4 Defining comorbidity - protocol

4.1 Approach to defining comorbidities

We used concomitant medications to identify comorbid conditions.

Previous studies have also used the WHO ATC criteria to define comorbid diseases, but usually in routine healthcare data with the goal being descriptive epidemiology, as in a recent paper by Huber et al. We are not aware of any previous study which has used this approach for individual-level participant data from clinical trials, or to examine heterogeneity of treatment effects. To reduce non-differential misclassification bias, we have chosen definitions which we think favour specificity over sensitivity. (Table S4.2).

For each of the drug-based definitions in Table S4.2, reported concomitant medications were eligible if they were started at any time on or before starting the trial drug (or comparator) regardless of when they were stopped. Topical drugs are not included in the definitions except for M02 or S01E. Nor drugs with an inhaled or nebulised route of administration, except for R03 drugs.

This approach to defining comorbidities is a compromise between sensitivity and specificity and some misclassification is inevitable, reflecting the difficulty of inferring diagnoses from drug-usage. We have attempted to minimise the misclassification based on our understanding of clinical practice, with an emphasis on specificity over sensitivity.

For example, rather than assuming all participants taking a drug in the A02B class have an acid-related disorder (Table S4.2) we have limited this definition to exclude participants also taking non-steroidal drugs or any drug with anti-thrombotic actions (aspirin, antiplatelets, warfarin etc.). Similarly, we have not used aspirin to define cardiovascular disease because it is in widespread use for primary prevention.

Moreover, while the ATC system is organised around therapeutic indications, not all indications are coded. This is because “A medicinal substance can be given more than one ATC code ONLY if it is available in two or more strengths or routes of administration with clearly different therapeutic uses.” For example, finasteride is classified as a dermatological drug if low-dose and as a drug for benign prostatic hyperplasia if high-dose. Therefore, for a single strength and route of administration, there is only “one code, the main indication being decided on the basis of the available information.”

Moreover, the “main” WHO ATC indication is not necessarily the commonest indication. For example, in a US study of drug “mentions” in a representative database, >80% of mentions for gabapentin and amitriptyline were for off-label indications, predominantly pain. Similarly, in a Canadian study of antidepressant use in primary care, amitriptyline was “almost exclusively prescribed for off-label indications” most commonly for pain, insomnia, and migraine. These published findings are consistent with the clinical observation of members of our steering committee (and an independent epileptologist), that these drugs are predominantly used for pain. Despite this, the WHO ATC scheme does not include pain as an indication for these drugs, classifying gabapentin and pregabalin exclusively as antiepileptics and amitriptyline exclusively as an antidepressant.

Nor is there necessarily a code where routes/strengths do differ. For example, prochlorperazine is defined solely as an antipsychotic, despite being available in a buccal preparation for nausea. In this case the accompanying note states that “The substances in this group are sometimes used for other indications in much lower doses”.

We had initially planned to add skin disease to the list of diagnoses, however we found that topical therapies were very poorly recorded in the trial data and so have opted to drop this from the comorbid disease definitions.
A tabular summary of the drug-based comorbidity definitions given in Table 4.2, the definitive description is contained in the R code R code for comorbid disease definitions.

### 4.2 Incomplete ATC coding

An additional complexity is caused by the fact that for certain trials, sponsors have only provided less specific codes (eg 3 or 4-character codes) and not 5-character codes which uniquely identify each class (7-character codes identifying each agent). Where this is the case, but the drug name with or without route and indication information have been provided, we assigned each potentially-relevant drug to a WHO ATC code using the US Government drug meta-thesaurus (RXNORM). Where neither the drug name, nor sufficiently detailed drug-class information is provided, we adopted a workaround suited to each definition (Table S4.2). This had a limited impact on the overall comorbidity totals, and only applied to inflammatory, pain, urological and erectile definitions (4.1). In the case of pain and inflammatory definitions, the broader categorisation was used. For urological and erectile definitions, the narrower categorisation was used.

#### Table S4.1: Proportion of participants with definition met on basis of 3/4 character ATC code

<table>
<thead>
<tr>
<th>Condition</th>
<th>YODA</th>
<th>CSDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trials</td>
<td>37</td>
<td>82</td>
</tr>
<tr>
<td>Pain</td>
<td>253 (5.7%)</td>
<td>832 (5%)</td>
</tr>
<tr>
<td>Inflammatory</td>
<td>148 (77.5%)</td>
<td>308 (8.1%)</td>
</tr>
<tr>
<td>Urological</td>
<td>40 (21.6%)</td>
<td>156 (12.7%)</td>
</tr>
<tr>
<td>Erectile</td>
<td>40 (27.3%)</td>
<td>181 (12.5%)</td>
</tr>
</tbody>
</table>

