Innovations in use of registry data (INOREG) – Design of a registry-based study analyzing care pathways and outcomes for chronic patients

TRON ANDERS MOGER[1*], OLAV AMUNDSEN[2], TROND TJERBO[1], RAGNHILD HELLESØ[3], JON H. HOLTE[1], NINA K. VØLLESTAD[2]

[1] Institute of Health and Society, Dept. of Health Management and Health Economics, University of Oslo, Norway
[2] Institute of Health and Society, Dept. for Interdisciplinary Health Sciences, University of Oslo, Norway
[3] Institute of Health and Society, Dept. of Public Health Science, University of Oslo, Norway

Abstract: In recent years there have been several political initiatives in Norway, requiring more research into how multimorbidity and health care pathways in the municipality affect outcomes such as work participation, hospital admissions, disability and quality of life for patients with chronic diseases. Most of the care is provided outside hospitals and has been difficult to capture in large, registry-based studies. Focusing on two important groups, patients with chronic obstructive pulmonary disease (COPD) and musculoskeletal disorders (MSD), the INOREG project aims to reduce these knowledge gaps. In the paper we present 1) the data that are used in the project, 2) the construction of samples, variables and possible methods for analysis and 3) an example on how the data and methods will be applied. The project database is constructed from a novel linkage of national health and welfare registries. The data cover social, primary and specialized care for all COPD and MSD patients in Norway, long-term care data from Oslo and Trondheim municipalities and functioning and quality of life for ca. 2,700 patients treated at physiotherapy clinics in the FYSIOPRIM project. This enables construction of care pathways and outcomes at the individual level from 2008 through 2019. The project will fill knowledge gaps regarding the patterns of care at different levels in the health care system, and the association to outcomes for chronic patient groups. If the project is successful, it will provide improved insight on how to further develop provision and coordination of services to the decision makers, and ideally reduce inequalities in health.

1 Introduction

In Norway, as in most countries, there is a high degree of specialization and differentiation within health care. As a result, many patients need to navigate in a highly complex web of care (Goodwin *et al.*, 2021). This particularly applies to patients with chronic diseases. Despite long-standing efforts to ensure continuity of care, problems caused by poor coordination are regarded as a major policy concern (Ministry of Health and Care Services, 2014). Numerous studies and evaluations show that problems often arise in the interface between different service providers and that lack of integration may negatively affect patients' outcomes (Amelung *et al.*, 2021). However, given the high policy interest in integrated care, surprisingly little is known about the patterns of health service use for individual patients. With some notable exceptions (e.g. Gershon *et al.*, 2012, Henri *et al.*, 2021, and McKay *et al.*, 2022), few studies examine how the use of health services for individual patients develop over time. Even less is known about the association between different types of care pathways (history of care across providers over time, types and contents of care) and patient outcomes. To the best of our knowledge no previous study captures the complete care pathways within and across primary and secondary care for chronic diseases. It is therefore a clear need for more information on variations in the patients' health service use over time, and how this may influence outcomes.

Chronic diseases are the leading cause for disease burden in terms of disability-adjusted life-years, and in particular when assessed as years lived with disability (Vos *et al.*, 2010, and Murray *et al.*, 2012). They are characterized by being long-lasting or having episodic flares over time. The patients use health services across levels of care, have increased risk of hospitalization and often suffer from multimorbidity (van den Bussche *et al.*, 2011). In addition, many patients are still active in the work force, and the diseases likely lead to increased risk of sick leaves and thus pose a burden on disability benefits. For example, chronic obstructive pulmonary disease (COPD) and two of the most prevalent musculoskeletal disorders (MSD), neck and low back pain, are ranked among top five of the causes for years lived with disability world-wide (Vos *et al.*, 2010, and Murray *et al.*, 2012) and in Norway (Institute for Health Metrics and Evaluation, 2016). They are among the most common diagnoses found in patients with multimorbidity (van den Bussche *et al.*, 2011). A Norwegian study estimated that the annual cost related to health care for the three most affected groups of COPD patients was €105million in 2009 (Nielsen *et al.*, 2011). Moreover, in a report from 2013 (Lærum *et al.*, 2013) the authors estimate the annual health care costs of MSD in Norway to be around €1.5 billion. For both patient groups, slightly more than 50 % of the costs were due to treatment in specialized health care. Hence, describing and understanding care pathways for patients with chronic diseases is particularly important. Improving service delivery for these patients has a high potential for improving the quality of life for the patient and at the same time reduce overall societal costs.

Numerous studies show that patients with chronic diseases may benefit from improved care integration. For example, studies on COPD patients show that better care integration prevents exacerbations, hospitalizations and readmissions (Casas *et al.*, 2006), and improves quality of life (Koff *et al.*, 2009) Improved care integration for patients with low back pain has been demonstrated to reduce sick leaves, disability and societal costs (Lambeek *et al.*, 2010, and Hill *et al.*, 2011). However, the studies are generally small-scaled and focus on specific interventions only. The evaluation of such effects in large, observational data is lacking in the literature.

The project INnovations in use Of REGistry data (INOREG) was established to reduce the knowledge gaps outlined above. The aims of the project are first to describe care pathways for MSD and COPD patients as observed in a population-based sample, and

second to identify health care factors in the pathway associated with improved outcomes. A third important result is to test the feasibility of using registry data in order to achieve the former two aims. This paper presents the database constructed in the project, and how care patterns and care variables particularly tailored for COPD and MSD can be identified. Section 2 gives an introduction to the INOREG project and the database. In Section 3 we discuss variable definitions and possible strategies for analysis, and show an illustrative example in Section 4. We present estimated pathways across general practitioner (GP), physiotherapist and specialist health care in MSD patients over time, some characteristics of patients included in the pathways, and discuss possible implications of the results. Throughout we discuss the possibilities and challenges when analyzing care pathways constructed from registry data. Our overall objective is to provide decision makers with improved insight on how to further develop provision and coordination of services, and ideally reduce inequalities in health.

2 The INOREG database

2.1 INOREG in a nutshell

INOREG is an interdisciplinary collaborative project across departments at the Institute of Health and Society, University of Oslo. The project will add to the current literature by providing a better understanding of the care pathways observed in a real life setting, and identify types and patterns of care associated with improved outcomes. We will use COPD and MSD as cases for chronic diseases, as they represent patient groups with a distinct disease (COPD) or more unspecific symptom-based diseases often with less clear biological foundation (MSD). Due to the long-term prospects of the diseases, they also share several relevant outcomes within the societal (work participation, disability pension) and personal areas (functioning, quality of life), as well as for complex health needs (hospitalizations, long-term care including home services and overall health care costs).

