# Assessing trial representativeness using Serious Adverse Events: An observational analysis using aggregate and individual-level data from clinical trials and routine healthcare data


**Authors**

Peter Hanlon[1], Elaine Butterly[1], Anoop SV Shah[2], Laurie J Hannigan[3,4,5], Sarah H Wild[6], Bruce Guthrie[6], Frances S Mair[1], Sofia Dias[7], Nicky J Welton[4], David A McAllister[1]

**Affiliations:**

1. Institute for Health and Wellbeing, University of Glasgow
2. London School of Hygiene and Tropical Medicine
3. Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway
4. Population Health Sciences, Bristol Medical School, University of Bristol
5. Department of Mental Disorders, Norwegian Institute of Public Health
6. Usher Institute, University of Edinburgh
7. Centre for Reviews and Dissemination, University of York

**Contact information:**

Dr Peter Hanlon
Institute of Health and Wellbeing
University of Glasgow
1 Horselethill Road
Glasgow, UK
G12 9LX

e-mail: peter.hanlon@glasgow.ac.uk
Telephone: +441413301663


Word count: 4407

# Abstract

**Background**: The applicability of randomised controlled trials of pharmacological agents to older people with frailty/multimorbidity is often uncertain, due to concerns that trials are not representative. However, assessing trial representativeness is challenging and complex. We explore an approach assessing trial representativeness by comparing rates of trial Serious Adverse Events (SAE) to rates of hospitalisation/death in routine care.

**Methods**: This was an observational analysis of individual (125 trials, n=122,069) and aggregate-level drug trial data (483 trials, n=636,267) for 21 index conditions compared to population-based routine healthcare data (routine care). Trials identified from clinicaltrials.gov. Routine care comparison from linked primary care and hospital data from Wales, UK (n=2.3M). Our outcome of interest was SAEs (routinely reported in trials). In routine care, SAEs were based on hospitalisations and deaths (which are SAEs by definition). We compared trial SAEs in trials to expected SAEs based on age/sex standardised routine care populations with the same index condition. Using IPD, we assessed the relationship between multimorbidity count and SAEs in both trials and routine care, and assessed the impact on the observed/expected SAE ratio additionally accounting for multimorbidity.

**Results:** For 12/21 index conditions the pooled observed/expected SAE ratio was <1, indicating fewer SAEs in trial participants than in routine care. A further 6/21 had point estimates <1 but the 95% CI included the null. The median pooled estimate of observed/expected SAE ratio was 0.60 (95% CI 0.55-0.64; COPD) and the interquartile range was 0.44 (0.34-0.55; Parkinson's disease) to 0.87 (0.58-1.29; inflammatory bowel disease). Higher multimorbidity count was associated with SAEs across all index conditions in both routine care and trials. For all trials, the observed/expected SAE ratio moved closer to 1 after additionally accounting for multimorbidity count, but it nonetheless remained below 1 for most.

**Conclusions**: Trial participants experience fewer SAEs than expected based on age/sex/condition hospitalisation and death rates in routine care, confirming the predicted lack of representativeness. This difference is only partially explained by differences in multimorbidity. Assessing observed/expected SAE may help assess applicability of trial findings to older populations in whom multimorbidity and frailty are common.

## Keywords

Randomised controlled trials, serious adverse events, multimorbidity, epidemiology, chronic disease

## Background

Randomised controlled trials (hereafter trials) provide the most robust and valid evidence about relative treatment efficacy. However, many patients treated in routine clinical care do not meet trial eligibility criteria; older patients and those with multimorbidity (the presence of two or more conditions) or frailty are often excluded or under-represented.[1] Although not always an explicit exclusion criterion, investigators also routinely exclude people where they have concerns over an individual's ability to manage the burdens of trial participation,[2] in order to minimise loss to follow-up.

Where such groups are under-represented within trials, the applicability of effect estimates to the wider clinical population is uncertain.[3] Relative treatment effects (e.g., odds ratios) might plausibly differ in older patients or those with frailty or multimorbidity.[4] Even where it is reasonable to assume that relative treatment effects are the same, absolute treatment effects (both benefits and harms), and therefore the balance between risk and benefit, may differ because baseline rates of relevant outcomes vary.[3] Additionally, people excluded from trials may be at greater risk of adverse effects or complications of treatment, particularly in the context of frailty, meaning that assessment of safety based on trials may not be applicable to people receiving treatment in routine care. Thus, it is important to consider the representativeness of trial participants.

Participant representativeness in terms of age, sex and the severity of the target condition can be readily assessed as these characteristics are routinely included in trial reports. However, this is not true for measures of frailty or multimorbidity. We previously examined representativeness in terms of multimorbidity and frailty across a range of industry-funded phase 3 trials.[5,6] However, this involved the analysis of individual-level participant data which is a complex and time-consuming process, not feasible in most contexts. Consequently, clinicians and guideline developers are usually unable to fully assess trial representativeness.

Trial serious adverse event (SAE) reporting may help address this problem. The primary purpose of collecting SAE data is to detect if the treatments being tested in the trial are harmful. However, any event that is life threatening, leads to death, causes or prolongs hospitalisation, results in serious or lasting impairment or disability, or causes a birth defect is defined as a SAE, regardless of causation, and must be reported.[7] Therefore, where a trial is representative, we would expect the SAE rate to reflect age-sex specific rates of hospitalisation and death among people with the same condition identified from routine care. SAE rates may therefore be utilised to help assess trial representativeness. In trials for hypertension, we tested this hypothesis, finding that the SAE rates were consistently lower than predicted based on hospitalisation and death rates among people with hypertension in routine care.[8] We also found that although SAE rates were higher in hypertension trials which aimed to be representative of older people, the rates were still lower than in routine care.

This study will examine SAEs in trials and in routine care across 21 index conditions. Using routine healthcare data and trial data, we aim to explore: (i) how observed SAE counts in trial populations compare to SAEs (defined as hospitalisations and deaths) for people with the same index condition in a clinical population, (ii) whether multimorbidity counts will predict hospitalisation and deaths and SAEs similarly in trial and clinical populations and (iii) whether any differences between expected and observed SAEs will be attenuated by accounting for different levels of multimorbidity between trial and clinical populations.

# Methods

## Study design

This observational analysis compares incident SAEs among people enrolled in randomised controlled trials of pharmacological therapies for a range of index conditions to SAEs (defined as deaths or hospitalisations) among people with similar index conditions in routine care. First, we use aggregate data from trials and routine care data from a clinical population with the same index condition and similar age-sex distribution to the trial population to generate an SAE ratio of observed to expected SAEs. Secondly, in trials for which individual participant data were available, we compare observed and expected SAEs in two ways; first where the expected SAEs are based solely on age and sex distribution and secondly where the expected SAEs are additionally estimated using the number of additional long-term conditions (multimorbidity count).

## Data sources and participants

### Trials – aggregate data

We identified trials registered with clinicaltrials.gov for 21 index conditions. Trials were selected according to a pre-specified protocol (Prospero CRD42018048202).[9] Trial selection is described in detail elsewhere.[6] Briefly, trials had to be registered with clinicaltrials.gov; start after 1st January 1990; be phase-2/3, phase-3, or phase-4; include ≥300 participants; have an upper age limit no younger than 60 years; and evaluate pharmacological treatments for one of a range of cardiovascular, respiratory, gastrointestinal, musculoskeletal, metabolic, autoimmune and connective tissue, urological and otolaryngological index conditions (listed in table 1).[6] We grouped trials by index condition, defined by the treatment indication as described in the trial registration. For this analysis of SAEs, we then restricted this set of trials to those registered after 2010 (range 2010-2017, mean 2012 and median 2012), as reporting of SAEs on clinicaltrials.gov was more complete after this date.