#### Table S4.2: Comorbidity Definitions

<table>
<thead>
<tr>
<th>Condition</th>
<th>ATC codes</th>
<th>ATC label</th>
<th>Exceptions/more specific definitions/notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid related disorders</td>
<td>A02A</td>
<td>ANTACIDS</td>
<td>Exclude where also taking M01A ANTIINFLAMMATORY AND ANTIRHEUMATIC PRODUCTS, NON-STEROIDS, or B01 (antithrombotic) drugs Note this includes insulins and analogues, other blood glucose lowering drugs etc. It does not include cardiovascular prevention drugs We do not exclude metformin, although this is used to treat Polycystic ovary syndrome (PCOS).</td>
</tr>
<tr>
<td></td>
<td>A02B</td>
<td>DRUGS FOR ACID RELATED DISORDERS</td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>A10</td>
<td>DRUGS USED IN DIABETES</td>
<td>Do not include if only 4-level codes are available.</td>
</tr>
<tr>
<td>Thromboembolic disease, atrial fibrillation or valvular heart disease</td>
<td>B01AA</td>
<td>Vitamin K antagonists</td>
<td>Do not include if only 4-level codes are available.</td>
</tr>
<tr>
<td></td>
<td>B01AE</td>
<td>Direct thrombin inhibitors</td>
<td>Do not include if only 4-level codes are available.</td>
</tr>
<tr>
<td></td>
<td>B01AF</td>
<td>Direct factor Xa inhibitors</td>
<td>Do not include if only 4-level codes are available.</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>C01</td>
<td>CARDIAC THERAPY DRUGS</td>
<td></td>
</tr>
<tr>
<td>Condition/ diseases</td>
<td>ATC codes</td>
<td>ATC label</td>
<td>Exceptions/more specific definitions/notes</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Urinary frequency and incontinence</td>
<td>C04</td>
<td>PERIPHERAL VASODILATORS</td>
<td></td>
</tr>
<tr>
<td>Erectile dysfunction</td>
<td>C02</td>
<td>ANTIHYPERTENSIVES</td>
<td></td>
</tr>
<tr>
<td>Benign prostatic hypertrophy</td>
<td>C07</td>
<td>beta-Adrenergic Blocking Agents</td>
<td></td>
</tr>
<tr>
<td>Glaucoma</td>
<td>C08</td>
<td>CALCIUM CHANNEL BLOCKERS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C09</td>
<td>AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM</td>
<td></td>
</tr>
<tr>
<td>Urinary frequency and incontinence</td>
<td>G04BD</td>
<td>Drugs for urinary frequency and incontinence</td>
<td>If only-4-character code exclude. When only 4-character code is provided, define as eye disease. If only 3-character code exclude. Accept any A07 for indications associated with impaired copper metabolism, e.g., Wilson's disease.</td>
</tr>
<tr>
<td>Erectile dysfunction</td>
<td>G04BE</td>
<td>Drugs used in erectile dysfunction</td>
<td>If only-4-character code exclude. When only 4-character code is provided, define as eye disease. If only 3-character code exclude. Accept any A07 for indications associated with impaired copper metabolism, e.g., Wilson's disease.</td>
</tr>
<tr>
<td>Benign prostatic hypertrophy</td>
<td>G04C</td>
<td>DRUGS USED IN BENIGN PROSTATIC HYPERTROPHY</td>
<td>If only-3-character code exclude. Accept any A07 for indications associated with impaired copper metabolism, e.g., Wilson's disease.</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>S01E</td>
<td>ANTIGLAUCOMA PREPARATIONS AND MIOTICS</td>
<td>If only-3-character code exclude. Accept any A07 for indications associated with impaired copper metabolism, e.g., Wilson's disease.</td>
</tr>
<tr>
<td>Arthritis and arthralgia</td>
<td>M01A</td>
<td>ANTIINFLAMMATORY AND ANTI RHEUMATIC PRODUCTS, NON-Steroids</td>
<td>If only-3-character code define as arthritis and arthralgia, but some misclassification possible as indications for penicillamine include conditions associated with impaired copper metabolism, e.g., Wilson's disease.</td>
</tr>
<tr>
<td>Osteoporosis (or risk factors for osteoporosis)</td>
<td>M01B</td>
<td>ANTIINFLAMMATORY/ANTIRHEUMATIC AGENTS IN COMBINATION</td>
<td>If only-3-character code define as arthritis and arthralgia, but some misclassification possible as indications for penicillamine include conditions associated with impaired copper metabolism, e.g., Wilson's disease.</td>
</tr>
<tr>
<td></td>
<td>M02</td>
<td>ANTIINFLAMMATORY/ANTIRHEUMATIC AGENTS IN COMBINATION</td>
<td>If only-3-character code define as arthritis and arthralgia, but some misclassification possible as indications for penicillamine include conditions associated with impaired copper metabolism, e.g., Wilson's disease.</td>
</tr>
<tr>
<td>Gout</td>
<td>M05</td>
<td>DRUGS FOR TREATMENT OF BONE DISEASES</td>
<td>Although allopurinol is being used for other indications, this is unlikely to be widespread.</td>
</tr>
<tr>
<td>Inflammatory arthropathies, inflammatory bowel disease, systemic lupus</td>
<td>M04</td>
<td>ANTIGOUT PREPARATIONS</td>
<td>Where only 3-character codes are provided, define as any A07.</td>
</tr>
<tr>
<td>Condition</td>
<td>ATC codes</td>
<td>ATC label</td>
<td>Exceptions/more specific definitions/notes</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>erythematous and connective tissue diseases</td>
<td>A07EC</td>
<td>Aminosalicylic acid and similar agents</td>
<td>Where only 3-character codes are provided, define as any A07</td>
</tr>
<tr>
<td></td>
<td>L04AB</td>
<td>Tumour necrosis factor alpha (TNF-) inhibitors</td>
<td>Where only 3-character codes are provided, define as any L04</td>
</tr>
<tr>
<td></td>
<td>L04AA</td>
<td>Selective immunosuppressants</td>
<td>Where only 3-character codes are provided, define as any L04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other immunosuppressants (includes L04AX methotrexate, azathioprine, and leflunomide)</td>
<td>Where only 3-character codes are provided, define as any L04</td>
</tr>
<tr>
<td></td>
<td>M01CB</td>
<td>Gold preparations</td>
<td>Do not define if only 3-character code is available. If only 4-character code available define, since only other agent is an obscure drug oxycinchofen</td>
</tr>
<tr>
<td></td>
<td>M01CC</td>
<td>Penicillamine and similar agents</td>
<td>Do not define if only 3-character code is available. If only 4-character code available define, since only other agent is an obscure drug oxycinchofen</td>
</tr>
<tr>
<td>D05</td>
<td>ANTIPSORIATICS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migraine</td>
<td>N02C</td>
<td>ANTIMIGRAINE PREPARATIONS</td>
<td>Do not define if only 3-character code is available</td>
</tr>
<tr>
<td>Pain</td>
<td>N02A</td>
<td>OPIOIDS</td>
<td>If only 3-character code available define as pain</td>
</tr>
<tr>
<td></td>
<td>N02B</td>
<td>OTHER ANALGESICS AND ANTIPYRETICS</td>
<td></td>
</tr>
<tr>
<td>Schizophrenia and delusional disorders</td>
<td>N05A</td>
<td>ANTIPSYCHOTICS</td>
<td>Prochlorperazine is included in this class. If individual drug data available exclude if prochlorperazine. If 5-character code available exclude N05AB Phenothiazines with piperazine structure?. If only 4-character code is available do not define.</td>
</tr>
<tr>
<td>Mood, neurotic and sleep disorders</td>
<td>N05B</td>
<td>ANXIOLYTICS</td>
<td>If only 3-character code available do not define.</td>
</tr>
<tr>
<td></td>
<td>N05C</td>
<td>HYPNOTICS AND SEDATIVES</td>
<td>If only 3-character code available do not define.</td>
</tr>
<tr>
<td></td>
<td>N06A</td>
<td>ANTIDEPRESSANTS</td>
<td>Except amitriptyline. If drug term not available</td>
</tr>
</tbody>
</table>
Table S4.2: Comorbidity Definitions

