

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Comorbidities in Osteoarthritis (CoMO): a multinational study in four European countries

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2021-052816
Article Type:	Protocol
Date Submitted by the Author:	26-Apr-2021
Complete List of Authors:	Swain, Subhashisa; University of Nottingham, Academic Rheumatology Kamps, Anne; Erasmus MC - University Medical Center Rotterdam, Department of General Practice Runhaar, Jos; Erasmus University Medical Center Rotterdam, Department of General Practice Dell'Isola, Andrea; Lunds University Faculty of Medicine, Turkiewicz, Aleksandra; Lund University, Dept of Orthopedics Robinson, Danielle ; University of Oxford Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, Strauss, V; University of Oxford Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, Mallen, Christian; Keele University, Arthritis Research UK Primary Care Centre Kuo, Chang-Fu; Chang Gung Memorial Hospital Linkou Branch, Coupland, Carol; University of Nottingham, Division of Primary Care Doherty, Michael ; University of Nottingham, Academic Rheumatology Sarmanova, Aliya ; University of Bristol, Prieto-Alhambra, Daniel; University of Oxford, Centre for Statistics in Medicine, NDORMS Englund, Martin; Lund University, Dept of Orthopedics Bierma-Zeinstra, Sita; Erasmus University Medical Centre, Department of General Practice Zhang, Weiya; University of Nottingham, Academic Rheumatology
Keywords:	EPIDEMIOLOGY, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, RHEUMATOLOGY

SCHOLARONE™
Manuscripts

Title- Comorbidities in Osteoarthritis (CoMO): a multinational study in four European countries

Authors

Subhashisa Swain^{1,10}, Anne Kamps², Jos Runhaar², Andrea Dell 'Isola³, Aleksandra Turkiewicz³, Danielle Robinson⁴, Victoria Y Strauss⁴, Christian Mallen⁵, Chang Fu Kuo^{1,6}, Carol Coupland⁷, Michael Doherty^{1,10}, Aliya Sarmanova⁸, Daniel Prieto Alhambra⁴, Martin Englund³, Sita Bierma-Zeinstra⁹, Weiya Zhang^{1,10}

1. Academic Rheumatology, School of Medicine, University of Nottingham, UK
2. Department of General Practice, Erasmus MC - University Medical Center Rotterdam, The Netherlands
3. Clinical Epidemiology Unit, Orthopaedics, Department of Clinical Sciences Lund, Lund University, Sweden
4. Centre for Statistics in Medicine, NDORMS, University of Oxford, UK
5. School of Medicine, Keele University, UK
6. Division of Rheumatology, Allergy and Immunology, Chang Gung Memorial Hospital, Taiwan.
7. Division of Primary Care, School of Medicine, University of Nottingham, UK
8. Musculoskeletal Research Unit, Bristol Medical School, Translational Health Sciences, University of Bristol, UK
9. Department of General Practice, Department of Orthopaedic Surgery, Erasmus MC - University Medical Center Rotterdam, The Netherlands
10. Pain Centre Versus Arthritis, University of Nottingham, UK

Corresponding Author:

Professor Weiya Zhang
Academic Rheumatology, School of Medicine
University of Nottingham
Clinical Sciences Building, Nottingham city Hospital
NG5 1PB, Nottingham, UK
Email: weiya.zhang@nottingham.ac.uk

Abstract

Introduction

Osteoarthritis (OA) is one of the leading chronic conditions in the older population. People with OA are more likely to have one or more other chronic conditions than those without. However, the temporal associations, clusters of the comorbidities, role of analgesics and the causality and variation between populations are yet to be investigated. This paper describes the protocol of a multinational study in four European countries (UK, Netherlands, Sweden, and Spain) exploring comorbidities in people with OA .

Methods and analysis

This multinational study will investigate i) the temporal associations of 61 identified comorbidities with OA, ii) the clusters and trajectories of comorbidities in people with OA, iii) the role of analgesics on incidence of comorbidities in people with OA, iv) the potential biomarkers and mechanistic pathways between OA and the comorbidities, and v) variations between countries.

A combined case-control and cohort study will be conducted to find the temporal association of OA with the comorbidities using the national or regional health databases. Latent class analysis will be performed to identify the clusters at baseline and joint latent class analysis will be used to examine trajectories during the follow-up. A cohort study will be undertaken to evaluate the role of non-steroidal anti-inflammatory drugs (NSAIDs), opioids, and paracetamol on the incidence of comorbidities. Mendelian randomisation will be performed to investigate the potential biomarkers for causality between OA and the comorbidities using the UK Biobank and the Rotterdam databases. Finally, a meta-analysis will be used to examine the variations and pool the results from different countries.

Ethics and Dissemination

Research ethics was obtained according to each database requirement. Results will be disseminated through the FOREUM website, scientific meetings, publications, and in partnership with patient organisations.

Key words: Osteoarthritis; Comorbidity; Multimorbidity; Primary care; Analgesics, Mendelian Randomisation, PheWAS

Strengths and limitations of this study

- This is first ever multicenter study on comorbidities in osteoarthritis in Europe involving nearly 27 million electronic health records
- More than 60 chronic conditions are being studied – representing a wider coverage of diseases
- We will examine the causal association using Mendelian Randomization -PheWAS methods with genetic data collected from two countries
- Same protocol with robust statistical methods will be used across all countries to replicate the findings and examine the variations.
- Possible biases may be introduced by the nature of electronic health records and length of data availability.

Background

Osteoarthritis (OA) affects 27% of people aged over 45 years at peripheral synovial joints such as knees, hips, hands and feet (1) and it is by far the most common arthritis and a leading cause of chronic joint pain and disability in older people (2,3). It is anticipated that the burden of OA will continue to rise in the coming decades because of aging and obesity – two major risk factors for OA (4,5). A recent systematic review has confirmed that people with OA are more likely to have other diseases, especially stroke, peptic ulcer, hypertension, and depression (6). However, whether these comorbidities just co-exist with OA, share common risk factors with OA, or are causes or consequences of OA remains largely unknown.

There was wide heterogeneity in definitions of OA and other chronic conditions, diagnosis and recording of diseases, sample sizes, and number of diseases studied (6). This diversity made the comparison and pooled estimation of comorbidity prevalence difficult. One of the major limitations of previous small studies was, they mostly focused on cardiovascular and musculoskeletal conditions (7–9). Various hypotheses have been used to explain the existence of comorbidities in general, the most accepted of which are the concordant (diseases sharing similar pathophysiological risk factors) and discordant (diseases not sharing similar pathophysiological risk factors) theories (10). Multiple factors are reported to be associated with comorbidity and multimorbidity, and prescription of drugs is one of these (11). That is the presence of multiple chronic conditions increases the chances of polypharmacy which further escalates the risk of other conditions. Especially in people with OA, the prescription of analgesics is common, and is associated with increased risk of other conditions such as cardiovascular, gastrointestinal and chronic renal diseases (12,13). The additional presence of disease has a significant impact on an individual. OA is one of the leading conditions reported in multimorbidity research and exploring the association with other diseases would help in further explaining the causality (14). Understanding the temporal association and disease trajectory is crucial for any chronic condition, and this is possible through studies using longitudinal databases (15). Observational studies have limitations of establishing causal association between OA and comorbidities, for which newly developed ‘Mendelian Randomisation’ method is more powerful.

Another major issue in comorbidity research is the number and types of conditions studied (16). Therefore, it is important to develop a consensus on both the count and typology of conditions to be studied to enable comparisons across populations and to derive pooled estimates as appropriate. Further, using uniform methods and definitions of diseases in

1
2
3 computing these estimates would reduce heterogeneity and make the comparison more
4 reliable. The burden of diseases in primary care often depends on the population structure,
5 health infrastructure and individual factors such as income and education. Such factors vary
6 between countries, but because of the heterogeneity mentioned above there are no robust data
7 comparing OA and its comorbidities between countries. Therefore, the aims of this study
8 were to explore the burden, pattern, and causal factors of comorbidities in people with OA
9 across four European countries using national registration databases in the UK, the
10 Netherlands, Sweden, and Spain.
11
12
13
14
15
16
17
18

19 Objectives

- 20 1. To estimate the prevalence, incidence, and time sequence of comorbidities in OA
- 21 2. To examine the clusters of comorbidities and trajectories of clusters in OA and
- 22 associations with death
- 23 3. To investigate the associations between commonly used OA drugs, such as non-
- 24 steroidal anti-inflammatory (NSAIDs) and opioids, and risk of comorbidities
- 25 4. To identify the potential biomarkers and mechanistic pathways between OA and the
- 26 comorbidities
- 27 5. To examine the variations of OA comorbidities and clusters across countries.
- 28
29
30
31
32
33
34
35

36 Methods

37 *Ethics*

38 The study has obtained the following ethics approvals: UK- ISAC 19/30R, Netherlands IPCI
39 registration no. 11/2019, Spain - 4R19/011, Sweden – Dnr 2011-432, Dnr 2014-276, and Dnr
40 2018-233.
41
42
43
44

45 *Databases*

46 Four routinely collected national (the UK and the Netherlands) or regional (Sweden and
47 Spain) health databases will be used for objectives 1-3. In addition, for objective 4, genomic
48 associations of OA with comorbidities will be examined using two cohort studies from the
49 UK (UK Biobank) (17) and the Netherlands (Rotterdam study) (18). The four national
50 representative and regional databases contain information about the population with primary
51 care consultations in four different countries. The longitudinal databases provide information
52 about the diagnosis of the diseases by the general practitioners and some diagnoses made in
53 secondary care, prescription of drugs, deaths, and other health utilisation indicators. The
54 details of the databases and their properties are given in Table 1 and 2.
55
56
57
58
59
60

Participants

People registered with the respective databases aged 18 year or above are eligible for the study.

Patient public involvement (PPI): Three PPI representatives (with OA and all with multiple chronic conditions) were involved in this study through group meetings. Difficulties of living with multiple conditions, lack of research in causal relationship and identification of diseases to be studied and the role of drugs in comorbidity were discussed. They are constantly in touch through providing their inputs at each steps of the study.

Definition of Osteoarthritis

OA will be defined as the at least one recorded physician diagnosis of OA for hip, knee, ankle/foot, wrist/hand, or site recorded as 'unspecified' during the study period for the respective database. People with any previous recording of the OA prior to the first diagnosis (index date) will be excluded.

Comorbidities

We defined comorbidity as the recording of diagnosis of predefined chronic conditions in individuals using either ICD-10 or Read or international classification of primary care (ICPC) code. An extensive list of 61 chronic conditions was prepared from the Quality Outcome Framework (QOF) (19), list of the US Department of Health and Human Services Initiative on Multiple Chronic Conditions (20), global burden of diseases (21) and the Charlson comorbidity index (22). The list has been updated with findings from our systematic review (6) and a previous UK community-based knee pain study (6,23) by including common and important morbidities not included in the above (24,25). A code mapping exercise was conducted to finalise the list of conditions available for all the research centres. The comparison of codes was made, and it was reviewed by four researchers including a clinician from the team. The detailed list of the conditions is given in Table 3.

Covariates

Age, and gender will be used in all centres as covariates to adjust in regression models. Additionally, information on body mass index (BMI), smoking, alcohol use, socioeconomic variables such as education level, income, place of birth (to identify those who immigrated to the country), and residential area, marriage (or registered partner) will be included when available. For calculating severity of the comorbidities in an individual, Elixhauser comorbidity index will be used to estimate the impact of comorbidities on death (26,27).

1
2
3 Missing data on covariates will be substituted using multiple imputation methods, provided
4 that the data is missing at random, if applicable.
5
6

7 *Study design and data analysis*
8

9 Summary of the study design and analysis is provided in Figure 1.
10
11

12
13
14 *Objective 1 – Prevalence, incidence, and time sequence of comorbidities in OA*
15

16 A combined retrospective and prospective study of OA cases and sex, age (+/-2 years), first
17 year of registration, and practice matched controls (1:1-4) without OA (28) will be used to
18 determine the prevalence, incidence, and time sequence of comorbidities in OA. Incident OA
19 cases will be identified, and the first diagnosis date will be used as the starting point (index
20 date). For controls the same index date as their matched case will be used. They will be both
21 retrospectively reviewed for prior diagnoses of comorbidities and prospectively followed-up
22 for posterior new comorbidities. In the retrospective analysis the prevalence and 95%
23 confidence interval (CI) of each specific comorbidity will be calculated separately in OA
24 cases and matched controls using the number of people diagnosed with the comorbidity
25 divided by the total number of OA cases or controls at the index date. The prevalence of each
26 comorbidity in OA cases and matched controls will be calculated for given time intervals
27 prior to the index date of 0-1, 0-5 and 0-10 years separately to assess observational bias (28).
28 Discrete time intervals of 1-5, and 5-10 years before will also be used to estimate the
29 prevalence to minimise consultation bias/misclassification bias of OA (if possible).
30 Conditional logistic regression will be used to calculate the odds ratios (OR) for each
31 comorbidity unadjusted and adjusted for BMI, smoking and alcohol consumption.
32 For the prospective analysis participants with incident OA but without the specific
33 comorbidity of interest at the index date (i.e., people at risk) and matched controls without
34 OA will be followed up until the date of the first diagnosis of the comorbidity, deregistration,
35 or death whichever comes first. The cumulative incidence will be calculated for each
36 comorbidity in OA cases and matched controls at 1, 3, 5, 10, 15, 20 years (based on the data
37 available) after the index date to examine the dynamic change of developing comorbidities
38 during follow-up. Kaplan-Meier survival curves will be used to display the cumulative
39 probability in OA and non-OA groups. Proportional hazard assumption will be tested using
40 Schoenfeld residual plots. The Cox regression model will be used to calculate hazard ratios
41 (HR) for each comorbidity unadjusted and adjusted for age, gender, practice, BMI, smoking
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 and alcohol consumption. This hybrid design has been previously used by us to examine the
4 temporality of associations between other rheumatic musculoskeletal diseases (RMDs) (e.g.,
5 gout and lupus) and comorbidities (28,29).
6
7

8
9
10 *Objective 2 – Clusters and impact of comorbidities in people with OA*

11 For each dataset, an 80%: 20% split into the training and testing data will be introduced. The
12 following analysis in objective 2 will be first employed into the training dataset and then
13 tested its generalisability in the testing dataset. At baseline, clusters of people based on 61
14 comorbidities will be identified using Latent class (i.e., Gaussian mixture models algorithms
15 of cluster) analysis (30). For each model, we will examine the association between clusters
16 and covariates using multinomial logistic regressions. The distinctness of clusters will be
17 examined by comparing covariates among clusters. The optimal model is the one where most
18 clusters found in the training data are also identified independently in the testing data and
19 clusters have most distinct patients' characteristics. We will then use both latent trajectory
20 analysis, such as joint latent class models (31), and unsupervised machine learning approach,
21 such as deep autoencoder or recurrent neural networks (32), to identify distinct clusters of
22 new comorbidity numbers development over time and their association with mortality with
23 adjustment for baseline covariates.
24
25
26
27
28
29
30
31
32
33

34
35 *Objective 3 - Association between OA drugs and incident comorbidities*

36 A cohort study will be undertaken for this objective to evaluate the contribution of common
37 analgesics for OA to the development of comorbidity such as NSAIDs, opioids and
38 paracetamol. We are interested in the interaction between OA and use of drugs on the
39 incidence of comorbidities, i.e. to evaluate if the drug use in persons with OA poses increased
40 or decreased risk of comorbidities compared to persons without OA and/or analgesics.
41 Individual comorbidity, as well as clusters of comorbidities identified from Objective 2 will
42 be examined as outcomes. The 61 comorbidities in our study will be further categorised into
43 eight groups, specifically: musculoskeletal (MSK), respiratory, neurodegenerative,
44 psychological/psychiatric, cancer, cardiovascular, metabolic, renal problem, liver diseases,
45 gastrointestinal (GI) and others. (Table 4) The prospective cohort established from Objective
46 1 will form the source population for this objective. Individuals with incident OA will be
47 identified from the database and the first diagnosis date will be used as index date for follow-
48 up. Individuals without OA during the study period will be selected and matched with cases
49 by age, sex, and practice. The same index date will be given from their matched OA cases.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Individuals with analgesics prescriptions prior to the index date will be excluded (or recorded
4 as a confounding factor to be adjusted as appropriate). Only analgesic prescriptions after the
5 index date will be considered for this analysis. Prescriptions will be quantified as number of
6 prescriptions within year 1 (initial use, primary analysis) (33), 2, 3, 4, 5 etc. It will also be
7 dichotomised as episodic (e.g., at least one gap of ≥ 90 days between prescriptions) and
8 continuous (no gap of more less than 90 days) users as appropriate (34,35). Analgesic use
9 will be included in the model as a risk factor together with OA diagnosis (yes/no, primary
10 exposure) to examine the independent risk of each variable (OA and analgesics), as well as
11 the interaction between the two to the development of comorbidity. Dose response
12 relationship will be examined using number of prescriptions during the exposure window
13 examined. The effect of stopping analgesics will also be examined by looking into the
14 patterns of analgesic prescriptions, e.g., stopping analgesics after initial use in year 1 versus
15 continuous use of analgesics afterwards. For the primary analysis (initial prescriptions within
16 year 1), a landmark analysis will be used to minimise the immortal time bias where the
17 follow up will start after 12 months from the index date (36). Participants at risk (i.e., without
18 a specific comorbidity of interest) at the landmark date will be followed up until the first
19 diagnosis of the comorbidity, deregistration, or death whichever comes first. For secondary
20 analyses, time varying covariate analysis will be used to examine the long-term,
21 episodic/continuous use of analgesics after the index date and interaction between OA and
22 analgesics in the development of the comorbidity. The propensity score matching or the
23 inverse probability weighing methods will be used to adjust for confounding by indication
24 during the follow-up as appropriate. Depending on the country-specific drug use patterns, we
25 may modify this definition to allow for short breaks in between the episodes. Cox-regression
26 model will be used to calculate the HR and 95% CI. We will use flexible parametric models
27 using restricted cubic splines (developed by Lambert, “stpm2” in Stata) to estimate the HRs
28 and differences in time to diagnosis of comorbidities (outcome) with drugs as time-varying to
29 account for non-proportional hazards (37).

50 *Objective 4 – Potential causal pathways between OA and the comorbidities*

51
52 We will perform a Mendelian Randomisation (MR) phenome-wide association (MR-
53 PheWAS) study(38) to examine the causal relationship between OA, its phenotypes,
54 biomarkers or risk factors and comorbidities using the UK Biobank and the Rotterdam Study
55 database.
56
57
58
59
60

1
2
3 We will use the Rotterdam Study and the UK Biobank jointly for this objective. This is
4 because that the Rotterdam Study is an OA cohort with deep phenotypes and biomarkers of
5 OA, whereas the UK Biobank is a primary cohort for cancer and multiple disease outcomes,
6 and both have detailed genetic variants. We will use two sample MR approach, i.e., to
7 establish an association between OA and genetic variants in the Rotterdam Cohort to identify
8 genetic instrumental variables (IV), e.g., a set of single nucleotide polymorphisms (SNPs)
9 associated with OA (or a deep phenotype, biomarker, or risk factor of OA). We will then
10 undertake the MR-PheWAS analysis to examine the causal effects of the OA IV on
11 comorbidities in the UK Biobank. The MR method has been widely used in real world data to
12 examine the causal relationship between IV and specific disease, under two assumptions: [1]
13 genetic variants are randomly assigned in the population; and [2] genetic variants can only be
14 the cause not consequence of disease (39). The PheWAS is a series of case control studies to
15 estimate the associations between the IV and multiple disease outcomes (38,40). The
16 combination of the two permits investigation of the causal effects of OA on multiple disease
17 outcomes.

18
19 The MR-PheWAS analysis includes three steps. Firstly, we will identify the genetic variants
20 that are associated with OA - IV. Secondly, we will undertake the PheWAS analysis – a
21 series of case control analyses to estimate the associations between the IV and other disease
22 outcomes (38,40), with an adjustment for multiple testing using the false discovery rate
23 (FDR) methods (41). Thirdly, we will implement conventional MR analysis to investigate the
24 causal effects of the OA IV on comorbidities (39). An inverse variance weighted (IVW)
25 method will be used to pool the associations (ORs) as appropriate (42). The MR-Egger
26 regression analysis will be used to count for the pleiotropic effect - the effects of one genetic
27 variant on multiple outcomes (43). The heterogeneity in dependent instruments (HEIDI) test
28 will be used to exclude the cross-phenotype associations caused by genetic linkage (44).

29
30 With the MR-PheWAS study, the OR can be interpreted as causal association. We are
31 primarily interested in the causality from OA to comorbidities. We are also interested in
32 inflammatory (e.g., CRP), metabolic (e.g., gut microbiome) and biomechanics (e.g., BMI)
33 biomarkers and deep phenotypes of OA such as knee, hip, and hand OA with and without
34 symptoms. This will be undertaken if it is feasible within 3 years of this funded project,
35 otherwise will be considered as our future research agenda.

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58 *Objective 5 – Variation of OA comorbidity patterns across countries*
59
60

1
2
3 We will use meta-analysis (MA) to examine the variation between countries and to pool the
4 data as appropriate. Estimates from first three objectives such as prevalence, incidence, OR,
5 HR and 95% CI for each specific comorbidity across different populations will be distributed
6 in a forest plot. Heterogeneity will be examined using the I^2 statistic and the Q test (45).
7
8 Results will be pooled if they are homogenous based on the I^2 value using the fixed effects
9 model, otherwise the reasons for the heterogeneity will be investigated. Random effects
10 models will be used to pool the results if the reasons for the heterogeneity cannot be
11 identified and if the overall pooling is appropriate. Individual patient data (IPD) meta-
12 analysis may be used to help identify the reasons for heterogeneity (46). Common clusters
13 and trajectories as well as burdens of comorbidities will also be compared between
14 populations.
15
16
17
18
19
20
21
22

23 Discussion

24
25 OA is the most common arthritis and the second most common musculoskeletal (the first
26 being back pain) in older people. However, unlike other RMDs, relatively little is known
27 about comorbidities in OA. OA previously was defined as a “wear and tear” joint-specific
28 degenerative condition, but recent research has found that it is a common complex disorder
29 which may increase the risk of other chronic conditions in other systems. This study will be
30 the largest epidemiological study on comorbidities of OA in primary care.
31
32
33

34 One of the key advantages of this multinational study is the use of the same protocol to
35 measure the burden of comorbidities in primary care settings in four European countries to
36 ensure reproducibility and comparison. There is scant evidence on the comorbidities in
37 people with OA, and this approach should help to identify the leading and most important
38 associations before and after presenting clinical OA (the index date). Further advantages of
39 this study are the large and representative populations studied and the same/similar extensive
40 list of chronic conditions for identifying comorbidity clusters. Often comorbidities
41 accumulate with age over time and the large primary care databases in this study have the
42 advantage of having long follow-up time which will enable us to detect the incidence of
43 comorbidities. Also, longer follow up would help to identify the picture of the trajectory of
44 the diseases (47). Both the incidence and the trajectories of comorbidity clusters are
45 highlighted as key elements needed in current research in multimorbidity, so findings from
46 this study should help to fill the knowledge gaps on multimorbidity in OA.
47
48
49
50
51
52
53
54
55
56
57

58 The relationship between chronic conditions and polypharmacy is a complex area of research.
59 The count of the medications and more importantly the nature of prescribed drugs may be
60

1
2
3 responsible for developing many new comorbidities in people with OA. We aim to explore
4 the associations of the most commonly prescribed drugs in OA, such as NSAIDs, with the
5 incidence of a wide range of comorbidities, which will be the first time that conditions other
6 than established comorbidities such as psychological conditions and endocrine diseases will
7 be examined. Finally, the causality study will further explore the associations at genetic
8 levels and phenotypes, which will be novel in OA research. Using a two sample MR
9 approach - one for OA deep phenotypes and the other for other chronic conditions maximises
10 the potentials of sample size, disease phenotypes and comorbidity spectrum to better explore
11 the causal pathways between OA and comorbidity.

12
13
14
15
16
17
18
19 There are some limitations to this study. Firstly, there are inherent issues in the nature of
20 electronic health records with respect to possible misdiagnosis, ascertainment biases, under-
21 or over-recording, and changes in databases due to change in coding structures. Also, the
22 analysis will be restricted to fewer covariates in some databases due to missing information
23 on lifestyle factors such as physical activities and diet. Even though the databases have
24 different durations of data available, if possible we will use a common follow-up time for
25 objective five.

26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Chronic conditions, especially comorbidities recorded in general practices, depend on
multiple factors such as population structure, health care facilities, health policies, and the
nature of the national databases. A major strength of this study is that it will include medical
records on approximately 27 million people in four European countries. Also, the study will
cover the sequence of research questions in comorbidity or multimorbidity starting from the
burden through to the causality and variation. Such a research model can be used for other
similar multimorbidity studies. The expected results should inform health professionals in
primary care settings with respect to management of people with OA and associated
comorbidities.