### Trials – individual participant data

From within the list of eligible trials, we identified and accessed individual participant data for trials available via one of two repositories: the Clinical Study Data Request (CSDR) and the Yale University Open Data Access (YODA) project as described in detail previously.[6]

### Routine care comparison

Data from the Secure Anonymised Information Linkage (SAIL) Databank were used to identify a routine care population for each of the trial index conditions. SAIL is a repository of health and administrative data from Wales, includes approximately 70% of the Welsh population, and is nationally representative in terms of distribution of age, sex, and socioeconomic status.[6,10] We identified people with each of the index conditions from a sample of 2.3 million people registered with a participating general practice between 1st January 2011 and 1st January 2012 (to match the median start date of the trials). Index conditions were identified using diagnostic codes used in UK primary care, as described in detail elsewhere.[6]

## Identifying outcomes

The outcome of interest was incident SAEs (in trials) and incident SAEs in the routine care population – defined as incidence hospitalisations/deaths. For the routine care population, SAEs were identified through linkage to the Patient Episode Database for Wales and the National Mortality Registry, respectively. We included all hospital episodes that were coded as 'urgent' (as opposed to 'elective'). For each participant, we assessed incident events occurring between 1st January 2012 and 1st August 2012 (first 6 months available following identification of the index condition, concurrent with the median time of trial registration), de-registration with a participating practice, or death, whichever happened first. This follow-up time was selected to be similar to the follow-up in the included trials (median 26 weeks; interquartile range 12 to 52 weeks). Total observation time was calculated for each individual.

For all registered trials we extracted the number of participants for whom SAEs were reported, number of persons at risk as reported on clinicaltrials.gov, and the timeframe for which the events were reported. For trial IPD, event information was identified from the standard adverse event data tables within the CSDR or YODA repositories, and follow-up time was calculated as the number of days from randomisation to end of follow-up. All IPD trials reported whether an event was classified as serious, however fewer trials provided details of classification (i.e. few trials specified what proportion of SAEs were hospitalisations and deaths versus other causes such as events resulting in impairment or disability). On examining 24 trials providing complete data for death and hospitalisation within YODA, the proportion of SAEs due to hospitalisation or death was generally high (for these 24 trials the proportions were 100% for ankylosing spondylitis, 82% for dementia, 97% for diabetes, 97% for IBD, 96% for psoriasis, 92% for psoriatic arthropathy and 87% for rheumatoid arthritis). SAE ascertainment in the routine care population is likely to be slightly lower than in the trial population.

## Assessing demographics and multimorbidity

For trials, age and sex information was obtained at a summary level from trial registration reporting and directly from individual participant data. For the routine care population, age and sex were obtained from primary care data.

In order to explore the relationship between multimorbidity and SAEs in the trial IPD, we identified twenty-one comorbidities (cardiovascular disease, chronic pain, arthritis, affective disorders, acid-related disorders, asthma/chronic obstructive pulmonary disease, diabetes mellitus, osteoporosis, thyroid disease, thromboembolic disease, inflammatory conditions, benign prostatic hyperplasia, gout, glaucoma, urinary incontinence, erectile dysfunction, psychotic disorders, epilepsy, migraine, parkinsonism and dementia). These were identified using concomitant medication data. Concomitant medication data were used as a surrogate for prevalent multimorbidity as to maintain patient confidentiality medical history was frequently redacted in the trial datasets. Medication-

based definitions were prespecified following clinical and epidemiological review, and are described in detail elsewhere.[6] Briefly, chronic conditions were grouped to allow identification of broad categories of conditions from medication use (e.g. the use of inhaled bronchodilators was taken to indicate 'obstructive airways disease', but we did not attempt to differentiate between asthma and chronic obstructive pulmonary disease). Furthermore, medications which were likely to be used for a diverse range of indications were not used to identify chronic conditions (e.g. we excluded tricyclic antidepressants from the list used to identify affective disorders as these are also used to treat chronic pain). Data from baseline recruitment were used to quantify a total multimorbidity count for each participant in each trial.

For the routine care population, prescription data from the primary care record were used to apply identical medication-based multimorbidity definitions. We applied these definitions to drugs prescribed during 2011, which was treated as 'baseline'.

## Statistical analysis

The routine care data and trial IPD were both held in different secure data analysis platforms with restrictions on what could be exported. Analysis therefore involved exporting summary statistics and model outputs from each platform. The analyses are presented below relating to the three main aims of the study. A more detailed description of the statistical analyses is given in the supplementary material.

### Comparison of SAEs in trials (aggregate data) and routine care

This analysis aimed to compare the observed SAEs in trials to SAEs for people with the same index condition in routine care. First, for each index condition, we modelled first hospitalisation or death in routine care using age-adjusted Poisson regression models, stratified by sex. To allow for non-linearity in age, up to two fractional polynomial terms were included. An offset was included to account for differences in person-time. This model therefore allowed us to calculate the predicted SAE rate for each index condition, conditional on age and sex. Second, for each trial with aggregate-

level data (n=483), we estimated the percentage of trial participants of each sex in one-year age bands based on the age (mean, variance, upper and lower bounds) and sex statistics as reported on clinicaltrial.gov. Third, for each one-year age/sex band, we calculated the expected number of SAEs given the trial size and follow-up time based on the routine care models summing these (weighting by the percentage of participants in each band) to obtain the expected SAEs for the whole trial. Finally, we compared the observed number of SAEs in each trial to the expected number of SAEs, expressed as a ratio (observed/expected SAE ratio). We calculated 95% credible intervals for each trial using sampling methods presented in the statistical appendix. We also estimated the pooled observed/expected ratio at the level of each index condition by fitting a random effects model with a Poisson likelihood treating the expected count as an offset term.

For these analyses we ignored treatment effects, combining SAEs across all arms, implicitly assuming that the effect of trial interventions on SAEs were ignorable for this set of trials. Following peer review, we conducted the following post-hoc exploratory analyses to test this assumption. Although SAE results and design information at the level of trial arms are held within CTG, these are not linked to one another by a unique identifier, so we first harmonised these manually before categorising each trial according to the type of comparison. For 12 trials the serious adverse event rates were not available at arm level but only as summaries, leaving 471 trials with arm-level SAE information. We then characterised the nature of the comparison in each trial. Of these, in 269 trials there was a placebo arm, in 110 trials all arms within the trial had the same designation (experimental or active but not both) and in 92 trials there were different designations across arms (i.e. experimental versus active). For each trial (for the available comparison) we estimated the log-rate ratio for the difference in SAE rate between arms. We did so by fitting a Poisson regression model with the log person-time (for each arm) as an offset. The person time was estimated similarly as in the main analysis (follow-up time x participants – 0.5 * follow-up time x incident events). We subsequently combined these log-rate ratios in random effects meta-analyses for each index condition according to the type of treatment comparison.

## Association between multimorbidity count and SAEs in trials (IPD) and routine care

This analysis aimed to compare the association between multimorbidity count and SAEs in trials (using trial IPD) and in routine care. This analysis was limited to trials with IPD and where the total number of SAEs per sex was ≥20 (n=60 trials for 11 index conditions). For each sex, we estimated the association between multimorbidity count and SAEs using Poisson regression models, adjusted for age. The log-transformed time to each event or the end of follow-up was included as an offset term. For the trials, the coefficients and standard errors for the comorbidity terms were then meta-analysed in random effects meta-analysis fitted using restricted maximum likelihood estimation. For each index condition, we then plotted the rate ratio (with 95% confidence intervals) for SAEs across the range of multimorbidity counts found in the trials (meta-analysed for each index condition) and in routine care.

## Comparison of observed and expected trial SAEs before and after accounting for multimorbidity

This final analysis aimed to assess if any differences between expected and observed SAE rates were attenuated by accounting for different levels of multimorbidity between trial and routine care populations. This analysis was based on trial IPD (n=125 trials). First, we estimated the expected SAE count for each trial based on age sex and index condition, using the same models as for the aggregate data (Unlike with the aggregate trial analysis, the percentage of participants of each age and sex were directly observed rather than estimated from summary statistics). Next, we fitted a further sex-specific models to the routine care dataset for each index condition including, in addition to age, multimorbidity count. Age and multimorbidity count were each modelled using up to two fractional polynomial terms. These models were used to calculate the expected number of SAEs per trial based on the age, sex and multimorbidity count of trial participants. Finally, we calculated the ratio of observed to expected SAEs based on age and sex alone, and based on age, sex and multimorbidity count. The two ratios were then compared for each trial.

All analyses were conducted using R statistical software. All trial-level data, including model outputs, as well as analysis code are provided on the project github repository (https://github.com/dmcalli2/sae_ctg_multicond_public).