The project consists of a quantitative part, which is the focus of the present paper, and a qualitative part. The qualitative part aims to gain a more comprehensive understanding and insight into the care pathways identified in the quantitative analyses, and make us able to uncover phenomena that may not be reflected in the registry data. This approach enables elaboration of the underlying care processes. For example, how to interpret variations in patients' pathways and use of health care services, in cases where we would expect these to be similar. Vignettes based on the pathways identified in the quantitative analysis will be used in interviews with health care professionals involved in the care for COPD and MSD patients. We also plan to interview patients to obtain in-depth information about their care pathways. Patients are recruited from collaborating hospitals in INOREG and from providers in the municipalities.

2.2 Data sources in INOREG

We utilize the unique opportunities for data linkage at the individual level in Norway. Our data include all levels of health care as far back as 2008. A summary of the data sources included in INOREG, reflecting the levels of care, is given in Table 1.

Table 1: **Registries included in INOREG according to care level/sector. Dates, and durations where relevant, is included for all health care and welfare/employment variables.**

Care level/sector	Data source and population	Owner/administrator	Health services covered	Variables extracted
Primary and municipality health care	KUHR - Control and Payment of Reimbursement to Health Service Providers *All COPD and MSD patients in Norway*	Directorate of Health/ Norwegian Health Economics Administration	General practitioners Emergency medical services Physiotherapists Chiropractors	Practitioner id Profession of practitioner Type of consultation, tests and images from fee codes ICPC-2 diagnosis
	Municipality electronic patient journal *COPD and MSD patients in Oslo/Trondheim*	Oslo and Trondheim municipality/ Norwegian Health Economics Administration	Home nursing Nursing homes Assisted living Rehabilitation at home Rehabilitation in institution	Type of service provided Need for assistance (individualized for relevant activities of daily living)
	Regular GP registry *All regular GPs in Norway*	Directorate of Health/ Norwegian Health Economics Administration	Regular GPs	GP id-number Age Gender Practice type (single/group) Specialization
	FYSIOPRIM database *COPD and MSD patients in included in FYSIOPRIM*	University of Oslo	Physiotherapists	Disability and function (EQ5D, Patient-Specific Functional Scale) Pain (intensity, duration, body regions affected) Emotional distress (Hopkins Symptom Check List – 10 items) Pain self-efficacy Comorbidity Treatments provided
Specialized health care	NPR - Norwegian Patient Registry *All COPD and MSD patients in Norway*	Directorate of Health/ Norwegian Health Economics Administration	Hospital services (in- and outpatient) Rehabilitation in institution	ICD-10 diagnoses (main and secondary) Procedures provided Costs as given by diagnosis related group weights
	KUHR - Control and Payment of Reimbursement to Health Service Providers *All COPD and MSD patients in Norway*	Directorate of Health/ Norwegian Health Economics Administration	Contract specialists (outpatient)	Practitioner id Type of consultation, tests and images from fee codes ICD-10 diagnosis

Table 1: **Registries included in INOREG according to care level/sector. Dates, and durations where relevant, is included for all health care and welfare/employment variables (continued).**

| Care level/sector	Data source and population	Owner/administrator	Health services covered	Variables extracted
Welfare and work participation	FD-trygd *All COPD and MSD patients in Norway*	Norwegian Labor and Welfare Service/ Statistics Norway	-	Registered sick days Registered unemployment days Periods of employment Disability pension
Other	Sociodemographics and –economics *All COPD and MSD patients in Norway*	Statistics Norway	-	Age Gender Income Education Wealth Marital status Employment status Municipality of residence
	Cause of Death Registry *All COPD and MSD patients in Norway*	National Institute of Public Health	-	Date of death

We have access to administrative hospital data on admission diagnosis, comorbidities, and type of specialist services from the Norwegian Patient Registry (NPR, using ICD-10 codes for diagnoses). We also have access to data from primary care on numbers and types of consultations, diagnosis (ICPC-2 and ICD-10), referrals, tests and images registered by GPs, contract specialists, emergency medical services, physiotherapists and chiropractors from the KUHR registry (Kontroll og Utbetaling av HelseRefusjoner). KUHR is the national registry for all primary care contacts who receive reimbursement from The Norwegian Health Economics Administration (HELFO). Data from NPR and KUHR cover COPD and MSD patients in all of Norway. In addition, municipalities in Norway record data for individuals receiving any type of long-term care services in an electronic patient journal (municipality EPJ). The municipality EPJ is developed to describe the level of resource use and need for the users. These data include the need for assistance and type of services provided. We have access to the data for individuals in the municipalities of Oslo and Trondheim, as we are collaborating with the Departments of Health in the two cities.

The abovementioned registries have limited data on functioning. The scores on need for assistance in the municipal EPJ can be a potential source we can use in the analyses, although the validity may be restricted by the variation in time points at which scores are updated across patients. For MSD, INOREG will take advantage of data from a large number of physiotherapy clinics collected in FYSIOPRIM. The database has one-year follow-up data EQ-5D and Patient-Specific Functional Scale for about 2,700 patients with MSD since 2015, mainly from Trondheim and Oslo. The database includes information on treatment and functioning for patients managed by physiotherapists in the participating municipalities.

Finally, we have access to data on welfare and work participation from FD-trygd, and socioeconomic and –demographic information from Statistics Norway. We also have date of death from the Cause of Death Registry. These data cover COPD and MSD patients in all of Norway.

Regarding completeness of the health and care data, the providers have an incentive to register all activity in NPR and KUHR in order to be reimbursed. Hence, these registries should be complete with regard to activity and services yielding reimbursements. In a previous evaluation close to 100% of patients in the Norwegian COPD quality registry were identified using the diagnoses in NPR. The same applied to between 85% and 98% for subgroups in the MSD quality registry (Norwegian Directorate of Health, 2012). Thus, most of the COPD and MSD patients can be identified from NPR alone. According to the guidelines in Municipality EPJ, assessment should be repeated when there is a change in need or in care delivery. However, there is no financial incentive for the municipality to do the registrations, hence the data on long-term care may be less complete and subject to coding errors. Still, for municipality EPJ nationally, close to 100% of the individuals receiving services had a valid score on the need for assistance in 2017. The same applied to the type of service received (Beyrer *et al.*, 2018). FYSIOPRIM has previously been shown to include a representative sample of MSD patients followed up by physiotherapists in primary health care (Evensen *et al.*, 2018).