Status of the study

All the centres have obtained the necessary approvals for using the database in 2020. A
consensus has been made on the code mapping exercise. The statistical analysis will be
explained in detail in each of the publications. The team is expected to produce results by
mid-2021.

Acknowledgements: We thank the Patient Research Participants (PRP) members Jenny
Cockshull, Stevie Vanhegan, and Irene Pitsillidou for their involvement since the beginning
of the project. We would like to thank the FOREUM for financially supporting the research.
The authors would like to acknowledge Keele University's Prognosis and Consultation

1
2
3 Epidemiology Research Group who have given us permission to utilise the Code Lists
4 (©2014).
5
6

7 **Contributor and guarantor information:**

8 WZ, MD, CC, SMA, ME, DA conceived and designed the study. SS, AK, AD, AT, DR and
9 will perform the analysis and CC, WZ, JR and AT will supervise the statistical analysis. All
10 authors contributed to the critical revision of the manuscript for important intellectual
11 content. The corresponding author attests that all listed authors meet authorship criteria and
12 that no others meeting the criteria have been omitted.
13
14

15 **Funding:** This work was supported by Foundation for Research in Rheumatology
16 (FOREUM) grant (2019-2022), The Swedish Research Council (2020-01103), Governmental
17 funding of clinical research within the national health services (ALF), and The Swedish
18 Rheumatism Association. CM is funded by the National Institute for Health Research (NIHR)
19 Applied Research Collaboration West Midlands, the National Institute for Health Research
20 (NIHR) School for Primary Care Research and a National Institute for Health Research (NIHR)
21 Research Professorship in General Practice (NIHR-RP-2014-04-026)
22
23

24 **Role of the funding sources:** The sponsors did not participate in the design and conduct of
25 the study; collection, management, analysis, and interpretation of the data; or preparation,
26 review, or approval of the manuscript and the decision to submit the manuscript for
27 publication.
28
29

30 **Competing interests:** WZ declares serving as an advisory board for Ely Lilly (Ixekizumab,
31 2020) and Regeneron (Fasinomab, 2020). ME declares serving as an advisory Panel Board
32 Member for Pfizer (Nov 2019, Tanezumab). CM provided advice to BMS on recruiting to a
33 non-pharmacological atrial fibrillation trial. The other authors report no financial competing
34 interests.
35
36

37 **Copy right:** The Corresponding Author has the right to grant on behalf of all authors and
38 does grant on behalf of all authors, an exclusive licence (or non-exclusive for government
39 employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if
40 accepted) to be published in BMJ editions and any other BMJ PGL products and sublicences
41 such use and exploit all subsidiary rights, as set out in our licence.
42
43
44

45 **Studies involving humans or animals:** No direct participant recruitment was done for the
46 study.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **Data sharing statement:** We used anonymised data on individual patients on which the
4 analysis, results, and conclusions reported in the paper are based. The used data is not
5 distributable under licence. However, the relevant data can be obtained directly from the
6 respective agencies. The codes developed for the analysis can be available upon a valid
7 request.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Reference

1. Turkiewicz A, Petersson IF, Björk J, Hawker G, Dahlberg LE, Lohmander LS, et al. Current and future impact of osteoarthritis on health care: a population-based study with projections to year 2032. *Osteoarthritis Cartilage*. 2014 Nov;22(11):1826–32.
2. Badley EM. The effect of osteoarthritis on disability and health care use in Canada. *J Rheumatol Suppl*. 1995 Feb;43:19–22.
3. McDonough CM, Jette AM. The Contribution of Osteoarthritis to Functional Limitations and Disability. *Clin Geriatr Med*. 2010 Aug;26(3):387–99.
4. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. 2020 Dec;396(10267):2006–17.
5. World Health Organization. Musculoskeletal conditions [Internet]. [cited 2021 Apr 6]. Available from: <https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions>
6. Swain S, Sarmanova A, Coupland C, Doherty M, Zhang W. Comorbidities in Osteoarthritis: A systematic review and meta-analysis of observational studies. 2019 Jun 17 [cited 2019 Jun 30]; Available from: <http://doi.wiley.com/10.1002/acr.24008>
7. Suri P, Morgenroth DC, Hunter DJ. Epidemiology of osteoarthritis and associated comorbidities. *PM R*. 2012 May;4(5 Suppl):S10-19.
8. Hall AJ, Stubbs B, Mamas MA, Myint PK, Smith TO. Association between osteoarthritis and cardiovascular disease: Systematic review and meta-analysis. *Eur J Prev Cardiol*. 2016;23(9):938–46.
9. Stubbs B, Aluko Y, Myint PK, Smith TO. Prevalence of depressive symptoms and anxiety in osteoarthritis: a systematic review and meta-analysis. *Age Ageing*. 2016 Mar 1;45(2):228–35.
10. Ricci-Cabello I, Stevens S, Kontopantelis E, Dalton ARH, Griffiths RI, Campbell JL, et al. Impact of the Prevalence of Concordant and Discordant Conditions on the Quality of Diabetes Care in Family Practices in England. *Ann Fam Med*. 2015 Nov 1;13(6):514–22.
11. Calderón-Larrañaga A, Poblador-Plou B, González-Rubio F, Gimeno-Feliu LA, Abad-Díez JM, Prados-Torres A. Multimorbidity, polypharmacy, referrals, and adverse drug events: Are we doing things well? *Br J Gen Pract*. 2012;62(605):e821–6.
12. Jüni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *The Lancet*. 2004 Dec;364(9450):2021–9.
13. Castellsague J, Riera-Guardia N, Calingaert B, Varas-Lorenzo C, Fourrier-Reglat A, Nicotra F, et al. Individual NSAIDs and Upper Gastrointestinal Complications: A Systematic Review and Meta-Analysis of Observational Studies (the SOS Project). *Drug Saf*. 2012 Dec;35(12):1127–46.
14. Lowe D.B., Taylor M.J., Hill S.J. Cross-sectional examination of musculoskeletal conditions and multimorbidity: influence of different thresholds and definitions on prevalence and association estimates. *BMC Res Notes*. 2017;10(1):51.
15. Alaeddini A., Jaramillo C.A., Faruqi S.H.A., Pugh M.J. Mining Major Transitions of Chronic Conditions in Patients with Multiple Chronic Conditions. *Methods Inf Med*. 2017;56(5):391–400.

16. John R, Kerby DS, Hagan Hennessy C. Patterns and Impact of Comorbidity and Multimorbidity Among Community-Resident American Indian Elders. *The Gerontologist*. 2003 Oct 1;43(5):649–60.
17. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med*. 2015 Mar 31;12(3):e1001779.
18. Engelhart MJ, Geerlings MI, Meijer J, Kiliaan A, Ruitenber A, van Swieten JC, et al. Inflammatory Proteins in Plasma and the Risk of Dementia: The Rotterdam Study. *Arch Neurol*. 2004 May 1;61(5):668.
19. Quality and Outcome Framework (QOF) [Internet]. [cited 2018 Mar 21]. Available from: <https://digital.nhs.uk/article/8910/Quality-and-Outcome-Framework-QOF-Indicators-No-Longer-In-QOF-INLIQ-Enhanced-Services-ES-Vaccinations-and-Immunisations-V-I-and-GMS-Core-Contract-CC-extraction-specifications-business-rules->
20. Medicare C for, Baltimore MS 7500 SB, Usa M. CC_Main [Internet]. 2017 [cited 2018 Oct 24]. Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/CC_Main.html
21. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. 2018 Nov;392(10159):1789–858.
22. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987 Jan 1;40(5):373–83.
23. Sarmanova A, Fernandes GS, Richardson H, Valdes AM, Walsh DA, Zhang W, et al. Contribution of central and peripheral risk factors to prevalence, incidence and progression of knee pain: a community-based cohort study. *Osteoarthritis Cartilage*. 2018 Nov;26(11):1461–73.
24. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol*. 1994 Nov 1;47(11):1245–51.
25. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. *Am J Epidemiol*. 2011 Mar 15;173(6):676–82.
26. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A Modification of the Elixhauser Comorbidity Measures Into a Point System for Hospital Death Using Administrative Data: *Med Care*. 2009 Jun;47(6):626–33.
27. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998 Jan;36(1):8–27.
28. Kuo C-F, Grainge MJ, Mallen C, Zhang W, Doherty M. Comorbidities in patients with gout prior to and following diagnosis: case-control study. *Ann Rheum Dis*. 2014 Nov 14;annrheumdis-2014-206410.
29. Swain S, Coupland C, Mallen C, Kuo CF, Sarmanova A, Bierma-Zeinstra SMA, et al. Temporal relationship between osteoarthritis and comorbidities: a combined case control and cohort study in the UK primary care setting. *Rheumatology*. 2021 Mar 1;keab067.

- 1
- 2
- 3 30. Collins LM, Lanza ST. Latent Class and Latent Transition Analysis [Internet]. Hoboken, NJ,
4 USA: John Wiley & Sons, Inc.; 2009 [cited 2020 Apr 5]. (Wiley Series in Probability and
5 Statistics). Available from: <http://doi.wiley.com/10.1002/9780470567333>
6
- 7 31. Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal
8 and time-to-event data: A review. *Stat Methods Med Res.* 2014 Feb;23(1):74–90.
9
- 10 32. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked Denoising Autoencoders:
11 Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J Mach*
12 *Learn Res.* 2010 Dec;11:3371–408.
13
- 14 33. Zeng C, Zhang W, Doherty M, Persson MSM, Mallen C, Swain S, et al. Initial analgesic
15 prescriptions for osteoarthritis in the United Kingdom, 2000–2016. *Rheumatology.* 2021 Jan
16 5;60(1):147–59.
17
- 18 34. Hippisley-Cox J, Coupland C. Unintended effects of statins in men and women in England and
19 Wales: population based cohort study using the QResearch database. *BMJ.* 2010 May
20 20;340(may19 4):c2197–c2197.
21
- 22 35. Sarmanova A, Doherty M, Kuo C, Wei J, Abhishek A, Mallen C, et al. Statin use and risk of joint
23 replacement due to osteoarthritis and rheumatoid arthritis: a propensity-score matched
24 longitudinal cohort study. *Rheumatology.* 2020 Oct 1;59(10):2898–907.
25
- 26 36. Kuo C-F, Grainge MJ, Mallen C, Zhang W, Doherty M. Effect of allopurinol on all-cause
27 mortality in adults with incident gout: propensity score–matched landmark analysis.
28 *Rheumatology.* 2015 Jul 12;kev246.
29
- 30 37. Royston P, Lambert PC. Flexible parametric survival analysis using Stata: beyond the Cox model.
31 College Station, TX: Stata Press; 2011. 347 p.
32
- 33 38. Millard LAC, Davies NM, Timpson NJ, Tilling K, Flach PA, Smith GD. MR-PheWAS:
34 hypothesis prioritization among potential causal effects of body mass index on many outcomes,
35 using Mendelian randomization. *Sci Rep [Internet].* 2015 Dec [cited 2020 Jan 26];5(1). Available
36 from: <http://www.nature.com/articles/srep16645>
37
- 38 39. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide,
39 glossary, and checklist for clinicians. *BMJ.* 2018 Jul 12;k601.
40
- 41 40. Li X, Meng X, Spiliopoulou A, Timofeeva M, Wei W-Q, Gifford A, et al. MR-PheWAS:
42 exploring the causal effect of SUA level on multiple disease outcomes by using genetic
43 instruments in UK Biobank. *Ann Rheum Dis.* 2018 Jul;77(7):1039–47.
44
- 45 41. Benjamini Y, Yekutieli D. The Control of the False Discovery Rate in Multiple Testing under
46 Dependency. *Ann Stat.* 2001;29(4):1165–88.
47
- 48 42. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple
49 Genetic Variants Using Summarized Data: Mendelian Randomization Using Summarized Data.
50 *Genet Epidemiol.* 2013 Nov;37(7):658–65.
51
- 52 43. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian
53 Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet*
54 *Epidemiol.* 2016 May;40(4):304–14.
55
- 56
- 57
- 58
- 59
- 60

- 1
- 2
- 3
- 4 44. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data
- 5 from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016
- 6 May;48(5):481–7.
- 7
- 8 45. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002 Jun
- 9 15;21(11):1539–58.
- 10
- 11 46. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic
- 12 reviews using individual patient data. *Eval Health Prof.* 2002 Mar;25(1):76–97.
- 13
- 14 47. Vetrano DL, Roso-Llorach A, Fernández S, Guisado-Clavero M, Violán C, Onder G, et al.
- 15 Twelve-year clinical trajectories of multimorbidity in a population of older adults. *Nat Commun*
- 16 [Internet]. 2020 Dec [cited 2020 Aug 12];11(1). Available from:
- 17 <http://www.nature.com/articles/s41467-020-16780-x>
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

Figure 1. Overview of the study design and statistical analysis plan for the Comorbidities in Osteoarthritis (CoMO) study.

	Objective 1	Objective 2	Objective 3	Objective 4
Study Design	Case-control and Cohort	Clustering and longitudinal	Cohort	Genomic association
Exposure	OA	OA	Analgesics (NSAIDs, Opioids, paracetamol)	OA
Outcome	Comorbidities	Clusters of comorbidities	Comorbidities	Comorbidities
Statistical methods	Conditional logistic regression, Cox regression	Latent class analysis, Latent class growth analysis, Joint latent class analysis	Cox regression, Time varying analysis, Flexible parametric method	Mendelian randomisation
Reported Outcome	Odds Ratio and Hazard Ratio	Clusters and groups	Hazard Ratio	Coefficients
Participating centres	ALL	ALL	ALL	UK and Netherlands

Objective 5	Meta Analysis
-------------	---------------

OA- osteoarthritis; NSAIDs- Nonsteroidal anti-inflammatory drugs

Table 1. Characteristics of the included databases

	Netherlands	Spain	Sweden	UK
Objectives 1-3				
Name of the database	Integrated Primary Care Information (IPCI)	The Spanish Sistema information del Deveolpment de l'Investigació a Atenció Primària (SIDIAP)	Skåne Healthcare Register	Clinical Practice Research Datalink (CPRD)
Settings	Routinely collected primary care database	Routinely collected primary care data	Swedish healthcare in Skane region, primary, specialist and in-patient care	Routinely collected primary care database with linkage database
Size and Coverage	2.2 million (Randomly distributed over the country)	6.5 million (> 85% of total Catalan region)	1.3 million (all residents of the Skane region)	17 million (country-wide, nearly 740 practices)
Start year	1998- (Better coding after 2000)	2006	1998	1993-
Age group	All	All	All	All
Gender	All	All	All	All
Coding system	ICPC	ICD 10	ICD 10	Read codes and ICD 10
Drug prescribed by	GP	GP	GP	GP
Death record (Either date of death and/or cause)	Both date and cause	Only date	Both date and cause (until year 2015)	Only date
Covariates/ additional variables	NA	BMI, Smoking, Alcohol, Social class, cholesterol, and other biomarkers	Education, Income, profession, and sick leave, residential area, region of birth	BMI, Smoking, Alcohol, Deprivation index, Ethnic group

BMI- Body mass index; GP- general practitioners; ICD- International classification of diseases; ICPC- International classification of primary care; NA- Not available

Table 2. Database for the mendelian randomisation

Name of the database	Rotterdam Cohort Study	UK Biobank
Population coverage	15000	500,000
Age group	>=40 years	40-69 years
Start year -till now	1989-onwards	2010-
Types of data	Radiographic data, joint pain, joint stiffness, of hip, knee, and hand, GWAS, biochemical markers	Genetic and phenotypes

GWAS- Genome Wide Association Studies

For peer review only

Table 3. List of chronic conditions across four databases

SI no	Conditions	ICD 10	ICPC	Read Code
1	Anaemia (All types)	D50-D64	B78, B80, B81, B82	D00..., D01...
2	Ankylosing spondylitis	M45.9	NA	N10...
3	Anxiety disorder	F41.0-F41.9	P74, P74.01, P74.02	E200..., Eu41...
4	Asthma	J450-J45.9, J46.9	R96...	66Y..., H33...
5	Benign prostatic hypertrophy (BPH)	N40.9	Y85	K20..., K21..., K22...
6	Cardiac arrhythmias (Atrial Fibrillation)	I47.0-I49.9	K78	Gyu..., G573...
7	Cataract	H25.0-H25.9, H26.1-H26.9	F92	F46...,
8	Chronic Back pain	M47-M48, M51-M54, M99, G54.4	L02, L03, L86	N12..., N14...,
9	Chronic kidney disease (any cause)	N02.0-N8.8, N11.0-N11.9, N12.9, N15.0-N18.9, N19.9	U99.01	I21..., K01..., K02...
10	Chronic neck pain	M54.2	L83...	Nyu..., N11..., N12..., N14...
11	Chronic sinusitis	J32 ...	R75...	H13...
12	Chronic obstructive pulmonary diseases	J41.0-J41.8, J42.9, J43.0-J43.9, J44.0-J44.9	R91, R95	H3...
13	Coronary Heart Disease (Including Acute Myocardial infarction, Valvular disease, Angina),	I20.0-I25.0, I34.0-I37.0	K74-K76	G11..., G30..., G31... - G38..
14	Dementia	F00.0-F00.9, F01.0-F03.9, G30.0-G30.9, G31.0-G31.9	P70-P70.02	E00..., Eu0..., F11...
15	Depression	F32.0-F33.9	P76	Eu..., E11...
16	Diabetes mellitus	E10.0-E14.9	T90..., F83.01	C10..., F32...,
17	Dyslipidaemia (Hyper)	E78.1	T93	C32...
18	Eating disorders (Both)			Eu5..., R03...
19	Eczema/ Skin disease	L20.0-L22.9, L26.9	S74, S87, S88	M11...
20	Epilepsy	G40.0-G41.9	N88	F25...
21	Fatigue	F48.0	A04.11	F286...
22	Fibromyalgia	M79.7	L18.01	N248..., N239...
23	Gall bladder stone	K80.0-K80.8	D98-D98.03	781..., J65..., 4G2...,
24	GERD (Gastritis, Oesophageal bleeding, duodenitis, peptic ulcer)	K21 ...	D840...	J12..., J13..., J15...
25	Gastrointestinal bleeding	K25.0-K28.9	D84-D87	J11...,
26	Gout	M10.0-M10.9	T92	C34..., N023...
27	Hearing impairment (All types)	H90.0-H91.9	H83-H86	F59..., ZE87...
28	Heart Failure	I50.0-I50.9	K77-K77.02	G58...,
29	Hepatitis	K73.0-K73.9	D72-D72.05	J61..., J63...
30	HIV/AIDS	B20-B24	B90	A788..., A789..., AyuC...
31	Hypertension	I10.9, I11.0-I13.9, I15.0-I15.9	K86-K87, F83.02	G20..., G24..., G25..., G26...
32	Hyperthyroidism	E05.0-E05.9	T85	C02...
33	Hypothyroidism	E02.9, E03.0-E03.9	T86	C03..., C04...
34	Inflammatory Bowel Disease (IBD)	K50.0-K52.9	D94-D94.02	J4...,
35	Irritable Bowel Symptoms (IBS)	K58.1-K58.8		J52...
36	Leukaemia, Lymphoma	C81.0-C86.6, C91.0-C96.9	B72-B73	B60..., B61..., B64...
37	Liver Cirrhosis	K70.0-K71.9, K74.0-K74.6	D97...	J615...
38	Migraine	G43.0-G43.9	N89	F26...,
39	Multiple sclerosis	G35.9	N86	F20...,
40	Osteoarthritis	M16.0-M16.9, M17.0-M17.9	L89-L91	N05...,
41	Osteoporosis	M80.0-M82.9	L95...	N33...,

42	Other blood vessel disease (Raynaud's disease, Burger's disease)	I73.0-I73.9	K92...	G73... ,
43	Parkinson's disease	G20.9	N87-N87.01	F12... ,
44	Peripheral vascular disease (Atherosclerosis)	I70.0-I70.9	K91	G70... , G71... , G72... ,
45	Polymyalgia	M35.3	L99.12	N20.11
46	Psoriasis	L40.0-L41.9	S91	M161...
47	Psoriatic arthritis	M07.0-M07.3	L99.13	M160...
48	Rheumatoid Arthritis	M05.0-M05.9	L88... , K71...	N04....
49	Renal stones	N20.0	U95	4G4... , 7B07... , KB12...
50	Schizophrenia and/or psychosis	F20.0-F20.9, F25.0- F25.9	P72	E10...
51	Severe allergy			H17... , SN5...
52	Sjögren's syndrome	M35.0	NA	N002...
53	Systemic Lupus Erythematosus	M32.0, M32.1, M32.8, M32.9	NA	N000...
54	Sleep disorder (Insomnia)	F51.0	P06	Fy0... , 1B1B...
55	Solid malignancy	C00.0-C80.9, D00.0-D09.9, C97.9	A29... , A79, B74 – Y78	B0... - B67... , Byu...
56	Stroke	G45.0-G46.8, I60.0-I63.9, I65.0-166.9, I69.0-I69.4	K89-K90.02	G60... - G68... , F22...
57	Substance abuse/ Drug addiction	F10.0-F19.9	P18, P19...	E24... , Eu1...
58	Thrombotic diseases	I74.0-I74.9	K93, K94... , W99.03	G80... , G81... , G74...
59	Tuberculosis	A15.0-A16.9, B90.9	A70, R70	A1... , A11...
60	Vertigo	H81.4 ...	N17-N17.02, H82-H82.03	R004... , F561...
61	Vision problem (Glaucoma and other)	H27.0-H27.9, H40.0-H40.9, H42.0-H42.8	F93... , F94	F45... , F49... ,

ICD- International Classification of Diseases; ICPC- International classification in primary care. All the codes are the primary code initials used in the database.