# Results

Trial selection is summarised in figure 1. Of the 2,173 eligible trials identified in our original search, 777 were registered after 1st January 2010. Of these, 578 reported SAE data. 14 were excluded because insufficient information was reported for calculation of SAE counts and a further 81 were excluded as the index condition was not included in our list. We therefore included 483 trials in our analysis of aggregate trial data (n=636,267 participants). We obtained IPD for 125 trials (n=122,069 participants) from the CSDR and YODA repositories (trials for which sponsors made IPD available to third party researchers, for which we did not apply a cut-off date of 2010 given that there are relatively few trials for which IPD are available), which were included in subsequent analyses of multimorbidity count. 42 trials were included in both the IPD and aggregate sets. Trials for each of the 21 index conditions are summarised in table 1, with individual trial summary data shown on the project Github repository ( https://github.com/dmcalli2/sae_ctg_multicond_public). This table also shows the total number and mean age of people with each of the index conditions in the routine care sample of 2.3M people registered with a SAIL practice during 2011.

## Comparison of SAEs in trials (aggregate data) and routine care

For each of the index conditions, the observed/expected SAE ratio for each index condition is shown in figure 2 pooled across trials. For 18 of the 21 index conditions, the SAEs were lower than that expected; for 12 of these the 95% confidence intervals did not include the null. COPD was the index condition with the median observed/expected SAE ratio (0.60; 95% CI 0.55-0.64), Parkinson's disease and inflammatory disease were at the 25th and 75th centiles respectively (0.44; 95% CI 0.34-0.55 and 0.87; 0.58-1.29 respectively). The most extreme ratio was seen for dementia (0.27; 95% confidence interval 0.17-0.44) indicating that the rate of SAEs in these trials was around a quarter of that seen in routine care. For three out of 21 conditions, the observed/expected SAE ratio was >1, and for each of these the confidence intervals were wide and included the null (pulmonary hypertension 1.12 (0.39-3.67), atrial fibrillation 1.13 (0.39-3.07), and thromboembolism 1.85 (0.51-5.80)).

Considerable variation in the observed/expected SAE ratio was apparent between trials within the same index condition. Trial level estimates are shown in figure 3 for the six index conditions with the greatest number of trials (all other index conditions are shown in the supplementary appendix, Figures S1-S21). Taking type 2 diabetes as an example, although the pooled ratio of observed/expected SAE was less than half (0.46 (95% CI 0.43-0.50)), for some trials it was close to unity.

These analyses pooled SAEs from treatment and control arms of each trial, assuming that SAE rates would be similar across arms. Post-hoc analyses comparing SAE rates across trials, conducted to explore this assumption, are shown in the supplementary appendix. For 471 trials for which trial arm-level data was available, the SAEs were generally similar across trial arms. Even where a treatment was compared with placebo – where we would expect to have the best chance of identifying a difference in SAE rates – there was rarely evidence of a statistically significant difference in SAE rates. This was true for both individual trials and the meta-analyses (Figures S22 and S23). Where the 95% confidence intervals excluded the null (eg in type 2 diabetes trials with a placebo comparator), the magnitude of such differences were much smaller than the differences in rates we observed between trial participants and individuals in the community.

## Association between multimorbidity count and SAEs in trials (IPD) and routine care

For all 21 index conditions the multimorbidity count predicted the SAE rate in routine care. In the 11 index conditions for which we had sufficient data, multimorbidity count also predicted the SAE rate in the trial data. These associations are shown in figure 4 for trials and routine care, respectively. The SAE rate did not differ across trial treatment arms (RR men 0.91; 95%CI 0.81-1.02, RR women 0.99; 95%CI 0.87-1.10), and multimorbidity predicted SAE rates similarly in trial treatment arms and control arms (rate ratio (RR)-interaction 1.02; 95%CI 0.97-1.06).

## Comparison of observed and expected trial SAE rates before and after accounting for multimorbidity

Since multimorbidity counts are lower in trial than in routine care populations (results reported previously)[6] and multimorbidity counts predict SAE rates in clinical populations (figure 4), the ratio of observed/expected SAE is inevitably higher when multimorbidity count is included in the standardisation than when age and sex alone are included. For most trials, for which the age/sex adjusted ratio was <1, this meant that additionally adjusting for multimorbidity attenuated the observed/expected ratio closer to one. Figure 5 shows the magnitude of this effect for the four index conditions with the greatest number of IPD trials (other conditions shown in supplementary material). The solid line indicates the ratio of observed to expected based on age and sex and the dotted line the ratio of observed to expected based on age, sex and multimorbidity count. In some cases, the impact of accounting for the multimorbidity count was sufficiently large for the ratios to move from being below one to being at or above one. However, for most trials the observed/expected SAE ratio remained <1 regardless of whether the expected count was also based on multimorbidity (figure 5 and supplementary appendix). This implies that differences in the multimorbidity count between trial and routine care populations may account for some, but not all, of the difference in event rates between trials and routine care.

320


321

322 *Figure 1: identification and inclusion of eligible trials*

*Table 1: Description of numbers of people in community and participants in trials for each index condition*

| | Routine care | | Aggregate data trials N=483 | | | IPD trials N=125 | | |
|---|---|---|---|---|---|---|---|---|
| Index condition | Total N | Mean age (sd) | Included trials | Total number of participants | Range of mean trial age | Included trials | Total number of participants | Range of mean trial age |
| Asthma | 191160 | 45.6 (22.9) | 47 | 74833 | 35.2-51.4 | 4 | 1084 | 32.0-50.2 |
| Atrial Fibrillation | 43330 | 74.7 (11.9) | 9 | 12539 | 59.3-75.0 | 1 | 18113 | 72.8 |
| Axial Spondyloarthritis | 1982 | 52.4 (15.3) | 8 | 2994 | 29.9-45.2 | 2 | 320 | 38.0-43.8 |
| Benign Prostatic Hyperplasia | 19906 | 72.0 (10.0) | 7 | 4617 | 60.9-66.5 | 5 | 1710 | 62.2-66.6 |
| Chronic Obstructive Pulmonary Disease | 57378 | 69.1 (11.6) | 94 | 131630 | 61.1-70.8 | 7 | 3376 | 61.0-66.1 |
| Dementia | 13871 | 82.1 (9.0) | 3 | 2506 | 73.8-74.4 | 6 | 4791 | 69.0-83.0 |
| Epilepsy | 29554 | 45.8 (21.0) | 8 | 3715 | 32.2-41.1 | 0 | - | |
| Hypertension | 310691 | 67.0 (12.9) | 14 | 10380 | 49.2-70.7 | 12 | 6863 | 51.4-70.9 |
| Inflammatory bowel disease | 12514 | 52.3 (17.8) | 7 | 4086 | 37.4-44.7 | 10 | 4352 | 36.0-41.9 |
| Myocardial infarction | 3510 | 70.7 (14.1) | 11 | 76425 | 58.2-67.0 | 0 | - | |
| Osteoarthritis | 124521 | 67.6 (12.7) | 4 | 1794 | 60.7-62.7 | 1 | 1321 | 63.9 |
| Osteoporosis | 38212 | 72.8 (12.2) | 5 | 5335 | 68.8-74.8 | 7 | 51204 | 53.6-73.2 |
| Parkinson's disease | 4998 | 74.9 (10.4) | 14 | 5754 | 61.9-67.5 | 4 | 1212 | 61.0-62.9 |
| Psoriasis | 52810 | 49.1 (19.0) | 24 | 19064 | 43.1-54.2 | 7 | 3609 | 43.6-46.0 |
| Psoriatic arthropathy | 3523 | 54.1 (14.0) | 13 | 5168 | 47.4-51.9 | 3 | 596 | 45.9-49.0 |
| Pulmonary fibrosis | 1465 | 73.3 (10.9) | 4 | 1962 | 66.6-70.3 | 2 | 1063 | 66.2-67.7 |
| Pulmonary Hypertension | 759 | 60.5 (27.0) | 2 | 1757 | 48.1-55.7 | 1 | 161 | 52.2-56.8 |
| Rheumatoid arthritis | 13809 | 62.2 (15.5) | 29 | 21545 | 46.6-60.7 | 11 | 5223 | 49.0-55.6 |
| Systemic Lupus Erythematosus | 1033 | 52.8 (15.5) | 3 | 1998 | 32.1-41.3 | 2 | 1129 | 33.6-39.8 |
| Thromboembolism | 9162 | 66.1 (15.7) | 4 | 8503 | 40.0-76.4 | 7 | 16959 | 53.3-55.7 |
| Type 2 Diabetes Mellitus | 82473 | 65.3 (13.0) | 173 | 239662 | 48.8-74.2 | 23 | 19830 | 53.5-64.1 |

