16.6)**
35–44 years[1]				
PTSL	84.7 (81.9 to 87.5)	68.1 (64.2 to 72.0)	−16.6 (−20.8 to −12.5)	2.0 (−2.2 to 6.2)
FTSL	78.4 (77.9 to 79.0)	59.8 (59.1 to 60.5)	−18.6 (−19.4 to −17.8)	6.7 (0.7 to 12.6)**
45-54 years[1]				
PTSL	86.9 (84.7 to 89.0)	65.7 (62.6 to 68.8)	−21.1 (−24.4 to −17.9)	4.7 (1.4 to 8.0) **
FTSL	77.6 (77.2 to 78.1)	51.8 (51.2 to 52.4)	−25.9 (−26.5 to −25.2)	11.1 (6.3 to 15.9)**
55-65 years[1]				
PTSL	89.6 (86.3 to 92.9)	57.0 (52.3 to 61.7)	−32.6 (−37.7 to −27.5)	0.40 (0.5 to 10.8) **
FTSL	78.5 (78.0 to 78.9)	40.2 (39.5 to 40.8)	−38.3 (−39.0 to −37.6)	12.9 (6.5 to 19.4) **
Musculoskeletal diseases[3]				
PTSL	87.0 (84.8 to 89.3)	60.3 (57.0 to 63.6)	−26.7 (−30.3 to −23.2)	0.7 (−2.9 to 4.3)
FTSL	79.7 (79.4 to 80.1)	52.3 (51.7 to 52.9)	−27.4 (−28.0 to −26.8)	6.3 (1.5 to 11.2)**
Mental disorders[3]				
PTSL	84.6 (82.2 to 87.1)	67.0 (63.8 to 70.3)	−17.6 (−21.3 to −13.9)	12.8 (9.0 to 16.5) **
FTSL	74.6 (74.0 to 75.1)	44.2 (43.5 to 44.9)	−30.4 (−31.1 to −29.6)	18.9 (14.2 to 23.5)**
Traumas[3]				
PTSL	86.7 (82.0 to 91.3)	68.1 (61.5 to 74.6)	−18.6 (−25.3 to −11.8)	−3.2 (−10.0 to 3.5)
FTSL	82.9 (82.3 to 91.3)	67.6 (66.7 to 68.4)	−15.3 (−16.2 to −14.5)	0.3 (-9.3 to 9.9)
Tumors[3]				
PTSL	90.6 (85.9 to 95.4)	75.0 (67.4 to 82.5)	−15.7 (−23.5 to −7.9)	0.40 (−2.6 to 13.2)
FTSL	87.2 (86.3 to 88.1)	66.2 (64.8 to 67.6)	−21.0 (−22.4 to −19.5)	12.5 (1.8 to 23.2) **
Other diagnostic categories[3]				
PTSL	87.4 (83.4 to 91.4)	63.6 (57.8 to 69.4)	−23.8 (−30.0 to −17.6)	6.2 (−0.05 to 12.5)
FTSL	80.2 (79.6 to 80.7)	50.1 (49.3 to 50.9)	−30.0 (−30.9 to −29.2)	11.1 (3.3 to 18.9)**

Confidence interval (CI); difference-in-differences (DID); general lineal model (GLM)

[1] Adjusted for age, sex, income, diagnosis, occupational group, insurance district

[2] Adjusted for age, income, diagnosis, occupational group, insurance district

[3] Adjusted for sex, income, occupational group, insurance district

*all post-pre differences were statistically significant at p<0.001

**p<0.05

Lie 2014 (43)		
Outcome: Work participation (reported as RTW)		
	Unadjusted analysis	Adjusted analysis
Covariates	NA	Age, sex, diagnosis, time in job, and doctor's data