Figure 1: **Overview of data sources used in the sample selection and construction. In the left column, samples are identified from ICD-10 codes in the Norwegian Patient Registry (NPR) and ICPC-2 codes in Control and Payment of Reimbursement to Health Service Providers (KUHR). In the middle column, the data sources are used to define and construct variables capturing care pathways, general health status (e.g. comorbidities), sociodemographic and -economic information for the patients. The right column shows data sources used to construct outcome variables.**
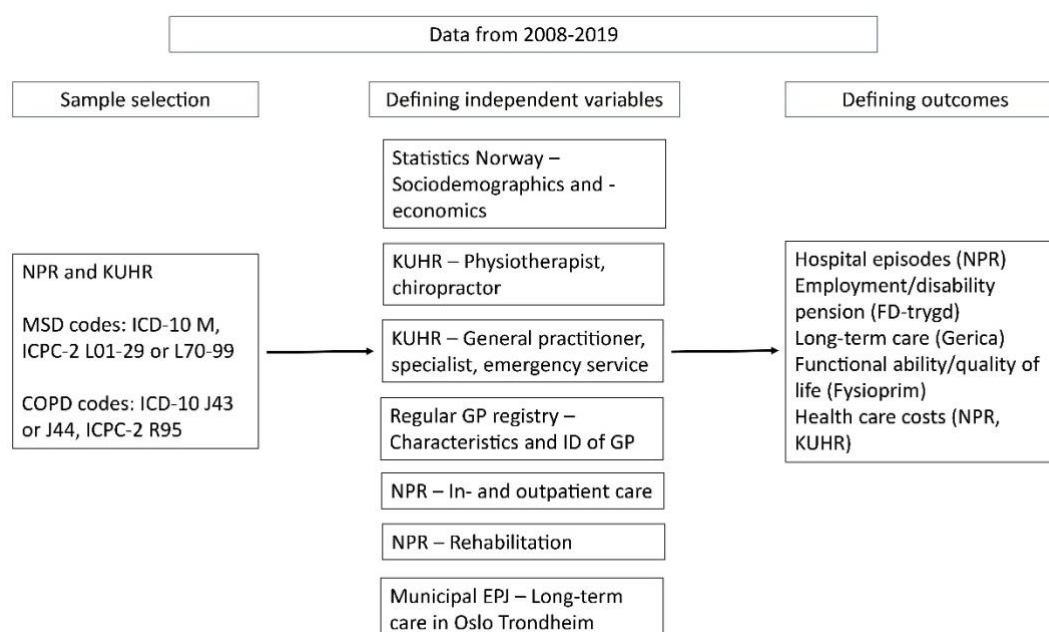
2.3 Sample selection

Figure 1 shows an overview of how the data sources contribute to the sample selection and construction. All data are linked at the patient level by national id-numbers. The sample is based on an earlier extraction of all contacts in KUHR during 2007-2020, with an ICPC-2 L (musculoskeletal system), R95, R96 or ICD-10 J43-46 (COPD and asthma) diagnosis and patient residence in Oslo or Trondheim at the time of contact. In addition, all contacts in NPR (somatic, specialist and rehabilitation services) with J43-46 as main diagnosis during 2008-2020 regardless of residential municipality. From the patient id-numbers of these contacts, we identify MSD patients from ICPC-2 codes L and ICD-10 codes M in the gross INOREG sample (Figure 1, left). COPD patients are selected from the ICD-10 codes J43-44, or the ICPC-2 code R95. This results in samples of around 800,000 MSD patients and 140,000 COPD patients. In KUHR, a primary diagnosis for the contact has to be registered in order to get reimbursed. Hence, most registrations in KUHR only have one diagnosis.

Further criteria for the sample selection are necessary depending on the analysis. For example, the full sample is suitable as a basis for analyzing outcomes such as hospital episodes, employment and disability pension for MSD patients. Many of these are still in the workforce and do not require long-term care services. The sample only needs to be restricted to individuals not in retirement, as identified from variables in FD-trygd. For analyses where either reception of long-term care is the outcome, or is important in the care of the patient for other outcomes (the case for COPD), the sample is restricted to patients residing in Oslo or Trondheim over time. This reduces the sample size to around 22,000 COPD patients. To further reduce heterogeneity, one may restrict the sample to patients having at least three contacts (GP, emergency room, contract specialist or physiotherapist) with a COPD or MSD diagnosis. For example, this reduces the sample to around 15,000 COPD patients. Finally, for analyses of MSD where functional ability is the outcome, the sample is restricted to patients included in FYSIOPRIM. This results in a gross sample of 2,700 patients.

2.4 Sample construction

A schematic illustration of the sample construction and patients as they appear in the data is shown in Figure 2. The first time a patient is registered with a MSD or COPD diagnosis is defined as the index date. There are three important time periods in the sample construction: First, a pre-index period, capturing the health status (including non-MSD/COPD related care), sociodemographics and –economics prior to the first MSD/COPD-diagnosis. Second, a follow-up period, where types and pathways of care are identified, and third an outcome period in which the outcomes are measured. Example 1 captures all individuals as they enter the data in 2008. Using a window of at least one year to capture the health and socioeconomic and –demographic status when entering, the index date is defined as the first contact in KUHR (GP, emergency room, contract specialist, physiotherapist or chiropractor) or NPR from 2009 with respectively a COPD or MSD diagnosis. The first patient in the example has an index contact shortly after Jan 2009, and has an early exit, possibly due to death. The second patient has a later index date, and is followed until the data ends in Jan 2020. The example is representative for COPD, as approximately 40% of patients in the data have a first contact with a COPD diagnosis in the first two years from 2008.

Example 2 captures new treatment spells, as only individuals without contacts with respectively COPD or MSD during a period of at least five years prior to the index date are included in the sample. A window of one year is used to capture health status when entering the data. The first patient has an index date shortly after Jan 2014, hence the one-year pre-index window extends into 2013. The second patient enters later. The example is

representative for MSD, as the sample size is much larger than for COPD. It is thus relevant to analyze care pathways and outcomes for new treatment spells. Note that in both cases, the length of the follow-up period may vary due to late entry and early exit (death).

Although the registries contain data from 2008, the contacts with physiotherapists are not fully complete until some years later. Our preliminary analyses show that physiotherapy data in KUHR appear complete from 2014. This may restrict the follow-up period, particularly for MSD. The period required to capture health status prior to the index date may also be extended. To identify MSD/COPD-related health care, a follow-up period of at least one year is required for a patient to be included in any analysis. Sensitivity analyses to assess the effect of different inclusion and exclusion criteria in the sample construction, and the length of pre-, follow-up and outcome periods, is vital in the project. In some analyses, it can be sufficient to use fixed lengths of the follow-up and outcome periods.