Table 4. Group of conditions/Outcome

Group	Conditions
Cardiovascular	Cardiac arrhythmias, Coronary Heart Disease (including AMI, valvular disease, angina), Heart failure, Hypertension, Peripheral vascular disease (claudication, Raynaud syndrome, Buerger's disease), Other blood vessel disease (atherosclerosis and aneurysm), Thrombotic diseases
Gastrointestinal	GERD (Esophageal diseases, gastritis, duodenitis), GI bleeding, Inflammatory bowel disease (IBD), Irritable bowel syndrome (IBS)
Musculoskeletal	Ankylosing spondylitis, Chronic Back pain, Chronic Neck pain, Fibromyalgia, Polymyalgia, Gout, Osteoporosis, Psoriatic arthritis, Rheumatoid arthritis, Sjögren's syndrome, Systemic lupus erythematosus (SLE)
Endocrine	Diabetes mellitus, Dyslipidemia (hyper), Hyperthyroidism, Hypothyroidism
Neurological	Dementia, Epilepsy, Fatigue, Migraine, Multiple sclerosis, Parkinson disease, Stroke
Psychological	Anxiety, Depression, Eating disorders (Anorexia / Bulimia nervosa), Schizophrenia, Sleep disorder (insomnia),
Kidney disease	Chronic kidney disease (any cause), Renal stones
Liver diseases	Gall bladder stone, Hepatitis, Liver cirrhosis
Respiratory	Asthma, Chronic obstructive pulmonary disease (COPD)
Cancer	Leukemia, Lymphoma, Solid Malignancy (any type)
Others	Anemia (all types), Benign prostate hypertrophy (BPH), Cataract, Chronic sinusitis, Eczema/Skin disease, Hearing impairment (all types), Psoriasis, Severe allergy (anaphylactic shock), angioneurotic oedema

BMJ Open

Comorbidities in Osteoarthritis (CoMO): a multinational study in four European countries

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2021-052816.R1
Article Type:	Protocol
Date Submitted by the Author:	10-Jan-2022
Complete List of Authors:	Swain, Subhashisa; University of Nottingham, Academic Rheumatology Kamps, Anne; Erasmus Medical Center Department of Rheumatology, Department of General Practice Runhaar, Jos; Erasmus Medical Center Department of Rheumatology, Department of General Practice Dell'Isola, Andrea; Lunds University Faculty of Medicine, Turkiewicz, Aleksandra; Lund University Faculty of Medicine, Dept of Orthopedics Robinson, Danielle ; University of Oxford Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, Strauss, V; University of Oxford Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, Mallen, Christian; Keele University, Arthritis Research UK Primary Care Centre Kuo, Chang-Fu; Chang Gung Memorial Hospital Linkou Branch, Coupland, Carol; University of Nottingham, Division of Primary Care Doherty, Michael ; University of Nottingham School of Medicine, Academic Rheumatology Sarmanova, Aliya ; University of Bristol, Prieto-Alhambra, Daniel; University of Oxford, Centre for Statistics in Medicine, NDORMS Englund, Martin; Lund University, Dept of Orthopedics Bierma-Zeinstra, Sita; Erasmus University Medical Centre, Department of General Practice Zhang, Weiya; University of Nottingham School of Medicine, Academic Rheumatology
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	Epidemiology, Health services research, Public health, Pharmacology and therapeutics, Rheumatology
Keywords:	EPIDEMIOLOGY, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, RHEUMATOLOGY

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



1
2
3 1 **Title-** Comorbidities in Osteoarthritis (CoMO): a multinational study in four European
4 countries

5
6
7 3 **Authors**

8
9 4 Subhashisa Swain^{1,10}, Anne Kamps², Jos Runhaar², Andrea Dell 'Isola³, Aleksandra
10
11 5 Turkiewicz³, Danielle Robinson⁴, Victoria Y Strauss⁴, Christian Mallen⁵, Chang Fu Kuo^{1,6},
12
13 6 Carol Coupland⁷, Michael Doherty^{1,10}, Aliya Sarmanova⁸, Daniel Prieto Alhambra⁴, Martin
14
15 7 Englund³, Sita Bierma-Zeinstra⁹, Weiya Zhang^{1,10}

- 16
17 8
18
19 9 1. Academic Rheumatology, School of Medicine, University of Nottingham, UK
20
21 10 2. Department of General Practice, Erasmus MC University Medical Center Rotterdam,
22
23 11 The Netherlands
24
25 12 3. Clinical Epidemiology Unit, Orthopaedics, Department of Clinical Sciences Lund,
26
27 13 Lund University, Sweden
28
29 14 4. Centre for Statistics in Medicine, NDORMS, University of Oxford, UK
30
31 15 5. School of Medicine, Keele University, UK
32
33 16 6. Division of Rheumatology, Allergy and Immunology, Chang Gung Memorial
34
35 17 Hospital, Taiwan.
36
37 18 7. Division of Primary Care, School of Medicine, University of Nottingham, UK
38
39 19 8. Musculoskeletal Research Unit, Bristol Medical School, Translational Health
40
41 20 Sciences, University of Bristol, UK
42
43 21 9. Department of General Practice, Department of Orthopaedic Surgery & Sports
44
45 22 Medicine, Erasmus MC University Medical Center Rotterdam, The Netherlands
46
47 23 10. Pain Centre Versus Arthritis, University of Nottingham, UK

48
49 24 **Corresponding Author:**

50
51 25 Professor Weiya Zhang
52
53 26 Academic Rheumatology, School of Medicine
54
55 27 University of Nottingham
56
57 28 Clinical Sciences Building, Nottingham city Hospital
58
59 29 NG5 1PB, Nottingham, UK
60
30 Email: weiya.zhang@nottingham.ac.uk

34 **Abstract**

35 *Introduction*

36 Osteoarthritis (OA) is one of the leading chronic conditions in the older population. People
37 with OA are more likely to have one or more other chronic conditions than those without.
38 However, the temporal associations, clusters of the comorbidities, role of analgesics and the
39 causality and variation between populations are yet to be investigated. This paper describes
40 the protocol of a multinational study in four European countries (UK, Netherlands, Sweden,
41 and Spain) exploring comorbidities in people with OA .

42 *Methods and analysis*

43 This multinational study will investigate i) the temporal associations of 61 identified
44 comorbidities with OA, ii) the clusters and trajectories of comorbidities in people with OA,
45 iii) the role of analgesics on incidence of comorbidities in people with OA, iv) the potential
46 biomarkers and causality between OA and the comorbidities, and v) variations between
47 countries.

48 A combined case-control and cohort study will be conducted to find the temporal association
49 of OA with the comorbidities using the national or regional health databases. Latent class
50 analysis will be performed to identify the clusters at baseline and joint latent class analysis
51 will be used to examine trajectories during the follow-up. A cohort study will be undertaken
52 to evaluate the role of non-steroidal anti-inflammatory drugs (NSAIDs), opioids, and
53 paracetamol on the incidence of comorbidities. Mendelian randomisation will be performed
54 to investigate the potential biomarkers for causality between OA and the comorbidities using
55 the UK Biobank and the Rotterdam Study databases. Finally, a meta-analyses will be used to
56 examine the variations and pool the results from different countries.

57 *Ethics and Dissemination*

58 Research ethics was obtained according to each database requirement. Results will be
59 disseminated through the FOREUM website, scientific meetings, publications, and in
60 partnership with patient organisations.

61

62 Key words: Osteoarthritis; Comorbidity; Multimorbidity; Primary care; Analgesics,
63 Mendelian Randomisation, PheWAS

64

65

66

1
2
3 67 **Strengths and limitations of this study**
4
5
6 68

- 7
8 69 • This is first ever multicenter study on comorbidities in osteoarthritis in Europe
9
10 70 involving nearly 27 million electronic health records
11
12 71 • More than 60 chronic conditions are being studied – representing a wider coverage of
13
14 72 diseases
15
16 73 • We will examine the causal association using Mendelian Randomization -PheWAS
17
18 74 methods with genetic data collected from two countries
19
20 75 • Same protocol with robust statistical methods will be used across all countries to
21
22 76 replicate the findings and examine the variations.
23
24 77 • Possible biases may be introduced by the nature of electronic health records and
25
26 78 length of data availability.
27
28 79
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

81 **Background**

82 Osteoarthritis (OA) affects 27% of people aged over 45 years at peripheral synovial joints
83 such as knees, hips, hands and feet (1). It is by far the most common form of arthritis, and a
84 leading cause of chronic joint pain and disability in older people (2,3). It is anticipated that
85 the burden of OA will continue to rise in the coming decades because of population ageing
86 and the increasing obesity prevalence – two major risk factors for OA (4,5). Co-occurrence of
87 multiple chronic conditions in an individual with ageing is becoming a norm and OA is not
88 an exception to this.

89 A recent systematic review has confirmed that people with OA are more likely to have other
90 diseases, especially stroke, peptic ulcer, hypertension, and depression (6). Vast majority of
91 these studies focused on additional presence (comorbidity) of cardiovascular and
92 musculoskeletal conditions only (7–9). Whether these comorbidities just co-exist with OA,
93 share common risk factors with OA, or are causes or consequences of OA remains largely
94 unknown. There was also reporting of wide heterogeneity in definitions of OA and other
95 chronic conditions, diagnosis and recording of diseases, sample sizes, and number of diseases
96 studied in previously published studies included in the review (6). This diversity made the
97 comparison and pooled estimation of comorbidity prevalence difficult.

98 Comorbidity in OA can occur due to multiple factors. Various hypotheses have been used to
99 explain the existence of comorbidities in general, the most accepted of which are the
100 concordant (diseases sharing similar pathophysiological risk factors) and discordant
101 (diseases not sharing similar pathophysiological risk factors) theories (10). Additionally,
102 prescription of drugs is also reported to be associated with comorbidity and multimorbidity
103 (11). Especially in people with OA, the prescription of analgesics is common, and is
104 associated with increased risk of other conditions such as cardiovascular, gastrointestinal and
105 chronic renal diseases (12,13). Also, having multiple chronic conditions increases the chances
106 of polypharmacy which further escalates the risk of other conditions.

107 OA is one of the leading conditions reported in multimorbidity research. Exploring the
108 association of OA with other diseases would help in further explaining the burden and pattern
109 of the comorbidity (14). However, the major issue in OA comorbidity research is the low
110 number and specific types of conditions studied (15). Therefore, it is important to develop a
111 consensus on both the count and typology of conditions to be studied to enable comparisons
112 across populations and to derive pooled estimates as appropriate. Further, using uniform

1
2
3 113 methods and definitions of diseases in computing these estimates would reduce heterogeneity
4 114 and make the comparison more reliable.

5
6 115 Understanding the temporal association with comorbidity and disease trajectory is crucial for
7 116 any chronic condition, and this is possible through studies using longitudinal databases (16).

8
9 117 However, one of the limitations of using observational data is that causal associations are
10 118 difficult to establish, due to the interference of known and unknown confounders. In this
11 119 study we have used the more recently developed method ‘Mendelian Randomization’, that
12 120 can determine causal estimates through combining the use of genetic data and instrumental
13 121 variable methods.”

14
15 122 The burden of diseases in primary care often depends on the population structure, health
16 123 infrastructure and individual factors such as income and education. Such factors vary
17 124 between countries, because of the heterogeneity mentioned above there are no robust data
18 125 comparing OA and its comorbidities between countries. Therefore, the aims of this study
19 126 were to explore the burden, pattern, and causal factors of comorbidities in people with OA
20 127 across four European countries using national registration databases in the UK, the
21 128 Netherlands, Sweden, and Spain.

22
23
24
25
26
27
28
29
30
31 129

32 130 **Objectives**

- 33
34 131 1. To estimate the prevalence, incidence, and time sequence of comorbidities in OA
35 132 2. To examine the clusters of comorbidities and trajectories of clusters in OA and
36 133 associations with death
37 134 3. To investigate the associations between commonly used OA drugs, such as non-
38 135 steroidal anti-inflammatory (NSAIDs) and opioids, and risk of comorbidities
39 136 4. To identify the potential biomarkers and causal pathways between OA and the
40 137 comorbidities
41 138 5. To examine the variations of OA comorbidities and clusters across countries.

42
43
44
45
46
47
48
49
50 139

51 140 **Methods**

52 141 *Databases*

53 142 Four routinely collected national (the UK and the Netherlands) or regional (Sweden and
54 143 Spain) health databases will be used for objectives 1-3. In addition, for objective 4, genomic
55 144 associations of OA with comorbidities will be examined using two cohort studies from the
56 145 UK (UK Biobank) (17) and the Netherlands (Rotterdam study) (18). The four national
57 146 representative and regional databases contain information about the population with primary

1
2
3 147 care consultations in four different countries. The longitudinal databases provide information
4
5 148 about the diagnosis of the diseases by the general practitioners and some diagnoses made in
6
7 149 secondary care, prescription of drugs, deaths, and other health utilisation indicators. The
8
9 150 details of the databases and their properties are given in Table 1 and 2.

10 151 *Participants*

11 152 People registered with the respective databases aged 18 year or above are eligible for the
12
13 153 study.

14 154
15 155 *Patient public involvement (PPI):* Three PPI representatives (with OA and all with multiple
16
17 156 chronic conditions) were involved in this study through group meetings. Difficulties of living
18
19 157 with multiple conditions, lack of research in causal relationship and identification of diseases
20
21 158 to be studied and the role of drugs in comorbidity were discussed. They are constantly in
22
23 159 touch through providing their inputs at each step of the study.

24 160 *Definition of Osteoarthritis*

25
26 161 OA will be defined as having at least one recorded physician diagnosis of OA for hip, knee,
27
28 162 ankle/foot, wrist/hand, or site recorded as ‘unspecified’ during the study period for the
29
30 163 respective database. People with any previous recording of the OA prior to the start date of
31
32 164 the study will be excluded.

33 165 *Comorbidities*

34 166 We defined comorbidity as the recording of diagnosis of predefined chronic conditions in
35
36 167 individuals using either ICD-10 or Read or international classification of primary care (ICPC)
37
38 168 code. An extensive list of 61 chronic conditions was prepared from the Quality Outcome
39
40 169 Framework (QOF) (19), list of the US Department of Health and Human Services Initiative
41
42 170 on Multiple Chronic Conditions (20), global burden of diseases (21) and the Charlson
43
44 171 comorbidity index (22). The list has been updated with findings from our systematic review
45
46 172 (6) and a previous UK community-based knee pain study (6,23) by including common and
47
48 173 important morbidities not included in the above (24,25). A code mapping exercise was
49
50 174 conducted to finalise the list of conditions available for all the research centres. The
51
52 175 comparison of codes was made, and it was reviewed by four researchers including a clinician
53
54 176 from the team. The detailed list of the conditions is given in Table 3.

54 177 *Covariates*

55 178 Age and sex will be used in all centres as covariates to adjust in regression models.
56
57 179 Additionally, information on body mass index (BMI), smoking, alcohol use, socioeconomic
58
59 180 variables such as education level, income, place of birth (to identify those who immigrated to
60

1
2
3 181 the country), and residential area, marriage (or registered partner) will be included when
4
5 182 available. For calculating severity of the comorbidities in an individual, Elixhauser
6
7 183 comorbidity index will be used to estimate the impact of comorbidities on death (26,27).
8
9 184 Missing data on covariates will be substituted using multiple imputation methods, provided
10
11 185 that the data is missing at random, if applicable.

12 186 *Data Harmonisation:*

13
14
15 187 Firstly, we carried out a code mapping exercise for identification of people with osteoarthritis
16
17 188 (OA) and other comorbidities. We developed a list of chronic conditions and each centres
18
19 189 shared the list of codes to be used for the conditions, such as Read code in CPRD (UK),
20
21 190 ICPC2 (Rotterdam) and ICD-10 for Lund and Spanish as per their database. The code lists
22
23 191 were compared and edited to maintain the uniformity. The list was screened by verified by
24
25 192 two researchers and two GPs. We also used uniform definition for inclusion of condition e.g.
26
27 193 at least one recording of the chronic conditions. Because all the centres did not have all the
28
29 194 listed comorbidities, a minimum number of chronic conditions and covariates common in all
30
31 195 the database were identified to be studied. Similarly, we decided to have a minimum follow-
32
33 196 up study of 5 years and centres with more registration period can use the entire length of data
34
35 197 available.

36 198 *Study design and data analysis*

37 199 Summary of the study design and analysis is provided in Figure 1. All the centres plan to use
38
39 200 same statistical analysis plan to investigate each objective.

40 201 **Ethics and Dissemination:**

41
42 202 The study has obtained the following ethics approvals: UK- Independent Scientific Advisory
43
44 203 Council (ISAC) 19/30R, The Netherlands- The Integrated Primary Care Information (IPCI)
45
46 204 registration no. 11/2019, Spain - the Information System for the Development of Research in
47
48 205 Primary Care (SIDIAP), 4R19/011 , Sweden –Ethical Review Authority, Skåne Healthcare
49
50 206 Register, ‘Dnr 2011-432, Dnr 2014-276, and Dnr 2018-233. The registry databases are made
51
52 207 available anonymised for the research purposes.

53 208 *Objective 1 – Prevalence, incidence, and time sequence of comorbidities in OA*

54 209
55 210
56 211 A combined retrospective and prospective study of OA cases and sex, age (+/-2 years), first
57
58 212 year of registration, and practice matched controls (1:1-4) without OA (28) will be used to
59
60 213 determine the prevalence, incidence, and time sequence of comorbidities in OA. Incident OA

1
2
3 214 cases will be identified, and the first diagnosis date will be used as the starting point (index
4 215 date). For controls the same index date as their matched case will be used. They will be both
5 216 retrospectively reviewed for prior diagnoses of comorbidities and prospectively followed-up
6 217 for posterior new comorbidities. In the retrospective analysis the prevalence and 95%
7 218 confidence interval (CI) of each specific comorbidity will be calculated separately in OA
8 219 cases and matched controls using the number of people diagnosed with the comorbidity
9 220 divided by the total number of OA cases or controls at the index date. The prevalence of each
10 221 comorbidity in OA cases and matched controls will be calculated for given time intervals
11 222 prior to the index date of 0-1, 0-5 and 0-10 years separately to assess observational bias (28).
12 223 Discrete time intervals of 1-5, and 5-10 years before will also be used to estimate the
13 224 prevalence to minimise consultation bias/misclassification bias of OA (if possible). Logistic
14 225 regression will be used to calculate the odds ratios (OR) for each comorbidity unadjusted and
15 226 adjusted for BMI, smoking and alcohol consumption.
16 227 For the prospective analysis participants with incident OA but without the specific
17 228 comorbidity of interest at the index date (i.e., people at risk) and matched controls without
18 229 OA will be followed up until the date of the first diagnosis of the comorbidity, deregistration,
19 230 or death whichever comes first. The cumulative incidence will be calculated for each
20 231 comorbidity in OA cases and matched controls at 1, 3, 5, 10, 15, 20 years (based on the data
21 232 available) after the index date to examine the dynamic change of developing comorbidities
22 233 during follow-up. Kaplan-Meier survival curves will be used to display the cumulative
23 234 probability in OA and non-OA groups. Proportional hazard assumption will be tested using
24 235 Schoenfeld residual plots. The Cox regression model will be used to calculate hazard ratios
25 236 (HR) for each comorbidity unadjusted and adjusted for age, sex, practice, BMI, smoking and
26 237 alcohol consumption. This hybrid design has been previously used by us to examine the
27 238 temporality of associations between other rheumatic musculoskeletal diseases (RMDs) (e.g.,
28 239 gout and lupus) and comorbidities (28,29).

240 241 *Objective 2 – Clusters and impact of comorbidities in people with OA*

242 For each dataset, an 80%: 20% split into the training and testing data will be introduced. The
243 following analysis in objective 2 will be first employed into the training dataset and then
244 tested its generalisability in the testing dataset. At baseline, clusters of people based on 61
245 comorbidities will be identified using Latent class (i.e., Gaussian mixture models algorithms
246 of cluster) analysis (30). For each model, we will examine the association between clusters
247 and covariates using multinomial logistic regressions. The distinctness of clusters will be

1
2
3 248 examined by comparing covariates among clusters. The optimal model is the one where most
4
5 249 clusters found in the training data are also identified independently in the testing data and
6
7 250 clusters have most distinct patients' characteristics. We will then use both latent trajectory
8
9 251 analysis, such as joint latent class models (31), and unsupervised machine learning approach,
10
11 252 such as deep autoencoder or recurrent neural networks (32), to identify distinct clusters of
12
13 253 new comorbidity numbers development over time and their association with mortality with
14
15 254 adjustment for baseline covariates.

16 255 *Objective 3 - Association between analgesics and incident comorbidities*

17 256
18 257 A cohort study will be undertaken for this objective to evaluate the contribution of common
19
20 258 analgesics for OA to the development of comorbidity such as NSAIDs, opioids and
21
22 259 paracetamol. We are interested in the interaction between OA and use of drugs on the
23
24 260 incidence of comorbidities, i.e. to evaluate if the drug use in persons with OA poses increased
25
26 261 or decreased risk of comorbidities compared to persons without OA and/or analgesics.
27
28 262 Individual comorbidity, as well as clusters of comorbidities identified from Objective 2 will
29
30 263 be examined as outcomes. The 61 comorbidities in our study will be further categorised into
31
32 264 eight groups, specifically: musculoskeletal (MSK), respiratory, neurodegenerative,
33
34 265 psychological/psychiatric, cancer, cardiovascular, metabolic, renal problem, liver diseases,
35
36 266 gastrointestinal (GI) and others. (Table 4) The prospective cohort established from Objective
37
38 267 1 will form the source population for this objective. Individuals with incident OA will be
39
40 268 identified from the database and the first diagnosis date will be used as index date for follow-
41
42 269 up. Individuals without OA during the study period will be selected and matched with cases
43
44 270 by age, sex, and practice. The same index date will be given from their matched OA cases.
45
46 271 Individuals with analgesics prescriptions prior to the index date will be excluded (or recorded
47
48 272 as a confounding factor to be adjusted as appropriate). Only analgesic prescriptions after the
49
50 273 index date will be considered for this analysis. Prescriptions will be quantified as number of
51
52 274 prescriptions within year 1 (initial use, primary analysis) (33), 2, 3, 4, 5 etc. It will also be
53
54 275 dichotomised as episodic (e.g., at least one gap of ≥ 90 days between prescriptions) and
55
56 276 continuous (no gap of more less than 90 days) users as appropriate (34,35). Analgesic use
57
58 277 will be included in the model as a risk factor together with OA diagnosis (yes/no, primary
59
60 278 exposure) to examine the independent risk of each variable (OA and analgesics), as well as
279
280 the interaction between the two to the development of comorbidity. Dose response
281
281 280 relationship will be examined using number of prescriptions during the exposure window
examined. The effect of stopping analgesics will also be examined by looking into the

1
2
3 282 patterns of analgesic prescriptions, e.g., stopping analgesics after initial use in year 1 versus
4 283 continuous use of analgesics afterwards. For the primary analysis (initial prescriptions within
5 284 year 1), a landmark analysis will be used to minimise the immortal time bias where the
6 285 follow up will start after 12 months from the index date (36). Participants at risk (i.e., without
7 286 a specific comorbidity of interest) at the landmark date will be followed up until the first
8 287 diagnosis of the comorbidity, deregistration, or death whichever comes first. For secondary
9 288 analyses, time varying covariate analysis will be used to examine the long-term,
10 289 episodic/continuous use of analgesics after the index date and interaction between OA and
11 290 analgesics in the development of the comorbidity. The propensity score matching or the
12 291 inverse probability weighing methods will be used to adjust for confounding by indication
13 292 during the follow-up as appropriate. Depending on the country-specific drug use patterns, we
14 293 may modify this definition to allow for short breaks in between the episodes. Cox-regression
15 294 model will be used to calculate the HR and 95% CI. We will use flexible parametric models
16 295 using restricted cubic splines (developed by Lambert, “stpm2” in Stata) to estimate the HRs
17 296 and differences in time to diagnosis of comorbidities (outcome) with drugs as time-varying to
18 297 account for non-proportional hazards (37).

19 298 *Objective 4 – Potential causal pathways between OA and the comorbidities*

20 299 We will perform a Mendelian Randomisation (MR) phenome-wide association (MR-
21 300 PheWAS) study(38) to examine the causal relationship between OA, its phenotypes,
22 301 biomarkers or risk factors and comorbidities using the UK Biobank and the Rotterdam Study
23 302 database.

24 303 We will use the Rotterdam Study and the UK Biobank jointly for this objective. This is
25 304 because that the Rotterdam Study is an OA cohort with deep phenotypes and biomarkers of
26 305 OA, whereas the UK Biobank is a primary cohort for cancer and multiple disease outcomes,
27 306 and both have detailed genetic variants. We will use two sample MR approach, i.e., to
28 307 establish an association between OA and genetic variants in the Rotterdam Cohort to identify
29 308 genetic instrumental variables (IV), e.g., a set of single nucleotide polymorphisms (SNPs)
30 309 associated with OA (or a deep phenotype, biomarker, or risk factor of OA). We will then
31 310 undertake the MR-PheWAS analysis to examine the causal effects of the OA IV on
32 311 comorbidities in the UK Biobank. The MR method has been widely used in real world data to
33 312 examine the causal relationship between IV and specific disease, under two assumptions: [1]
34 313 genetic variants are randomly assigned in the population; and [2] genetic variants can only be
35 314 the cause not consequence of disease (39). The PheWAS is a series of case control studies to

1
2
3 315 estimate the associations between the IV and multiple disease outcomes (38,40). The
4
5 316 combination of the two permits investigation of the causal effects of OA on multiple disease
6
7 317 outcomes.

8
9 318 The MR-PheWAS analysis includes three steps. Firstly, we will identify the genetic variants
10
11 319 that are associated with OA - IV. Secondly, we will undertake the PheWAS analysis – a
12
13 320 series of case control analyses to estimate the associations between the IV and other disease
14
15 321 outcomes (38,40) , with an adjustment for multiple testing using the false discovery rate
16
17 322 (FDR) methods (41). Thirdly, we will implement conventional MR analysis to investigate the
18
19 323 causal effects of the OA IV on comorbidities (39). An inverse variance weighted (IVW)
20
21 324 method will be used to pool the associations (ORs) as appropriate (42). The MR-Egger
22
23 325 regression analysis will be used to count for the pleiotropic effect - the effects of one genetic
24
25 326 variant on multiple outcomes (43). The heterogeneity in dependent instruments (HEIDI) test
26
27 327 will be used to exclude the cross-phenotype associations caused by genetic linkage (44).

28
29 328 With the MR-PheWAS study, the OR can be interpreted as causal association. We are
30
31 329 primarily interested in the causality from OA to comorbidities. We are also interested in
32
33 330 inflammatory (e.g., CRP), metabolic (e.g., gut microbiome) and biomechanics (e.g., BMI)
34
35 331 biomarkers and deep phenotypes of OA such as knee, hip, and hand OA with and without
36
37 332 symptoms. This will be undertaken if it is feasible within 3 years of this funded project,
38
39 333 otherwise will be considered as our future research agenda.