	HR	(95% CI)	P value	HR	(95% CI)	P value
0 weeks (start)	0.27	(0.27 to 0.28)	<0.001	0.37	(0.36 to 0.38)	<0.001
0-2 weeks	0.71	(0.70 to 0.72)	<0.001	0.79	(0.77 to 0.80)	<0.001
2 weeks	0.77	(0.76 to 0.78)	<0.001	0.70	(0.68 to 0.71)	<0.001
4 weeks	0.82	(0.81 to 0.83)	<0.001	0.78	(0.76 to 0.79)	<0.001
8 weeks	0.83	(0.81 to 0.84)	<0.001	0.84	(0.82 to 0.85)	<0.001
12 weeks	0.84	(0.83 to 0.86)	<0.001	0.90	(0.88 to 0.92)	<0.001
16 weeks	0.92	(0.90 to 0.94)	<0.001	0.99	(0.96 1.02)	0.42
20 weeks	0.98	(0.95 to 1.01)	0.11	1.03	(1.00 to 1.06)	0.079
24 weeks	1.07	(1.03 to 1.10)	<0.001	1.13	(1.09 to 1.17)	<0.001
28 weeks	1.17	(1.13 to 1.22)	<0.001	1.24	(1.19 to 1.29)	<0.001
32 weeks	1.32	(1.27 to 1.38)	<0.001	1.38	(1.31 to 1.44)	<0.001
36 weeks	1.34	(1.28 to 1.40)	<0.001	1.36	(1.30 to 1.43)	<0.001
40 weeks	1.26	(1.20 to 1.31)	<0.001	1.25	(1.19 to 1.31)	<0.001
44 weeks	0.82	(0.78 to 0.85)	<0.001	0.83	(0.79 to 0.86)	<0.001

Separated data for women and men neither data on recurrent sick leaves (i.e., first, second and third leave) are not presented here and can be found in the original publication (43).
Confidence interval (CI); hazard ratio (HR); return-to-work (RTW)

Lie 2014 (43)						
Outcome: Allowance of social benefits (reported as rehabilitation)						
Covariates	Unadjusted analysis			Adjusted analysis		
	NA			age, sex, diagnosis, time in job, and doctor's data		
	HR	(95% CI)	P value	HR	(95% CI)	P value
0 weeks (start)	1.33	(1.22 to 1.44)	<0.001	1.11	(1.01 to 1.22)	0.033
2 weeks	0.92	(0.83 to 1.03)	0.16	0.98	(0.87 to 1.09)	0.67
4 weeks	0.90	(0.82 to 1.00)	0.044	0.95	(0.85 to 1.06)	0.34
8 weeks	1.03	(0.92 to 1.15)	0.60	1.05	(0.94 to 1.18)	0.38
12 weeks	0.81	(0.72 to 0.91)	<0.001	0.81	(0.72 to 0.91)	0.001
16 weeks	0.87	(0.77 to 0.97)	0.013	0.84	(0.74 to 0.95)	0.006
20 weeks	0.97	(0.86 to 1.09)	0.58	0.99	(0.87 to 1.12)	0.85
24 weeks	0.85	(0.76 to 0.96)	0.009	0.89	(0.78 to 1.01)	0.081
28 weeks	0.83	(0.74 to 0.94)	0.003	0.80	(0.69 to 0.93)	0.003
32 weeks	0.86	(0.76 to 0.97)	0.013	0.89	(0.78 to 1.02)	0.090
36 weeks	0.71	(0.63 to 0.80)	<0.001	0.72	(0.63 to 0.81)	<0.001
40 weeks	0.72	(0.64 to 0.81)	<0.001	0.73	(0.65 to 0.83)	<0.001
44 weeks	1.15	(1.00 to 1.32)	0.053	1.18	(1.02 to 1.38)	0.027

Separated data for women and men neither data on recurrent sick leaves (i.e., first, second and third leave) are not presented here and can be found in the original publication (43).