Figure 2: **Examples of the general design for constructing samples: The example for COPD includes all individuals based on the first contact (index date) with a COPD diagnosis (index diagnosis). The example for MSD includes individuals given a pre-index date period of five years without the index diagnosis, hence individuals with new treatment spells. In the one-year pre-index period, socioeconomic and – demographic information, non-MSD/COPD related health care use and comorbidities are captured. In the follow-up period, care pathways and specific care factors are identified. Outcomes are measured in a period succeeding the follow-up period. Length of follow-up period may vary across individuals due to late entry and early exit (death). The outcome period is fixed within an analysis, but the length may vary across analyses depending on the outcome.**
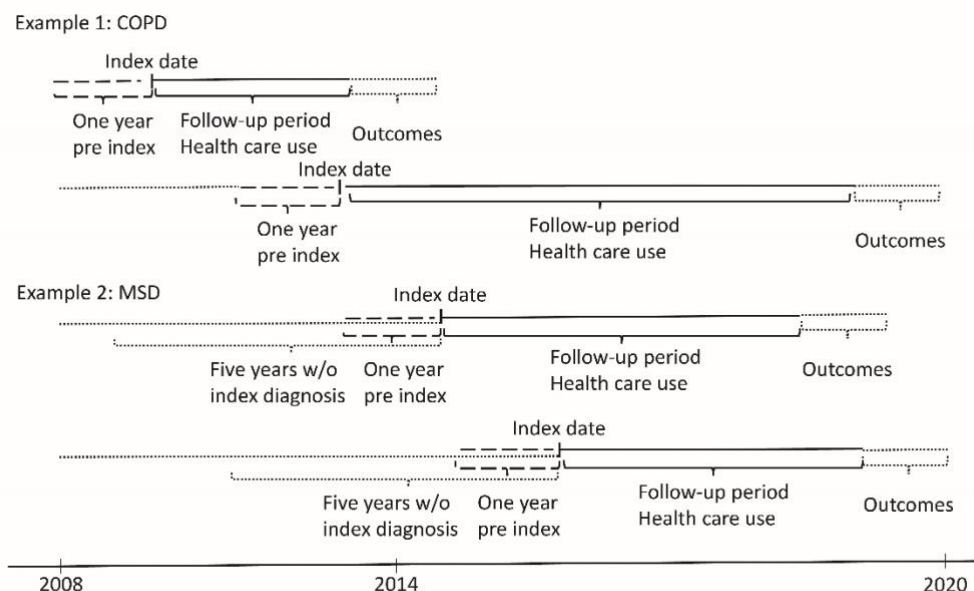
3   Methods

3.1   Defining independent variables

We distinguish between care pathways, describing the frequency and patterns of care across types of providers, and care indicators capturing specific factors of care during the follow-up period. Both will be used as independent variables in analyses of associations to outcomes, in addition to non-MSD/COPD related health care use, comorbidities, sociodemographic and –economic factors (Figure 1, middle).

The clinical course of patients with COPD and MSD differ in many respects, and the analyses will thus apply different care indicators. Using COPD as an example, we construct independent variables based on knowledge gaps mentioned in the national treatment guidelines (Norwegian Directorate of Health, 2013). Specifically, patients with stable COPD should have yearly spirometry measurements and GP consultations. A hospital admission with COPD as the main diagnosis should be followed by a check-up with a GP within four weeks. We can estimate to which extent this is observed in the data. When in need of rehabilitation following a hospital stay for COPD, the patient should start the treatment shortly after discharge. From the data, an indicator on whether rehabilitation is provided within four weeks or later can be constructed. Assuming that all patients in need of rehabilitation will receive it, the effect of early vs late rehabilitation on outcomes may be estimated. Patients with stable moderate to severe COPD should be referred to physiotherapist for exercise training. We can observe how often physiotherapists are involved in the treatment of COPD patients in the data.

Other examples of care indicators are variables associated with care interaction and continuity. Examples are indicators on interaction between physiotherapist and the municipality, between GP and municipality or specialist, and home visits by practitioners as identified from specific fee codes in KUHR. From KUHR we have a unique id-number for the performing practitioner per contact, and the data from the regular GP registry includes the id-number of the regular GP per patient. Hence, one may construct specific care indicators capturing continuity of care within GPs (or physiotherapists) over time, whether the regular GP frequently or rarely involve physiotherapists/specialists in the care pathway of his/her patients, is in a group practice vs individual practice (possible interaction between GPs), the number of patients on the regular GP patient list (having sufficient time for each patient), being a specialist in general practice, and also GP age and gender. These variables may capture aspects of the GP's role in the care. We provide more specific examples of care indicators and other variables we construct from the data in the Appendix. Hence, there is a range of specific care indicators that are available, which can be used either for descriptive purposes, for testing the effect on outcomes, or both. Still, some important variables are missing in the data. The most important is clinical information on severity of the MSD or COPD diagnosis, other examples are level of obesity and smoking (relevant for COPD). For the former, indicators of severity is only available in the FYSIOPRIM subsample, which is primarily intended for analyzing functioning and quality of life outcomes. For the latter, we can only indicate the presence of either from specific diagnosis codes for lifestyle counselling or obesity in NPR and KUHR.

3.2   Identifying care pathways

A more complex issue is how to operationalize care pathways. Dates are included in the data for all contacts with health care providers, thus enabling flexibility in the construction of care pathways. Several approaches will be considered in the project. As multimorbidity is expected in the patient groups, it is important to separate health care contacts likely due

to COPD or MSD from those due to other diagnoses, using codes for respiratory or MSD diagnoses. Only the former are considered in the care pathways, while the latter are independent variables related to overall health status, similar to comorbidity indicators. Describing pathways according to observed combinations of contact frequency across types of providers (for instance GP, physiotherapist, contract specialist, outpatient hospital visits) during the follow-up period (Figure 2) is the first aim of the project.

As a simple approach, individuals with similar contact frequency across providers are grouped into categories. We will then get an overview on the combinations of providers involved in the care pathway, and how frequent these are. A more detailed approach is to use group trajectory modelling (Nagin, 2005), for which a wide range of different methods exist (Nguena Nguefack *et al.*, 2020). Trajectory modelling is based on grouping patients on contact frequency across several dimensions: The combination of practitioners involved, the number of contacts, and the course of contacts over the follow-up period. Some patients will have trajectories indicating improvement in health over time by reduced frequency of contacts, others will have constant high or low use, but with differences in contact frequency across the practitioners. Yet another approach is to use machine learning in order to group patients based on similar patterns of health care use (Brnabic and Hess, 2021). Here there are also several possible methods, such as basic cluster analysis, classification and regression trees, random forests and neural networks. In all approaches, the resulting grouped pathways need validation by clinicians in the project group. We plan to pursue these approaches further. We will study the characteristics of patients in the pathways, with respect to sociodemographic and -economic variables, comorbidities, and the specific care indicators mentioned above.