<table>
<thead>
<tr>
<th>Condition</th>
<th>ATC codes</th>
<th>ATC label</th>
<th>Exceptions/more specific definitions/notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epilepsy</td>
<td>N03</td>
<td>ANTIEPILEPTICS</td>
<td>Except where drug is pregabalin, gabapentin or valproic acid. If specific drug is not given, and if indication for drug is not stated proceed as follows. If only a 5-character code is provided exclude N03AX (which includes gabapentin and pregabalin) and N03AG (includes valproic acid). This will reduce sensitivity, but improve specificity. If only a 4-character code is provided do not attempt to define.</td>
</tr>
<tr>
<td>Parkinson’s disease and Parkinsonism</td>
<td>N04</td>
<td>ANTI-PARKINSON DRUGS</td>
<td></td>
</tr>
<tr>
<td>Dementia</td>
<td>N06D</td>
<td>ANTI-DEMENTIA DRUGS</td>
<td>Do not define if only 3-character code is available</td>
</tr>
<tr>
<td>Chronic lower respiratory disease</td>
<td>R03</td>
<td>DRUGS FOR OBSTRUCTIVE AIRWAY DISEASES</td>
<td></td>
</tr>
<tr>
<td>Thyroid disorders</td>
<td>H03</td>
<td>THYROID THERAPY DRUGS</td>
<td></td>
</tr>
<tr>
<td>Skin diseases</td>
<td>D02A</td>
<td>EMOLLIENTS AND PROTECTIVES, ANTI-PRURITICS, INCL.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D04</td>
<td>ANTIHISTAMINES, ANESTHETICS, ETC., ANTIBIOTICS AND</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D06</td>
<td>CHEMOTHERAPEUTICS FOR DERMATOLOGICAL USE, CORTICOSTEROIDS,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D07</td>
<td>DERMATOLOGICAL PREPARATIONS</td>
<td></td>
</tr>
</tbody>
</table>

4.3 Mapping ATC codes to READ codes

We mapped the ATC codes to Read codes (which are used in the SAIL data) off-line, using the NHS Business Authority mappings and, for some more recent drugs such as novel antidiabetic drugs, by manually mapping between ATC and Read codes. The mapping was very good, even retaining information on route. For example the READ code for topical beclomethasone preparations mapped to different ATC codes to those for oral preparations.

4.4 Suppression of comorbid conditions

For all three data sources, comorbid diseases were excluded if these were considered to be identical to the main condition (Table S4.3). For example, a patient/participant with asthma could not be considered to have airways disease as a comorbidity. This inevitably involved clinical judgements, for example pain was not suppressed for rheumatoid arthritis as we considered that where inflammation was fully controlled, pain may not be present, while the patient would nonetheless have clear evidence of the disease. In contrast, given that osteoarthritis is essentially a degenerative process, we
concluded that the diagnosis was of doubtful validity in the absence of at least some pain. These exclusions were nonetheless applied identically across all three cohorts.