| Index condition | Total N in routine care | Number of trials (Total N with SAE) | Expected SAEs | Observed/expected SAE ratio | Sigma |
|---|---|---|---|---|---|
| Asthma | 191160 | 47 (2816) | 5531 | 0.44 (0.35-0.54) | 0.42 |
| Atrial Fibrillation | 43330 | 9 (2261) | 3870 | 1.13 (0.39-3.07) | 2.58 |
| Axial Spondyloarthritis | 1982 | 8 (177) | 251 | 0.72 (0.55-0.91) | 0.05 |
| Benign Prostatic Hyperplasia | 19906 | 7 (240) | 524 | 0.42 (0.29-0.58) | 0.16 |
| Chronic Obstructive Pulmonary Disease | 57378 | 94 (17658) | 27057 | 0.60 (0.55-0.64) | 0.13 |
| Dementia | 13871 | 3 (140) | 520 | 0.27 (0.17-0.44) | 0.16 |
| Epilepsy | 29554 | 8 (167) | 298 | 0.59 (0.33-1.11) | 0.61 |
| Hypertension | 310691 | 14 (83) | 188 | 0.43 (0.26-0.72) | 0.63 |
| Inflammatory bowel disease | 12514 | 7 (207) | 249 | 0.87 (0.58-1.29) | 0.22 |
| Myocardial infarction | 3510 | 11 (16550) | 51097 | 0.57 (0.35-0.96) | 0.71 |
| Osteoarthritis | 124521 | 4 (38) | 79 | 0.58 (0.27-1.34) | 0.46 |
| Osteoporosis | 38212 | 5 (644) | 1418 | 0.41 (0.24-0.66) | 0.28 |
| Parkinsons disease | 4998 | 14 (310) | 709 | 0.44 (0.34-0.55) | 0.15 |
| Psoriasis | 52810 | 24 (509) | 569 | 0.91 (0.78-1.05) | 0.08 |
| Psoriatic arthropathy | 3523 | 13 (390) | 573 | 0.71 (0.49-1.02) | 0.4 |
| Pulmonary fibrosis | 1465 | 4 (471) | 660 | 0.71 (0.53-0.90) | 0.05 |
| Pulmonary Hypertension | 759 | 2 (779) | 679 | 1.12 (0.39-3.67) | 0.71 |
| Rheumatoid arthritis | 13809 | 29 (2088) | 4098 | 0.56 (0.42-0.74) | 0.58 |
| Systemic Lupus Erythematosus | 1033 | 3 (277) | 290 | 0.97 (0.65-1.34) | 0.08 |
| Thromboembolism | 9162 | 4 (1300) | 388 | 1.85 (0.51-5.80) | 1.6 |
| Type 2 Diabetes Mellitus | 82473 | 173 (51472) | 74968 | 0.46 (0.43-0.50) | 0.23 |

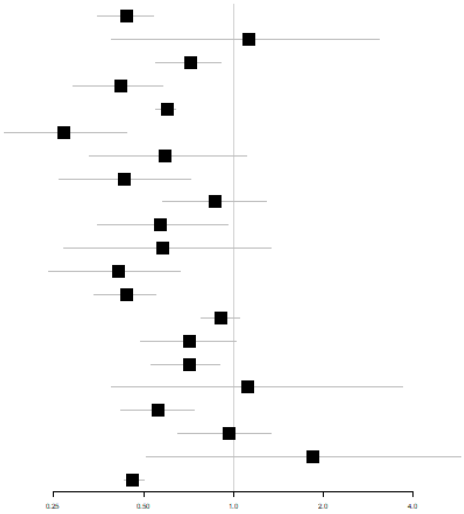
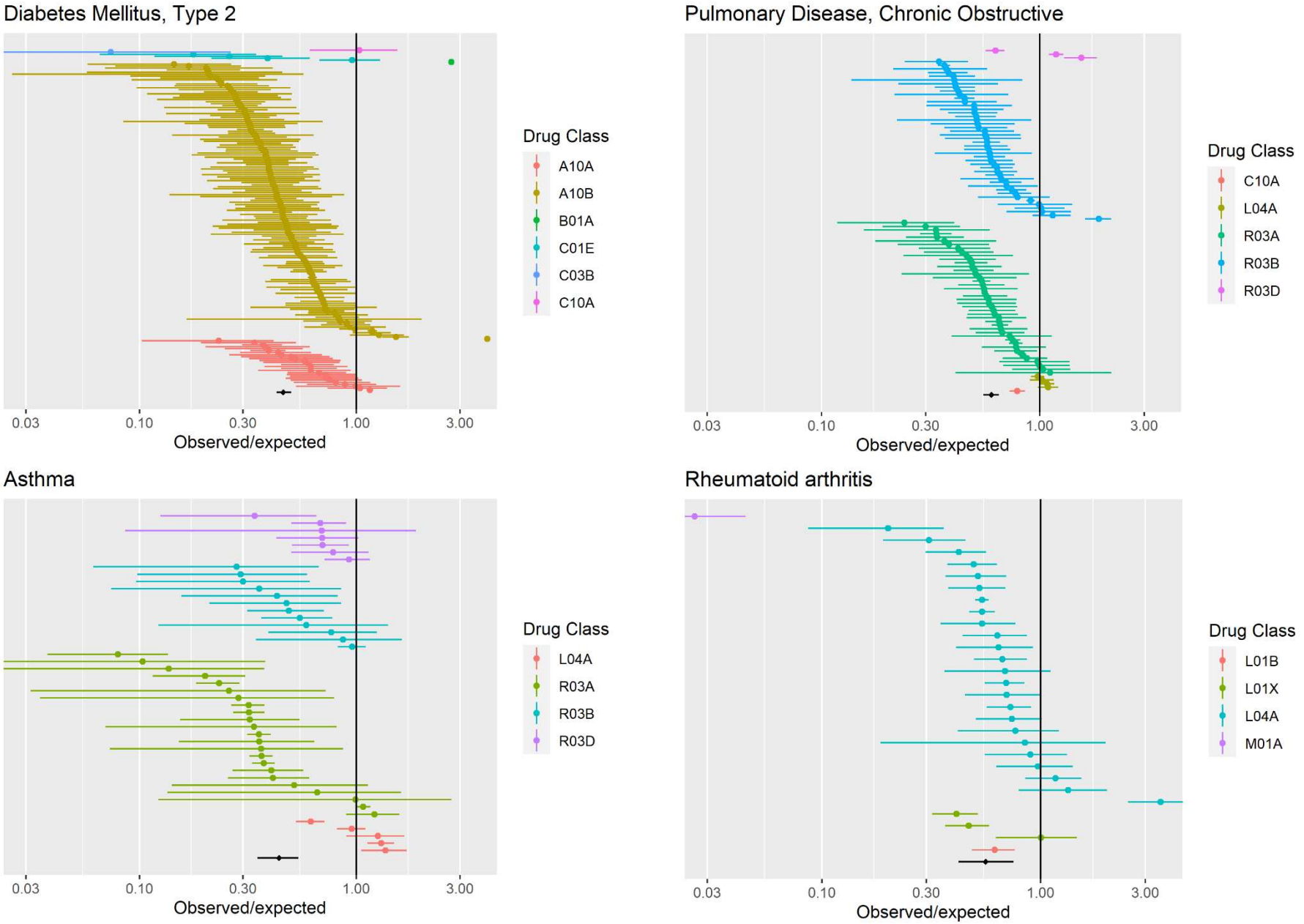



Figure 2 This figure displayed the pooled observed/expected SAE ratio for each of the index conditions. It also shows the number of people in the routine care cohort with each index condition, the number of trials with aggregate data and the total number of SAEs.