3.3 Defining outcomes

Outcomes are measured in time periods succeeding the follow-up periods (Figure 2). These include the number of respiratory/musculoskeletal hospital episodes, number of sick days for those in full- or part-time employment, receiving permanent disability pension, total MSD- or COPD-related costs in primary and specialist care, quality of life (measured by EQ-5D) and Patient-Specific Functional Scale for the FYSIOPRIM subsample (MSD), and use of long term care services (Figure 1, right). Both COPD and several MSD subgroups are considered among ambulatory care–sensitive conditions (Purdy *et al.*, 2009), and a central goal of outpatient care is to reduce the number of hospital episodes, reduce the health care costs and enable work participation. For MSD patients work participation is a particularly relevant outcome, as a large proportion of the patients are rather young. It is also an important outcome for the society. Work participation is extracted from FD-trygd, together with social benefits associated with absence from work. The total MSD- or COPD-related health care costs can be estimated by adding primary care fees and costs based on DRGs, as registered in respectively KUHR and NPR, during the outcome period. Patient profiles in terms of care pathways and comorbidity patterns prior to receiving long-term care services, is of interest to decision makers in the municipalities.

3.4 Identifying effects of care pathways and care indicators on outcomes

In order to achieve the ultimate aim of suggesting ways to improve care, effects of the variables defined above needs to be assessed. This is challenging, and we will describe some preliminary approaches. Due to the sample size and high number of time-dependent variables, a useful simplification is to aggregate data per year of follow-up. This reduces the complexity in the data structure, while still allowing for yearly updates of variables such as

care indicators and comorbidities. Because of the chronic nature of the diagnoses, effects that persist over time are more important than what happens in the short term.

One possible approach is motivated from the variables defined above. It is based on regression analyses, using grouped care pathways and care indicators as independent variables (see e.g. Nylund *et al.*, 2019, for methods within group trajectory modelling). A challenge is the lack of data on severity in the registries. However, the frequencies and pathways of COPD/MSD-related health care contacts are intuitively expected to be correlated with severity. As patients in similar pathways across providers are grouped, assume for instance that two groups have similar pathways across GP, physiotherapist, and outpatient hospital visits, but differ for contract specialists. If specialist visits are associated with better outcomes (e.g. fewer hospital episodes) after adjusting for socioeconomic and -demographic factors, municipality of residence, comorbidities and non-MSD/COPD health care use, it could indicate an effect of access to specialists. In particular, if the result is consistent across several groups with similar pathways in the other providers. For example, both in patients with increasing, decreasing or stable high frequency of GP consultations over time. Second, assuming that the grouped pathways are highly correlated to severity, protective effects of care indicators may be causal. An example is if higher continuity with GP, early vs. late rehabilitation and more frequent interaction between GP and specialist are associated with fewer hospital episodes for COPD patients, after adjusting for the other variables. Although it is crucial to discuss the validity of any findings with clinicians in the project, we believe that this approach is useful.

3.5 Statistical analyses

Following the set-up of aggregating data per year, we can present possible regression strategies. Negative binomial models clustered by patient can be used for count outcomes such as MSD/COPD-related hospital episodes and sick days. We can analyze the association between current year care pathways and care indicators to next year outcome, adjusting for current year hospital episodes/sick days, comorbidities, socioeconomic and –demographic factors. Thus, expanding the examples in Figure 2 to include repeated follow-up and outcome periods each of length one year. Discrete time survival models may be used to analyze patient profiles and their association to receiving long term care services or disability pension by the end of follow-up. For cost outcomes, it is of interest to identify care pathways and multimorbidity associated with consistently high costs over time, or a change from high to low use and vice versa from the follow-up period to the outcome period. Here, variants of logistic regression may be used.

4 Example: Health care trajectories for patients with musculoskeletal disorders and associations to future health care costs

Introduction: Musculoskeletal disorders (MSD) comprise some of the most prevalent conditions both worldwide and in Norway (Vos *et al.*, 2012, and Kinge *et al.*, 2015). Most persons with MSD have a low health care use, while a small proportion have a very high use over several years (Lentz *et al.*, 2019, and Mose *et al.*, 2021). The aim is to describe health care trajectories and assess associations between combinations of health care use for MSD in the first three years and future health care costs.

Methods: This was a registry-based study, using KUHR, NPR, Statistics Norway and FD-trygd as data sources. We included patients with a health care contact with MSD registered in 2013-2015 and *no history* of MSD in the previous three years. Group based multi-trajectory modelling was used to model combinations of health care services over

several years, based on frequency of consultations per year at the GP, physiotherapist and hospital. The goodness-of-fit from different models was compared using Akaike's and Bayesian information criterion, and the probability of belonging to each group, to decide the number of groups and functional forms of the trajectories identified in the data. Characteristics of patients in the resulting groups were compared on age, gender, education, income, non-Nordic background, Charlson Comorbidity Index and some main MSD diagnosis groups. In addition, we studied the likelihood of being a future high-cost patient, defined as being in the top 5% for MSD-related health care costs in year 4 to 6 after the initial diagnosis.

Results: We identified six trajectories, Figure 3. The largest group (group 1, 74% of the sample) had only one GP consultation the first year, and thereafter no use of the services. Three groups (groups 2-4, 3-8%) had a high use of physiotherapy in one of the three years and some consultations at the GP and hospital. One group (group 5, 9%) used GP and hospital services only. The smallest group (group 6, 2%) had high use of all three services across the three years, and the highest likelihood of being a high-cost user in years 4-6. Patients with non-Nordic background and secondary school or lower were overrepresented in the group not using physiotherapy (group 5, Table 2).