Table S4.3: Indication/condition pairs where the condition is not considered a comorbidity

<table>
<thead>
<tr>
<th>Indication</th>
<th>Not defined as comorbid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s Disease</td>
<td>Dementia</td>
</tr>
<tr>
<td>Alzheimer’s Disease</td>
<td>Schizophrenia and delusional disorders</td>
</tr>
<tr>
<td>Ankylosing Spondylitis</td>
<td>Inflammatory arthropathies, inflammatory bowel disease,</td>
</tr>
<tr>
<td></td>
<td>systemic lupus erythematosus and connective tissue diseases</td>
</tr>
<tr>
<td>Ankylosing Spondylitis</td>
<td>Arthritis and arthralgia</td>
</tr>
<tr>
<td>Asthma</td>
<td>Chronic lower respiratory disease</td>
</tr>
<tr>
<td>Atrial Fibrillation, Stroke</td>
<td>Cardiovascular diseases</td>
</tr>
<tr>
<td>Atrial Fibrillation, Stroke</td>
<td>Thromboembolic disease, atrial fibrillation or valvular heart disease</td>
</tr>
<tr>
<td>Benign Prostatic Hyperplasia</td>
<td>Benign prostatic hypertrophy</td>
</tr>
<tr>
<td>Benign Prostatic Hyperplasia</td>
<td>Urinary frequency and incontinence</td>
</tr>
<tr>
<td>Chronic Idiopathic Urticaria (Ciu)</td>
<td>Inflammatory arthropathies, inflammatory bowel disease,</td>
</tr>
<tr>
<td></td>
<td>systemic lupus erythematosus and connective tissue diseases</td>
</tr>
<tr>
<td>Diabetes Mellitus</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Diabetes Mellitus, Type 2</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Diabetes Mellitus, Type 2; Hypertension</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Diabetes Mellitus, Type 2; Hypertension</td>
<td>Hypertension</td>
</tr>
<tr>
<td>Diabetes Mellitus, Type 2; Renal Insufficiency</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Diabetes Mellitus, Type 2; Renal Insufficiency</td>
<td>Renal</td>
</tr>
<tr>
<td>Erectile Dysfunction, Benign Prostatic Hyperplasia</td>
<td>Erectile dysfunction</td>
</tr>
<tr>
<td>Erectile Dysfunction, Benign Prostatic Hyperplasia</td>
<td>Benign prostatic hypertrophy</td>
</tr>
<tr>
<td>Erectile Dysfunction, Benign Prostatic Hyperplasia</td>
<td>Urinary frequency and incontinence</td>
</tr>
<tr>
<td>Hypertension, Pulmonary</td>
<td>Thromboembolic disease, atrial fibrillation or valvular heart disease</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>Osteoporosis (or risk factors for osteoporosis)</td>
</tr>
<tr>
<td>Osteoporosis, Male</td>
<td>Osteoporosis (or risk factors for osteoporosis)</td>
</tr>
<tr>
<td>Osteoporosis; Hip Fracture</td>
<td>Osteoporosis (or risk factors for osteoporosis)</td>
</tr>
<tr>
<td>Parkinson Disease</td>
<td>Parkinsons disease and Parkinsonism</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>Inflammatory arthropathies, inflammatory bowel disease,</td>
</tr>
<tr>
<td></td>
<td>systemic lupus erythematosus and connective tissue diseases</td>
</tr>
<tr>
<td>Pulmonary Disease, Chronic Obstructive</td>
<td>Chronic lower respiratory disease</td>
</tr>
<tr>
<td>Rheumatoid Arthritis</td>
<td>Inflammatory arthropathies, inflammatory bowel disease,</td>
</tr>
<tr>
<td></td>
<td>systemic lupus erythematosus and connective tissue diseases</td>
</tr>
<tr>
<td></td>
<td>Arthritis and arthralgia</td>
</tr>
<tr>
<td>Systemic Lupus Erythematosus</td>
<td>Inflammatory arthropathies, inflammatory bowel disease,</td>
</tr>
<tr>
<td></td>
<td>systemic lupus erythematosus and connective tissue diseases</td>
</tr>
<tr>
<td></td>
<td>Thromboembolic disease, atrial fibrillation or valvular heart disease</td>
</tr>
<tr>
<td>Thromboembolism</td>
<td>Thromboembolic disease, atrial fibrillation or valvular heart disease</td>
</tr>
<tr>
<td>Thromboprophylaxis</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Type 2 Diabetes Mellitus</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Ulcerative Colitis; Crohn’s Disease</td>
<td>Inflammatory arthropathies, inflammatory bowel disease,</td>
</tr>
</tbody>
</table>
Table S4.3: Indication/condition pairs where the condition is not considered a comorbidity

