This is a repository copy of *Improving the Estimation of Subgroup Effects for Clinical Trial Participants with Multimorbidity by Incorporating Drug Class-Level Information in Bayesian Hierarchical Models : A Simulation Study*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/175536/

Version: Accepted Version

Article:

Hannigan, Laurie, Phillippo, David, Hanlon, Peter et al. (6 more authors) (Accepted: 2021) Improving the Estimation of Subgroup Effects for Clinical Trial Participants with Multimorbidity by Incorporating Drug Class-Level Information in Bayesian Hierarchical Models : A Simulation Study. Medical Decision Making. ISSN 1552-681X (In Press)

https://doi.org/10.1177/0272989X211029556

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Improving the estimation of subgroup effects for clinical trial participants with multimorbidity by incorporating drug class-level information in Bayesian hierarchical models: a simulation study

Running head

Improving estimation of subgroup effects using drug class information

Word count

4785

Author list

Laurie J. Hannigan (PhD) [1, 2, 3], David M. Phillippo (PhD) [2], Peter Hanlon (MSc) [3], Laura Moss (PhD) [4, 5], Elaine W. Butterly (MD) [3], Neil Hawkins (PhD) [3], Sofia Dias (PhD) [6], Nicky J. Welton (PhD) [2], David. A. McAllister (MD) [3]

[1] Nic Waals Institute, Lovisenberg Diaconal Hospital, Norway
[2] Population Health Sciences, Bristol Medical School, University of Bristol, UK
[3] Institute of Health and Wellbeing, University of Glasgow, UK
[4] NHS Greater Glasgow & Clyde, UK
[5] School of Medicine, University of Glasgow, UK
[6] Centre for Reviews and Dissemination, University of York, UK

Correspondence

Dr David A McAllister
Institute of Health and Wellbeing
University of Glasgow
1 Lilybank Gardens
Glasgow, UK
G12 8RZ
e-mail: David.McAllister@glasgow.ac.uk
Telephone: +441413301663

Financial support for this study was provided entirely by a grant from the Wellcome Trust [201492/Z/16/Z]. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Abstract (275/275 words)

Background

There is limited guidance for using common drug therapies in the context of multimorbidity. In part, this is because their effectiveness for patients with specific comorbidities cannot easily be established using subgroup analyses in clinical trials. Here, we use simulations to explore the feasibility and implications of concurrently estimating effects of related drug treatments in patients with multimorbidity by partially pooling subgroup efficacy estimates across trials.

Methods

We performed simulations based on the characteristics of 161 real clinical trials of non-insulin glucose lowering drugs for diabetes, estimating subgroup effects for patients with a hypothetical comorbidity across related trials in different scenarios using Bayesian hierarchical generalised linear models. We structured models according to an established ontology – the World Health Organisation Anatomic Chemical Therapeutic Classifications (WHO-ATC) – allowing us to nest all trials within drugs and all drugs within ATC classes, with effects partially pooled at each level of the hierarchy. In a range of scenarios, we compared the performance of this model to random effects meta-analyses of all drugs individually.

Results

Hierarchical, ontology-based Bayesian models were unbiased and accurately recovered simulated comorbidity-drug interactions. Compared to single drug meta-analyses, they offered a relative increase in precision of up to 250% in some scenarios due to information sharing across the hierarchy. Due to the relative precision of the approaches, a high proportion of small subgroup effects were only detectable using the hierarchical model.

Conclusions

By assuming that similar drugs may have similar subgroup effects, Bayesian hierarchical models based on structures defined by existing ontologies can be used to improve the precision of treatment efficacy estimates in patients with multimorbidity – with potential implications for clinical decision-making.

Keywords

Subgroup analysis; individual-patient data meta-analysis; multimorbidity; hierarchical modelling; medical ontologies

Introduction

Multimorbidity, which is defined as the presence of two or more chronic conditions within an individual, is common and increasing. More than half of patients with any chronic disease have multimorbidity.(1) This represents a challenge because the applicability of clinical trial results to patients with multimorbidity is uncertain. Consequently, several clinical guideline bodies have urged caution in applying trial results to patients with multimorbidity,(2) while in practice patients with multimorbidity are less likely to receive drug-treatments shown to be effective in clinical trials, even where there is no contraindication to therapy.(3–6)

One reason for this uncertainty is that multimorbidity is under-represented in clinical trials.(7,8) For this reason, some researchers have used observational data – particularly administrative data, in which multimorbidity is common – to estimate treatment effects. However, such "pharmaco-epidemiological" approaches are subject to confounding by indication (9) despite methodological advances, (10,11) and so remain restricted in terms of their utility to support medical decision making in this regard.