Diabetes Mellitus, Type 2
0.03
0.10
0.30
1.00
3.00
Observed/expected
Drug Class
A10A
A10B
B01A
C01E
C03B
C10A
Pulmonary Disease, Chronic Obstructive
0.03
0.10
0.30
1.00
3.00
Observed/expected
Drug Class
C10A
L04A
R03A
R03B
R03D
Asthma
0.03
0.10
0.30
1.00
3.00
Observed/expected
Drug Class
L04A
R03A
R03B
R03D
Rheumatoid arthritis
0.03
0.10
0.30
1.00
3.00
Observed/expected
Drug Class
L01B
L01X
L04A
M01A

329

*Figure 3: This figure shows the ratio of observed/expected serious adverse event rates in aggregate data trials. Four selected index conditions with the largest number of eligible trials are displayed here, with the remaining conditions displayed in the supplementary appendix. The point-estimate and 95% confidence interval for the ratio for each trial is shown by the coloured points and bars, respectively. Different drug classes are separated by colour (full key displayed in supplementary appendix). The pooled ratio and 95% confidence intervals meta-analysed across all trials within each index condition is shown by the black point and line at the bottom of each plot. Findings are based on analysis of aggregate trial data from clinicaltrials.gov (index condition, trial drug, age, sex, SAEs and follow-up) for the observed rate and individual patient data from SAIL was used to calculate the expected rate.*

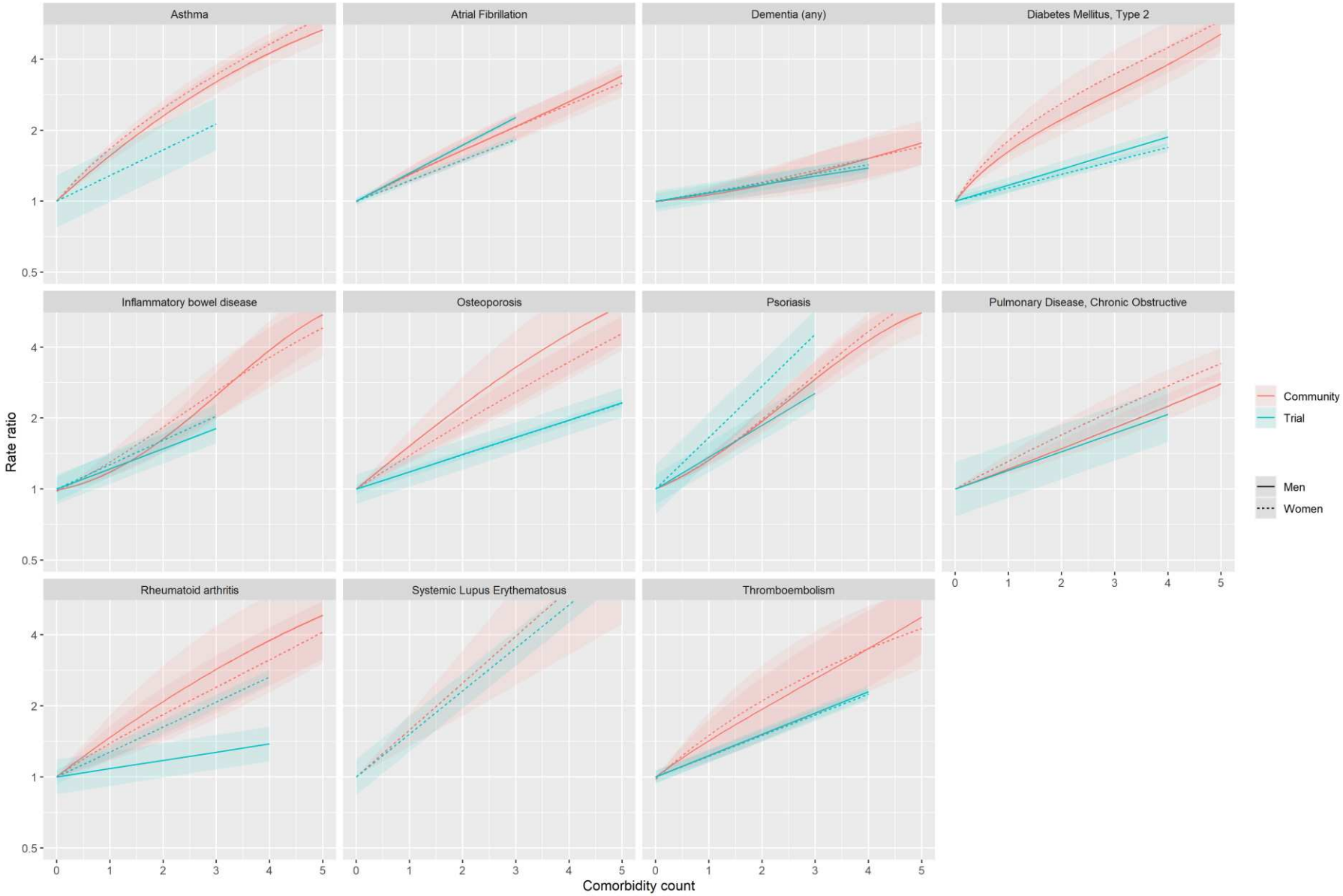



*Figure 4: This figure shows the relationship between multimorbidity count and SAE rate in men and women meta-analysed for trials of each index condition (blue) and for each index condition in routine care (red). Shaded areas indicate 95% confidence intervals for the meta-analysis and routine care estimates.*