<table>
<thead>
<tr>
<th>Indication</th>
<th>Not defined as comorbid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venous Thromboembolism</td>
<td>systemic lupus erythematosus and connective tissue diseases</td>
</tr>
<tr>
<td></td>
<td>Thromboembolic disease, atrial fibrillation or valvular heart disease</td>
</tr>
<tr>
<td>Crohn’s Disease</td>
<td>Inflammatory arthropathies, inflammatory bowel disease, systemic lupus erythematosus</td>
</tr>
<tr>
<td></td>
<td>and connective tissue diseases</td>
</tr>
<tr>
<td>Ulcerative Colitis</td>
<td>Inflammatory arthropathies, inflammatory bowel disease, systemic lupus erythematosus</td>
</tr>
<tr>
<td></td>
<td>and connective tissue diseases</td>
</tr>
<tr>
<td>Psoriatic Arthritis</td>
<td>Inflammatory arthropathies, inflammatory bowel disease, systemic lupus erythematosus</td>
</tr>
<tr>
<td></td>
<td>and connective tissue diseases</td>
</tr>
<tr>
<td>Psoriatic Arthritis</td>
<td>Arthritis and arthralgia</td>
</tr>
<tr>
<td>Migraine</td>
<td>Pain</td>
</tr>
<tr>
<td>Migraine</td>
<td>Migraine</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>Pain</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>Arthritis and arthralgia</td>
</tr>
<tr>
<td>Restless Legs Syndrome</td>
<td>Parkinsons disease and Parkinsonism</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Hypertension</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Cardiovascular diseases</td>
</tr>
<tr>
<td>Diabetic Nephropathies</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Diabetic Nephropathies</td>
<td>Renal</td>
</tr>
<tr>
<td>Arthroplasty, Replacement, Knee; Thromboembolism</td>
<td>Pain</td>
</tr>
<tr>
<td>Thromboembolism; Arthroplasty, Replacement, Hip</td>
<td>Pain</td>
</tr>
<tr>
<td>Diabetes Mellitus, Type 2; Hyperglycemia</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Rhinitis, Allergic, Perennial</td>
<td>Chronic lower respiratory disease</td>
</tr>
</tbody>
</table>

4.5 Comorbidity counts

For the comorbidity count calculation, these concomitant medication definition pairs were also collapsed into a single condition (Table S4.4).

Table S4.4: Conditions as recorded on CSDR site

<table>
<thead>
<tr>
<th>Definition 1</th>
<th>Definition 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td>Migraine</td>
</tr>
<tr>
<td>Pain</td>
<td>Rheumatologic conditions</td>
</tr>
</tbody>
</table>

4.6 R code for comorbid disease definitions

Comorbid diseases were implemented, consistently across all datasets using the following R code.

```R
library(tidyverse)
library(stringr)

# read in conmed data
conmed <- readRDS("Data/conmed_randomised_cleaned_no_contradictions.Rds") %>%
  filter(!is.na(drug_class)) %>%
  distinct()
rxnorm_bnf <- read_csv("Supporting/bnf_rxnorm_atc_codes.zip")
rxnorm_bnf <- rxnorm_bnf %>%
  filter(!is.na(str))

##### Functions
PrintDrugChoices <- function (incld, excld = FALSE, mydf = conmed){
```
mydf2 <- mydf %>%
  mutate(incl = incld, excl = excld) %>%
  group_by(trial, id) %>%
  mutate(excl = any(excld)) %>%
  ungroup() %>%
  filter(incl & !excl) %>%
  group_by(term, route_classify, atc_code) %>%
  count(sort = TRUE)
mydf2

ApplyMedCriteria <- function(incl, excl = FALSE, print = TRUE, mydf = conmed){
  PrintDrugChoices(incl, excl, mydf) %>%
  head(19) %>%
  print()

  mydf2 <- mydf %>%
  mutate(incl = incld, excl = excld) %>%
  group_by(trial, id) %>%
  summarise(present = any(incl) & !any(excl))
  print(paste0(round(100 * mean(mydf2$present), 1), "%"))
mydf2
}

## Rename variables in dataset so do not have to rename code
conmed <- conmed %>%
  rename(atc_code = drug_class,
         trial = trial_id_trunc,
         route_classify = route)

############ Define comorbidities based on concomitant medications
## First remove aspirin from all analyses as is used widely as prophylaxis
## Similarly remove amitriptyline as most frequently used for PAIN not for depression, anxiety etc
## Updated this code to be more specific, particularly for aspirin

DropDrugs <- function(mystring = "acetylsalicylic acid|aspirin") {
  x <- rxnorm_bnf %>%
    filter(str_detect(str, mystring))
  x <- rxnorm_bnf %>%
    filter(code %in% x$code | str %in% x$str) %>%
    transmute(term_lower = str, atc_code = str_sub(code, 1, 5))
  x
}

aspirin_codes <- c("A01DA05", "B01AC06", "N02BA01")
aspirin <- rxnorm_bnf %>%
  filter(code %in% aspirin_codes) %>%
  mutate(term_lower = str_to_lower(str)) %>%
  distinct(term_lower, code)
amitriptyline <- DropDrugs("amitriptyl")
pre_gab_val <- DropDrugs("gabapentin|pregabalin|valrpoate|valproic acid")

conmed <- conmed %>%
  mutate(term_lower = str_to_lower(term)) %>%
  str_replace("/\[0-9]{8,8}/", "") %>%
  str_trim() %>%
  anti_join(aspirin %>%
            select(term_lower))

conmed <- conmed %>%
  anti_join(amitriptyline %>%
            select(term_lower))

conmed <- conmed %>%
  anti_join(pre_gab_val %>%
            select(term_lower))