34 334 *Objective 5 – Variation of OA comorbidity patterns across countries*

40 335 We will use meta-analyses (MA) to examine the variation between countries and to pool the
41
42 336 data as appropriate. Estimates from first three objectives such as prevalence, incidence, OR,
43
44 337 HR and 95% CI for each specific comorbidity across different populations will be distributed
45
46 338 in a forest plot. Heterogeneity will be examined using the I^2 statistic and the Q test (45).
47
48 339 Results will be pooled if they are homogenous based on the I^2 value using the fixed effects
49
50 340 model, otherwise the reasons for the heterogeneity will be investigated. Random effects
51
52 341 models will be used to pool the results if the reasons for the heterogeneity cannot be
53
54 342 identified and if the overall pooling is appropriate. Individual patient data (IPD) meta-
55
56 343 analysis may be used to help identify the reasons for heterogeneity (46). Common clusters
57
58 344 and trajectories as well as burdens of comorbidities will also be compared between
59
60 345 populations.

59 346 *Feasibility and sample size*

60 347

1
2
3 348 To detect minimum incidence of 1% comorbidity (required for cluster analysis) with a
4
5 349 minimum clinical important difference of hazard ratio (HR)1.2, and 90% power of the study,
6
7 350 the estimated sample size was 197561 for 1581 events. It was calculated using STATA, with
8
9 351 a correlation= 0.2, standard deviation of 0.5, proportion of withdrawal= 0.20, alpha=0.05.
10
11 352 The initial check with the registry database revealed to have minimum required sample size
12
13 353 for the study.

14 354 15 355 **Discussion**

16 356
17 357 This study will be the largest epidemiological study on comorbidities of OA in primary care.
18
19 358 One of the key advantages of this multinational study is the use of the same protocol to
20
21 359 measure the burden of comorbidities in primary care settings in four European countries to
22
23 360 ensure reproducibility and comparison. There is scant evidence on the comorbidities in
24
25 361 people with OA, and this approach should help to identify the leading and most important
26
27 362 associations before and after presenting clinical OA (the index date). Further advantages of
28
29 363 this study are the large and representative populations studied and the same/similar extensive
30
31 364 list of chronic conditions for identifying comorbidity clusters. Often comorbidities
32
33 365 accumulate with age over time and the large primary care databases in this study have the
34
35 366 advantage of having long follow-up time which will enable us to detect the incidence of
36
37 367 comorbidities. Also, longer follow up would help to identify the picture of the trajectory of
38
39 368 the diseases (47). Both the incidence and the trajectories of comorbidity clusters are
40
41 369 highlighted as key elements needed in current research in multimorbidity, so findings from
42
43 370 this study should help to fill the knowledge gaps on multimorbidity in OA.
44
45 371 The relationship between chronic conditions and polypharmacy is a complex area of research.
46
47 372 The count of the medications and more importantly the nature of prescribed drugs may be
48
49 373 responsible for developing many new comorbidities in people with OA. We aim to explore
50
51 374 the associations of the most commonly prescribed drugs in OA, such as NSAIDs, with the
52
53 375 incidence of a wide range of comorbidities, which will be the first time that conditions other
54
55 376 than established comorbidities such as psychological conditions and endocrine diseases will
56
57 377 be examined. Finally, the causality study will further explore the associations at genetic
58
59 378 levels and phenotypes, which will be novel in OA research. Using a two sample MR
60
379 approach - one for OA deep phenotypes and the other for other chronic conditions maximises
380
381 the potentials of sample size, disease phenotypes and comorbidity spectrum to better explore
the causal pathways between OA and comorbidity.

1
2
3 382 There are some limitations to this study. Firstly, there are inherent issues in the nature of
4
5 383 electronic health records with respect to possible misdiagnosis, ascertainment biases, under-
6
7 384 or over-recording, and changes in databases due to change in coding structures. Also, the
8
9 385 analysis will be restricted to fewer covariates in some databases due to missing information
10
11 386 on lifestyle factors such as physical activities and diet. Even though the databases have
12
13 387 different durations of data available, if possible we will use a common follow-up time for
14
15 388 objective five. Another important limitation is that we do not have information on quality of
16
17 389 life and other outcomes to measure functional limitations recorded in the database.

17 390 Chronic conditions, especially comorbidities recorded in general practices, depend on
18
19 391 multiple factors such as population structure, health care facilities, health policies, and the
20
21 392 nature of the national databases. A major strength of this study is that it will include medical
22
23 393 records on approximately 27 million people in four European countries. Also, the study will
24
25 394 cover the sequence of research questions in comorbidity or multimorbidity starting from the
26
27 395 burden through to the causality and variation. Such a research model can be used for other
28
29 396 similar multimorbidity studies. The expected results should inform health professionals in
30
31 397 primary care settings with respect to management of people with OA and associated
32
33 398 comorbidities.

32 399 *Status of the study*

34 400 All the centres have obtained the necessary approvals for using the database in 2020. A
35
36 401 consensus has been made on the code mapping exercise. The statistical analysis will be
37
38 402 explained in detail in each of the publications. The team is expected to produce results by
39
40 403 mid-2021.

41 404 **Acknowledgements:** We thank the Patient Research Participants (PRP) members Jenny
42
43 405 Cockshull, Stevie Vanhegan, and Irene Pitsillidou for their involvement since the beginning
44
45 406 of the project. We would like to thank the FOREUM for financially supporting the research.
46
47 407 The authors would like to acknowledge Keele University's Prognosis and Consultation
48
49 408 Epidemiology Research Group who have given us permission to utilise the Code Lists
50
51 409 (©2014).

50 410 **Contributor and guarantor information:**

51 411 WZ, MD, CC, SMA, ME, DA conceived and designed the study. SS, AK, AD, AT, DR, and
52
53 412 VS developed the methods and will perform the analysis, and interpretation of the results.
54
55 413 CC, WZ, JR, AS, CFK, VS, and AT will supervise the statistical analysis. CM and MD will
56
57 414 guide with clinical interpretations of the results. SS drafted this manuscript and all authors
58
59 415 contributed to the critical revision of the manuscript for important intellectual content. The
60
416 corresponding author attests that all listed authors meet authorship criteria and that no others
417 meeting the criteria have been omitted.

1
2
3 418 **Funding:** This work was supported by Foundation for Research in Rheumatology
4 419 (FOREUM) grant (2019-2022), The Swedish Research Council (2020-01103), Governmental
5 420 funding of clinical research within the national health services (ALF), and The Swedish
6 421 Rheumatism Association. CM is funded by the National Institute for Health Research (NIHR)
7 422 Applied Research Collaboration West Midlands, the National Institute for Health Research
8 423 (NIHR) School for Primary Care Research and a National Institute for Health Research
9 424 (NIHR) Research Professorship in General Practice (NIHR-RP-2014-04-026)

12 425 **Role of the funding sources:** The sponsors did not participate in the design and conduct of
13 426 the study; collection, management, analysis, and interpretation of the data; or preparation,
14 427 review, or approval of the manuscript and the decision to submit the manuscript for
15 428 publication.

18 429 **Competing interests:** WZ declares serving as an advisory board for Ely Lilly (Ixekizumad,
19 430 2020) and Regeneron (Fasinomab, 2020). ME declares serving as an advisory Panel Board
20 431 Member for Pfizer (Nov 2019, Tanezumab). CM provided advice to BMS on recruiting to a
21 432 non-pharmacological atrial fibrillation trial. The other authors report no financial competing
22 433 interests.

25 434 **Copy right:** The Corresponding Author has the right to grant on behalf of all authors and
26 435 does grant on behalf of all authors, an exclusive licence (or non-exclusive for government
27 436 employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if
28 437 accepted) to be published in BMJ editions and any other BMJ PGL products and sublicences
29 438 such use and exploit all subsidiary rights, as set out in our licence.

32 439 **Studies involving humans or animals:** No direct participant recruitment was done for the
33 440 study.

36 441 **Data sharing statement:** We used anonymised data on individual patients on which the
37 442 analysis, results, and conclusions reported in the paper are based. The used data is not
38 443 distributable under licence. However, the relevant data can be obtained directly from the
39 444 respective agencies. The codes developed for the analysis can be available upon a valid
40 445 request.

43 446 **Figure legend**

45 447 Figure 1. Overview of the study design and statistical analysis plan for the Comorbidities in
46 448 Osteoarthritis (CoMO) study.

Reference

1. Turkiewicz A, Petersson IF, Björk J, Hawker G, Dahlberg LE, Lohmander LS, et al. Current and future impact of osteoarthritis on health care: a population-based study with projections to year 2032. *Osteoarthritis Cartilage*. 2014 Nov;22(11):1826–32.
2. Badley EM. The effect of osteoarthritis on disability and health care use in Canada. *J Rheumatol Suppl*. 1995 Feb;43:19–22.
3. McDonough CM, Jette AM. The Contribution of Osteoarthritis to Functional Limitations and Disability. *Clin Geriatr Med*. 2010 Aug;26(3):387–99.
4. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. 2020 Dec;396(10267):2006–17.
5. World Health Organization. Musculoskeletal conditions [Internet]. [cited 2021 Apr 6]. Available from: <https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions>
6. Swain S, Sarmanova A, Coupland C, Doherty M, Zhang W. Comorbidities in Osteoarthritis: A systematic review and meta-analysis of observational studies. 2019 Jun 17 [cited 2019 Jun 30]; Available from: <http://doi.wiley.com/10.1002/acr.24008>
7. Suri P, Morgenroth DC, Hunter DJ. Epidemiology of osteoarthritis and associated comorbidities. *PM R*. 2012 May;4(5 Suppl):S10-19.
8. Hall AJ, Stubbs B, Mamas MA, Myint PK, Smith TO. Association between osteoarthritis and cardiovascular disease: Systematic review and meta-analysis. *Eur J Prev Cardiol*. 2016;23(9):938–46.
9. Stubbs B, Aluko Y, Myint PK, Smith TO. Prevalence of depressive symptoms and anxiety in osteoarthritis: a systematic review and meta-analysis. *Age Ageing*. 2016 Mar 1;45(2):228–35.
10. Ricci-Cabello I, Stevens S, Kontopantelis E, Dalton ARH, Griffiths RI, Campbell JL, et al. Impact of the Prevalence of Concordant and Discordant Conditions on the Quality of Diabetes Care in Family Practices in England. *Ann Fam Med*. 2015 Nov 1;13(6):514–22.
11. Calderón-Larrañaga A, Poblador-Plou B, González-Rubio F, Gimeno-Feliu LA, Abad-Díez JM, Prados-Torres A. Multimorbidity, polypharmacy, referrals, and adverse drug events: Are we doing things well? *Br J Gen Pract*. 2012;62(605):e821–6.
12. Jüni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *The Lancet*. 2004 Dec;364(9450):2021–9.
13. Castellsague J, Riera-Guardia N, Calingaert B, Varas-Lorenzo C, Fourrier-Reglat A, Nicotra F, et al. Individual NSAIDs and Upper Gastrointestinal Complications: A Systematic Review and Meta-Analysis of Observational Studies (the SOS Project). *Drug Saf*. 2012 Dec;35(12):1127–46.
14. Lowe D.B., Taylor M.J., Hill S.J. Cross-sectional examination of musculoskeletal conditions and multimorbidity: influence of different thresholds and definitions on prevalence and association estimates. *BMC Res Notes*. 2017;10(1):51.
15. John R, Kerby DS, Hagan Hennessy C. Patterns and Impact of Comorbidity and Multimorbidity Among Community-Resident American Indian Elders. *The Gerontologist*. 2003 Oct 1;43(5):649–60.

16. Alaeddini A., Jaramillo C.A., Faruqi S.H.A., Pugh M.J. Mining Major Transitions of Chronic Conditions in Patients with Multiple Chronic Conditions. *Methods Inf Med.* 2017;56(5):391–400.
17. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med.* 2015 Mar 31;12(3):e1001779.
18. Engelhart MJ, Geerlings MI, Meijer J, Kiliaan A, Ruitenberg A, van Swieten JC, et al. Inflammatory Proteins in Plasma and the Risk of Dementia: The Rotterdam Study. *Arch Neurol.* 2004 May 1;61(5):668.
19. Quality and Outcome Framework (QOF) [Internet]. [cited 2018 Mar 21]. Available from: <https://digital.nhs.uk/article/8910/Quality-and-Outcome-Framework-QOF-Indicators-No-Longer-In-QOF-INLIQ-Enhanced-Services-ES-Vaccinations-and-Immunisations-V-I-and-GMS-Core-Contract-CC-extraction-specifications-business-rules->
20. Medicare C for, Baltimore MS 7500 SB, Usa M. CC_Main [Internet]. 2017 [cited 2018 Oct 24]. Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/CC_Main.html
21. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet.* 2018 Nov;392(10159):1789–858.
22. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis.* 1987 Jan 1;40(5):373–83.
23. Sarmanova A, Fernandes GS, Richardson H, Valdes AM, Walsh DA, Zhang W, et al. Contribution of central and peripheral risk factors to prevalence, incidence and progression of knee pain: a community-based cohort study. *Osteoarthritis Cartilage.* 2018 Nov;26(11):1461–73.
24. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol.* 1994 Nov 1;47(11):1245–51.
25. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. *Am J Epidemiol.* 2011 Mar 15;173(6):676–82.
26. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A Modification of the Elixhauser Comorbidity Measures Into a Point System for Hospital Death Using Administrative Data: *Med Care.* 2009 Jun;47(6):626–33.
27. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care.* 1998 Jan;36(1):8–27.
28. Kuo C-F, Grainge MJ, Mallen C, Zhang W, Doherty M. Comorbidities in patients with gout prior to and following diagnosis: case-control study. *Ann Rheum Dis.* 2014 Nov 14;annrheumdis-2014-206410.
29. Swain S, Coupland C, Mallen C, Kuo CF, Sarmanova A, Bierma-Zeinstra SMA, et al. Temporal relationship between osteoarthritis and comorbidities: a combined case control and cohort study in the UK primary care setting. *Rheumatology.* 2021 Mar 1;keab067.

- 1
- 2
- 3 30. Collins LM, Lanza ST. Latent Class and Latent Transition Analysis [Internet]. Hoboken, NJ,
4 USA: John Wiley & Sons, Inc.; 2009 [cited 2020 Apr 5]. (Wiley Series in Probability and
5 Statistics). Available from: <http://doi.wiley.com/10.1002/9780470567333>
6
- 7 31. Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal
8 and time-to-event data: A review. *Stat Methods Med Res.* 2014 Feb;23(1):74–90.
9
- 10 32. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked Denoising Autoencoders:
11 Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J Mach*
12 *Learn Res.* 2010 Dec;11:3371–408.
13
- 14 33. Zeng C, Zhang W, Doherty M, Persson MSM, Mallen C, Swain S, et al. Initial analgesic
15 prescriptions for osteoarthritis in the United Kingdom, 2000–2016. *Rheumatology.* 2021 Jan
16 5;60(1):147–59.
17
- 18 34. Hippisley-Cox J, Coupland C. Unintended effects of statins in men and women in England and
19 Wales: population based cohort study using the QResearch database. *BMJ.* 2010 May
20 20;340(may19 4):c2197–c2197.
21
- 22 35. Sarmanova A, Doherty M, Kuo C, Wei J, Abhishek A, Mallen C, et al. Statin use and risk of joint
23 replacement due to osteoarthritis and rheumatoid arthritis: a propensity-score matched
24 longitudinal cohort study. *Rheumatology.* 2020 Oct 1;59(10):2898–907.
25
- 26 36. Kuo C-F, Grainge MJ, Mallen C, Zhang W, Doherty M. Effect of allopurinol on all-cause
27 mortality in adults with incident gout: propensity score–matched landmark analysis.
28 *Rheumatology.* 2015 Jul 12;kev246.
29
- 30 37. Royston P, Lambert PC. Flexible parametric survival analysis using Stata: beyond the Cox model.
31 College Station, TX: Stata Press; 2011. 347 p.
32
- 33 38. Millard LAC, Davies NM, Timpson NJ, Tilling K, Flach PA, Smith GD. MR-PheWAS:
34 hypothesis prioritization among potential causal effects of body mass index on many outcomes,
35 using Mendelian randomization. *Sci Rep [Internet].* 2015 Dec [cited 2020 Jan 26];5(1). Available
36 from: <http://www.nature.com/articles/srep16645>
37
- 38 39. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide,
39 glossary, and checklist for clinicians. *BMJ.* 2018 Jul 12;k601.
40
- 41 40. Li X, Meng X, Spiliopoulou A, Timofeeva M, Wei W-Q, Gifford A, et al. MR-PheWAS:
42 exploring the causal effect of SUA level on multiple disease outcomes by using genetic
43 instruments in UK Biobank. *Ann Rheum Dis.* 2018 Jul;77(7):1039–47.
44
- 45 41. Benjamini Y, Yekutieli D. The Control of the False Discovery Rate in Multiple Testing under
46 Dependency. *Ann Stat.* 2001;29(4):1165–88.
47
- 48 42. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple
49 Genetic Variants Using Summarized Data: Mendelian Randomization Using Summarized Data.
50 *Genet Epidemiol.* 2013 Nov;37(7):658–65.
51
- 52 43. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian
53 Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet*
54 *Epidemiol.* 2016 May;40(4):304–14.
55
- 56
- 57
- 58
- 59
- 60

- 1
2
3 44. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data
4 from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016
5 May;48(5):481–7.
6
7 45. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002 Jun
8 15;21(11):1539–58.
9
10 46. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic
11 reviews using individual patient data. *Eval Health Prof.* 2002 Mar;25(1):76–97.
12
13 47. Vetrano DL, Roso-Llorach A, Fernández S, Guisado-Clavero M, Violán C, Onder G, et al.
14 Twelve-year clinical trajectories of multimorbidity in a population of older adults. *Nat Commun*
15 [Internet]. 2020 Dec [cited 2020 Aug 12];11(1). Available from:
16 <http://www.nature.com/articles/s41467-020-16780-x>
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Characteristics of the included databases

	Netherlands	Spain	Sweden	UK
Objectives 1-3				
Name of the database	Integrated Primary Care Information (IPCI)	The Spanish Sistema information del Deveolpment de l'Investigació a Atenció Primària (SIDIAP)	Skåne Healthcare Register	Clinical Practice Research Datalink (CPRD)
Settings	Routinely collected primary care database	Routinely collected primary care data	Swedish healthcare in Skane region, primary, specialist and in-patient care	Routinely collected primary care database with linkage database
Size and Coverage	2.2 million (Randomly distributed over the country)	6.5 million (> 85% of total Catalan region)	1.3 million (all residents of the Skane region)	17 million (country-wide, nearly 740 practices)
Start year	1998- (Better coding after 2000)	2006	1998	1993-
Age group	All	All	All	All
Gender	All	All	All	All
Coding system	ICPC	ICD 10	ICD 10	Read codes and ICD 10
Drug prescribed by	GP	GP	GP	GP
Death record (Either date of death and/or cause)	Both date and cause	Only date	Both date and cause (until year 2015)	Only date
Covariates/ additional variables	NA	BMI, Smoking, Alcohol, Social class, cholesterol, and other biomarkers	Education, Income, profession, and sick leave, residential area, region of birth	BMI, Smoking, Alcohol, Deprivation index, Ethnic group

BMI- Body mass index; GP- general practitioners; ICD- International classification of diseases; ICPC- International classification of primary care; NA- Not available

Table 2. Database for the mendelian randomisation

Name of the database	Rotterdam Cohort Study	UK Biobank
Population coverage	15000	500,000
Age group	>=40 years	40-69 years
Start year -till now	1989-onwards	2010-
Types of data	Radiographic data, joint pain, joint stiffness, of hip, knee, and hand, GWAS, biochemical markers	Genetic and phenotypes

GWAS- Genome Wide Association Studies

For peer review only

Table 3. List of chronic conditions across four databases

SI no	Conditions	ICD 10	ICPC	Read Code
1	Anaemia (All types)	D50-D64	B78, B80, B81, B82	D00..., D01...
2	Ankylosing spondylitis	M45.9	NA	N10...
3	Anxiety disorder	F41.0-F41.9	P74, P74.01, P74.02	E200..., Eu41...
4	Asthma	J450-J45.9, J46.9	R96...	66Y..., H33...
5	Benign prostatic hypertrophy (BPH)	N40.9	Y85	K20..., K21..., K22...
6	Cardiac arrhythmias (Atrial Fibrillation)	I47.0-I49.9	K78	Gyu..., G573...
7	Cataract	H25.0-H25.9, H26.1-H26.9	F92	F46...,
8	Chronic Back pain	M47-M48, M51-M54, M99, G54.4	L02, L03, L86	N12..., N14...,
9	Chronic kidney disease (any cause)	N02.0-N8.8, N11.0-N11.9, N12.9, N15.0-N18.9, N19.9	U99.01	I21..., K01..., K02...
10	Chronic neck pain	M54.2	L83...	Nyu..., N11..., N12..., N14...
11	Chronic sinusitis	J32 ...	R75...	H13...
12	Chronic obstructive pulmonary diseases	J41.0-J41.8, J42.9, J43.0-J43.9, J44.0-J44.9	R91, R95	H3...
13	Coronary Heart Disease (Including Acute Myocardial infarction, Valvular disease, Angina),	I20.0-I25.0, I34.0-I37.0	K74-K76	G11..., G30..., G31... - G38..
14	Dementia	F00.0-F00.9, F01.0-F03.9, G30.0-G30.9, G31.0-G31.9	P70-P70.02	E00..., Eu0..., F11...
15	Depression	F32.0-F33.9	P76	Eu..., E11...
16	Diabetes mellitus	E10.0-E14.9	T90..., F83.01	C10..., F32...,
17	Dyslipidaemia (Hyper)	E78.1	T93	C32...
18	Eating disorders (Both)			Eu5..., R03...
19	Eczema/ Skin disease	L20.0-L22.9, L26.9	S74, S87, S88	M11...
20	Epilepsy	G40.0-G41.9	N88	F25...
21	Fatigue	F48.0	A04.11	F286...
22	Fibromyalgia	M79.7	L18.01	N248..., N239...
23	Gall bladder stone	K80.0-K80.8	D98-D98.03	781..., J65..., 4G2...,
24	GERD (Gastritis, Oesophageal bleeding, duodenitis, peptic ulcer)	K21 ...	D840...	J12..., J13..., J15...
25	Gastrointestinal bleeding	K25.0-K28.9	D84-D87	J11...,
26	Gout	M10.0-M10.9	T92	C34..., N023...
27	Hearing impairment (All types)	H90.0-H91.9	H83-H86	F59..., ZE87...
28	Heart Failure	I50.0-I50.9	K77-K77.02	G58...,
29	Hepatitis	K73.0-K73.9	D72-D72.05	J61..., J63...
30	HIV/AIDS	B20-B24	B90	A788..., A789..., AyuC...
31	Hypertension	I10.9, I11.0-I13.9, I15.0-I15.9	K86-K87, F83.02	G20..., G24..., G25..., G26...
32	Hyperthyroidism	E05.0-E05.9	T85	C02...
33	Hypothyroidism	E02.9, E03.0-E03.9	T86	C03..., C04...
34	Inflammatory Bowel Disease (IBD)	K50.0-K52.9	D94-D94.02	J4...,
35	Irritable Bowel Symptoms (IBS)	K58.1-K58.8		J52...
36	Leukaemia, Lymphoma	C81.0-C86.6, C91.0-C96.9	B72-B73	B60..., B61..., B64...
37	Liver Cirrhosis	K70.0-K71.9, K74.0-K74.6	D97...	J615...
38	Migraine	G43.0-G43.9	N89	F26...,
39	Multiple sclerosis	G35.9	N86	F20...,
40	Osteoarthritis	M16.0-M16.9, M17.0-M17.9	L89-L91	N05...,
41	Osteoporosis	M80.0-M82.9	L95...	N33...,

42	Other blood vessel disease (Raynaud's disease, Burger's disease)	I73.0-I73.9	K92...	G73... ,
43	Parkinson's disease	G20.9	N87-N87.01	F12... ,
44	Peripheral vascular disease (Atherosclerosis)	I70.0-I70.9	K91	G70... , G71... , G72... ,
45	Polymyalgia	M35.3	L99.12	N20.11
46	Psoriasis	L40.0-L41.9	S91	M161...
47	Psoriatic arthritis	M07.0-M07.3	L99.13	M160...
48	Rheumatoid Arthritis	M05.0-M05.9	L88... , K71...	N04....
49	Renal stones	N20.0	U95	4G4... , 7B07... , KB12...
50	Schizophrenia and/or psychosis	F20.0-F20.9, F25.0- F25.9	P72	E10...
51	Severe allergy			H17... , SN5...
52	Sjögren's syndrome	M35.0	NA	N002...
53	Systemic Lupus Erythematosus	M32.0, M32.1, M32.8, M32.9	NA	N000...
54	Sleep disorder (Insomnia)	F51.0	P06	Fy0... , 1B1B...
55	Solid malignancy	C00.0-C80.9, D00.0-D09.9, C97.9	A29... , A79, B74 – Y78	B0... - B67... , Byu...
56	Stroke	G45.0-G46.8, I60.0-I63.9, I65.0-I66.9, I69.0-I69.4	K89-K90.02	G60... - G68... , F22...
57	Substance abuse/ Drug addiction	F10.0-F19.9	P18, P19...	E24... , Eu1...
58	Thrombotic diseases	I74.0-I74.9	K93, K94... , W99.03	G80... , G81... , G74...
59	Tuberculosis	A15.0-A16.9, B90.9	A70, R70	A1... , A11...
60	Vertigo	H81.4 ...	N17-N17.02, H82-H82.03	R004... , F561...
61	Vision problem (Glaucoma and other)	H27.0-H27.9, H40.0-H40.9, H42.0-H42.8	F93... , F94	F45... , F49... ,

ICD- International Classification of Diseases; ICPC- International classification in primary care. All the codes are the primary code initials used in the database.