Confidence interval (CI); hazard ratio (HR)

Lie 2014 (43)						
Outcome: Disability						
	Unadjusted analysis			Adjusted analysis		
Covariates	NA			age, sex, diagnosis, time in job, and doctor's data		
	HR	(95% CI)	P value	HR	(95% CI)	P value
0 weeks (start)	1.89	(1.53 to 2.34)	<0.001	1.53	(1.23 to 1.92)	<0.001
2 weeks	0.88	(0.66 to 1.18)	0.39	1.02	(0.75 to 1.40)	0.90
4 weeks	0.91	(0.68 to 1.23)	0.55	1.01	(0.75 to 1.37)	0.93
8 weeks	1.24	(0.91 to 1.68)	0.17	1.34	(0.98 to 1.84)	0.070
12 weeks	0.87	(0.62 to 1.22)	0.42	0.94	(0.66 to 1.34)	0.73
16 weeks	1.05	(0.73 to 1.50)	0.81	1.06	(0.74 to 1.53)	0.75
20 weeks	0.84	(0.60 to 1.19)	0.33	0.84	(0.59 to 1.20)	0.34
24 weeks	0.98	(0.68 to 1.43)	0.93	1.02	(0.68 to 1.54)	0.91
28 weeks	0.72	(0.48 to 1.07)	0.11	0.73	(0.49 to 1.09)	0.12
32 weeks	0.86	(0.59 to 1.24)	0.41	0.85	(0.58 to 1.25)	0.42
36 weeks	0.95	(0.67 to 1.35)	0.77	0.92	(0.64 to 1.31)	0.63
40 weeks	0.72	(0.51 to 1.00)	0.050	0.63	(0.45 to 0.89)	0.010
44 weeks	0.92	(0.66 to 1.27)	0.60	0.80	(0.57 to 1.11)	0.18

Separated data for women and men neither data on recurrent sick leaves (i.e., first, second and third leave) are not presented here and can be found in the original publication (43).
Confidence interval (CI); hazard ratio (HR)

Markussen 2012 (49)						
Outcome: Sickness absence, work participation, and allowance of social welfare benefits (data from spells exceeding 8 weeks)						
	Baseline model (The grading decision) Estimates (standard error)		Full model Estimates (standard error)		Full model/physician char. Estimates (standard error)	
	OLS[1]	SLS[2]	OLS[1]	SLS[2]	OLS[1]	SLS[2]
Covariates	Sickness diagnosis, patient/job characteristics (age, marital status, nationality, education, industry, work-hours, and earnings), and calendar time.		This model also controls for patient's geographical location (neighborhood), patient history in terms of employment, earnings, and sick leave during the past three years prior to the spell used in our analysis.		This model adds a vector of physician characteristics[3]	
Sickness absence duration (days)	−59.6 (0.5)	−58.8 (8.0)	−60.2 (0.5)	−66.1 (10.8)	−60.1(0.5)	−62.6 (11.3)
Work participation (# fulltime-	−63.9 (0.3)	−93.8 (4.7)	−63.9 (0.3)	−97.8 (6.5)	−63.9 (0.3)	−96.7 (7.0)

equiv. lost working days in abs. spell)						
Allowance of social welfare benefits (# Fulltime-equiv. soc. Sec. days next two years)	−26.8 (0.5)	−102.3 (8.2)	−26.7 (0.5)	−91.0 (9.7)	−26.5 (0.5)	−79.7 (11.8)
Employment probability	0.10 (0.0)	0.21 (0.0)	0.10 (0.0)	0.19 (0.0)	0.10 (0.0)	0.16 (0.0)

Ordinary least squares (OLS); semiparametric least squares (SLS)

[1]OLS captures the interaction of both observed covariates and physician's grading propensity on the probability of obtaining a PTSL.

[2]In addition to OLS, the SLS model removes sorting bias, basing the estimates only in the participants whose grading outcomes were influenced by the physician's grading propensity.

[3]Physician characteristics: age; specialist education; gender; sharing office with other physicians; fixed or variable salary; taking part in emergency service; number of patients on capitation list; desired number of patients relative to actual number; physician leniency (centiles in the distribution of estimated leniency for issuing short-term and long-term absence certificates); one scalar indicator for treatment quality.