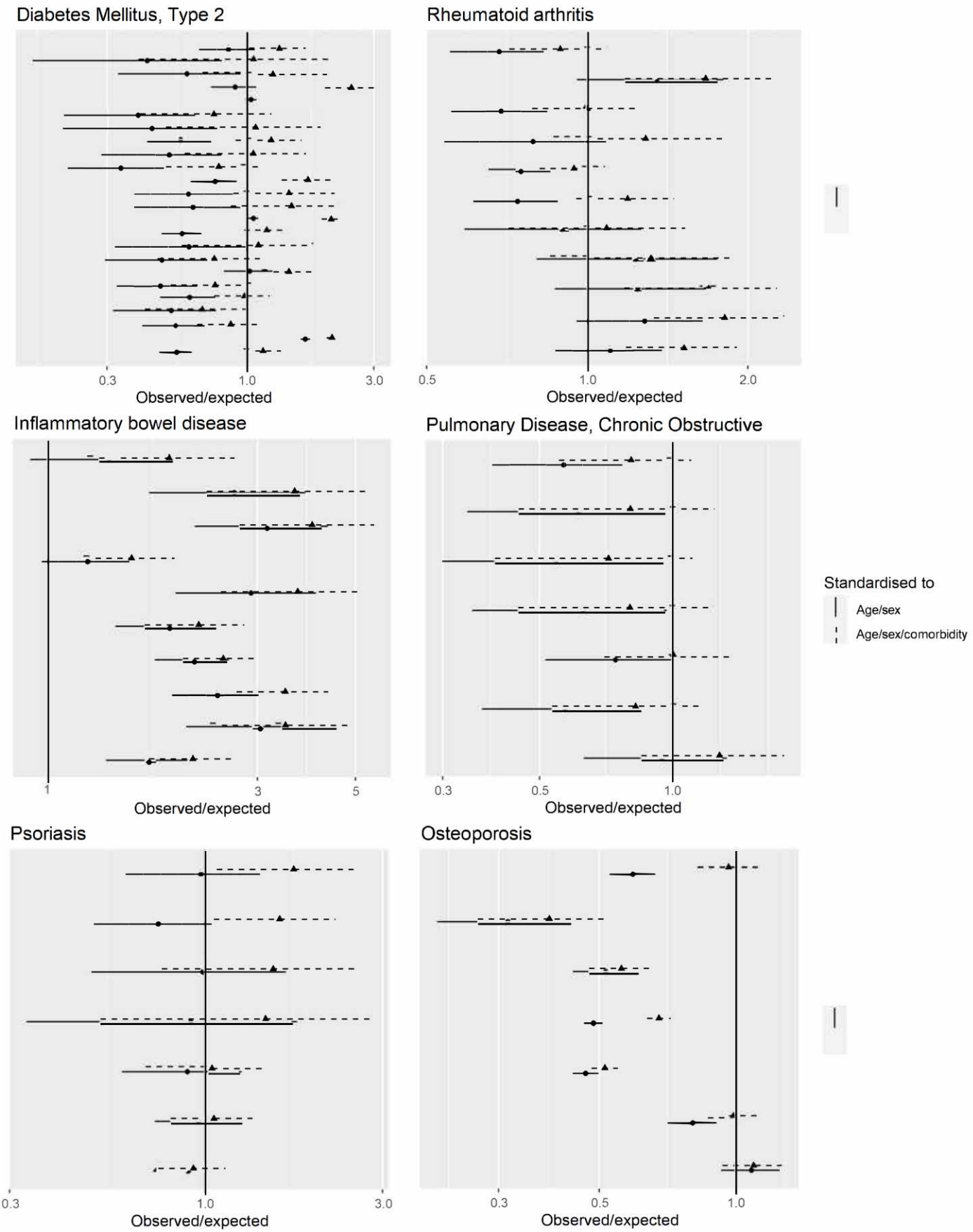



*Figure 5: Ratio of observed/expected SAE based on age and sex (square points with solid lines indicating 95% confidence interval), and based on age, sex and multimorbidity count (triangle points with broken lines indicating 95% confidence intervals) for six selected index conditions. Each pair of points correspond to a single trial. Ratios for all other conditions are shown in the supplementary appendix.*

# Discussion

We compared SAEs in trials to the expected number of SAEs based on age/sex adjusted hospitalisation and death rates among people with the same index conditions in routine care. For most trials, and the majority of index conditions, trials had fewer than expected SAEs. On meta-analysing 483 trials with aggregate-level data, across 21 index conditions, we found that the point estimate for this ratio was below one for 18 index conditions and that for 12 of the 18 index conditions the confidence intervals did not include one. Secondly, we assessed the relationship between multimorbidity (which is known to be less common in clinical trial populations) and SAEs. Multimorbidity count was associated with increased SAE rates in trials as well as in routine care. Finally, we found that, where the expected SAE count in trials was derived from age, sex and multimorbidity (rather than age and sex alone), the observed/expected SAE ratio for most trials in most index conditions moved closer to one. Despite this, the observed SAEs remained lower than expected in most trials, even after additionally accounting for multimorbidity. These findings show (i) that age/sex standardised observed rates of SAEs are lower than expected in trial populations; (ii) that some of this difference is explained by lower levels of multimorbidity within trial populations; and (iii) that many trials are not representative even after age, sex and multimorbidity standardisation, suggesting that trial populations differ systematically from those treated in routine care in aspects not fully accounted for by a simple multimorbidity count.

Our findings suggest that trial populations are on average experience fewer SAEs than people with the same index conditions in routine care. While this suggests trials are often under-representative, there could be several other factors contributing to differences between trials and routine care, and to heterogeneity in the observed/expected ratio between trials. It is possible that trial participation led to better quality care for some participants, which in turn may reduce the likelihood of hospitalisation or death. Furthermore, under-reporting of SAEs by trial sponsors would also result in the differences seen. However, SAE reporting is a regulatory requirement for drug approval and our

aggregate data analysis was limited to the time period in which this requirement was in place. While these factors may account for some difference between trials and routine care, for many trials, the difference between trials and routine care was large and unlikely to be explained by these other factors alone. Furthermore, taking type 2 diabetes as an example, many of the trials for which the observed/expected SAE ratio was closer to one were trials in which the population of interest was likely to be at higher risk of SAEs (e.g. people with nephropathy or at increased cardiovascular risk). Our findings are therefore consistent with previous literature suggesting that many clinical trial populations are often not representative, and that trial participants are on average healthier than patients encountered in routine care, including having lower prevalence of multimorbidity. It further suggests that examination of expected and observed SAEs may offer additional insights into trial representativeness.

## Comparison with other studies

It is widely recognised that many trials exclude a large proportion of people with the condition of interest.[1] Specifically, older people and those with multimorbidity or frailty are often excluded from trials (either explicitly or implicitly).[5,6,11] It has been argued that this lack of representativeness limits the generalisability and applicability of trial findings,[12,13] and more recently this uncertainty has been reflected in clinical guidelines for managing multimorbidity.[14] However, assessment of representativeness is also challenging as reporting of exclusion criteria and participant characteristics is variable and often limited.[15]

Two previous studies used psoriasis registry data to compare rates of SAEs in trial ineligible vs trial eligible patients, with lower rates observed in the population eligible for trial participation.[16,17] Another study found higher rates of SAEs in a UK psoriasis registry than in IPD from two psoriasis trials, even after re-weighting the register data to more closely resemble the trial populations.[18] Our own previous study also compared SAE rates in trial participants and patients in routine care finding higher rates in the routine care population after age-sex standardisation, but did so solely for trials

of agents acting on the renin-angiotensin-aldosterone system in order to treat hypertension.[8] The current study adds to this somewhat sparse literature by examining the trial age-sex standardised observed/expected SAE ratio across a wide range of index conditions. For many conditions, particularly those predominantly affecting older people and in which multimorbidity is common, we showed that most trials have substantially lower rates of SAEs than expected, suggesting that most trials are unrepresentative.

Our observation that the multimorbidity count was associated with SAEs and with hospitalisation and death similarly across index conditions adds to the previous literature showing that multimorbidity predicts death and hospitalisations in the general population,[19-23] and that SAEs are associated with the frailty index in trial participants.[5] We are not aware of any previous study exploring the relationship between multimorbidity and SAEs in trials.

## Implications

Our findings suggest that trials systematically select people at lower risk of SAEs. As a result, even if the relative benefit of the trial treatment were the same for all patients, the overall net benefit of treatment may be different for people at higher risk of SAEs who are more likely to be excluded from trials. It is therefore important to assess trial representativeness in order to judge the extent to which trial-derived estimates of relative and absolute treatment effects (eg odds ratios and absolute risk reductions, respectively), net overall treatment benefits (balancing the effects of treatments on target and adverse events) and cost-effectiveness can be applied to clinical practice.[24] By design, trials exclude many people with the condition being treated. Even accounting for explicit eligibility criteria, it seems likely that, even in the absence of specific exclusions addressing this, trial investigators may be less likely to recruit patients they suspect of being liable to early withdrawal or to experiencing adverse events due to multimorbidity or frailty. Furthermore, trial descriptions of baseline characteristics rarely capture all relevant characteristics of trial participants; frailty and multimorbidity, for example, are rarely included in such summaries. Currently, approaches to assess

trial representativeness are based on detailed consideration of the trial design (recruitment strategies, eligibility criteria etc), participant flow (numbers screened, enrolled etc) and participant characteristics. There are useful tools to guide this process, but it remains complex, time consuming and impractical for widespread, rapid use. Furthermore, detailed assessment of trial participant characteristics relies on these characteristics being reported in trials, which is not always the case for important measures such as socioeconomic status, multimorbidity or frailty.

In this context, assessing the observed/expected SAE ratio may be a useful metric to aid assessment of trial representativeness. This may be used to supplement more complex methods of assessing representativeness. A possible advantage of the observed/expected SAE rate ratio is a single number which provides an integrated measure of the susceptibility of trial participants to SAEs relative to general populations with the relevant index condition. Moreover, it is based on a measure which is a fundamental component of current clinical trial reporting.[25,26] At present however, because there is no benchmark against which to judge the observed SAE count, this potentially useful information on representativeness is opaque . We show that after defining a notional target population to whom trial findings may apply, one can use age and sex specific hospitalisation and death rates for that population to estimate the observed/expected SAEs ratio for individual trials.