Table 4. Group of conditions/Outcome

Group	Conditions
Cardiovascular	Cardiac arrhythmias, Coronary Heart Disease (including AMI, valvular disease, angina), Heart failure, Hypertension, Peripheral vascular disease (claudication, Raynaud syndrome, Buerger's disease), Other blood vessel disease (atherosclerosis and aneurysm), Thrombotic diseases
Gastrointestinal	GERD (Esophageal diseases, gastritis, duodenitis), GI bleeding, Inflammatory bowel disease (IBD), Irritable bowel syndrome (IBS)
Musculoskeletal	Ankylosing spondylitis, Chronic Back pain, Chronic Neck pain, Fibromyalgia, Polymyalgia, Gout, Osteoporosis, Psoriatic arthritis, Rheumatoid arthritis, Sjögren's syndrome, Systemic lupus erythematosus (SLE)
Endocrine	Diabetes mellitus, Dyslipidemia (hyper), Hyperthyroidism, Hypothyroidism
Neurological	Dementia, Epilepsy, Fatigue, Migraine, Multiple sclerosis, Parkinson disease, Stroke
Psychological	Anxiety, Depression, Eating disorders (Anorexia / Bulimia nervosa), Schizophrenia, Sleep disorder (insomnia),
Kidney disease	Chronic kidney disease (any cause), Renal stones
Liver diseases	Gall bladder stone, Hepatitis, Liver cirrhosis
Respiratory	Asthma, Chronic obstructive pulmonary disease (COPD)
Cancer	Leukemia, Lymphoma, Solid Malignancy (any type)
Others	Anemia (all types), Benign prostate hypertrophy (BPH), Cataract, Chronic sinusitis, Eczema/Skin disease, Hearing impairment (all types), Psoriasis, Severe allergy (anaphylactic shock), angioneurotic oedema

6/bmjopen-2021-00816 on 6 April 2022. Downloaded from <http://bmjopen.bmj.com/> on April 18, 2024 by guest. Protected by copyright.

Figure 1. Overview of the study design and statistical analysis plan for the Comorbidities in Osteoarthritis (CoMO) study.

	Objective 1	Objective 2	Objective 3	Objective 4
Study Design	Case-control and Cohort	Clustering and longitudinal	Cohort	Genomic association
Exposure	OA	OA	Analgesics (NSAIDs, Opioids, paracetamol)	OA
Outcome	Comorbidities	Clusters of comorbidities	Comorbidities	Comorbidities
Statistical methods	Conditional logistic regression, Cox regression	Latent class analysis, Latent class growth analysis, Joint latent class analysis	Cox regression, Time varying analysis, Flexible parametric method	Mendelian randomisation
Reported Outcome	Odds Ratio and Hazard Ratio	Clusters and groups	Hazard Ratio	Coefficients
Participating centres	ALL	ALL	ALL	UK and Netherlands

Objective 5	Meta Analysis
-------------	---------------

OA- osteoarthritis; NSAIDs- Nonsteroidal anti-inflammatory drugs

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*
Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract Comorbidities in Osteoarthritis (CoMO): a multinational study in four European countries	1-2
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found <i>Introduction</i> Osteoarthritis (OA) is one of the leading chronic conditions in the older population. People with OA are more likely to have one or more other chronic conditions than those without. However, the temporal associations, clusters of the comorbidities, role of analgesics and the causality and variation between populations are yet to be investigated. This paper describes the protocol of a multinational study in four European countries (UK, Netherlands, Sweden, and Spain) exploring comorbidities in people with OA . <i>Methods and analysis</i> This multinational study will investigate i) the temporal associations of 61 identified comorbidities with OA, ii) the clusters and trajectories of comorbidities in people with OA, iii) the role of analgesics on incidence of comorbidities in people with OA, iv) the potential biomarkers and causality between OA and the comorbidities, and v) variations between countries. A combined case-control and cohort study will be conducted to find the temporal association of OA with the comorbidities using the national or regional health databases. Latent class analysis will be performed to identify the clusters at baseline and joint latent class analysis will be used to examine trajectories during the follow-up. A cohort study will be undertaken to evaluate the role of non-steroidal anti-inflammatory drugs (NSAIDs), opioids, and paracetamol on the incidence of comorbidities. Mendelian randomisation will be performed to investigate the potential biomarkers for causality between OA and the comorbidities using the UK Biobank and the Rotterdam	1

6/bmjopen-2022-052816 on 6 April 2022. Downloaded from <http://bmjopen.bmj.com/> on April 18, 2024 by guest. Protected by copyright.

		<p>Study databases. Finally, a meta-analyses will be used to examine the variations and pool the results from different countries.</p> <p><i>Ethics and Dissemination</i></p> <p>Research ethics was obtained according to each database requirement. Results will be disseminated through the FOREUM website, scientific meetings, publications, and in partnership with patient organisations.</p>	
Introduction			
Background/rationale	2	<p>Explain the scientific background and rationale for the investigation being reported</p> <p>Osteoarthritis (OA) affects 27% of people aged over 45 years at peripheral synovial joints such as knees, hips, hands and feet (1). It is by far the most common form of arthritis, and a leading cause of chronic joint pain and disability in older people (2,3). It is anticipated that the burden of OA will continue to rise in the coming decades because of population ageing and the increasing obesity prevalence – two major risk factors for OA (4,5). Co-occurrence of multiple chronic conditions in an individual with ageing is becoming a norm and OA is not an exception to this.</p> <p>A recent systematic review has confirmed that people with OA are more likely to have other diseases, especially stroke, peptic ulcer, hypertension, and depression (6). Vast majority of these studies focused on additional presence (comorbidity) of cardiovascular and musculoskeletal conditions only (7–9). Whether these comorbidities just co-exist with OA, share common risk factors with OA, or are causes or consequences of OA remains largely unknown. There was also reporting of wide heterogeneity in definitions of OA and other chronic conditions, diagnosis and recording of diseases, sample sizes, and number of diseases studied in previously published studies included in the review (6). This diversity made the comparison and pooled estimation of comorbidity prevalence difficult.</p> <p>Comorbidity in OA can occur due to multiple factors. Various hypotheses have been used to explain the existence of comorbidities in general, the most accepted of which are the concordant (disease sharing similar</p>	4-5

pathophysiological risk factors) and discordant (diseases not sharing similar pathophysiological risk factors) theories (10). Additionally, prescription of drugs is also reported to be associated with comorbidity and multimorbidity (11). Especially in people with OA, the prescription of analgesics is common, and is associated with increased risk of other conditions such as cardiovascular, gastrointestinal and chronic renal diseases (12,13). Also, having multiple chronic conditions increases the chances of polypharmacy which further escalates the risk of other conditions.

OA is one of the leading conditions reported in multimorbidity research. Exploring the association of OA with other diseases would help in further explaining the burden and pattern of the comorbidity (14). However, the major issue in OA comorbidity research is the low number and specific types of conditions studied (16). Therefore, it is important to develop a consensus on both the count and typology of conditions to be studied to enable comparisons across populations and to derive pooled estimates as appropriate. Further, using uniform methods and definitions of diseases in computing these estimates would reduce heterogeneity and make the comparison more reliable.

Understanding the temporal association with comorbidity and disease trajectory is crucial for any chronic condition, and this is possible through studies using longitudinal databases (15). However, one of the limitations of using observational data is that causal associations are difficult to establish, due to the interference of known and unknown confounders. In this study we have used the more recently developed method ‘Mendelian Randomization’, that can determine causal estimates through combining the use of genetic data and instrumental variable methods.”

The burden of diseases in primary care often depends on the population structure, health infrastructure and individual factors such as income and education. Such factors vary between countries, because of the heterogeneity mentioned above there are no robust data comparing OA and its comorbidities between countries. Therefore, the aims of this study were to explore the burden, pattern, and causal factors of comorbidities in people with OA across four European countries using national registration databases in the UK, the Netherlands, Sweden, and

		Spain.	
Objectives	3	<p>State specific objectives, including any pre-specified hypotheses</p> <ol style="list-style-type: none"> 1. To estimate the prevalence, incidence, and time sequence of comorbidities in OA 2. To examine the clusters of comorbidities and trajectories of clusters in OA and associations with death 3. To investigate the associations between commonly used OA drugs, such as non-steroidal anti-inflammatory (NSAIDs) and opioids, and risk of comorbidities 4. To identify the potential biomarkers and causal pathways between OA and the comorbidities 5. To examine the variations of OA comorbidities and clusters across countries 	5
Methods			
Study design	4	<p>Present key elements of study design early in the paper</p> <p>Figure 1</p>	Figure 1
Setting	5	<p>Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection</p> <p><i>Databases</i></p> <p>Four routinely collected national (the UK and the Netherlands) or regional (Sweden and Spain) health databases will be used for objectives 1-3. In addition, for objective 4, genomic associations of OA with comorbidities will be examined using two cohort studies from the UK (UK Biobank) (17) and the Netherlands (Rotterdam study) (18). The four national representative and regional databases contain information about the population with primary care consultations in four different countries. The longitudinal databases provide information about the diagnosis of the diseases by the general practitioners and some diagnoses made in secondary care, prescription of drugs, deaths, and other health utilisation indicators. The details of the databases and their properties are given in Table 1 and 2.</p> <p><i>Participants</i></p> <p>People registered with the respective databases aged 18 year or above are eligible for the study.</p>	5-6

<p>Participants</p>	<p>6</p>	<p>(a) <i>Cohort study</i>—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i>—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls</p>	<p>n/a</p>
		<p>(b) <i>Cohort study</i>—For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i>—For matched studies, give matching criteria and the number of controls per case</p> <p><i>Objective 1 – Prevalence, incidence, and time sequence of comorbidities in OA</i></p> <p>A combined retrospective and prospective study of OA cases and sex, age (+/-2 years), first year of registration, and practice matched controls (1:1-4) without OA (28) will be used to determine the prevalence, incidence, and time sequence of comorbidities in OA. Incident OA cases will be identified, and the first diagnosis date will be used as the starting point (index date). For controls the same index date as their matched case will be used. They will be both retrospectively reviewed for prior diagnoses of comorbidities and prospectively followed-up for posterior new comorbidities.</p> <p>For the prospective analysis participants with incident OA but without the specific comorbidity of interest at the index date (i.e., people at risk) and matched controls without OA will be followed up until the date of the first diagnosis of the comorbidity, deregistration, or death whichever comes first. The cumulative incidence will be calculated for each comorbidity in OA cases and matched controls at 1, 3, 5, 10, 15, 20 years (based on the data available) after the index date to examine the dynamic change of developing comorbidities during follow-up.</p> <p><i>Objective 3</i></p> <p>A cohort study will be undertaken for this objective to evaluate the contribution of common analgesics for OA to the development of comorbidity such as NSAIDs, opioids and paracetamol. We are interested in the interaction between OA and use of drugs on the incidence of comorbidities, i.e. to evaluate if the drug use in persons with OA poses increased or decreased risk of comorbidities compared to persons without OA and/or analgesics. Individual</p>	<p>7-8</p> <p>9</p>

comorbidity, as well as clusters of comorbidities identified from Objective 2 will be examined as outcomes. The 61 comorbidities in our study will be further categorised into eight groups, specifically musculoskeletal (MSK), respiratory, neurodegenerative, psychological/psychiatric, cancer, cardiovascular, metabolic, renal problem, liver diseases, gastrointestinal (GI) and others. (Table 4) The prospective cohort established from Objective 1 will form the source population for this objective. Individuals with incident OA will be identified from the database and the first diagnosis date will be used as index date for follow-up. Individuals without OA during the study period will be selected and matched with cases by age, sex, and practice. The same index date will be given from their matched OA cases. Individuals with analgesics prescriptions prior to the index date will be excluded (or recorded as a confounding factor to be adjusted as appropriate). Only analgesic prescriptions after the index date will be considered for this analysis. Prescriptions will be quantified as number of prescriptions within year 1 (initial use, primary analysis) (33), 2, 3, 4, 5 etc. It will also be dichotomised as episodic (e.g., at least one gap of ≥ 90 days between prescriptions) and continuous (no gap of more less than 90 days) users as appropriate (34,35). Analgesic use will be included in the model as a risk factor together with OA diagnosis (yes/no, primary exposure) to examine the independent risk of each variable (OA and analgesics), as well as the interaction between the two to the development of comorbidity. Dose response relationship will be examined using number of prescriptions during the exposure window examined. The effect of stopping analgesics will also be examined by looking into the patterns of analgesic prescriptions, e.g., stopping analgesics after initial use in year 1 versus continuous use of analgesics afterwards.

Objective 4

We will use the Rotterdam Study and the UK Biobank jointly for this objective. This is because that the Rotterdam Study is an OA cohort with deep phenotypes and biomarkers of OA, whereas the UK Biobank is a primary cohort for cancer and multiple disease outcomes, and both have detailed genetic variants. We will use two sample MR

		<p>approach, i.e., to establish an association between OA and genetic variants in the Rotterdam Cohort to identify genetic instrumental variables (IV), e.g., a set of single nucleotide polymorphisms (SNPs) associated with OA (or a deep phenotype, biomarker, or risk factor of OA). We will then undertake the MR-PheWAS analysis to examine the causal effects of the OA IV on comorbidities in the UK Biobank. The MR method has been widely used in real world data to examine the causal relationship between IV and specific disease, under two assumptions: [1] genetic variants are randomly assigned in the population; and [2] genetic variants can only be the cause not consequence of disease (39). The PheWAS is a series of case control studies to estimate the associations between the IV and multiple disease outcomes (38,40). The combination of the two permits investigation of the causal effects of OA on multiple disease outcomes.</p>	
Variables	7	<p>Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable</p> <p><i>Definition of Osteoarthritis</i></p> <p>OA will be defined as having at least one recorded physician diagnosis of OA for hip, knee, ankle/foot, wrist/hand, or site recorded as ‘unspecified’ during the study period for the respective database. People with any previous recording of the OA prior to the start date of the study will be excluded.</p> <p><i>Comorbidities</i></p> <p>We defined comorbidity as the recording of diagnosis of predefined chronic conditions in individuals using either ICD-10 or Read or international classification of primary care (ICPC) code. An extensive list of 61 chronic conditions was prepared from the Quality Outcome Framework (QOF) (19), list of the US Department of Health and Human Services Initiative on Multiple Chronic Conditions (20), global burden of diseases (21) and the Charlson comorbidity index (22). The list has been updated with findings from our systematic review (6) and a previous UK community-based knee pain study (6,23) by including common and important morbidities not included in the above (24,25). A code mapping exercise was conducted to finalise the list of conditions available for all the research centres. The comparison of codes was made, and it was reviewed by four researchers including</p>	6-7

		two researchers and two GPs. We also used uniform definition for inclusion of condition e.g. at least one recording of the chronic conditions. Because all the centres did not have all the listed comorbidities, a minimum number of chronic conditions and covariates common in all the database were identified to be studied. Similarly, we decided to have a minimum follow-up study of 5 years and centres with more registration period can use the entire length of data available.	
Bias	9	Describe any efforts to address potential sources of bias	Explained under each objective
Study size	10	<p>Explain how the study size was arrived at <i>Feasibility and sample size</i></p> <p>To detect minimum incidence of 1% comorbidity (required for cluster analysis) with a minimum clinically important difference of hazard ratio (HR) 1.2, and 90% power of the study, the estimated sample size was 197561 for 1581 events. It was calculated using STATA, with a correlation= 0.2, standard deviation of 0.5, proportion of withdrawal= 0.20, alpha=0.05. The initial check with the registry database revealed to have minimum required sample size for the study.</p>	11-12
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Explained under each objective
Statistical methods	12	<p>(a) Describe all statistical methods, including those used to control for confounding</p> <p><i>Objective 1</i></p> <p>In the retrospective analysis the prevalence and 95% confidence interval (CI) of each specific comorbidity will be calculated separately in OA cases and matched controls using the number of people diagnosed with the comorbidity divided by the total number of OA cases or controls at the index date. The prevalence of each</p>	8

comorbidity in OA cases and matched controls will be calculated for given time intervals prior to the index date of 0-1, 0-5 and 0-10 years separately to assess observational bias (28). Discrete time intervals of 1-5, and 5-10 years before will also be used to estimate the prevalence to minimise consultation bias/misclassification bias of OA (if possible). Logistic regression will be used to calculate the odds ratios (OR) for each comorbidity unadjusted and adjusted for BMI, smoking and alcohol consumption.

Kaplan-Meier survival curves will be used to display the cumulative probability in OA and non-OA groups.

Proportional hazard assumption will be tested using Schoenfeld residual plots. The Cox regression model will be used to calculate hazard ratios (HR) for each comorbidity unadjusted and adjusted for age, sex, practice, BMI, smoking and alcohol consumption. This hybrid design has been previously used by us to examine the temporality of associations between other rheumatic musculoskeletal diseases (RMDs) (e.g., gout and lupus) and comorbidities (28,29).

Objective 2

For each dataset, an 80%: 20% split into the training and testing data will be introduced. The following analysis in objective 2 will be first employed into the training dataset and then tested its generalisability in the testing dataset. At baseline, clusters of people based on 61 comorbidities will be identified using Latent class (i.e., Gaussian mixture models algorithms of cluster) analysis (30). For each model, we will examine the association between clusters and covariates using multinomial logistic regressions. The distinctness of clusters will be examined by comparing covariates among clusters. The optimal model is the one where most clusters found in the training data are also identified independently in the testing data and clusters have most distinct patients' characteristics. We will then use both latent trajectory analysis, such as joint latent class models (31), and unsupervised machine learning approach, such as deep autoencoder or recurrent neural networks (32), to identify distinct clusters of new

8

9

	<p>comorbidity numbers development over time and their association with mortality with adjustment for baseline covariates.</p> <p><i>Objective 3</i></p> <p>For the primary analysis (initial prescriptions within year 1), a landmark analysis will be used to minimise the immortal time bias where the follow up will start after 12 months from the index date (36). Participants at risk (i.e., without a specific comorbidity of interest) at the landmark date will be followed up until the first diagnosis of the comorbidity, deregistration, or death whichever comes first. For secondary analyses, time varying covariate analysis will be used to examine the long-term, episodic/continuous use of analgesics after the index date and interaction between OA and analgesics in the development of the comorbidity. The propensity score matching or the inverse probability weighing methods will be used to adjust for confounding by indication during the follow-up as appropriate. Depending on the country-specific drug use patterns, we may modify this definition to allow for short breaks in between the episodes. Cox-regression model will be used to calculate the HR and 95% CI. We will use flexible parametric models using restricted cubic splines (developed by Lambert, “stpm2” in Stata) to estimate the HRs and differences in time to diagnosis of comorbidities (outcome) with drugs as time-varying to account for non-proportional hazards (37).</p> <p><i>Objective 4</i></p> <p>The MR-PheWAS analysis includes three steps. Firstly, we will identify the genetic variants that are associated with OA - IV. Secondly, we will undertake the PheWAS analysis – a series of case-control analyses to estimate the associations between the IV and other disease outcomes (38,40), with an adjustment for multiple testing using the false discovery rate (FDR) methods (41). Thirdly, we will implement conventional MR analysis to investigate the causal effects of the OA IV on comorbidities (39). An inverse variance weighted (IVW) method will be used to pool the associations (ORs) as appropriate (42). The MR-Egger regression analysis will be used to count for the pleiotropic effect - the effects of one genetic variant on multiple outcomes (43). The heterogeneity in dependent</p>	<p>10</p> <p>10-11</p>
--	--	------------------------

	<p>instruments (HEIDI) test will be used to exclude the cross-phenotype associations caused by genetic linkage (44).</p> <p>With the MR-PheWAS study, the OR can be interpreted as causal association. We are primarily interested in the causality from OA to comorbidities. We are also interested in inflammatory (e.g., CRP), metabolic (e.g., gut microbiome) and biomechanics (e.g., BMI) biomarkers and deep phenotypes of OA such as knee, hip, and hand OA with and without symptoms. This will be undertaken if it is feasible within 3 years of this funded project, otherwise will be considered as our future research agenda.</p> <p><i>Objective 5</i></p> <p>We will use meta-analyses (MA) to examine the variation between countries and to pool the data as appropriate. Estimates from first three objectives such as prevalence, incidence, OR, HR and 95% CI for each specific comorbidity across different populations will be distributed in a forest plot. Heterogeneity will be examined using the I^2 statistic and the Q test (45). Results will be pooled if they are homogenous based on the I^2 value using the fixed effects model, otherwise the reasons for the heterogeneity will be investigated. Random effects models will be used to pool the results if the reasons for the heterogeneity cannot be identified and if the overall pooling is appropriate. Individual patient data (IPD) meta-analysis may be used to help identify the reasons for heterogeneity (46). Common clusters and trajectories as well as burdens of comorbidities will also be compared between populations.</p>	11
	<p>(b) Describe any methods used to examine subgroups and interactions Please see method sections</p>	n/a
	<p>(c) Explain how missing data were addressed Missing data on covariates will be substituted using multiple imputation methods, provided that the data is missing at random, if applicable.</p>	7
	<p><i>(d) Cohort study</i>—If applicable, explain how loss to follow-up was addressed <i>Case-control study</i>—If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i>—If applicable, describe analytical methods taking account of sampling strategy</p>	n/a
	<p>(e) Describe any sensitivity analyses</p>	n/a

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	n/a
		(b) Give reasons for non-participation at each stage	n/a
		(c) Consider use of a flow diagram	n/a
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	n/a
		(b) Indicate number of participants with missing data for each variable of interest	n/a
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	n/a
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	n/a
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	n/a
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	n/a
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	n/a
		(b) Report category boundaries when continuous variables were categorized	n/a
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n/a
Discussion			
Key results	18	Summarise key results with reference to study objectives	n/a
Limitations	19	<p>Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias</p> <p>There are some limitations to this study. Firstly, there are inherent issues in the nature of electronic health records with respect to possible misdiagnosis, ascertainment biases, under- or over-recording, and changes in databases due to change in coding structures. Also, the analysis will be restricted to fewer covariates in some databases due to missing information on lifestyle factors such as physical activities and diet. Even though the databases have different durations of data available, if possible we will use a common follow-up time for objective five. Another important limitation is that we do not have information on quality of life and other outcomes to measure functional limitations recorded in the database.</p> <p>Chronic conditions, especially comorbidities recorded in general practices, depend on multiple factors such as population structure, health care facilities, health policies, and the nature of the national databases. A major</p>	13

		strength of this study is that it will include medical records on approximately 27 million people in four European countries. Also, the study will cover the sequence of research questions in comorbidity or multimorbidity starting from the burden through to the causality and variation. Such a research model can be used for other similar multimorbidity studies. The expected results should inform health professionals in primary care settings with respect to management of people with OA and associated comorbidities.	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	n/a
Generalisability	21	Discuss the generalisability (external validity) of the study results	n/a
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based This work was supported by Foundation for Research in Rheumatology (FOREUM) grant (2019-2022), The Swedish Research Council (2020-01103), Governmental funding of clinical research within the national health services (ALF), and The Swedish Rheumatism Association. CM is funded by the National Institute for Health Research (NIHR) Applied Research Collaboration West Midlands, the National Institute for Health Research (NIHR) School for Primary Care Research and a National Institute for Health Research (NIHR) Research Professorship in General Practice (NIHR-RP-2014-04-026)	14

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Comorbidities in Osteoarthritis (ComOA): a combined cross-sectional, case-control and cohort study using large electronic health records in four European countries

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2021-052816.R2
Article Type:	Protocol
Date Submitted by the Author:	15-Feb-2022
Complete List of Authors:	Swain, Subhashisa; University of Nottingham, Academic Rheumatology; University of Oxford, Nuffield Department of Primary Care Health Sciences Kamps, Anne; Erasmus Medical Center Department of Reumatology, Department of General Practice Runhaar, Jos; Erasmus Medical Center Department of Reumatology, Department of General Practice Dell'Isola, Andrea; Lunds University Faculty of Medicine, Turkiewicz, Aleksandra; Lund University Faculty of Medicine, Dept of Orthopedics Robinson, Danielle ; University of Oxford Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, Strauss, V; University of Oxford Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, Mallen, Christian; Keele University, Arthritis Research UK Primary Care Centre Kuo, Chang-Fu; Chang Gung Memorial Hospital Linkou Branch, Coupland, Carol; University of Nottingham, Division of Primary Care Doherty, Michael ; University of Nottingham School of Medicine, Academic Rheumatology Sarmanova, Aliya ; University of Bristol, Prieto-Alhambra, Daniel; University of Oxford, Centre for Statistics in Medicine, NDORMS Englund, Martin; Lund University, Dept of Orthopedics Bierma-Zeinstra, Sita; Erasmus University Medical Centre, Department of General Practice Zhang, Weiya; University of Nottingham School of Medicine, Academic Rheumatology
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	Epidemiology, Health services research, Public health, Pharmacology and therapeutics, Rheumatology
Keywords:	EPIDEMIOLOGY, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, RHEUMATOLOGY

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Title- Comorbidities in Osteoarthritis (ComOA): a combined cross-sectional, case-control and cohort study using large electronic health records in four European countries

Authors

Subhashisa Swain^{1,2}, Anne Kamps³, Jos Runhaar³, Andrea Dell 'Isola⁴, Aleksandra Turkiewicz⁴, Danielle Robinson⁵, Victoria Y Strauss⁵, Christian Mallen⁶, Chang Fu Kuo^{1,7}, Carol Coupland⁸, Michael Doherty^{1,11}, Aliya Sarmanova⁹, Daniel Prieto Alhambra⁵, Martin Englund⁴, Sita Bierma-Zeinstra¹⁰, Weiya Zhang^{1,11}

1. Academic Rheumatology, School of Medicine, University of Nottingham, UK
2. Nuffield Department of Primary Care Health Sciences, University of Oxford, UK
3. Department of General Practice, Erasmus MC University Medical Center Rotterdam, The Netherlands
4. Clinical Epidemiology Unit, Orthopaedics, Department of Clinical Sciences Lund, Lund University, Sweden
5. Centre for Statistics in Medicine, NDORMS, University of Oxford, UK
6. School of Medicine, Keele University, UK
7. Division of Rheumatology, Allergy and Immunology, Chang Gung Memorial Hospital, Taiwan.
8. Division of Primary Care, School of Medicine, University of Nottingham, UK
9. Musculoskeletal Research Unit, Bristol Medical School, Translational Health Sciences, University of Bristol, UK
10. Department of General Practice, Department of Orthopaedic Surgery & Sports Medicine, Erasmus MC University Medical Center Rotterdam, The Netherlands
11. Pain Centre Versus Arthritis, University of Nottingham, UK

Corresponding Author:

Professor Weiya Zhang
Academic Rheumatology, School of Medicine
University of Nottingham
Clinical Sciences Building, Nottingham city Hospital
NG5 1PB, Nottingham, UK

Email: weiya.zhang@nottingham.ac.uk

34

Abstract*Introduction*

Osteoarthritis (OA) is one of the leading chronic conditions in the older population. People with OA are more likely to have one or more other chronic conditions than those without. However, the temporal associations, clusters of the comorbidities, role of analgesics and the causality and variation between populations are yet to be investigated. This paper describes the protocol of a multinational study in four European countries (UK, Netherlands, Sweden, and Spain) exploring comorbidities in people with OA .