Figure 3:   Group-based trajectory models for the number of MSD-consultations per year at the GP, physiotherapist and hospital.
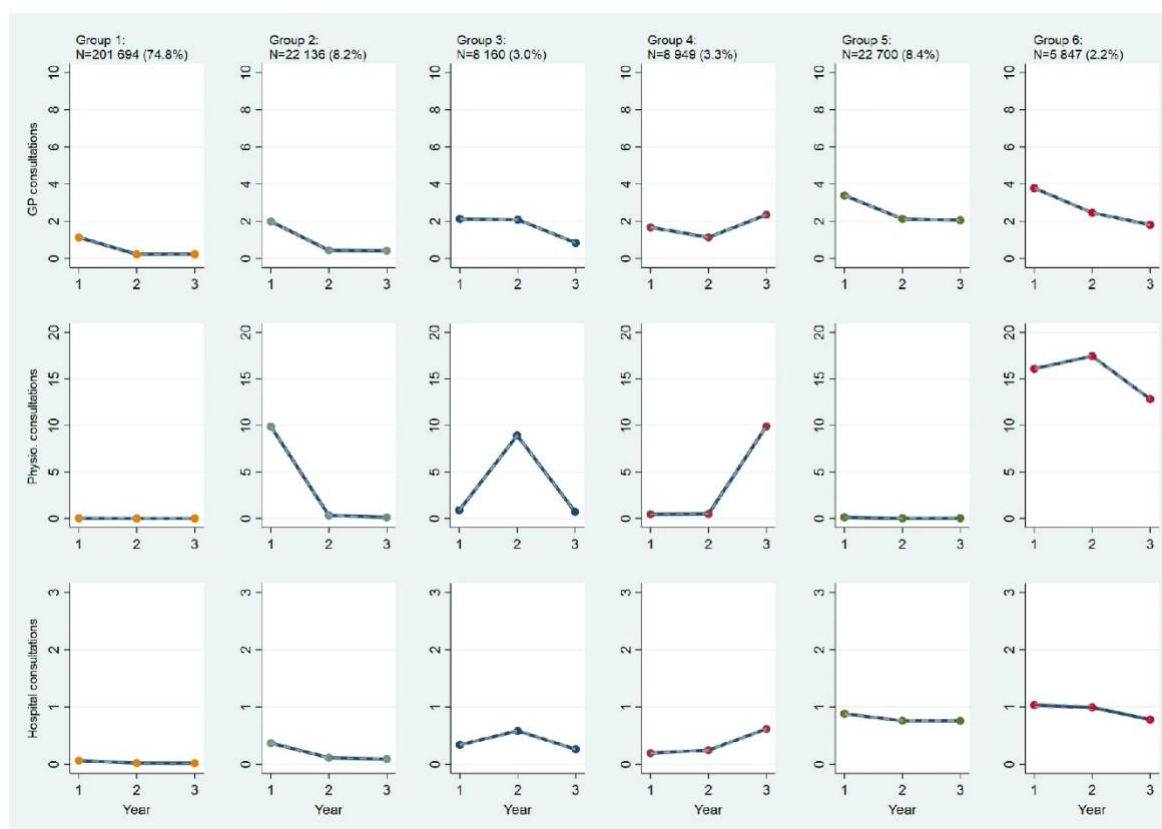
Discussion: Interestingly, there were differences depending on when or whether physiotherapy was used. Groups showing similar trajectories across GPs and hospitals (groups 2 to 5, Table 2), had marked differences in the likelihood of being a future high-cost user. The percentage of future high-cost users in groups 4 and 5 was more than twice as high as in group 2. We will study further whether adjusting for variables on patient

characteristics, diagnoses or municipal fixed effects explain these differences. Another approach is using the results to construct vignettes for interviews and group discussions with care providers to further reveal explanations for the various pathways and likelihood of being a future high-cost user.

Table 2: **Characteristics of patients in the identified trajectory groups. SD=Standard deviation.**

	Group 1:	Group 2:	Group 3:	Group 4:	Group 5:	Group 6:
Number of patients (% of total)	201,694 (74.8)	22,136 (8.2)	8,160 (3.0)	8,949 (3.3)	22,700 (8.4)	5,847 (2.2)
Mean age (SD)	33.3 (21.2)	37.3 (21.9)	38.0 (19.8)	38.3 (19.7)	42.6 (19.5)	47.9 (19.4)
% female	47.6	53.6	56.3	58.0	50.9	60.6
% 2+ Charlson Comorbidity Index	4.8	5.1	4.6	4.4	6.4	6.5
% non-Nordic	17.3	13.1	16.5	17.2	25.1	17.4
% secondary education or less	47.3	44.8	50.3	51.7	63.5	56.4
% less than NOK 400 000 gross income	73.5	65.8	69.0	69.7	72.4	65.3
Index diagnosis, %:						
Back and neck (Columna)	28.7	34.9	29.9	30.8	30.3	28.9
Shoulder	6.2	11.1	8.4	7.6	7.7	11.2
Lower extremity	15.8	17.6	18.4	16.5	15.3	19.8
Widespread/unspecific	12.4	10.2	11.5	11.6	13.3	11.3
Osteoarthritis	1.1	2.9	2.3	2.3	2.9	8.0
% high-cost patients (top 5% year 4-6)	2.7	5.4	8.8	17.0	12.9	27.6

5   Discussion

To our knowledge, this is the first time data on primary, long-term and specialist health care are linked at the individual level in a large, population-based sample of MSD and COPD patients. Using existing registry data as a basis for identifying care pathways and factors associated with improved outcomes is a real novelty of the project, representing new and cost-effective ways of utilizing registries for health services research. There are several strengths to this approach. First, using data from 2008 and onwards will allow us to ensure inclusion of a large number of patients at their first-time consultation for the disease with a sufficiently long control prior to the index date. Furthermore, we are able to analyse the use of different health care services and outcomes over a long period. Importantly, INOREG adds data from primary care. Due to limitations in data availability, previous initiatives for registry-based research on chronic patients have mainly focused on how specialist care influences outcomes. The combined use of existing registries as well as the comprehensive information on functional ability that is present in FYSIOPRIM adds to the possibilities for risk adjustment; this is a crucial component in identifying effects of health care that cannot be evaluated with controlled trials. Many findings may be hypothesis generating rather than causal. Specific suggestions on how to improve care from results in the project need validation in additional studies before implementation. However, we believe this applies regardless of the methods used in analyses of registry data.

The work done in INOREG will benefit other researchers studying care pathway variations in absence of specific interventions. Despite the range of data sources available, some aspects of care are difficult to capture. The registry data do not directly show

cooperation between services. We therefore identify the various services used, as a proxy for interaction and cooperation. Aggregating data per year is useful in order to simplify the data structure, while still being able to retain the overall trends in the health care pathways and yearly updates of the specific care indicators. On the other hand, details in the short-term trends and sequence of health care contacts are lost, which may cause us to miss some factors in the analyses. The completeness of the registries is generally very good, and no systematic differences in coding practice between health care providers in NPR and KUHR that we are aware of. There may be differences in coding of long-term care services across municipalities, but this is less relevant as only data from Oslo and Trondheim are included. Although at the possible cost of less generalizability, there are also important advantages in restricting analyses to the two cities. Heterogeneity in access to both primary and specialist care is reduced and this increases the possibility of identifying important factors. Still, the choices we make when we define and construct variables for care pathways and care indicators are to some extent subjective. Validating variables and findings with the clinicians involved in the project is hence critical.