## Antacids
# Note only antacid codes are A02A, A02B, A02X and the last is an empty category
antacids_included <- conmed$atc_code %>%
  str_sub(1, 4) %in% c("A02A", "A02B")
  conmed$atc_code %>%
  str_sub(1, 3) == "A02"
antacids_excluded <- (conmed$atc_code %>% str_sub(1, 4) %in% "M01A") |
  (conmed$atc_code %>% str_sub(1, 3) %in% "B01")
conmed_ant <- ApplyMedCriteria(antacids_included, antacids_excluded)

## Diabetes
diabetes_included <- conmed$atc_code %>% str_sub(1, 3) %in% c("A10")
conmed_diab <- ApplyMedCriteria(diabetes_included)

## Thromboembolic
tbe_included <- str_sub(conmed$atc_code, 1, 5) %in% c("B01AA", "B01AE", "B01AF")
conmed_tbe <- ApplyMedCriteria(tbe_included)

## Cardiovascular (CV)
cv_codes <- paste0("C0", c(1, 2, 4, 7, 8, 9))
cv_included <- str_sub(conmed$atc_code, 1, 3) %in% cv_codes
cv_excluded <- str_sub(conmed$atc_code, 1, 4) == "N02C" &
  str_sub(conmed$atc_code, 1, 5) == "C07AA"
# Where has only 4 digits cannot exclude beta-blockers, where only has 3 cannot exclude
  antimigraine
conmed_cv <- ApplyMedCriteria(cv_included, cv_excluded)

## Urinary incontinence
ur_included <- str_sub(conmed$atc_code, 1, 5) == "G04BD"
conmed_ur <- ApplyMedCriteria(ur_included)

## Erectile dysfunction (ED)
ed_included <- str_sub(conmed$atc_code, 1, 5) == "G04BE"
conmed_ed <- ApplyMedCriteria(ed_included)

## Urinary incontinence or ED
## If only 4 character code is available define as urinary incontinence or ED
ur_ed_included <- str_sub(conmed$atc_code, 1, 4) == "G04B" &
  str_length(conmed$atc_code) == 4
conmed_ur_ed <- ApplyMedCriteria(ur_ed_included)

## Benign prostatic hypertrophy (BPH)
bph_included <- str_sub(conmed$atc_code, 1, 4) == "G04C"
conmed_bph <- ApplyMedCriteria(bph_included)

## Urinary incontinence or ED or BPH, these are tiny proportions, probably ignore
## If only 3 character code is available define as urinary incontinence or ED or BPH
ur_ed_bph_included <- str_sub(conmed$atc_code, 1, 3) == "G04" &
  str_length(conmed$atc_code) == 3
conmed_ur_ed_bph <- ApplyMedCriteria(ur_ed_bph_included)

## Glaucoma
# drop isosorbide even though classified as an eye drug as is not a contemporary drug for glaucoma

glaucoma <- str_sub(conmed$atc_code, 1, 4) == "S01E" &
  (is.na(conmed$term) | conmed$term != "ISOSORBIDE")
conmed_gl <- ApplyMedCriteria(glaucoma)

## Arthritis and arthralgia
# Exclude FOLIC ACID from this as is very common and (despite class here)
# is not an M01AX code in WHO ATC
art <- ((str_sub(conmed$atc_code, 1, 4) %in% c("M01A", "M01B") |
  str_sub(conmed$atc_code, 1, 3) == "M02") &
  (is.na(conmed$term) |
  conmed$term != "FOLIC ACID")
conmed_art <- ApplyMedCriteria(art)

## Osteoporosis
ost <- str_sub(conmed$atc_code, 1, 3) == "M05"
conmed_ost <- ApplyMedCriteria(ost)

## Gout
gou <- str_sub(conmed$atc_code, 1, 3) == "M04"
conmed_gou <- ApplyMedCriteria(gou)

## Inflammatory arthropathies, psoriasis, inflammatory bowel disease or connective tissue diseases
inf <- str_sub(conmed$atc_code, 1, 5) %in% c("A07EA", "A07EC", "L04AB", "L04AA",  
                  "L04AX", "M01CB", "M01CC")
conmed_inf <- ApplyMedCriteria(inf)

inf3 <- (!inf) & (str_length(conmed$atc_code == 3) &  
                  str_sub(conmed$atc_code, 1, 3) %in% c("A07", "L04", "D05"))
inf4 <- (!inf) & (str_length(conmed$atc_code == 4) &  
                  str_sub(conmed$atc_code, 1, 4) %in% c("M01C"))
conmed_inf4 <- ApplyMedCriteria(inf4 | inf3)

## Migraine
mig <- str_sub(conmed$atc_code, 1, 4) == "N02C"
conmed_mig <- ApplyMedCriteria(mig)

## Pain
pai <- str_sub(conmed$atc_code, 1, 4) %in% c("N02A", "N02B") &  
      (!str_sub(conmed$atc_code, 1, 5) == "N02BA")
# Already excluded aspirin above
conmed_pai <- ApplyMedCriteria(pai)

pai3 <- !pai & (str_length(conmed$atc_code == 3) &  
                  str_sub(conmed$atc_code, 1, 3) == "N02")
# already excluded aspirin
conmed_pai3 <- ApplyMedCriteria(pai3)