Methods and analysis

This multinational study will investigate i) the temporal associations of 61 identified comorbidities with OA, ii) the clusters and trajectories of comorbidities in people with OA, iii) the role of analgesics on incidence of comorbidities in people with OA, iv) the potential biomarkers and causality between OA and the comorbidities, and v) variations between countries.

A combined case-control and cohort study will be conducted to find the temporal association of OA with the comorbidities using the national or regional health databases. Latent class analysis will be performed to identify the clusters at baseline and joint latent class analysis will be used to examine trajectories during the follow-up. A cohort study will be undertaken to evaluate the role of non-steroidal anti-inflammatory drugs (NSAIDs), opioids, and paracetamol on the incidence of comorbidities. Mendelian randomisation will be performed to investigate the potential biomarkers for causality between OA and the comorbidities using the UK Biobank and the Rotterdam Study databases. Finally, a meta-analysis will be used to examine the variations and pool the results from different countries.

Ethics and Dissemination

Research ethics was obtained according to each database requirement. Results will be disseminated through the FOREUM website, scientific meetings, publications, and in partnership with patient organisations.

Key words: Osteoarthritis; Comorbidity; Multimorbidity; Primary care; Analgesics, Mendelian Randomisation, PheWAS

65

66

1
2
3 674
5 68 **Strengths and limitations of this study**6
7 69

- 8
-
- 9 70 • This is first ever multicenter study on comorbidities in osteoarthritis in Europe
-
- 10 involving nearly 27 million electronic health records
-
- 11 71
-
- 12 72 • More than 60 chronic conditions are being studied – representing a wider coverage of
-
- 13 diseases
-
- 14 73
-
- 15 74 • We will examine the causal association using Mendelian Randomization -PheWAS
-
- 16 methods with genetic data collected from two countries
-
- 17 75
-
- 18 76 • Same protocol with robust statistical methods will be used across all countries to
-
- 19 replicate the findings and examine the variations.
-
- 20 77
-
- 21 78 • Possible biases may be introduced by the nature of electronic health records and
-
- 22 length of data availability.
-
- 23 79
-
- 24 80
-
- 25 81
-
- 26
-
- 27
-
- 28
-
- 29
-
- 30
-
- 31
-
- 32
-
- 33
-
- 34
-
- 35
-
- 36
-
- 37
-
- 38
-
- 39
-
- 40
-
- 41
-
- 42
-
- 43
-
- 44
-
- 45
-
- 46
-
- 47
-
- 48
-
- 49
-
- 50
-
- 51
-
- 52
-
- 53
-
- 54
-
- 55
-
- 56
-
- 57
-
- 58
-
- 59
-
- 60

82 **Background**

83 Osteoarthritis (OA) affects 27% of people aged over 45 years at peripheral synovial joints
84 such as knees, hips, hands and feet (1). It is by far the most common form of arthritis, and a
85 leading cause of chronic joint pain and disability in older people (2,3). It is anticipated that
86 the burden of OA will continue to rise in the coming decades because of population ageing
87 and the increasing obesity prevalence – two major risk factors for OA (4,5). Co-occurrence of
88 multiple chronic conditions in an individual with ageing is becoming a norm and OA is not
89 an exception to this.

90 A recent systematic review has confirmed that people with OA are more likely to have other
91 diseases, especially stroke, peptic ulcer, hypertension, and depression (6). Vast majority of
92 these studies focused on additional presence (comorbidity) of cardiovascular and
93 musculoskeletal conditions only (7–9). Whether these comorbidities just co-exist with OA,
94 share common risk factors with OA, or are causes or consequences of OA remains largely
95 unknown. There was also reporting of wide heterogeneity in definitions of OA and other
96 chronic conditions, diagnosis and recording of diseases, sample sizes, and number of diseases
97 studied in previously published studies included in the review (6). This diversity made the
98 comparison and pooled estimation of comorbidity prevalence difficult.

99 Comorbidity in OA can occur due to multiple factors. Various hypotheses have been used to
100 explain the existence of comorbidities in general, the most accepted of which are the
101 concordant (diseases sharing similar pathophysiological risk factors) and discordant
102 (diseases not sharing similar pathophysiological risk factors) theories (10). Additionally,
103 prescription of drugs is also reported to be associated with comorbidity and multimorbidity
104 (11). Especially in people with OA, the prescription of analgesics is common, and is
105 associated with increased risk of other conditions such as cardiovascular, gastrointestinal and
106 chronic renal diseases (12,13). Also, having multiple chronic conditions increases the chances
107 of polypharmacy which further escalates the risk of other conditions.

108 OA is one of the leading conditions reported in multimorbidity research. Exploring the
109 association of OA with other diseases would help in further explaining the burden and pattern
110 of the comorbidity (14). However, the major issue in OA comorbidity research is the low
111 number and specific types of conditions studied (15). Therefore, it is important to develop a
112 consensus on both the count and typology of conditions to be studied to enable comparisons
113 across populations and to derive pooled estimates as appropriate. Further, using uniform

1
2
3 114 methods and definitions of diseases in computing these estimates would reduce heterogeneity
4
5 115 and make the comparison more reliable.

6
7 116 Understanding the temporal association with comorbidity and disease trajectory is crucial for
8
9 117 any chronic condition, and this is possible through studies using longitudinal databases (16).

10
11 118 However, one of the limitations of using observational data is that causal associations are
12
13 119 difficult to establish, due to the interference of known and unknown confounders. In this
14
15 120 study we have used the more recently developed method ‘Mendelian Randomization’, that
16
17 121 can determine causal estimates through combining the use of genetic data and instrumental
18
19 122 variable methods.”

20
21 123 The burden of diseases in primary care often depends on the population structure, health
22
23 124 infrastructure and individual factors such as income and education. Such factors vary
24
25 125 between countries, because of the heterogeneity mentioned above there are no robust data
26
27 126 comparing OA and its comorbidities between countries. Therefore, the aims of this study
28
29 127 were to explore the burden, pattern, and causal factors of comorbidities in people with OA
30
31 128 across four European countries using national registration databases in the UK, the
32
33 129 Netherlands, Sweden, and Spain.

34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

131 **Objectives**

- 132 1. To estimate the prevalence, incidence, and time sequence of comorbidities in OA
- 133 2. To examine the clusters of comorbidities and trajectories of clusters in OA and
134 associations with death
- 135 3. To investigate the associations between commonly used OA drugs, such as non-
136 steroidal anti-inflammatory (NSAIDs) and opioids, and risk of comorbidities
- 137 4. To identify the potential biomarkers and causal pathways between OA and the
138 comorbidities
- 139 5. To examine the variations of OA comorbidities and clusters across countries.

140

141 **Methods**

142 *Databases*

143 Four routinely collected national (the UK and the Netherlands) or regional (Sweden and
144 Spain) health databases will be used for objectives 1-3. In addition, for objective 4, genomic
145 associations of OA with comorbidities will be examined using two cohort studies from the
146 UK (UK Biobank) (17) and the Netherlands (Rotterdam study) (18). The four national
147 representative and regional databases contain information about the population with primary

1
2
3 148 care consultations in four different countries. The longitudinal databases provide information
4
5 149 about the diagnosis of the diseases by the general practitioners and some diagnoses made in
6
7 150 secondary care, prescription of drugs, deaths, and other health utilisation indicators. The
8
9 151 details of the databases and their properties are given in Table 1 and 2.

10 152 *Participants*

11 153 People registered with the respective databases aged 18 year or above are eligible for the
12
13 154 study.

14 155
15 156 *Patient public involvement (PPI)*: Three PPI representatives (with OA and all with multiple
16
17 157 chronic conditions) were involved in this study through group meetings. Difficulties of living
18
19 158 with multiple conditions, lack of research in causal relationship and identification of diseases
20
21 159 to be studied and the role of drugs in comorbidity were discussed. They are constantly in
22
23 160 touch through providing their inputs at each step of the study.

24 161 *Definition of Osteoarthritis*

25 162 OA will be defined as having at least one recorded physician diagnosis of OA for hip, knee,
26
27 163 ankle/foot, wrist/hand, or site recorded as ‘unspecified’ during the study period for the
28
29 164 respective database. People with any previous recording of the OA prior to the start date of
30
31 165 the study will be excluded.

32 166 *Comorbidities*

33 167 We defined comorbidity as the recording of diagnosis of predefined chronic conditions in
34
35 168 individuals using either ICD-10 or Read or international classification of primary care (ICPC)
36
37 169 code. An extensive list of 61 chronic conditions was prepared from the Quality Outcome
38
39 170 Framework (QOF) (19), list of the US Department of Health and Human Services Initiative
40
41 171 on Multiple Chronic Conditions (20), global burden of diseases (21) and the Charlson
42
43 172 comorbidity index (22). The list has been updated with findings from our systematic review
44
45 173 (6) and a previous UK community-based knee pain study (6,23) by including common and
46
47 174 important morbidities not included in the above (24,25). A code mapping exercise was
48
49 175 conducted to finalise the list of conditions available for all the research centres. The
50
51 176 comparison of codes was made, and it was reviewed by four researchers including a clinician
52
53 177 from the team. The detailed list of the conditions is given in Table 3.

54 178 *Covariates*

55 179 Age and sex will be used in all centres as covariates to adjust in regression models.
56
57 180 Additionally, information on body mass index (BMI), smoking, alcohol use, socioeconomic
58
59 181 variables such as education level, income, place of birth (to identify those who immigrated to
60

1
2
3 182 the country), and residential area, marriage (or registered partner) will be included when
4
5 183 available. For calculating severity of the comorbidities in an individual, Elixhauser
6
7 184 comorbidity index will be used to estimate the impact of comorbidities on death (26,27).
8
9 185 Missing data on covariates will be substituted using multiple imputation methods, provided
10
11 186 that the data is missing at random, if applicable.

12 187 *Data Harmonisation:*

13
14
15 188 Firstly, we carried out a code mapping exercise for identification of people with osteoarthritis
16
17 189 (OA) and other comorbidities. We developed a list of chronic conditions and each centres
18
19 190 shared the list of codes to be used for the conditions, such as Read code in CPRD (UK),
20
21 191 ICPC2 (Rotterdam) and ICD-10 for Lund and Spanish as per their database. The code lists
22
23 192 were compared and edited to maintain the uniformity. The list was screened by verified by
24
25 193 two researchers and two GPs. We also used uniform definition for inclusion of condition e.g.
26
27 194 at least one recording of the chronic conditions. Because all the centres did not have all the
28
29 195 listed comorbidities, a minimum number of chronic conditions and covariates common in all
30
31 196 the database were identified to be studied. Similarly, we decided to have a minimum follow-
32
33 197 up study of 5 years and centres with more registration period can use the entire length of data
34
35 198 available.

36 199 *Study design and data analysis*

37 200 Summary of the study design and analysis is provided in Figure 1. All the centres plan to use
38
39 201 same statistical analysis plan to investigate each objective.

40 202 **Ethics and Dissemination:**

41 203 The study has obtained the following ethics approvals: UK- Independent Scientific Advisory
42
43 204 Council (ISAC) 19/30R, The Netherlands- The Integrated Primary Care Information (IPCI)
44
45 205 registration no. 11/2019, Spain - the Information System for the Development of Research in
46
47 206 Primary Care (SIDIAP), 4R19/011 , Sweden –Ethical Review Authority, Skåne Healthcare
48
49 207 Register, ‘Dnr 2011-432, Dnr 2014-276, and Dnr 2018-233. The registry databases are made
50
51 208 available anonymised for the research purposes. Each centre will follow the data privacy
52
53 209 policy of respective countries.

54
55 210 We plan to publish all the results as manuscripts in peer reviewed journals and present the
56
57 211 findings in relevant conferences. The results would be circulated as ‘lay-person’ language
58
59 212 and would be available at least four different international languages such as English, Dutch,
60

1
2
3 213 Swedish and Spanish. These will be shared on appropriate patient-public forum and involved
4 214 institution's websites.

5 215
6
7 216 *Objective 1 – Prevalence, incidence, and time sequence of comorbidities in OA*

8 217
9 218 A combined retrospective and prospective study of OA cases and sex, age (+/-2 years), first
10 219 year of registration, and practice matched controls (1:1-4) without OA (28) will be used to
11 220 determine the prevalence, incidence, and time sequence of comorbidities in OA. Incident OA
12 221 cases will be identified, and the first diagnosis date will be used as the starting point (index
13 222 date). For controls the same index date as their matched case will be used. They will be both
14 223 retrospectively reviewed for prior diagnoses of comorbidities and prospectively followed-up
15 224 for posterior new comorbidities. In the retrospective analysis the prevalence and 95%
16 225 confidence interval (CI) of each specific comorbidity will be calculated separately in OA
17 226 cases and matched controls using the number of people diagnosed with the comorbidity
18 227 divided by the total number of OA cases or controls at the index date. The prevalence of each
19 228 comorbidity in OA cases and matched controls will be calculated for given time intervals
20 229 prior to the index date of 0-1, 0-5 and 0-10 years separately to assess observational bias (28).
21 230 Discrete time intervals of 1-5, and 5-10 years before will also be used to estimate the
22 231 prevalence to minimise consultation bias/misclassification bias of OA (if possible). Logistic
23 232 regression will be used to calculate the odds ratios (OR) for each comorbidity unadjusted and
24 233 adjusted for BMI, smoking and alcohol consumption.
25 234 For the prospective analysis participants with incident OA but without the specific
26 235 comorbidity of interest at the index date (i.e., people at risk) and matched controls without
27 236 OA will be followed up until the date of the first diagnosis of the comorbidity, deregistration,
28 237 or death whichever comes first. The cumulative incidence will be calculated for each
29 238 comorbidity in OA cases and matched controls at 1, 3, 5, 10, 15, 20 years (based on the data
30 239 available) after the index date to examine the dynamic change of developing comorbidities
31 240 during follow-up. Kaplan-Meier survival curves will be used to display the cumulative
32 241 probability in OA and non-OA groups. Proportional hazard assumption will be tested using
33 242 Schoenfeld residual plots. The Cox regression model will be used to calculate hazard ratios
34 243 (HR) for each comorbidity unadjusted and adjusted for age, sex, practice, BMI, smoking and
35 244 alcohol consumption. This hybrid design has been previously used by us to examine the
36 245 temporality of associations between other rheumatic musculoskeletal diseases (RMDs) (e.g.,
37 246 gout and lupus) and comorbidities (28,29).
38 247

1
2
3 248 *Objective 2 – Clusters and impact of comorbidities in people with OA*

4
5 249 For each dataset, an 80%: 20% split into the training and testing data will be introduced. The
6
7 250 following analysis in objective 2 will be first employed into the training dataset and then
8
9 251 tested its generalisability in the testing dataset. At baseline, clusters of people based on 61
10
11 252 comorbidities will be identified using Latent class (i.e., Gaussian mixture models algorithms
12
13 253 of cluster) analysis (30). For each model, we will examine the association between clusters
14
15 254 and covariates using multinomial logistic regressions. The distinctness of clusters will be
16
17 255 examined by comparing covariates among clusters. The optimal model is the one where most
18
19 256 clusters found in the training data are also identified independently in the testing data and
20
21 257 clusters have most distinct patients' characteristics. We will then use both latent trajectory
22
23 258 analysis, such as joint latent class models (31), and unsupervised machine learning approach,
24
25 259 such as deep autoencoder or recurrent neural networks (32), to identify distinct clusters of
26
27 260 new comorbidity numbers development over time and their association with mortality with
28
29 261 adjustment for baseline covariates.

28 262 *Objective 3 - Association between analgesics and incident comorbidities*

29 263
30 264 A cohort study will be undertaken for this objective to evaluate the contribution of common
31
32 265 analgesics for OA to the development of comorbidity such as NSAIDs, opioids and
33
34 266 paracetamol. We are interested in the interaction between OA and use of drugs on the
35
36 267 incidence of comorbidities, i.e. to evaluate if the drug use in persons with OA poses increased
37
38 268 or decreased risk of comorbidities compared to persons without OA and/or analgesics.
39
40 269 Individual comorbidity, as well as clusters of comorbidities identified from Objective 2 will
41
42 270 be examined as outcomes. The 61 comorbidities in our study will be further categorised into
43
44 271 eight groups, specifically: musculoskeletal (MSK), respiratory, neurodegenerative,
45
46 272 psychological/psychiatric, cancer, cardiovascular, metabolic, renal problem, liver diseases,
47
48 273 gastrointestinal (GI) and others. (Table 4) The prospective cohort established from Objective
49
50 274 1 will form the source population for this objective. Individuals with incident OA will be
51
52 275 identified from the database and the first diagnosis date will be used as index date for follow-
53
54 276 up. Individuals without OA during the study period will be selected and matched with cases
55
56 277 by age, sex, and practice. The same index date will be given from their matched OA cases.
57
58 278 Individuals with analgesics prescriptions prior to the index date will be excluded (or recorded
59
60 279 as a confounding factor to be adjusted as appropriate). Only analgesic prescriptions after the
280
281 280 index date will be considered for this analysis. Prescriptions will be quantified as number of
prescriptions within year 1 (initial use, primary analysis) (33), 2, 3, 4, 5 etc. It will also be

1
2
3 282 dichotomised as episodic (e.g., at least one gap of ≥ 90 days between prescriptions) and
4
5 283 continuous (no gap of more less than 90 days) users as appropriate (34,35). Analgesic use
6
7 284 will be included in the model as a risk factor together with OA diagnosis (yes/no, primary
8
9 285 exposure) to examine the independent risk of each variable (OA and analgesics), as well as
10
11 286 the interaction between the two to the development of comorbidity. Dose response
12
13 287 relationship will be examined using number of prescriptions during the exposure window
14
15 288 examined. The effect of stopping analgesics will also be examined by looking into the
16
17 289 patterns of analgesic prescriptions, e.g., stopping analgesics after initial use in year 1 versus
18
19 290 continuous use of analgesics afterwards. For the primary analysis (initial prescriptions within
20
21 291 year 1), a landmark analysis will be used to minimise the immortal time bias where the
22
23 292 follow up will start after 12 months from the index date (36). Participants at risk (i.e., without
24
25 293 a specific comorbidity of interest) at the landmark date will be followed up until the first
26
27 294 diagnosis of the comorbidity, deregistration, or death whichever comes first. For secondary
28
29 295 analyses, time varying covariate analysis will be used to examine the long-term,
30
31 296 episodic/continuous use of analgesics after the index date and interaction between OA and
32
33 297 analgesics in the development of the comorbidity. The propensity score matching or the
34
35 298 inverse probability weighing methods will be used to adjust for confounding by indication
36
37 299 during the follow-up as appropriate. Depending on the country-specific drug use patterns, we
38
39 300 may modify this definition to allow for short brakes in between the episodes. Cox-regression
40
41 301 model will be used to calculate the HR and 95% CI. We will use flexible parametric models
42
43 302 using restricted cubic splines (developed by Lambert, “stpm2” in Stata) to estimate the HRs
44
45 303 and differences in time to diagnosis of comorbidities (outcome) with drugs as time-varying to
46
47 304 account for non-proportional hazards (37).

43 305 *Objective 4 – Potential causal pathways between OA and the comorbidities*

45 306 We will perform a Mendelian Randomisation (MR) phenome-wide association (MR-
46
47 307 PheWAS) study(38) to examine the causal relationship between OA, its phenotypes,
48
49 308 biomarkers or risk factors and comorbidities using the UK Biobank and the Rotterdam Study
50
51 309 database.

52
53 310 We will use the Rotterdam Study and the UK Biobank jointly for this objective. This is
54
55 311 because that the Rotterdam Study is an OA cohort with deep phenotypes and biomarkers of
56
57 312 OA, whereas the UK Biobank is a primary cohort for cancer and multiple disease outcomes,
58
59 313 and both have detailed genetic variants. We will use two sample MR approach, i.e., to
60
314 establish an association between OA and genetic variants in the Rotterdam Cohort to identify

1
2
3 315 genetic instrumental variables (IV), e.g., a set of single nucleotide polymorphisms (SNPs)
4 316 associated with OA (or a deep phenotype, biomarker, or risk factor of OA). We will then
5 317 undertake the MR-PheWAS analysis to examine the causal effects of the OA IV on
6
7 318 comorbidities in the UK Biobank. The MR method has been widely used in real world data to
8 319 examine the causal relationship between IV and specific disease, under two assumptions: [1]
9 320 genetic variants are randomly assigned in the population; and [2] genetic variants can only be
10 321 the cause not consequence of disease (39). The PheWAS is a series of case control studies to
11 322 estimate the associations between the IV and multiple disease outcomes (38,40). The
12 323 combination of the two permits investigation of the causal effects of OA on multiple disease
13 324 outcomes.

14
15 325 The MR-PheWAS analysis includes three steps. Firstly, we will identify the genetic variants
16 326 that are associated with OA - IV. Secondly, we will undertake the PheWAS analysis – a
17 327 series of case control analyses to estimate the associations between the IV and other disease
18 328 outcomes (38,40), with an adjustment for multiple testing using the false discovery rate
19 329 (FDR) methods (41). Thirdly, we will implement conventional MR analysis to investigate the
20 330 causal effects of the OA IV on comorbidities (39). An inverse variance weighted (IVW)
21 331 method will be used to pool the associations (ORs) as appropriate (42). The MR-Egger
22 332 regression analysis will be used to count for the pleiotropic effect - the effects of one genetic
23 333 variant on multiple outcomes (43). The heterogeneity in dependent instruments (HEIDI) test
24 334 will be used to exclude the cross-phenotype associations caused by genetic linkage (44).