Markussen 2012 (49)		
SLS data from the full model/physician characteristics on alternative sick leave duration[1] (spells exceeding 4 and 12 weeks)		
Outcome: Sickness absence, work participation, and allowance of social welfare benefits (data from spells exceeding 4 and 12 weeks)		
Outcomes	Spells exceeding 4 weeks SLS estimate (standard error)	Spells exceeding 12 weeks SLS estimate (standard error)
Sickness absence duration (days)	-47.1 (8.8)	-74.8 (10.3)
Work participation (# fulltime-equiv. lost working days in abs. spell)	-70.1 (5.2)	-117.1 (8.2)
Allowance of social welfare benefits (# Fulltime-equiv. soc. Sec. days next two years)	-67.7 (8.6)	-99.8 (15.6)
Employment probability	0.13 (0.0)	0.20 (0.0)

Semiparametric least squares (SLS)

[1] In addition to the ordinary least squares (OLS) analysis, which captures the interaction of both observed covariates and physician's grading propensity on the probability of obtaining a PTSL, the SLS model removes sorting bias, basing the estimates only in the participants whose grading outcomes were influenced by the physician's grading propensity.

Streibelt 2017 (50)			
Outcomes: Full RTW, permanent disability pension and unemployment			
	PTSL	FTSL	RR (95% CI)
Covariates			None

Results			
Full RTW	88.4%	72.6%	1.22 (1.13 to 1.31)
Permanent disability pension	5.0%	11.3%	0.40 (0.23 to 0.70)
Unemployment	6.6%	16.1%	0.41 (0.26 to 0.65)

Confidence interval (CI); full-time sick leave (FTSL); partial sick leave (PTSL); relative risk (RR); return-to-work (RTW)

Streibelt 2017 (50)				
Outcomes: sick leave duration, physical role, emotional role, depression and anxiety, work ability				
	PTSL Mean (SD) at follow-up	FTSL Mean (SD) at follow-up	Linear regression coefficient (95% CI)	P value
Covariates			Disability score, stress, RTW beliefs and expectations, job satisfaction and security, sociodemographic data, and work-related information, risk of disability pension[1]	
Results				
Sick leave duration (weeks)	7.1 (13.2)	13.4 (18.3)	−6.3 (4.0 to 8.6)	<0.001
Physical role	49.2 (40.8)	41.3 (40.2)	7.9 (1.9 to 13.9)	0.01
Emotional role	41.0 (40.3)	34.2 (39.4)	6.8 (0.8 to 12.7)	0.02
Depression and anxiety	5.1 (3.3)	5.7 (3.5)	−0.6 (−1.1 to −0.1)	0.030
Work ability	2.9 (0.9)	2.8 (1.0)	0.1 (−0.1 to 0.3)	0.058

Confidence interval (CI); full-time sick leave (FTSL); partial sick leave (PTSL); standard deviation (SD)

[1] Covariates used in the linear regression model: A 9-item measure disability score (SIBAR) assessed the risk of permanent work disability, unspecified measures on work stress, return-to-work beliefs and expectations as well as job satisfaction and job security, sociodemographic data, work-related information (job status, time of sick leave prior to rehabilitation), and the risk index disability pension (RI-DP)

Streibelt 2017 (50)							
Outcomes: Full RTW and sick leave duration							
	Full RTW			Sick leave duration (weeks)			
Subjective prognosis of RTW[1]	%	OR[2] (95%CI)	P value	Baseline	Follow-up	Linear regression coefficient (95% CI)	P value
Positive							

PTSL	92.9	1.79 (0.67 to 4.77)	0.35	16.9	6.0	−1.5 (−4.7 to 1.6)	0.33
FTSL	87.9			14.9	7.5		
Unsure							
PTSL	90.1	3.42 (1.97 to 5.91)	<0.001	22.7	7.8	−6.9 (−10.3 to −3.6)	<0.001
FTSL	72.6			2.7	14.7		
Negative							
PTSL	73.6	2.87 (1.31 to 6.27)	0.008	28.6	6.5	−11.7 (−18.3 to −5.2)	<0.001
FTSL	49.2			29.6	18.2		