## schizophrenia and delusional disorders
sch_include <- str_sub(conmed$atc_code, 1, 4) == "N05A"

sch_exclude <- str_sub(conmed$atc_code, 1, 5) == "N05B"
conmed_sch <- ApplyMedCriteria(sch_include, sch_exclude)

## Anxiety and mood disorders
anx <- str_sub(conmed$atc_code, 1, 4) %in% c("N05B", "N05A", "N06A")
# already excluded amitriptyline
conmed_anx <- ApplyMedCriteria(anx)

## Epilepsy
epi <- str_sub(conmed$atc_code, 1, 3) == "N03"
# Already excluded GABAPENTIN, PREGABALIN AND VALPROATE terms
# Will also exclude 5-digit code when text is missing
pre_gab_val5 <- str_length(conmed$atc_code == 5) & is.na(conmed$term) & conmed$atc_code == "N03AX"
epi_excld <- pre_gab_val5
conmed_epi <- ApplyMedCriteria(epi, epi_excld)

## Parkinson's disease and Parkinsonism
pd <- str_sub(conmed$atc_code, 1, 3) == "N04"
conmed_pd <- ApplyMedCriteria(pd)

## Dementia
dem <- str_sub(conmed$atc_code, 1, 4) == "N06D"
conmed_dem <- ApplyMedCriteria(dem)

## Chronic lower respiratory disease (predominantly asthma and/or COPD)
resp <- str_sub(conmed$atc_code, 1, 3) == "R03"
conmed_resp <- ApplyMedCriteria(resp)

## Thyroid disease (hyper and hypothyroidism included)
thy <- str_sub(conmed$atc_code, 1, 3) == "H03"
conmed_thy <- ApplyMedCriteria(thy)

## Skin diseases
skn <- str_sub(conmed$atc_code, 1, 4) == "D02A" |  
      str_sub(conmed$atc_code, 1, 3) %in% c("D04", "D06", "D07")
conmed_skn <- ApplyMedCriteria(skn)

## Combine all conditions into a single dataset

conmed_all <- map(a, get) %>% distinct(company, trial, id) %>% nrow()

names(conmed_all) <- str_replace(a, "conmed_", "")

conmed_all <- map2(conmed_all, names(conmed_all), ~ set_names(.x, c("trial", "id", .y)))

conmed_all <- reduce(conmed_all, inner_join)

## Create more informative labels for conditions

b_lbl <- c('antacids', 'anxiety', 'arthritis', 'prostate', 'CV', 'dementia', 'diabetes', 'erectile', 'epilepsy', 'glaucoma', 'gout', 'inflammatory', 'inflammatory4', 'migraine', 'ostoporosis', 'pain', 'pain3', 'parkinsons', 'asthma COPD', 'schizophrenia', 'thromboembolic', 'thyroid', 'urological', 'urological or ed', 'urological or ed or bph', 'skin')

smrs_all <- map(b, function(x) tapply(conmed_all[[x]], conmed_all$trial, mean))

names(smrs_all) <- b_lbl

b_lkp <- b_lbl

names(b_lkp) <- b

other_names <- setdiff(names(conmed_all), names(b_lkp))

b_lkp <- c(other_names, b_lkp)

names(conmed_all) <- b_lkp[names(conmed_all)]

conmed_all <- trial_indic_drug %>%
  select(nct_id, medicine, condition, trial) %>%
  inner_join(conmed_all)

## Collapse additional conditions
## Clearly very uncommon to have urological 3-level codes, so drop for simplicity
conmed_all <- conmed_all %>%
  mutate(pain = pain|pain3, inflammatory = inflammatory|inflammatory4) %>%
  select(-pain3, -inflammatory4, -urological_or_ed, -urological_or_ed_or_bph)

## set condition to null where it corresponds to the indication condition
## Rename and update now adding in YODA trials
## Solely YODA condition terms are "Crohn's disease", "Ulcerative colitis", "Psoriatic arthritis", # and "Migraine" the others appear in CSDR

condition_match <- c("Alzheimer's Disease", "ankylosing spondylitis", "Asthma", "Atrial Fibrillation, Stroke", "Benign Prostatic Hyperplasia", "Chronic Idiopathic Urticaria (CIU)", "Diabetes Mellitus", "Diabetes Mellitus, Type 2", "Diabetes Mellitus, Type 2; Hypertension", "Diabetes Mellitus, Type 2; Renal Insufficiency", "Erectile Dysfunction, Benign Prostatic Hyperplasia", "Hypertension, Pulmonary", "Osteoporosis", "Osteoporosis, Male", "Osteoporosis; Hip Fracture", ...)
conmed_all <- conmed_all %>%
  mutate(condition = case_when(
    condition == "Rheumatoid arthritis" ~ "rheumatoid arthritis",
    condition == "Ankylosing spondylitis" ~ "ankylosing spondylitis",
    condition %in% 
      c("Type 2 diabetes", "Diabetes Mellitus, Type 2; Hyperglycemia") ~ "Diabetes Mellitus, Type 2; Hyperglycemia",
    condition == "Alzheimer's" ~ "Alzheimer's Disease",
    TRUE ~ condition
  ))