25
26 335 With the MR-PheWAS study, the OR can be interpreted as causal association. We are
27 336 primarily interested in the causality from OA to comorbidities. We are also interested in
28 337 inflammatory (e.g., CRP), metabolic (e.g., gut microbiome) and biomechanics (e.g., BMI)
29 338 biomarkers and deep phenotypes of OA such as knee, hip, and hand OA with and without
30 339 symptoms. This will be undertaken if it is feasible within 3 years of this funded project,
31 340 otherwise will be considered as our future research agenda.

32 341 *Objective 5 – Variation of OA comorbidity patterns across countries*

33 342 We will use meta-analyses (MA) to examine the variation between countries and to pool the
34 343 data as appropriate. Estimates from first three objectives such as prevalence, incidence, OR,
35 344 HR and 95% CI for each specific comorbidity across different populations will be distributed
36 345 in a forest plot. Heterogeneity will be examined using the I^2 statistic and the Q test (45).
37 346 Results will be pooled if they are homogenous based on the I^2 value using the fixed effects

1
2
3 347 model, otherwise the reasons for the heterogeneity will be investigated. Random effects
4 348 models will be used to pool the results if the reasons for the heterogeneity cannot be
5 349 identified and if the overall pooling is appropriate. Individual patient data (IPD) meta-
6 350 analysis may be used to help identify the reasons for heterogeneity (46). Common clusters
7 351 and trajectories as well as burdens of comorbidities will also be compared between
8 352 populations.

353 *Feasibility and sample size*

354
355 To detect minimum incidence of 1% comorbidity (required for cluster analysis) with a
356 minimum clinically important difference of hazard ratio (HR)1.2, and 90% power of the study,
357 the estimated sample size was 197561 for 1581 events. It was calculated using STATA, with
358 a correlation= 0.2, standard deviation of 0.5, proportion of withdrawal= 0.20, alpha=0.05.
359 The initial check with the registry database revealed to have minimum required sample size
360 for the study.

361 **Discussion**

362
363 This study will be the largest epidemiological study on comorbidities of OA in primary care.
364 One of the key advantages of this multinational study is the use of the same protocol to
365 measure the burden of comorbidities in primary care settings in four European countries to
366 ensure reproducibility and comparison. There is scant evidence on the comorbidities in
367 people with OA, and this approach should help to identify the leading and most important
368 associations before and after presenting clinical OA (the index date). Further advantages of
369 this study are the large and representative populations studied and the same/similar extensive
370 list of chronic conditions for identifying comorbidity clusters. Often comorbidities
371 accumulate with age over time and the large primary care databases in this study have the
372 advantage of having long follow-up time which will enable us to detect the incidence of
373 comorbidities. Also, longer follow up would help to identify the picture of the trajectory of
374 the diseases (47). Both the incidence and the trajectories of comorbidity clusters are
375 highlighted as key elements needed in current research in multimorbidity, so findings from
376 this study should help to fill the knowledge gaps on multimorbidity in OA.
377 The relationship between chronic conditions and polypharmacy is a complex area of research.
378 The count of the medications and more importantly the nature of prescribed drugs may be
379 responsible for developing many new comorbidities in people with OA. We aim to explore
380 the associations of the most commonly prescribed drugs in OA, such as NSAIDs, with the
381

1
2
3 382 incidence of a wide range of comorbidities, which will be the first time that conditions other
4
5 383 than established comorbidities such as psychological conditions and endocrine diseases will
6
7 384 be examined. Finally, the causality study will further explore the associations at genetic
8
9 385 levels and phenotypes, which will be novel in OA research. Using a two sample MR
10
11 386 approach - one for OA deep phenotypes and the other for other chronic conditions maximises
12
13 387 the potentials of sample size, disease phenotypes and comorbidity spectrum to better explore
14
15 388 the causal pathways between OA and comorbidity.

15 389 There are some limitations to this study. Firstly, there are inherent issues in the nature of
16
17 390 electronic health records with respect to possible misdiagnosis, ascertainment biases, under-
18
19 391 or over-recording, and changes in databases due to change in coding structures. Also, the
20
21 392 analysis will be restricted to fewer covariates in some databases due to missing information
22
23 393 on lifestyle factors such as physical activities and diet. Even though the databases have
24
25 394 different durations of data available, if possible we will use a common follow-up time for
26
27 395 objective five. Another important limitation is that we do not have information on quality of
28
29 396 life and other outcomes to measure functional limitations recorded in the database.
30
31 397 Chronic conditions, especially comorbidities recorded in general practices, depend on
32
33 398 multiple factors such as population structure, health care facilities, health policies, and the
34
35 399 nature of the national databases. A major strength of this study is that it will include medical
36
37 400 records on approximately 27 million people in four European countries. Also, the study will
38
39 401 cover the sequence of research questions in comorbidity or multimorbidity starting from the
40
41 402 burden through to the causality and variation. Such a research model can be used for other
42
43 403 similar multimorbidity studies. The expected results should inform health professionals in
44
45 404 primary care settings with respect to management of people with OA and associated
46
47 405 comorbidities.

46 406 *Status of the study*

46 407 All the centres have obtained the necessary approvals for using the database in 2020. A
47
48 408 consensus has been made on the code mapping exercise. The statistical analysis will be
49
50 409 explained in detail in each of the publications. The team is expected to produce results by
51
52 410 mid-2021.

53 411 **Acknowledgements:** We thank the Patient Research Participants (PRP) members Jenny
54
55 412 Cockshull, Stevie Vanhegan, and Irene Pitsillidou for their involvement since the beginning
56
57 413 of the project. We would like to thank the FOREUM for financially supporting the research.
58
59 414 The authors would like to acknowledge Keele University's Prognosis and Consultation
60
415 Epidemiology Research Group who have given us permission to utilise the Code Lists
416 (©2014).

1
2
3 417 **Contributor and guarantor information:**

4 418 WZ, MD, CC, SMA, ME, DA conceived and designed the study. SS, AK, AD, AT, DR, and
5 419 VS developed the methods and will perform the analysis, and interpretation of the results.
6
7 420 CC, WZ, JR, AS, CFK, VS, and AT will supervise the statistical analysis. CM and MD will
8 421 guide with clinical interpretations of the results. SS drafted this manuscript and all authors
9 422 contributed to the critical revision of the manuscript for important intellectual content. The
10 423 corresponding author attests that all listed authors meet authorship criteria and that no others
11 424 meeting the criteria have been omitted.

12
13
14 425 **Funding:** This work was supported by Foundation for Research in Rheumatology
15 426 (FOREUM) grant (2019-2022), The Swedish Research Council (2020-01103), Governmental
16 427 funding of clinical research within the national health services (ALF), and The Swedish
17 428 Rheumatism Association. CM is funded by the National Institute for Health Research (NIHR)
18 429 Applied Research Collaboration West Midlands, the National Institute for Health Research
19 430 (NIHR) School for Primary Care Research and a National Institute for Health Research
20 431 (NIHR) Research Professorship in General Practice (NIHR-RP-2014-04-026)

21 432 **Role of the funding sources:** The sponsors did not participate in the design and conduct of
22 433 the study; collection, management, analysis, and interpretation of the data; or preparation,
23 434 review, or approval of the manuscript and the decision to submit the manuscript for
24 435 publication.

25 436 **Competing interests:** WZ declares serving as an advisory board for Ely Lilly (Ixekizumab,
26 437 2020) and Regeneron (Fasinomab, 2020). ME declares serving as an advisory Panel Board
27 438 Member for Pfizer (Nov 2019, Tanezumab). CM provided advice to BMS on recruiting to a
28 439 non-pharmacological atrial fibrillation trial. The other authors report no financial competing
29 440 interests.

30 441 **Copy right:** The Corresponding Author has the right to grant on behalf of all authors and
31 442 does grant on behalf of all authors, an exclusive licence (or non-exclusive for government
32 443 employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if
33 444 accepted) to be published in BMJ editions and any other BMJ PGL products and sublicences
34 445 such use and exploit all subsidiary rights, as set out in our licence.

35
36 446 **Studies involving humans or animals:** No direct participant recruitment was done for the
37 447 study.

38
39 448 **Data sharing statement:** We used anonymised data on individual patients on which the
40 449 analysis, results, and conclusions reported in the paper are based. The used data is not
41 450 distributable under licence. However, the relevant data can be obtained directly from the
42 451 respective agencies. The codes developed for the analysis can be available upon a valid
43 452 request.

44
45
46
47 453 **Figure legend**

48 454 Figure 1. Overview of the study design and statistical analysis plan for the Comorbidities in
49 455 Osteoarthritis (CoMO) study.

Reference

1. Turkiewicz A, Petersson IF, Björk J, Hawker G, Dahlberg LE, Lohmander LS, et al. Current and future impact of osteoarthritis on health care: a population-based study with projections to year 2032. *Osteoarthritis Cartilage*. 2014 Nov;22(11):1826–32.
2. Badley EM. The effect of osteoarthritis on disability and health care use in Canada. *J Rheumatol Suppl*. 1995 Feb;43:19–22.
3. McDonough CM, Jette AM. The Contribution of Osteoarthritis to Functional Limitations and Disability. *Clin Geriatr Med*. 2010 Aug;26(3):387–99.
4. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. 2020 Dec;396(10267):2006–17.
5. World Health Organization. Musculoskeletal conditions [Internet]. [cited 2021 Apr 6]. Available from: <https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions>
6. Swain S, Sarmanova A, Coupland C, Doherty M, Zhang W. Comorbidities in Osteoarthritis: A systematic review and meta-analysis of observational studies. 2019 Jun 17 [cited 2019 Jun 30]; Available from: <http://doi.wiley.com/10.1002/acr.24008>
7. Suri P, Morgenroth DC, Hunter DJ. Epidemiology of osteoarthritis and associated comorbidities. *PM R*. 2012 May;4(5 Suppl):S10-19.
8. Hall AJ, Stubbs B, Mamas MA, Myint PK, Smith TO. Association between osteoarthritis and cardiovascular disease: Systematic review and meta-analysis. *Eur J Prev Cardiol*. 2016;23(9):938–46.
9. Stubbs B, Aluko Y, Myint PK, Smith TO. Prevalence of depressive symptoms and anxiety in osteoarthritis: a systematic review and meta-analysis. *Age Ageing*. 2016 Mar 1;45(2):228–35.
10. Ricci-Cabello I, Stevens S, Kontopantelis E, Dalton ARH, Griffiths RI, Campbell JL, et al. Impact of the Prevalence of Concordant and Discordant Conditions on the Quality of Diabetes Care in Family Practices in England. *Ann Fam Med*. 2015 Nov 1;13(6):514–22.
11. Calderón-Larrañaga A, Poblador-Plou B, González-Rubio F, Gimeno-Feliu LA, Abad-Díez JM, Prados-Torres A. Multimorbidity, polypharmacy, referrals, and adverse drug events: Are we doing things well? *Br J Gen Pract*. 2012;62(605):e821–6.
12. Jüni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *The Lancet*. 2004 Dec;364(9450):2021–9.
13. Castellsague J, Riera-Guardia N, Calingaert B, Varas-Lorenzo C, Fourrier-Reglat A, Nicotra F, et al. Individual NSAIDs and Upper Gastrointestinal Complications: A Systematic Review and Meta-Analysis of Observational Studies (the SOS Project). *Drug Saf*. 2012 Dec;35(12):1127–46.
14. Lowe D.B., Taylor M.J., Hill S.J. Cross-sectional examination of musculoskeletal conditions and multimorbidity: influence of different thresholds and definitions on prevalence and association estimates. *BMC Res Notes*. 2017;10(1):51.
15. John R, Kerby DS, Hagan Hennessy C. Patterns and Impact of Comorbidity and Multimorbidity Among Community-Resident American Indian Elders. *The Gerontologist*. 2003 Oct 1;43(5):649–60.

16. Alaeddini A., Jaramillo C.A., Faruqi S.H.A., Pugh M.J. Mining Major Transitions of Chronic Conditions in Patients with Multiple Chronic Conditions. *Methods Inf Med.* 2017;56(5):391–400.
17. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med.* 2015 Mar 31;12(3):e1001779.
18. Engelhart MJ, Geerlings MI, Meijer J, Kiliaan A, Ruitenberg A, van Swieten JC, et al. Inflammatory Proteins in Plasma and the Risk of Dementia: The Rotterdam Study. *Arch Neurol.* 2004 May 1;61(5):668.
19. Quality and Outcome Framework (QOF) [Internet]. [cited 2018 Mar 21]. Available from: <https://digital.nhs.uk/article/8910/Quality-and-Outcome-Framework-QOF-Indicators-No-Longer-In-QOF-INLIQ-Enhanced-Services-ES-Vaccinations-and-Immunisations-V-I-and-GMS-Core-Contract-CC-extraction-specifications-business-rules->
20. Medicare C for, Baltimore MS 7500 SB, Usa M. CC_Main [Internet]. 2017 [cited 2018 Oct 24]. Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/CC_Main.html
21. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet.* 2018 Nov;392(10159):1789–858.
22. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis.* 1987 Jan 1;40(5):373–83.
23. Sarmanova A, Fernandes GS, Richardson H, Valdes AM, Walsh DA, Zhang W, et al. Contribution of central and peripheral risk factors to prevalence, incidence and progression of knee pain: a community-based cohort study. *Osteoarthritis Cartilage.* 2018 Nov;26(11):1461–73.
24. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol.* 1994 Nov 1;47(11):1245–51.
25. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. *Am J Epidemiol.* 2011 Mar 15;173(6):676–82.
26. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A Modification of the Elixhauser Comorbidity Measures Into a Point System for Hospital Death Using Administrative Data: *Med Care.* 2009 Jun;47(6):626–33.
27. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care.* 1998 Jan;36(1):8–27.
28. Kuo C-F, Grainge MJ, Mallen C, Zhang W, Doherty M. Comorbidities in patients with gout prior to and following diagnosis: case-control study. *Ann Rheum Dis.* 2014 Nov 14;annrheumdis-2014-206410.
29. Swain S, Coupland C, Mallen C, Kuo CF, Sarmanova A, Bierma-Zeinstra SMA, et al. Temporal relationship between osteoarthritis and comorbidities: a combined case control and cohort study in the UK primary care setting. *Rheumatology.* 2021 Mar 1;keab067.

- 1
- 2
- 3 30. Collins LM, Lanza ST. Latent Class and Latent Transition Analysis [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2009 [cited 2020 Apr 5]. (Wiley Series in Probability and Statistics). Available from: <http://doi.wiley.com/10.1002/9780470567333>
- 4
- 5
- 6
- 7 31. Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: A review. *Stat Methods Med Res*. 2014 Feb;23(1):74–90.
- 8
- 9
- 10 32. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J Mach Learn Res*. 2010 Dec;11:3371–408.
- 11
- 12
- 13 33. Zeng C, Zhang W, Doherty M, Persson MSM, Mallen C, Swain S, et al. Initial analgesic prescriptions for osteoarthritis in the United Kingdom, 2000–2016. *Rheumatology*. 2021 Jan 5;60(1):147–59.
- 14
- 15 34. Hippisley-Cox J, Coupland C. Unintended effects of statins in men and women in England and Wales: population based cohort study using the QResearch database. *BMJ*. 2010 May 20;340(may19 4):c2197–c2197.
- 16
- 17 35. Sarmanova A, Doherty M, Kuo C, Wei J, Abhishek A, Mallen C, et al. Statin use and risk of joint replacement due to osteoarthritis and rheumatoid arthritis: a propensity-score matched longitudinal cohort study. *Rheumatology*. 2020 Oct 1;59(10):2898–907.
- 18
- 19 36. Kuo C-F, Grainge MJ, Mallen C, Zhang W, Doherty M. Effect of allopurinol on all-cause mortality in adults with incident gout: propensity score–matched landmark analysis. *Rheumatology*. 2015 Jul 12;kev246.
- 20
- 21 37. Royston P, Lambert PC. Flexible parametric survival analysis using Stata: beyond the Cox model. College Station, TX: Stata Press; 2011. 347 p.
- 22
- 23 38. Millard LAC, Davies NM, Timpson NJ, Tilling K, Flach PA, Smith GD. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. *Sci Rep* [Internet]. 2015 Dec [cited 2020 Jan 26];5(1). Available from: <http://www.nature.com/articles/srep16645>
- 24
- 25 39. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018 Jul 12;k601.
- 26
- 27 40. Li X, Meng X, Spiliopoulou A, Timofeeva M, Wei W-Q, Gifford A, et al. MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank. *Ann Rheum Dis*. 2018 Jul;77(7):1039–47.
- 28
- 29 41. Benjamini Y, Yekutieli D. The Control of the False Discovery Rate in Multiple Testing under Dependency. *Ann Stat*. 2001;29(4):1165–88.
- 30
- 31 42. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data: Mendelian Randomization Using Summarized Data. *Genet Epidemiol*. 2013 Nov;37(7):658–65.
- 32
- 33 43. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016 May;40(4):304–14.
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- 1
2
3 44. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data
4 from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016
5 May;48(5):481–7.
6
7 45. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002 Jun
8 15;21(11):1539–58.
9
10 46. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic
11 reviews using individual patient data. *Eval Health Prof.* 2002 Mar;25(1):76–97.
12
13 47. Vetrano DL, Roso-Llorach A, Fernández S, Guisado-Clavero M, Violán C, Onder G, et al.
14 Twelve-year clinical trajectories of multimorbidity in a population of older adults. *Nat Commun*
15 [Internet]. 2020 Dec [cited 2020 Aug 12];11(1). Available from:
16 <http://www.nature.com/articles/s41467-020-16780-x>
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Characteristics of the included databases

	Netherlands	Spain	Sweden	UK
Objectives 1-3				
Name of the database	Integrated Primary Care Information (IPCI)	The Spanish Sistema information del Deveolpment de l'Investigació a Atenció Primària (SIDIAP)	Skåne Healthcare Register	Clinical Practice Research Datalink (CPRD)
Settings	Routinely collected primary care database	Routinely collected primary care data	Swedish healthcare in Skane region, primary, specialist and in-patient care	Routinely collected primary care database with linkage database
Size and Coverage	2.2 million (Randomly distributed over the country)	6.5 million (> 85% of total Catalan region)	1.3 million (all residents of the Skane region)	17 million (country-wide, nearly 740 practices)
Start year	1998- (Better coding after 2000)	2006	1998	1993-
Age group	All	All	All	All
Gender	All	All	All	All
Coding system	ICPC	ICD 10	ICD 10	Read codes and ICD 10
Drug prescribed by	GP	GP	GP	GP
Death record (Either date of death and/or cause)	Both date and cause	Only date	Both date and cause (until year 2015)	Only date
Covariates/ additional variables	NA	BMI, Smoking, Alcohol, Social class, cholesterol, and other biomarkers	Education, Income, profession, and sick leave, residential area, region of birth	BMI, Smoking, Alcohol, Deprivation index, Ethnic group

BMI- Body mass index; GP- general practitioners; ICD- International classification of diseases; ICPC- International classification of primary care; NA- Not available

Table 2. Database for the mendelian randomisation

Name of the database	Rotterdam Cohort Study	UK Biobank
Population coverage	15000	500,000
Age group	>=40 years	40-69 years
Start year -till now	1989-onwards	2010-
Types of data	Radiographic data, joint pain, joint stiffness, of hip, knee, and hand, GWAS, biochemical markers	Genetic and phenotypes

GWAS- Genome Wide Association Studies

For peer review only

Table 3. List of chronic conditions across four databases

SI no	Conditions	ICD 10	ICPC	Read Code
1	Anaemia (All types)	D50-D64	B78, B80, B81, B82	D00..., D01...
2	Ankylosing spondylitis	M45.9	NA	N10...
3	Anxiety disorder	F41.0-F41.9	P74, P74.01, P74.02	E200..., Eu41...
4	Asthma	J450-J45.9, J46.9	R96...	66Y..., H33...
5	Benign prostatic hypertrophy (BPH)	N40.9	Y85	K20..., K21..., K22...
6	Cardiac arrhythmias (Atrial Fibrillation)	I47.0-I49.9	K78	Gyu..., G573...
7	Cataract	H25.0-H25.9, H26.1-H26.9	F92	F46...,
8	Chronic Back pain	M47-M48, M51-M54, M99, G54.4	L02, L03, L86	N12..., N14...,
9	Chronic kidney disease (any cause)	N02.0-N8.8, N11.0-N11.9, N12.9, N15.0-N18.9, N19.9	U99.01	I21..., K01..., K02...
10	Chronic neck pain	M54.2	L83...	Nyu..., N11..., N12..., N14...
11	Chronic sinusitis	J32 ...	R75...	H13...
12	Chronic obstructive pulmonary diseases	J41.0-J41.8, J42.9, J43.0-J43.9, J44.0-J44.9	R91, R95	H3...
13	Coronary Heart Disease (Including Acute Myocardial infarction, Valvular disease, Angina),	I20.0-I25.0, I34.0-I37.0	K74-K76	G11..., G30..., G31... - G38..
14	Dementia	F00.0-F00.9, F01.0-F03.9, G30.0-G30.9, G31.0-G31.9	P70-P70.02	E00..., Eu0..., F11...
15	Depression	F32.0-F33.9	P76	Eu..., E11...
16	Diabetes mellitus	E10.0-E14.9	T90..., F83.01	C10..., F32...,
17	Dyslipidaemia (Hyper)	E78.1	T93	C32...
18	Eating disorders (Both)			Eu5..., R03...
19	Eczema/ Skin disease	L20.0-L22.9, L26.9	S74, S87, S88	M11...
20	Epilepsy	G40.0-G41.9	N88	F25...
21	Fatigue	F48.0	A04.11	F286...
22	Fibromyalgia	M79.7	L18.01	N248..., N239...
23	Gall bladder stone	K80.0-K80.8	D98-D98.03	781..., J65..., 4G2...,
24	GERD (Gastritis, Oesophageal bleeding, duodenitis, peptic ulcer)	K21 ...	D840...	J12..., J13..., J15...
25	Gastrointestinal bleeding	K25.0-K28.9	D84-D87	J11...,
26	Gout	M10.0-M10.9	T92	C34..., N023...
27	Hearing impairment (All types)	H90.0-H91.9	H83-H86	F59..., ZE87...
28	Heart Failure	I50.0-I50.9	K77-K77.02	G58...,
29	Hepatitis	K73.0-K73.9	D72-D72.05	J61..., J63...
30	HIV/AIDS	B20-B24	B90	A788..., A789..., AyuC...
31	Hypertension	I10.9, I11.0-I13.9, I15.0-I15.9	K86-K87, F83.02	G20..., G24..., G25..., G26...
32	Hyperthyroidism	E05.0-E05.9	T85	C02...
33	Hypothyroidism	E02.9, E03.0-E03.9	T86	C03..., C04...
34	Inflammatory Bowel Disease (IBD)	K50.0-K52.9	D94-D94.02	J4...,
35	Irritable Bowel Symptoms (IBS)	K58.1-K58.8		J52...
36	Leukaemia, Lymphoma	C81.0-C86.6, C91.0-C96.9	B72-B73	B60..., B61..., B64...
37	Liver Cirrhosis	K70.0-K71.9, K74.0-K74.6	D97...	J615...
38	Migraine	G43.0-G43.9	N89	F26...,
39	Multiple sclerosis	G35.9	N86	F20...,
40	Osteoarthritis	M16.0-M16.9, M17.0-M17.9	L89-L91	N05...,
41	Osteoporosis	M80.0-M82.9	L95...	N33...,

42	Other blood vessel disease (Raynaud's disease, Burger's disease)	I73.0-I73.9	K92...	G73... ,
43	Parkinson's disease	G20.9	N87-N87.01	F12... ,
44	Peripheral vascular disease (Atherosclerosis)	I70.0-I70.9	K91	G70... , G71... , G72... ,
45	Polymyalgia	M35.3	L99.12	N20.11
46	Psoriasis	L40.0-L41.9	S91	M161...
47	Psoriatic arthritis	M07.0-M07.3	L99.13	M160...
48	Rheumatoid Arthritis	M05.0-M05.9	L88... , K71...	N04...
49	Renal stones	N20.0	U95	4G4... , 7B07... , KB12...
50	Schizophrenia and/or psychosis	F20.0-F20.9, F25.0- F25.9	P72	E10...
51	Severe allergy			H17... , SN5...
52	Sjögren's syndrome	M35.0	NA	N002...
53	Systemic Lupus Erythematosus	M32.0, M32.1, M32.8, M32.9	NA	N000...
54	Sleep disorder (Insomnia)	F51.0	P06	Fy0... , 1B1B...
55	Solid malignancy	C00.0-C80.9, D00.0-D09.9, C97.9	A29... , A79, B74 – Y78	B0... - B67... , Byu...
56	Stroke	G45.0-G46.8, I60.0-I63.9, I65.0-166.9, I69.0-I69.4	K89-K90.02	G60... - G68... , F22...
57	Substance abuse/ Drug addiction	F10.0-F19.9	P18, P19...	E24... , Eu1...
58	Thrombotic diseases	I74.0-I74.9	K93, K94... , W99.03	G80... , G81... , G74...
59	Tuberculosis	A15.0-A16.9, B90.9	A70, R70	A1... , A11...
60	Vertigo	H81.4 ...	N17-N17.02, H82-H82.03	R004... , F561...
61	Vision problem (Glaucoma and other)	H27.0-H27.9, H40.0-H40.9, H42.0-H42.8	F93... , F94	F45... , F49... ,

ICD- International Classification of Diseases; ICPC- International classification in primary care. All the codes are the primary code initials used in the database.