Using the observed/expected SAE ratio to assess representativeness of a given trial will require careful consideration of both the population identified from routine care and the arm of the trial used to assess the SAE count. In this analysis, having found no difference on average between treatment and control arms, we used the total SAE count from each trial (across all trial arms). However, for some trials, such an approach may lead to biased estimates if the SAE count was higher or lower in the treatment arm. For example, in trials where the treatment itself is likely to influence SAE rates (e.g. of potentially toxic treatments such as chemotherapy) it may be more appropriate to restrict the trial data to the control arm where the control treatment is more comparable to routine care. To facilitate such comparisons, reporting guidelines should encourage trialists to report the age

distribution, total follow-up time and SAE counts stratified by study arm and by sex. Where there is a difference in SAEs between treatment and control arms, it is likely that the control arm would provide the most meaningful comparison with routine care (particularly where the control involved active treatment reflecting 'usual care'). Selection of the routine care population is also important when considering representativeness of a specific trial. It may be more appropriate to select a routine care comparison to which the trial treatment is likely to be indicated (rather than the broad approach or including all those with the condition of interest). Furthermore, when treatment and control arms both include active treatment (which may influence SAE rates) is may be preferable to compare SAEs with patients receiving comparable treatments in routine care. We did not attempt to make such nuanced judgements for each trial assessed in this analysis, given the broad range of index conditions, treatment indications, and the large number of trials. However, future applications of this approach to assessing representativeness will need to judge the appropriate routine care population and trial arm comparison in the context of the trial(s) being assessed.

Although, for individual trials, combining SAE counts across arm will increase precision of the observed/expected SAEs ratios, particular caution should be employed (i) where the magnitude of difference between the community and trial participants is small (and hence small differences between the arms could have important implications for interpretation), (ii) where there is empirical evidence of differences in rates of SAEs between arms, or (iii) where is reason to believe from external evidence (e.g. biological plausibility or findings from other studies) suggesting that treatment arm is likely to have an effect on SAE rates.

In the hope that other groups will adopt our approach we have provided analysis code, data and a detailed description in the supplementary appendix section.

## Strengths and limitations of this study

Strengths of this analysis include the inclusion of many trials and multiple index conditions and the use of a UK representative routine care population in which expected SAE rates were calculated. However, this broad approach meant that we could not incorporate all characteristics which could affect the risk SAEs (such as socioeconomic status, ethnicity, or severity of the index condition). Had the routine care population been closer to the intended target population for each trial (eg by excluding patients with absolute contra-indications for the drug under study, or selecting only those suitable for second-line therapy where this was the trial indication) it is possible that the heterogeneity in the observed/expected ratio would have been lower.

We used the SAE count for all trial participants, not solely those in the usual care arm, in order to increase the statistical precision with which the observed/expected SAE ratio could be estimated. This means that investigational products not yet widely used in routine care, may have increased the SAE rate within trials. However, we found that SAEs did not differ by treatment arm and that there was no multimorbidity count-treatment interaction. This suggests that, at least for this set of trials we studied, underlying participant characteristics rather than investigational-product related effects were the major driver of SAEs. However, this may not be true of all trials (e.g. those of potentially toxic treatments such as chemotherapy or immunosuppressive treatments, in which a greater proportion of SAEs in the treatment arm are likely to be directly linked to treatment).

The use of individual participant data allowed us to examine associations between multimorbidity and SAEs within trials, and to explore the extent to which the age-sex standardised observed/expected SAE ratio was impacted by accounting for differences in multimorbidity count by trial and target populations. However, as we reported previously,[5,6] our measures of multimorbidity were based on re-analyses of trial data originally collected for purposes other than measuring multimorbidity. As such, the impact of additionally accounting for multimorbidity may have been larger if better measures were available. Furthermore this analysis was limited to trials with IPD

(which, while comparable in terms of size, start year and exclusion criteria, contain fewer phase 4 trials than the wider body of eligible trials) and with a sufficient number of SAEs per trial to allow estimates of associations (meaning that this relationship could not be assessed for some index conditions). A further limitation is that other events than hospitalisation and death, such as prolongation of hospitalisation, also qualify as SAEs. While the contribution of such events was low (between 0% and 13%), this over-counting would nonetheless tend to bias observed/expected SAE ratios upwards, in most cases giving the impression that trial populations were more similar to clinical populations. Finally, the community population was from Wales, UK, whereas the trials were multinational. Some of the difference between trial and routine care rates may therefore reflect differences in population characteristics or health systems. Differences between trials within a given index condition may also reflect differences in their respective healthcare settings. However, the observed/expected ratio did not appear to differ depending on whether trials did or did not have a site in the UK (supplementary appendix) and rates of hospitalisations and deaths in the UK are comparable to other high-income countries where most of the trials were conducted.

## Conclusions

SAE rates in trials are consistently lower than expected. Multimorbidity is associated with SAEs in both trials and routine care, and is less prevalent within trial populations. However, the lower prevalence of multimorbidity in trials only partially explains the difference in SAE rates between trials and routine care, suggesting additional systematic differences between trial and routine care populations. Conventional approaches to assessing trial representativeness are complex, time-consuming, and partial. Our findings suggest that the observed/expected SAE ratio has potential to be a useful metric of trial representativeness to aid interpreting the applicability of trials.

# Additional files

Additional file 1: Supplementary figures S1-S21 - Trial level estimates of observed/expected SAEs

Figure S1: Observed/expected SAEs in asthma trials
Figure S2: Observed/expected SAEs in atrial fibrillation trials
Figure S3: Observed/expected SAEs in axial spondyloarthritis trials
Figure S4: Observed/expected SAEs in benign prostatic hyperplasia trials
Figure S5: Observed/expected SAEs in dementia trials
Figure S6: Observed/expected SAEs in type 2 diabetes mellitus trials
Figure S7: Observed/expected SAEs in epilepsy trials
Figure S8: Observed/expected SAEs in hypertension trials
Figure S9: Observed/expected SAEs in pulmonary hypertension trials
Figure S10: Observed/expected SAEs in inflammatory bowel disease trials
Figure S11: Observed/expected SAEs in myocardial infarction trials
Figure S12: Observed/expected SAEs in osteoarthritis trials
Figure S13: Observed/expected SAEs in osteoporosis trials
Figure S14: Observed/expected SAEs in Parkinson's disease trials
Figure S15: Observed/expected SAEs in psoriasis trials
Figure S16: Observed/expected SAEs in psoriatic arthropathy trials
Figure S17: Observed/expected SAEs in chronic obstructive pulmonary disease trials
Figure S18: Observed/expected SAEs in pulmonary fibrosis trials
Figure S19: Observed/expected SAEs in rheumatoid arthritis trials
Figure S20: Observed/expected SAEs in systemic lupus erythematosus trials
Figure S21: Observed/expected SAEs in thromboembolism trials

Additional file 2: Supplementary figures S22-23 - comparison of SAEs between trial arms
Figure S22: Trial-level comparison of SAE rate between trial arms
Figure S23: Index condition-level meta-analyses of comparison of SAE rate between trial arms

Additional file 3: Supplementary figures S24-S44 - Observed/expected SAEs before and after standardization by multimorbidity count
Figure S24: observed/expected SAEs before and after standardization by multimorbidity count: asthma trials
Figure S25: observed/expected SAEs before and after standardization by multimorbidity count: atrial fibrillation trials
Figure S26: observed/expected SAEs before and after standardization by multimorbidity count: axial spondyloarthritis trials
Figure S27: observed/expected SAEs before and after standardization by multimorbidity count: benign prostatic hyperplasia trials
Figure S28: observed/expected SAEs before and after standardization by multimorbidity count: dementia trials
Figure S29: observed/expected SAEs before and after standardization by multimorbidity count: type 2 diabetes mellitus trials
Figure S30: observed/expected SAEs before and after standardization by multimorbidity count: hypertension trials
Figure S31: observed/expected SAEs before and after standardization by multimorbidity count: pulmonary hypertension trials
Figure S32: observed/expected SAEs before and after standardization by multimorbidity count: inflammatory bowel disease trials

Figure S33: observed/expected SAEs before and after standardization by multimorbidity count: migraine trials
Figure S34: observed/expected SAEs before and after standardization by multimorbidity count: osteoarthritis trials
Figure S35: observed/expected SAEs before and after standardization by multimorbidity count: osteoporosis trials
Figure S36: observed/expected SAEs before and after standardization by multimorbidity count: Parkinson's disease trials
Figure S37: observed/expected SAEs before and after standardization by multimorbidity count: psoriasis trials
Figure S3: observed/expected SAEs before and after standardization by multimorbidity count: psoriatic arthropathy trials
Figure S39: observed/expected SAEs before and after standardization by multimorbidity count: chronic obstructive pulmonary disease trials
Figure S40: observed/expected SAEs before and after standardization by multimorbidity count: pulmonary fibrosis trials
Figure S41: observed/expected SAEs before and after standardization by multimorbidity count: restless legs syndrome trials
Figure S42: observed/expected SAEs before and after standardization by multimorbidity count: rheumatoid arthritis trials
Figure S43: observed/expected SAEs before and after standardization by multimorbidity count: systemic lupus erythematosus trials
Figure S44: observed/expected SAEs before and after standardization by multimorbidity count: thromboembolism trials