Confidence interval (CI); full-time sick leave (FTSL); odds ratio (OR); partial sick leave (PTSL); return-to-work (RTW)

[1]Do you believe that you will go back to work after rehabilitation? (Answers: negative: no/probably not, unsure; positive: probably/yes)

[2]Logistic regression model

Viikari-Juntura 2017 (51)						
Outcomes: Work participation (reported as sustained return-to-work) and both partial and full disability retirement						
	Sustained return to work		Full disability retirement		Partial disability retirement	
	PTSL %	FTSL %	PTSL %	FTSL %	PTSL %	FTSL %
Covariates	Proportions are adjusted for imbalanced covariates: age, employment sector and industrial sector					
Total sample	81.9	73.2	2.6	6.8	8.4	1.3
Men	84.1	74.0	4.1	8.3	5.5	0.5
Women	81.2	73.0	2.2	6.3	9.2	1.5
20-44 years old	81.9	75.8	0	0.5*	0.5	0
45-54 years old	84.3	74.8	1.7	4.6	8.3	1.9
55-64 years old	76.6	68.2	8.0	18.9	22.3	3.8
Mental disorders	86.5	66.3	2.0	7.2	4.7	0.9
Musculoskeletal diseases	78.1	78.9*	3.1	6.5	11.3	1.7
Private sector	81.2	71.1	2.1	6.0	3.9	0.7
Public sector	82.4	75.5	3.1	7.7	14.1	2.3
Manufacturing	79.7	64.0	2.3	7.3	3.8	1.0*
Wholesale and retail trade; repair of motor vehicles and motorcycles	82.3	76.7*	2.9	3.3*	2.2	0
Technical and scientific work etc.	83.3	71.6	1.7	7.0	5.8	0.9
Human health and social work activities	81.4	76.5	2.5	6.7	14.7	3.2

Full-time sick leave (FTSL); partial sick leave (PTSL)

*non statistically significant difference (p >0.05)

Appendix 5. Definition of the statistical analyses in the registry-based studies

Analysis	Definition
Average treatment effect (ATE) (85)	This estimates the difference between the probability that the individual will fully recover after part-time sick leave and the probability that the individual will fully recover after full-time sick leave.
Difference-in-differences (DID) (86;87)	This method is used to estimate the difference in pre-post, within participant, differences between the intervention and the comparison group.
Distributional treatment parameter (28)	This analysis incorporates the distribution of the effect parameters (i.e., ATE and TT) to calculate the probability of four different events: successful, positive indifference, negative indifference, and unsuccessful
General lineal model (GLM) (88)	This is a general formal regression model, which is achieved by transformation of the dependent variable (response).
Generalized estimating equation (89)	This analysis is used to estimate the parameters of a generalized linear model with a possible unknown correlation between outcomes
Hazard ratio (HR) (90)	"The hazard ratio (HR) is the main, and often the only, effect measure reported in many epidemiologic studies. For dichotomous, non–time-varying exposures, the HR is defined as the hazard in the exposed groups divided by the hazard in the unexposed groups. For all practical purposes, hazards can be thought of as incidence rates and thus the HR can be roughly interpreted as the incidence rate ratio. The HR is commonly and conveniently estimated via a Cox proportional hazards model, which can include potential confounders as covariates."
Linear regression (88)	This analysis aims to determine the relationships between two continuous (quantitative) variables.
Logistic regression (91)	"Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes)".
Multinomial regression analysis (92)	This regression analysis aims to establish the relationship between one nominal dependent variable and one or more independent variables when the dependent variable is nominal with more than two levels.
Number needed to treat (NNT) (93)	"NNT is an absolute effect measure which is interpreted as the number of patients needed to be treated with one therapy versus another for one patient to encounter an additional outcome of interest within a defined period of time. The computation of NNT is founded on the cumulative incidence of the outcome per number of patients followed over a given period of time, being classically calculated by inverting absolute risk reduction (ARR) (also called risk difference [RD]) between two treatment options".
Ordinary least squares (OLS) (94)	The OLS is a linear least squares method, which aims to determine the unknown parameters in a linear regression model. The principle of