Table 4. Group of conditions/Outcome

Group	Conditions
Cardiovascular	Cardiac arrhythmias, Coronary Heart Disease (including AMI, valvular disease, angina), Heart failure, Hypertension, Peripheral vascular disease (claudication, Raynaud syndrome, Buerger's disease), Other blood vessel disease (atherosclerosis and aneurysm), Thrombotic diseases
Gastrointestinal	GERD (Esophageal diseases, gastritis, duodenitis), GI bleeding, Inflammatory bowel disease (IBD), Irritable bowel syndrome (IBS)
Musculoskeletal	Ankylosing spondylitis, Chronic Back pain, Chronic Neck pain, Fibromyalgia, Polymyalgia, Gout, Osteoporosis, Psoriatic arthritis, Rheumatoid arthritis, Sjögren's syndrome, Systemic lupus erythematosus (SLE)
Endocrine	Diabetes mellitus, Dyslipidemia (hyper), Hyperthyroidism, Hypothyroidism
Neurological	Dementia, Epilepsy, Fatigue, Migraine, Multiple sclerosis, Parkinson disease, Stroke
Psychological	Anxiety, Depression, Eating disorders (Anorexia / Bulimia nervosa), Schizophrenia, Sleep disorder (insomnia),
Kidney disease	Chronic kidney disease (any cause), Renal stones
Liver diseases	Gall bladder stone, Hepatitis, Liver cirrhosis
Respiratory	Asthma, Chronic obstructive pulmonary disease (COPD)
Cancer	Leukemia, Lymphoma, Solid Malignancy (any type)
Others	Anemia (all types), Benign prostate hypertrophy (BPH), Cataract, Chronic sinusitis, Eczema/Skin disease, Hearing impairment (all types), Psoriasis, Severe allergy (anaphylactic shock), angioneurotic oedema

6/bmjopen-2021-00816 on 6 April 2022. Downloaded from <http://bmjopen.bmj.com/> on April 18, 2024 by guest. Protected by copyright.

Figure 1. Overview of the study design and statistical analysis plan for the Comorbidities in Osteoarthritis (CoMO) study.

	Objective 1	Objective 2	Objective 3	Objective 4
Study Design	Case-control and Cohort	Clustering and longitudinal	Cohort	Genomic association
Exposure	OA	OA	Analgesics (NSAIDs, Opioids, paracetamol)	OA
Outcome	Comorbidities	Clusters of comorbidities	Comorbidities	Comorbidities
Statistical methods	Conditional logistic regression, Cox regression	Latent class analysis, Latent class growth analysis, Joint latent class analysis	Cox regression, Time varying analysis, Flexible parametric method	Mendelian randomisation
Reported Outcome	Odds Ratio and Hazard Ratio	Clusters and groups	Hazard Ratio	Coefficients
Participating centres	ALL	ALL	ALL	UK and Netherlands

Objective 5	Meta Analysis
-------------	---------------

OA- osteoarthritis; NSAIDs- Nonsteroidal anti-inflammatory drugs

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*
Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract Comorbidities in Osteoarthritis (CoMO): a multinational study in four European countries	1-2
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found <i>Introduction</i> Osteoarthritis (OA) is one of the leading chronic conditions in the older population. People with OA are more likely to have one or more other chronic conditions than those without. However, the temporal associations, clusters of the comorbidities, role of analgesics and the causality and variation between populations are yet to be investigated. This paper describes the protocol of a multinational study in four European countries (UK, Netherlands, Sweden, and Spain) exploring comorbidities in people with OA . <i>Methods and analysis</i> This multinational study will investigate i) the temporal associations of 61 identified comorbidities with OA, ii) the clusters and trajectories of comorbidities in people with OA, iii) the role of analgesics on incidence of comorbidities in people with OA, iv) the potential biomarkers and causality between OA and the comorbidities, and v) variations between countries. A combined case-control and cohort study will be conducted to find the temporal association of OA with the comorbidities using the national or regional health databases. Latent class analysis will be performed to identify the clusters at baseline and joint latent class analysis will be used to examine trajectories during the follow-up. A cohort study will be undertaken to evaluate the role of non-steroidal anti-inflammatory drugs (NSAIDs), opioids, and paracetamol on the incidence of comorbidities. Mendelian randomisation will be performed to investigate the potential biomarkers for causality between OA and the comorbidities using the UK Biobank and the Rotterdam	1

6/bmjopen-2022-052816 on 6 April 2022. Downloaded from <http://bmjopen.bmj.com/> on April 18, 2024 by guest. Protected by copyright.

		<p>Study databases. Finally, a meta-analyses will be used to examine the variations and pool the results from different countries.</p> <p><i>Ethics and Dissemination</i></p> <p>Research ethics was obtained according to each database requirement. Results will be disseminated through the FOREUM website, scientific meetings, publications, and in partnership with patient organisations.</p>	
Introduction			
Background/rationale	2	<p>Explain the scientific background and rationale for the investigation being reported</p> <p>Osteoarthritis (OA) affects 27% of people aged over 45 years at peripheral synovial joints such as knees, hips, hands and feet (1). It is by far the most common form of arthritis, and a leading cause of chronic joint pain and disability in older people (2,3). It is anticipated that the burden of OA will continue to rise in the coming decades because of population ageing and the increasing obesity prevalence – two major risk factors for OA (4,5). Co-occurrence of multiple chronic conditions in an individual with ageing is becoming a norm and OA is not an exception to this.</p> <p>A recent systematic review has confirmed that people with OA are more likely to have other diseases, especially stroke, peptic ulcer, hypertension, and depression (6). Vast majority of these studies focused on additional presence (comorbidity) of cardiovascular and musculoskeletal conditions only (7–9). Whether these comorbidities just co-exist with OA, share common risk factors with OA, or are causes or consequences of OA remains largely unknown. There was also reporting of wide heterogeneity in definitions of OA and other chronic conditions, diagnosis and recording of diseases, sample sizes, and number of diseases studied in previously published studies included in the review (6). This diversity made the comparison and pooled estimation of comorbidity prevalence difficult.</p> <p>Comorbidity in OA can occur due to multiple factors. Various hypotheses have been used to explain the existence of comorbidities in general, the most accepted of which are the concordant (disease sharing similar</p>	4-5

pathophysiological risk factors) and discordant (diseases not sharing similar pathophysiological risk factors) theories (10). Additionally, prescription of drugs is also reported to be associated with comorbidity and multimorbidity (11). Especially in people with OA, the prescription of analgesics is common, and is associated with increased risk of other conditions such as cardiovascular, gastrointestinal and chronic renal diseases (12,13). Also, having multiple chronic conditions increases the chances of polypharmacy which further escalates the risk of other conditions.

OA is one of the leading conditions reported in multimorbidity research. Exploring the association of OA with other diseases would help in further explaining the burden and pattern of the comorbidity (14). However, the major issue in OA comorbidity research is the low number and specific types of conditions studied (16). Therefore, it is important to develop a consensus on both the count and typology of conditions to be studied to enable comparisons across populations and to derive pooled estimates as appropriate. Further, using uniform methods and definitions of diseases in computing these estimates would reduce heterogeneity and make the comparison more reliable.

Understanding the temporal association with comorbidity and disease trajectory is crucial for any chronic condition, and this is possible through studies using longitudinal databases (15). However, one of the limitations of using observational data is that causal associations are difficult to establish, due to the interference of known and unknown confounders. In this study we have used the more recently developed method ‘Mendelian Randomization’, that can determine causal estimates through combining the use of genetic data and instrumental variable methods.”

The burden of diseases in primary care often depends on the population structure, health infrastructure and individual factors such as income and education. Such factors vary between countries, because of the heterogeneity mentioned above there are no robust data comparing OA and its comorbidities between countries. Therefore, the aims of this study were to explore the burden, pattern, and causal factors of comorbidities in people with OA across four European countries using national registration databases in the UK, the Netherlands, Sweden, and

		Spain.	
Objectives	3	<p>State specific objectives, including any pre-specified hypotheses</p> <ol style="list-style-type: none"> 1. To estimate the prevalence, incidence, and time sequence of comorbidities in OA 2. To examine the clusters of comorbidities and trajectories of clusters in OA and associations with death 3. To investigate the associations between commonly used OA drugs, such as non-steroidal anti-inflammatory (NSAIDs) and opioids, and risk of comorbidities 4. To identify the potential biomarkers and causal pathways between OA and the comorbidities 5. To examine the variations of OA comorbidities and clusters across countries 	5
Methods			
Study design	4	<p>Present key elements of study design early in the paper</p> <p>Figure 1</p>	Figure 1
Setting	5	<p>Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection</p> <p><i>Databases</i></p> <p>Four routinely collected national (the UK and the Netherlands) or regional (Sweden and Spain) health databases will be used for objectives 1-3. In addition, for objective 4, genomic associations of OA with comorbidities will be examined using two cohort studies from the UK (UK Biobank) (17) and the Netherlands (Rotterdam study) (18). The four national representative and regional databases contain information about the population with primary care consultations in four different countries. The longitudinal databases provide information about the diagnosis of the diseases by the general practitioners and some diagnoses made in secondary care, prescription of drugs, deaths, and other health utilisation indicators. The details of the databases and their properties are given in Table 1 and 2.</p> <p><i>Participants</i></p> <p>People registered with the respective databases aged 18 year or above are eligible for the study.</p>	5-6

Participants	6	<p><i>(a) Cohort study</i>—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i>—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls</p>	n/a
		<p>(b) Cohort study—For matched studies, give matching criteria and number of exposed and unexposed Case-control study—For matched studies, give matching criteria and the number of controls per case</p> <p><i>Objective 1 – Prevalence, incidence, and time sequence of comorbidities in OA</i></p> <p>A combined retrospective and prospective study of OA cases and sex, age (+/-2 years), first year of registration, and practice matched controls (1:1-4) without OA (28) will be used to determine the prevalence, incidence, and time sequence of comorbidities in OA. Incident OA cases will be identified, and the first diagnosis date will be used as the starting point (index date). For controls the same index date as their matched case will be used. They will be both retrospectively reviewed for prior diagnoses of comorbidities and prospectively followed-up for posterior new comorbidities.</p> <p>For the prospective analysis participants with incident OA but without the specific comorbidity of interest at the index date (i.e., people at risk) and matched controls without OA will be followed up until the date of the first diagnosis of the comorbidity, deregistration, or death whichever comes first. The cumulative incidence will be calculated for each comorbidity in OA cases and matched controls at 1, 3, 5, 10, 15, 20 years (based on the data available) after the index date to examine the dynamic change of developing comorbidities during follow-up.</p> <p><i>Objective 3</i></p> <p>A cohort study will be undertaken for this objective to evaluate the contribution of common analgesics for OA to the development of comorbidity such as NSAIDs, opioids and paracetamol. We are interested in the interaction between OA and use of drugs on the incidence of comorbidities, i.e. to evaluate if the drug use in persons with OA poses increased or decreased risk of comorbidities compared to persons without OA and/or analgesics. Individual</p>	7-8
		<p>A cohort study will be undertaken for this objective to evaluate the contribution of common analgesics for OA to the development of comorbidity such as NSAIDs, opioids and paracetamol. We are interested in the interaction between OA and use of drugs on the incidence of comorbidities, i.e. to evaluate if the drug use in persons with OA poses increased or decreased risk of comorbidities compared to persons without OA and/or analgesics. Individual</p>	9

6/bmjopen-2024-02816 on 6 April 2024. Downloaded from <http://bmjopen.bmj.com/> on April 18, 2024 by guest. Protected by copyright.

comorbidity, as well as clusters of comorbidities identified from Objective 2 will be examined as outcomes. The 61 comorbidities in our study will be further categorised into eight groups, specifically musculoskeletal (MSK), respiratory, neurodegenerative, psychological/psychiatric, cancer, cardiovascular, metabolic, renal problem, liver diseases, gastrointestinal (GI) and others. (Table 4) The prospective cohort established from Objective 1 will form the source population for this objective. Individuals with incident OA will be identified from the database and the first diagnosis date will be used as index date for follow-up. Individuals without OA during the study period will be selected and matched with cases by age, sex, and practice. The same index date will be given from their matched OA cases. Individuals with analgesics prescriptions prior to the index date will be excluded (or recorded as a confounding factor to be adjusted as appropriate). Only analgesic prescriptions after the index date will be considered for this analysis. Prescriptions will be quantified as number of prescriptions within year 1 (initial use, primary analysis) (33), 2, 3, 4, 5 etc. It will also be dichotomised as episodic (e.g., at least one gap of ≥ 90 days between prescriptions) and continuous (no gap of more less than 90 days) users as appropriate (34,35). Analgesic use will be included in the model as a risk factor together with OA diagnosis (yes/no, primary exposure) to examine the independent risk of each variable (OA and analgesics), as well as the interaction between the two to the development of comorbidity. Dose response relationship will be examined using number of prescriptions during the exposure window examined. The effect of stopping analgesics will also be examined by looking into the patterns of analgesic prescriptions, e.g., stopping analgesics after initial use in year 1 versus continuous use of analgesics afterwards.

Objective 4

We will use the Rotterdam Study and the UK Biobank jointly for this objective. This is because that the Rotterdam Study is an OA cohort with deep phenotypes and biomarkers of OA, whereas the UK Biobank is a primary cohort for cancer and multiple disease outcomes, and both have detailed genetic variants. We will use two sample MR

		<p>approach, i.e., to establish an association between OA and genetic variants in the Rotterdam Cohort to identify genetic instrumental variables (IV), e.g., a set of single nucleotide polymorphisms (SNPs) associated with OA (or a deep phenotype, biomarker, or risk factor of OA). We will then undertake the MR-PheWAS analysis to examine the causal effects of the OA IV on comorbidities in the UK Biobank. The MR method has been widely used in real world data to examine the causal relationship between IV and specific disease, under two assumptions: [1] genetic variants are randomly assigned in the population; and [2] genetic variants can only be the cause not consequence of disease (39). The PheWAS is a series of case control studies to estimate the associations between the IV and multiple disease outcomes (38,40). The combination of the two permits investigation of the causal effects of OA on multiple disease outcomes.</p>	
Variables	7	<p>Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable</p> <p><i>Definition of Osteoarthritis</i></p> <p>OA will be defined as having at least one recorded physician diagnosis of OA for hip, knee, ankle/foot, wrist/hand, or site recorded as ‘unspecified’ during the study period for the respective database. People with any previous recording of the OA prior to the start date of the study will be excluded.</p> <p><i>Comorbidities</i></p> <p>We defined comorbidity as the recording of diagnosis of predefined chronic conditions in individuals using either ICD-10 or Read or international classification of primary care (ICPC) code. An extensive list of 61 chronic conditions was prepared from the Quality Outcome Framework (QOF) (19), list of the US Department of Health and Human Services Initiative on Multiple Chronic Conditions (20), global burden of diseases (21) and the Charlson comorbidity index (22). The list has been updated with findings from our systematic review (6) and a previous UK community-based knee pain study (6,23) by including common and important morbidities not included in the above (24,25). A code mapping exercise was conducted to finalise the list of conditions available for all the research centres. The comparison of codes was made, and it was reviewed by four researchers including</p>	6-7

		two researchers and two GPs. We also used uniform definition for inclusion of condition e.g. at least one recording of the chronic conditions. Because all the centres did not have all the listed comorbidities, a minimum number of chronic conditions and covariates common in all the database were identified to be studied. Similarly, we decided to have a minimum follow-up study of 5 years and centres with more registration period can use the entire length of data available.	
Bias	9	Describe any efforts to address potential sources of bias	Explained under each objective
Study size	10	<p>Explain how the study size was arrived at <i>Feasibility and sample size</i></p> <p>To detect minimum incidence of 1% comorbidity (required for cluster analysis) with a minimum clinically important difference of hazard ratio (HR) 1.2, and 90% power of the study, the estimated sample size was 197561 for 1581 events. It was calculated using STATA, with a correlation= 0.2, standard deviation of 0.5, proportion of withdrawal= 0.20, alpha=0.05. The initial check with the registry database revealed to have minimum required sample size for the study.</p>	11-12
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Explained under each objective
Statistical methods	12	<p>(a) Describe all statistical methods, including those used to control for confounding</p> <p><i>Objective 1</i></p> <p>In the retrospective analysis the prevalence and 95% confidence interval (CI) of each specific comorbidity will be calculated separately in OA cases and matched controls using the number of people diagnosed with the comorbidity divided by the total number of OA cases or controls at the index date. The prevalence of each</p>	8

comorbidity in OA cases and matched controls will be calculated for given time intervals prior to the index date of 0-1, 0-5 and 0-10 years separately to assess observational bias (28). Discrete time intervals of 1-5, and 5-10 years before will also be used to estimate the prevalence to minimise consultation bias/misclassification bias of OA (if possible). Logistic regression will be used to calculate the odds ratios (OR) for each comorbidity unadjusted and adjusted for BMI, smoking and alcohol consumption.

Kaplan-Meier survival curves will be used to display the cumulative probability in OA and non-OA groups.

Proportional hazard assumption will be tested using Schoenfeld residual plots. The Cox regression model will be used to calculate hazard ratios (HR) for each comorbidity unadjusted and adjusted for age, sex, practice, BMI, smoking and alcohol consumption. This hybrid design has been previously used by us to examine the temporality of associations between other rheumatic musculoskeletal diseases (RMDs) (e.g., gout and lupus) and comorbidities (28,29).

Objective 2

For each dataset, an 80%: 20% split into the training and testing data will be introduced. The following analysis in objective 2 will be first employed into the training dataset and then tested its generalisability in the testing dataset. At baseline, clusters of people based on 61 comorbidities will be identified using Latent class (i.e., Gaussian mixture models algorithms of cluster) analysis (30). For each model, we will examine the association between clusters and covariates using multinomial logistic regressions. The distinctness of clusters will be examined by comparing covariates among clusters. The optimal model is the one where most clusters found in the training data are also identified independently in the testing data and clusters have most distinct patients' characteristics. We will then use both latent trajectory analysis, such as joint latent class models (31), and unsupervised machine learning approach, such as deep autoencoder or recurrent neural networks (32), to identify distinct clusters of new

8

9

	<p>comorbidity numbers development over time and their association with mortality with adjustment for baseline covariates.</p> <p><i>Objective 3</i></p> <p>For the primary analysis (initial prescriptions within year 1), a landmark analysis will be used to minimise the immortal time bias where the follow up will start after 12 months from the index date (36). Participants at risk (i.e., without a specific comorbidity of interest) at the landmark date will be followed up until the first diagnosis of the comorbidity, deregistration, or death whichever comes first. For secondary analyses, time varying covariate analysis will be used to examine the long-term, episodic/continuous use of analgesics after the index date and interaction between OA and analgesics in the development of the comorbidity. The propensity score matching or the inverse probability weighing methods will be used to adjust for confounding by indication during the follow-up as appropriate. Depending on the country-specific drug use patterns, we may modify this definition to allow for short breaks in between the episodes. Cox-regression model will be used to calculate the HR and 95% CI. We will use flexible parametric models using restricted cubic splines (developed by Lambert, “stpm2” in Stata) to estimate the HRs and differences in time to diagnosis of comorbidities (outcome) with drugs as time-varying to account for non-proportional hazards (37).</p> <p><i>Objective 4</i></p> <p>The MR-PheWAS analysis includes three steps. Firstly, we will identify the genetic variants that are associated with OA - IV. Secondly, we will undertake the PheWAS analysis – a series of case-control analyses to estimate the associations between the IV and other disease outcomes (38,40), with an adjustment for multiple testing using the false discovery rate (FDR) methods (41). Thirdly, we will implement conventional MR analysis to investigate the causal effects of the OA IV on comorbidities (39). An inverse variance weighted (IVW) method will be used to pool the associations (ORs) as appropriate (42). The MR-Egger regression analysis will be used to count for the pleiotropic effect - the effects of one genetic variant on multiple outcomes (43). The heterogeneity in dependent</p>	<p>10</p> <p>10-11</p>
--	--	------------------------

	<p>instruments (HEIDI) test will be used to exclude the cross-phenotype associations caused by genetic linkage (44).</p> <p>With the MR-PheWAS study, the OR can be interpreted as causal association. We are primarily interested in the causality from OA to comorbidities. We are also interested in inflammatory (e.g., CRP), metabolic (e.g., gut microbiome) and biomechanics (e.g., BMI) biomarkers and deep phenotypes of OA such as knee, hip, and hand OA with and without symptoms. This will be undertaken if it is feasible within 3 years of this funded project, otherwise will be considered as our future research agenda.</p> <p><i>Objective 5</i></p> <p>We will use meta-analyses (MA) to examine the variation between countries and to pool the data as appropriate. Estimates from first three objectives such as prevalence, incidence, OR, HR and 95% CI for each specific comorbidity across different populations will be distributed in a forest plot. Heterogeneity will be examined using the I^2 statistic and the Q test (45). Results will be pooled if they are homogenous based on the I^2 value using the fixed effects model, otherwise the reasons for the heterogeneity will be investigated. Random effects models will be used to pool the results if the reasons for the heterogeneity cannot be identified and if the overall pooling is appropriate. Individual patient data (IPD) meta-analysis may be used to help identify the reasons for heterogeneity (46). Common clusters and trajectories as well as burdens of comorbidities will also be compared between populations.</p>	11
	<p>(b) Describe any methods used to examine subgroups and interactions Please see method sections</p>	n/a
	<p>(c) Explain how missing data were addressed Missing data on covariates will be substituted using multiple imputation methods, provided that the data is missing at random, if applicable.</p>	7
	<p><i>(d) Cohort study</i>—If applicable, explain how loss to follow-up was addressed <i>Case-control study</i>—If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i>—If applicable, describe analytical methods taking account of sampling strategy</p>	n/a
	<p>(e) Describe any sensitivity analyses</p>	n/a

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	n/a
		(b) Give reasons for non-participation at each stage	n/a
		(c) Consider use of a flow diagram	n/a
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	n/a
		(b) Indicate number of participants with missing data for each variable of interest	n/a
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	n/a
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	n/a
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	n/a
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	n/a
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	n/a
		(b) Report category boundaries when continuous variables were categorized	n/a
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n/a
Discussion			
Key results	18	Summarise key results with reference to study objectives	n/a
Limitations	19	<p>Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias</p> <p>There are some limitations to this study. Firstly, there are inherent issues in the nature of electronic health records with respect to possible misdiagnosis, ascertainment biases, under- or over-recording, and changes in databases due to change in coding structures. Also, the analysis will be restricted to fewer covariates in some databases due to missing information on lifestyle factors such as physical activities and diet. Even though the databases have different durations of data available, if possible we will use a common follow-up time for objective five. Another important limitation is that we do not have information on quality of life and other outcomes to measure functional limitations recorded in the database.</p> <p>Chronic conditions, especially comorbidities recorded in general practices, depend on multiple factors such as population structure, health care facilities, health policies, and the nature of the national databases. A major</p>	13

		strength of this study is that it will include medical records on approximately 27 million people in four European countries. Also, the study will cover the sequence of research questions in comorbidity or multimorbidity starting from the burden through to the causality and variation. Such a research model can be used for other similar multimorbidity studies. The expected results should inform health professionals in primary care settings with respect to management of people with OA and associated comorbidities.	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	n/a
Generalisability	21	Discuss the generalisability (external validity) of the study results	n/a
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based This work was supported by Foundation for Research in Rheumatology (FOREUM) grant (2019-2022), The Swedish Research Council (2020-01103), Governmental funding of clinical research within the national health services (ALF), and The Swedish Rheumatism Association. CM is funded by the National Institute for Health Research (NIHR) Applied Research Collaboration West Midlands, the National Institute for Health Research (NIHR) School for Primary Care Research and a National Institute for Health Research (NIHR) Research Professorship in General Practice (NIHR-RP-2014-04-026)	14

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.