There are also limitations due to the lack of some highly relevant data. The registries do not include data on smoking, which is relevant in the analyses of COPD. In the full sample, we do not have information on the disease severity at index date or during follow-up. For COPD, lack of data on GOLD grade (Global Initiative for Chronic Obstructive Lung Disease) add to the difficulty in identifying homogenous groups. We at best observe severity indirectly by the frequency of health care contacts with COPD as the primary diagnosis. For MSD, this problem is reduced when analyzing the subset of patients included in FYSIOPRIM, but for the full sample, the challenge remains. To better capture severity, we plan to add medication data from the Norwegian Prescribed Drug Registry (e.g. medication for obstructive airway diseases). This has recently become available for linkage to other registries in Norway.

Finally, the care pathways for COPD and MSD are complex, both due to the number of care providers involved over time, the variability in affliction and the multimorbidity of the patients. This makes it particularly challenging to find the ideal analyses and variable definitions for identifying the most important factors associated to outcomes. Examining how far we can get in this regard by using administrative registry data has great value in itself. Even without causal interpretations, findings may be influential. For instance, a care pathway both showing low use of health care services during the follow-up period, and a relatively high probability of receiving disability pension. Studying the characteristics of the patients in this pathway, and discussing possible explanations for the apparent underuse of services, is important. The qualitative part of INOREG provides an opportunity to identify care practices and prioritizing processes that we cannot find from the registry-based analyses alone. This can inform us on important data that are missing in the registries, data that perhaps could be included. The qualitative part may further contribute to the interpretations of the quantitative findings, and to new hypotheses that can be analyzed in the data. In conclusion, the project is of great interest to the municipality health sector and health care workers in general by providing a better understanding of the process of care. As a result, there is a potential for developing more targeted and better services for the MSD and COPD patients, leading to better functioning and health.

Ethical approval
The project is approved by the Regional Ethics Committee, reference number 118725.

Disclaimer

Data from the Norwegian Patient Registry, Statistics Norway, and the Cause of Death Registry are used in the project. The interpretation and reporting of these data are the sole responsibility of the authors, and no endorsement by the registry owners is intended, nor should it be inferred.

References

Amelung, V., Stein, V., Suter, E., Goodwin, N., Nolte, E. and Balicer, R. (eds.) (2021). *Handbook Integrated Care*. Springer, Cham.

Beyrer, S., Otnes, B. and Karlsen, T.B. (2018). *Kvalitet i IPLOS-registeret 2017* [Quality in the IPLOS-registry 2017]. Statistics Norway, Document 2018/41.

Brnabic, A. and Hess, L.M. (2021). Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. *BMC Medical Informatics and Decision Making*, 21(54).

Casas, A., Troosters, T., Garcia-Aymerich, J., Roca, J., Hernández, C., Alonso, A., del Pozo, F., de Toledo, P., Antó, J.M., Rodríguez-Roisín, R., Decramer, M., members of the CHRONIC Project (2006). Integrated care prevents hospitalisations for exacerbations in COPD patients. *European Respiratory Journal,* 28(1), 123-130.

Evensen, K.A.I., Robinson, H.S., Meisingset, I., Woodhouse, A., Thielemann, M., Bjorbækmo, W.S., Myhre, G., Hansen, A.E., Vasseljen, O. and Vøllestad, N.K. (2018). Characteristics, course and outcome of patients receiving physiotherapy in primary health care in Norway: design of a longitudinal observational project. *BMC Health Service Research*, 18(1), 936.

Gershon, A. S., Guan, J., Victor, J. C., Goldstein, R. and To, T. (2013). Quantifying health services use for chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care medicine*, 187(6), 596–601.

Goodwin, N., Stein, V. and Amelung, V. (2021). *What is Integrated Care?* In: Amelung, V., Stein, V., Suter, E., Goodwin, N., Nolte, E. and Balicer, R. (eds.) *Handbook Integrated Care*. Springer, Cham.

Henri, S., Herrera, R., Vanasse, A., Forget, A. and Blais, L. (2022) Trajectories of care in patients with chronic obstructive pulmonary disease: A sequence analysis. *Canadian Journal of Respiratory, Critical Care, and Sleep Medicine*, 6(4), 237-247.

Hill, J.C., Whitehurst, D.G., Lewis, M., Bryan, S., Dunn, K.M., Foster, N.E., Konstantinou, K., Main, C.J., Mason, E., Somerville, S., Sowden, G., Vohora, K. and Hay, E.M. (2011). Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet*, 378(9802), 1560-1571.

Institute for Health Metrics and Evaluation (2016). *Norway: State of the Nation's Health: Findings from the Global Burden of Disease.* Seattle, Washington.

Kinge, J.M., Knudsen, A.K., Skirbekk, V. and Vollset, S.E. (2015). Musculoskeletal disorders in Norway: prevalence of chronicity and use of primary and specialist health care services. *BMC Musculoskeletal Disorders*, 16(75).

Koff, P.B., Jones, R.H., Cashman, J.M., Voelkel, N.F. and Vandivier, R.W. (2009). Proactive integrated care improves quality of life in patients with COPD. *European Respiratory Journal*, 33(5), 1031-1038.

Lambeek, L.C., Bosmans, J.E., Van Royen, B.J., Van Tulder, M.W., Van Mechelen, W. and Anema, J.R. (2010). Effect of integrated care for sick listed patients with chronic low back pain: economic evaluation alongside a randomised controlled trial. *BMJ*, 341, c6414.

Lærum, E., Brage, S., Johnsen, K., Natvig, B. and Aas, E. (2013). *Et muskel- og skjelettregnskap. Forekomst og kostnader knyttet til skader, sykdommer og plager i muskel- og skjelettsystemet* [Accounting musculoskeletal disorders. Prevalence and costs]. Oslo University Hospital: Muskel og Skjelett Tiåret [The Bone and Joint Decade] 2013.

McKay, R., Letarte, L., Lebel, A., Quesnel-Vallée, A. and the TORSADE Cohort Working Group (2022). Exploring social inequalities in healthcare trajectories following diagnosis of diabetes: a state sequence analysis of linked survey and administrative data. *BMC Health Services Research*, 22(1), 131.