names(condition_match) <- condition_match
condition_match <- as.list(condition_match)
condition_match$'Alzheimer's Disease' <- c("dementia", "schizophrenia")
condition_match$'Asthma' <- "asthma_COPD"
condition_match$'Atrial Fibrillation, Stroke' <- c("CV", "thromboembolic")
condition_match$'Benign Prostatic Hyperplasia' <- c("prostate", "urological")
condition_match$'Chronic Idiopathic Urticaria (CIU)' <- "inflammatory"
condition_match$'Diabetes Mellitus' <- "diabetes"
condition_match$'Diabetes Mellitus, Type 2' <- "diabetes"
condition_match$'Diabetes Mellitus, Type 2; Hypertension' <- c("diabetes", "hypertension")
condition_match$'Diabetes Mellitus, Type 2; Renal Insufficiency' <- c("diabetes", "renal")
condition_match$'Erectile Dysfunction, Benign Prostatic Hyperplasia' <- c("erectile", "prostate", "urological")
condition_match$'Hypertension, Pulmonary' <- "thromboembolic"
condition_match[c("Osteoporosis", "Osteoporosis, Male", "Osteoporosis; Hip Fracture")] <- "osteoarthritis"
condition_match$'Parkinson Disease' <- "parkinsons"
condition_match$'Psoriasis' <- c("inflammatory", "skin")
condition_match$'Pulmonary Disease, Chronic Obstructive' <- "asthma_COPD"
condition_match[c("rheumatoid arthritis", "ankylosing spondylitis", "Psoriatic arthritis", "rheumatoid arthritis", "Psoriatic arthritis")] <- map(
  function(x) c("inflammatory", "arthritis"))
condition_match$'Systemic Lupus Erythematosus' <- "inflammatory"
condition_match$'Thromboprophylaxis' <- "thromboembolic"
condition_match$'Type 2 Diabetes Mellitus' <- "diabetes"
condition_match[c("Ulcereative Colitis; Crohn's Disease", "Crohn's disease", "Ulcereative colitis")] <- "inflammatory"
condition_match$'Venous Thromboembolism' <- "thromboembolic"
condition_match$'Migraine' <- c("pain", "migraine")
condition_match$'Osteoarthritis' <- c("pain", "arthritis")
condition_match$'Restless Legs Syndrome' <- c("parkinsons")
condition_match$Hypertension <- c("hypertension", "CV")
condition_match$`Diabetic Nephropathies` <- c("diabetes", "renal")
condition_match$Arthroplasty, Replacement, Knee; Thromboembolism` <- c("pain")
condition_match$Thromboembolism; Arthroplasty, Replacement, Hip` <- c("pain")
condition_match$Diabetes Mellitus, Type 2; Hyperglycemia` <- "diabetes"
condition_match$`Rhinitis, Allergic, Perennial` <- "asthma_COPD"

saveRDS(condition_match, "Scratch_data/index_conditions_matching_comorbidities.Rds")

## For sail create as dataframe
# condition_match_df <- stack(condition_match)
# write_csv(condition_match_df, "Outputs/index_comorbid_suppress.csv")

## Apply exclusion by setting each concomitant disease to FALSE if is also the indication
conmed_all_final <- conmed_all
for(condition_slct in names(condition_match)) {
  print(paste0(condition_slct, " - ", condition_match[[condition_slct]]))
  conmed_all_final[conmed_all_final$condition %in% condition_slct, names(conmed_all_final2) %in% condition_match[[condition_slct]]] <- FALSE
}
conmed_all_final2 <- conmed_all_final
rm(conmed_all_final2)
saveRDS(conmed_all_final, "Data/comorbidity_based_on_conmeds.Rds")

## Count conditions after have excluded index disease
# Avoid double-counting of pain
# Drop skin as concerns about completeness of recording
conmed_all_final_count <- conmed_all_final2
mutate(skin = FALSE,
       pain = if_else(migraine|arthritis, FALSE, pain),
       arthritis = if_else(inflammatory, FALSE, arthritis))

## Create a disease count variable for each participant
conmed_all_final_count <- conmed_all_final_count
mutate(disease_count = NA)
conmed_all_final_count$disease_count <- rowSums(conmed_all_final_count %>%
  ungroup() %>%
  select(antacids:urological))
a <- table(conmed_all_final_count$disease_count)
round(100*cumsum(rev(a))/nrow(conmed_all_final_count),2)

## Aggregate the disease count for each trial
conmed_all_final_count_agg <- conmed_all_final_count
mutate(group_by(nct_id, disease_count) %>%
       summarise(x = length(disease_count)) %>%
       group_by(nct_id) %>%
       mutate(n = sum(x)) %>%
       ungroup() %>%
       mutate(prop = round(100*x/n,1)))

## Spread this to wide for easier examination
conmed_all_final_count_agg_wide <- conmed_all_final_count_agg
select(nct_id, n, disease_count, prop) %>%
  spread(disease_count, prop, fill = 0)

## Summarise the proportion of participants with each comorbidity within each trial
conmed_all_trial_lvl <- conmed_all_final
select(-id) %>%
  group_by(trial, nct_id, medicine, condition) %>%
  summarise_all(function(x) round(100*mean(x),1))
write_csv(conmed_all_trial_lvl, "Outputs/Trial level counts of people with each disease based on concomittant medicines.csv")
write_csv(conmed_all_final_count_agg_wide,
"Outputs/Trial level counts of people with disease counts based on concomitant medicines.csv"