	least squares implies minimizing the sum of square differences between the observed and predicted parameters.
Propensity-score matching analysis (95)	"The propensity score is the probability of treatment assignment conditional on observed baseline characteristics. The propensity score allows one to design and analyze an observational (nonrandomized) study so that it mimics some of the particular characteristics of a randomized controlled trial. In particular, the propensity score is a balancing score: conditional on the propensity score, the distribution of observed baseline covariates will be similar between treated and untreated subjects".
Recursive bivariate probit model (96)	This analysis estimates the treatment effect that a binary endogenous variable (PTSL) exerts on a binary outcome (full recovery), after controlling for a set of covariates.
Semiparametric least squares (SLS) (97)	"Semiparametric regression includes regression models that combine parametric and nonparametric models. They are often used in situations where the fully nonparametric model may not perform well or when the researcher wants to use a parametric model but the functional form with respect to a subset of the regressors or the density of the errors is not known".
Treatment on the treated effect (TT) (85)	This indicates the average effect of treatment only on those who have been treated.

Appendix 6. Quality appraisal of the randomized controlled trial

Quality appraisal of the included randomized controlled trial

Cochrane tool for assessment of risk of bias	**Shiri et al. 2013 (47)** **Viikari-Juntura et al. 2012 (25)**
Was the allocation sequence adequately generated?	Yes
Was the allocation adequately concealed?	Yes
Were baseline outcome measurements similar?	Yes
Were baseline outcome characteristics similar?	Yes
Were incomplete outcome data adequately addressed?	Yes
Was knowledge of the allocated intervention adequately prevented during the study?	No
Was the study adequately protected against contamination?	Yes
Are reports of the study free of suggestion of selective outcome reporting?	No
Is the study free of other sources of bias?	Yes
Overall quality assessment	Moderate

Appendix 7. Quality appraisal of the registry-based studies

The table below gives the quality appraisal of the 12 registry-based studies

Quality criteria	Andrén 2012 (48)	Andrén 2014 (28)	Bethge 2016 (17)	Grødem 2015 (15)	Høge-lund 2010 (52)	Kausto 2012 (24)	Kausto 2014 (1)	Lie 2014 (43)	Markus-sen 2012 (49)	Nossen 2013 (44)	Streibelt 2017 (50)	Viikari-Juntura 2017 (51)
Were the groups comparable for important background factors?	No	Yes	No	Yes	Unclear	No	Yes	Unclear	No	No	Yes	Yes
Were the exposed individuals representative of a defined population?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was the control group(s) selected from the same population as the exposed group(s)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Was the study prospective?	No	No	No	No	No	No	No	No	No	No	Yes	No
Was exposure and outcome measured equally and reliably in the groups?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were many enough people in the cohort followed-up?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

An analysis of attrition was done to explain whether those who have abandoned the study differ from those who have been followed-up?	No	No	No	No	No	No	No	No	No	No	No	No
Was the follow-up time long enough to show positive and/or negative outcomes?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were known, possible confounding factors taken into account in the design and/or analysis of the study?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes
Was the person who assessed the results (endpoints) blinded to who was exposed and who was not exposed?	No	No	No	No	No	No	No	No	No	No	No	No
Overall assessment	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	High	Moderate

Published by the Norwegian Institute of Public Health
Desember 2018
P.O.B 4404 Nydalen
NO-0403 Oslo
Telefon: + 47-21 07 70 00
The report can be downloaded as pdf
at www.fhi.no/en/publ/