Ministry of Health and Care Services (2014). *HelseOmsorg21 - Et kunnskapssystem for bedre folkehelse - Nasjonal forsknings- og innovasjonsstrategi for helse og omsorg* [National strategy for improved health and care]. The Norwegian Ministry of Health and Care Services, Report 2014.

Mose, S., Kent, P., Smith, A., Andersen, J.H. and Christiansen, D.H. (2021). Trajectories of Musculoskeletal Healthcare Utilization of People with Chronic Musculoskeletal Pain – A Population-Based Cohort Study. *Clinical Epidemiology*, 13, 825-843.

Murray, C.J., Vos, T., Lozano, R., et al. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380(9859), 2197-2223.

Nagin, D.S. (2005). *Group-based modeling of development over the life course*. Cambridge, MA: Harvard University Press.

Nguena Nguefack, H.L., Pagé, M.G., Katz, J., Choinière, M., Vanasse, A., Dorais, M., Samb, O.M. and Lacasse, A. (2020). Trajectory modelling techniques useful to epidemiological research: A comparative narrative review of approaches. *Clinical Epidemiology*, 30(12), 1205-1222.

Lentz, T.A., Harman, J.S., Marlow, N.M., Beneciuk, J.M., Fillingim, R.B. and George, S.Z. (2019). Factors associated with persistently high-cost health care utilization for musculoskeletal pain. *PLoS One*, 14(11): e0225125.

Nielsen, R., Johannessen, A., Omenaas, E.R., Bakke, P.S., Askildsen, J.E. and Gulsvik, A. (2011). Excessive costs of COPD in ever-smokers. A longitudinal community study. *Respiratory medicine*, 105(3), 485-493.

Norwegian Directorate of Health (2012). Kols. *Nasjonal faglig retningslinje og veileder for forebygging, diagnostisering og oppfølging* [COPD. National guideline for prevention, diagnosis and follow-up]. IS-2029. Updated in 2022.

Norwegian Directorate of Health (2019). *Beskrivelse av innhold i NPR* [Description of contents in NPR]. Oslo: Norwegian Directorate of Health. Available from www.helsedirektoratet.no/tema/statistikk-registre-og-rapporter/helsedata-og-helseregistre/norsk-pasientregister-npr/innhold-og-kvalitet-i-npr Accessed August 15th 2022.

Nylund-Gibson, K., Grimm, R.P. and Masyn, K.E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural Equation Modeling*, 26(6), 967-985,

Purdy, S., Griffin, T., Salisbury, C. and Sharp, D. (2009). Ambulatory care sensitive conditions: terminology and disease coding need to be more specific to aid policy makers and clinicians. *Public Health*, 123(2), 169-73.

van den Bussche, H., Koller, D., Kolonko, T., Hansen, H., Wegscheider, K., Glaeske, G., von Leitner, E.C., Schäfer, I. and Schön, G. (2011). Which chronic diseases and disease combinations are specific to multimorbidity in the elderly? Results of a claims data based cross-sectional study in Germany. *BMC Public Health*, 11, 101.

Vos, T., Flaxman, A.D., Naghavi, M., et al. (2010). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380(9859), 2163-2196.

Appendix

Table A1 gives examples on how specific care indicators and other variables can be defined in the data. The purpose is to show that the variables mentioned in **Methods** may be identified from diagnosis codes, fees, dates, information on the GP and dates of registration in the registries. Note that most of the variables are time-dependent.

Table A1: Examples of variables constructed from the data, the source registries and the variable definition. NPR=Norwegian Patient Registry, KUHR=Control and Payment of Reimbursement to Health Service Providers, RGP=Regular General Practitioner registry, Municipality EPJ=municipality electronic patient journal.

Variables	Registries	Definition
Specific care indicators		
Interaction between physiotherapist and municipality or GP and between GP and municipality or specialist	KUHR	Fee codes GP: 1J, 1F, 14 Physiotherapist: E50, E51A, E51B
Continuity of care	GP registry and KUHR	Estimated for each patient using the Bice–Boxerman Continuity of Care Index (Bice and Boxerman, 1977), with GP contacts and the GP id-number as inputs
Change of GP	RGP	Number of times patient has changed regular GP during follow-up
Referral GP and contract specialist, respiratory diagnoses or musculoskeletal diagnoses	KUHR	Indicator of regular GP having high/low rate of referrals with ICPC-2 code R (COPD) or L (MSD) to contract specialist compared to average
Face-to-face consultations with GP or physiotherapist, respiratory diagnosis or musculoskeletal diagnoses	KUHR	Fee codes 2AD, 2AE, 2AK, 2FK, 11AD, 11AK (GP), A1, A3, A8, A9 (physiotherapist) and ICPC-2 code R (COPD) or L (MSD)
Home visits by GP or physiotherapist	KUHR	Fee codes GP: 11A, 11B, 11C, 11D, 11E Physiotherapist: F1, F21
GP specialist	RGP	Indicator of regular GP being a specialist in general practice
Referral GP and physiotherapist, respiratory diagnoses or musculoskeletal diagnoses	KUHR	Indicator of regular GP having high/low rate of referrals with ICPC-2 code R (COPD) or L (MSD) to physiotherapist compared to average
Comorbidities		
Indicators for individual diagnoses in Charlson comorbidity index and common multimorbidites in the literature. Examples for COPD: Dementia, depression, obesity	KUHR, NPR	ICD-10: F00-03, G30, F32-34, E65-66 ICPC-2: P70, P03, P73, P76, T82
Long-term care services and needs		
Indicator on receiving any service from the municipality, variables for home nursing (duration and quantity is possible), or long-term stay	Municipality EPJ	Date of first registration in municipality EPJ, dates and quantity for service code 15 (home nursing), respectively registration date for code 21 (long term stay) in municipality EPJ
Most recent need-score at index date	Municipality EPJ	Calculated from latest ADL-score (activities of daily living) prior to

		index date. Sum of ADL-items if receiving services, 0 otherwise. Assuming that services are provided if needed.
Health outcomes		
Hospital episode	NPR	All day stays and overnight stays less than a day apart
Sick days, unemployment, indicator of permanent disability pension when entering sample	Statistics Norway	Duration of each sick leave and period of unemployment registered since 2008, receiving disability pension prior to index date
Health care costs		
Costs in primary care	KUHR	Reimbursed amount per contact
Costs use in specialist care	NPR	Diagnosis related group weights and cost per DRG-point

References

Bice, T.W. and Boxerman, S.B. (1977). Quantitative measure of continuity of care. *Medical Care*, 15, 347-349.