Additional file4: Statistical methods

# Abbreviations

CI: Credible interval

CSDR: Clinical Study Data Request

IPD: Individual participant data

SAE: serious adverse event

SAIL: Secure Anonymised Information Linkage

YODA: Yale Open Data Access

# Declarations

## Acknowledgements

This study, carried out under YODA Project # 2017-1746, used data obtained from the Yale University Open Data Access Project, which has an agreement with JANSSEN RESEARCH & DEVELOPMENT, L.L.C.. The interpretation and reporting of research using this data are solely the responsibility of the authors and does not necessarily represent the official views of the Yale University Open Data Access Project or JANSSEN RESEARCH & DEVELOPMENT, L.L.C.. This study was also carried out under ClinicalStudyDataRequest.com project number 1732, used data from the ClinicalStudyDataRequest.com repository, who provided data from Boehringer-Ingelheim, GSK, Lilly, Roche, Takeda, and Sanofi. The interpretation and reporting of research using these data are solely the responsibility of the authors and does not necessarily represent the official views of ClinicalStudyDataRequest.com or Boehringer-Ingelheim, GSK, Lilly, Roche, Takeda or Sanofi.

## Consent for publication

Not applicable

## Funding

David McAllister is funded via an Intermediate Clinical Fellowship and Beit Fellowship from the Wellcome Trust, who also supported other costs related to this project such as data access costs and database licences ("Treatment effectiveness in multimorbidity: Combining efficacy estimates from clinical trials with the natural history obtained from large routine healthcare databases to determine net overall treatment Benefits." - 201492/Z/16/Z). Peter Hanlon is funded through a Clinical Research Training Fellowship from the Medical Research Council (Grant reference: MR/S021949/1). None of the funders had any influence over the study design, analysis or decision to submit for publication.

## Availability of data and materials

Aggregated data and code required to run these models, along with full model descriptions, are available at https://github.com/dmcalli2/sae_ctg_multicond_public

## Authors' contributions

DM, PH, AS, BG, SD and NW conceived the study. DM acquired the data from trials and SAIL. PH, DM, ED and LH processed the data. DM and PH conducted the statistical analysis and interpretation of the data. NW and SD advised on the statistical analysis. PH wrote the first draft with support from DM. EB, AS, LH, SW, BG, FSM, SD, NW and DM reviewed the manuscript and made critical changes for intellectual content. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Ethical approval and consent to participate

This project had approval from the University of Glasgow, College of Medicine, Veterinary and Life Sciences ethics committee (200160070). SAIL analyses were approved by SAIL Information Governance Review Panel (Project 0830).

# References


1. He J, Morales DR, Guthrie B. Exclusion rates in randomized controlled trials of treatments for physical conditions: a systematic review. *Trials* 2020; **21**(1): 1-11.

2. Naidoo N, Nguyen VT, Ravaud P, et al. The research burden of randomized controlled trial participation: a systematic thematic synthesis of qualitative evidence. *BMC Med* 2020; **18**(1): 6.

3. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *Jama* 2007; **298**(10): 1209-12.

4. Boyd CM, Kent DM. Evidence-based medicine and the hard problem of multimorbidity. *Journal of general internal medicine* 2014; **29**(4): 552.

5. Hanlon P, Butterly E, Lewsey J, Siebert S, Mair FS, McAllister DA. Identifying frailty in trials: an analysis of individual participant data from trials of novel pharmacological interventions. *BMC Med* 2020; **18**(1): 1-12.

6. Hanlon P, Hannigan L, Rodriguez-Perez J, et al. Representation of people with comorbidity and multimorbidity in clinical trials of novel drug therapies: an individual-level participant data analysis. *BMC Med* 2019; **17**(1): 201.

7. US Food and Drug Administration Authority. What is a serious adverse event? . https://www.fda.gov/safety/reporting-serious-problems-fda/what-serious-adverse-event

8. Hanlon P, Corcoran N, Rughani G, et al. Observed and expected serious adverse event rates in randomised clinical trials for hypertension: an observational study comparing trials that do and do not focus on older people. *The Lancet Healthy Longevity* 2021.

9. Assessing heterogeneity in treatment efficacy by age, sex and multimorbidity. PROSPERO 2018 CRD42018048202. 2018. http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018048202.

10. Lyons RA, Jones KH, John G, et al. The SAIL databank: linking multiple health and social care datasets. *BMC medical informatics and decision making* 2009; **9**(1): 1-8.

11. Zulman DM, Sussman JB, Chen X, Cigolle CT, Blaum CS, Hayward RA. Examining the evidence: a systematic review of the inclusion and analysis of older adults in randomized controlled trials. *Journal of general internal medicine* 2011; **26**(7): 783-90.

12. Humphreys K, Weisner C. Use of exclusion criteria in selecting research subjects and its effect on the generalizability of alcohol treatment outcome studies. *American Journal of Psychiatry* 2000; **157**(4): 588-94.

13. Fortin M, Dionne J, Pinho G, Gignac J, Almirall J, Lapointe L. Randomized controlled trials: do they have external validity for patients with multiple comorbidities? *The Annals of Family Medicine* 2006; **4**(2): 104-8.

14. National Institute for Health and Care Excellence. Multimorbidity: clinical assessment and management. NICE guideline [NG56]. 2016. https://www.nice.org.uk/guidance/conditions-and-diseases/multiple-long-term-conditions.

15. Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *Jama* 2007; **297**(11): 1233-40.

16. Garcia-Doval I, Carretero G, Vanaclocha F, et al. Risk of Serious Adverse Events Associated With Biologic and Nonbiologic Psoriasis Systemic Therapy: Patients Ineligible vs Eligible for Randomized Controlled Trials. *Archives of Dermatology* 2012; **148**(4): 463-70.

17. Mason KJ, Barker J, Smith CH, et al. Comparison of Drug Discontinuation, Effectiveness, and Safety Between Clinical Trial Eligible and Ineligible Patients in BADBIR. *JAMA Dermatol* 2018; **154**(5): 581-8.

18. Yiu ZZN, Mason KJ, Barker JNWN, et al. A standardization approach to compare treatment safety and effectiveness outcomes between clinical trials and real-world populations in psoriasis. *Br J Dermatol* 2019; **181**(6): 1265-71.

19. Jani BD, Hanlon P, Nicholl BI, et al. Relationship between multimorbidity, demographic factors and mortality: findings from the UK Biobank cohort. *BMC Med* 2019; **17**(1): 74.

20. Nunes BP, Flores TR, Mielke GI, Thumé E, Facchini LA. Multimorbidity and mortality in older adults: a systematic review and meta-analysis. *Archives of gerontology and geriatrics* 2016; **67**: 130-8.

21. Cassell A, Edwards D, Harshfield A, et al. The epidemiology of multimorbidity in primary care: a retrospective cohort study. *British Journal of General Practice* 2018; **68**(669): e245-e51.

22. Payne RA, Mendonca SC, Elliott MN, et al. Development and validation of the Cambridge Multimorbidity Score. *CMAJ* 2020; **192**(5): E107-E14.

23. Hanlon P, Jani BD, Nicholl B, Lewsey J, McAllister DA, Mair FS. Associations between multimorbidity and adverse health outcomes in UK Biobank and the SAIL Databank: A comparison of longitudinal cohort studies. *PLos Med* 2022; **19**(3): e1003931.

24. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *bmj* 2015; **350**.

25. Ioannidis JP, Evans SJ, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Annals of internal medicine* 2004; **141**(10): 781-8.

26. US Food and Drug Administration Authority. Guidance for Industry and Investigators: Safety reporting requirements for INDs and BA/BE studies. 2012. https://www.fda.gov/files/drugs/published/Safety-Reporting-Requirements-for-INDs-%28Investigational-New-Drug-Applications%29-and-BA-BE-%28Bioavailability-Bioequivalence%29-Studies.pdf.