Moreover, while multimorbidity may not be present at the same rate in clinical trials compared to in the community, it is nonetheless common. For half of 22 medical conditions, we found that at least a third of trial participants – in standard industry-funded clinical trials – had multimorbidity. Furthermore, similar comorbidities were common in the trial and community settings.(8) Consequently, there is both a need and an opportunity to determine whether treatment effects in clinical trials differ for sub-groups of patients with and without multimorbidity, and for different patterns of multimorbidity.

Reliably estimating treatment effects for sub-groups in individual clinical trials is notoriously difficult.(12–14) Claims of sub-group effects made in clinical trial reports are frequently unsupported by appropriate statistical evidence (15). While pre-specified sub-group analyses can be adequately powered, there are often insufficient numbers of participants to estimate differences in effects across sub-groups (especially for specific comorbidities) with adequate precision to inform clinical decision-making.(13) Moreover, simple methods to reduce the risk of false positives (i.e., by asserting that there is heterogeneity when none exists) do so at the expense of precision and increase in Type 2 errors.(16) Consequently, attempts to estimate treatment effects for patients with multimorbidity are likely to suffer from both poor sensitivity and poor specificity.

Meta-analyses pool findings across trials to improve precision,(17) and individual patient data (IPD) meta-analyses can be used to pool treatment effect estimates for participants with specific characteristics such as particular comorbidities.(18) Even for meta-analyses, however, estimating sub-group effects with sufficient precision to inform clinical decision making is challenging because, compared to the overall trial, data on particular sub-groups can be limited.

One approach to dealing with limited data is to use hierarchical modelling.(19) Within a Bayesian framework, hierarchical modelling is straightforward,(20) and has previously been shown to be useful for analysing clinical trial data. Examples include performing subgroup analyses,(19) and estimating adverse treatment effects.(21) Such approaches rely on the assumption that information can be shared between parameters. In an information-sharing approach to sub-group analysis, Dixon and Simon(20) assumed that treatment-covariate interactions came from a common prior distribution. Similarly, in estimating effects of treatments on adverse events, Berry and Berry(21) assumed that events occurring within specific body systems (e.g., the gastrointestinal system) were related. In both examples, separate estimates were "partially pooled", increasing precision and attenuating extreme values towards the group-level mean (shrinkage). Partial pooling and shrinkage are established features of hierarchical models.(22) These are desirable features for subgroup analyses as, where the assumption that information can be shared holds, they are likely to improve our ability to detect true subgroup effects, while reducing false positives.

Despite these desirable properties, the use of hierarchical modelling in subgroup analyses has thus far has been limited. One reason for the limited adoption may be uncertainty in how to allow sharing of information between different trials – that is, how should hierarchical models be structured. Using established drug-related ontologies such as the World Health Organisation Anatomic Chemical Therapeutic Classifications (WHO ATC), which is a tree-like classification scheme based on therapeutic indications and chemical forms,(23) and MED-RT, a US-based ontology which provides finer granularity for mechanisms of action,(24) may help overcome this barrier. Such ontologies represent expert knowledge about similarities, differences, and relationships between different drugs in terms of indications, chemical structures, and other features, providing a starting point from which to define a hierarchical structure for modelling.

In other fields, relationships within ontologies have been used to predict protein–protein interactions, diagnoses and the classification of chemicals.(25) Ontologies have also been exploited to support the management and execution of clinical trials.(26–29) Since WHO-ATC, MED-RT and other ontologies are publicly available, they provide a transparent starting point for analyses. This aspect of ontologies is appealing in the field of clinical trial meta-analysis where transparency, consistency and pre-specification are highly prized.(30)

The current study

In this study, we address the question of whether partial-pooling of subgroup effects in existing clinical trial data – using structures borrowed from established drug classification ontologies – is feasible and has the potential to support clinical decision making. To do this, we first simulate datasets with interactions between a group of non-insulin glucose lowering drugs for diabetes and a single hypothetical comorbidity, based on the characteristics of real trials. Next, we apply Bayesian hierarchical generalized linear models, with individual trials nested within drugs nested within ATC drug classes, to these data. Our use of an established ontology to structure a hierarchical meta-analysis is based on the simple assumption that drugs that are similar may behave similarly in subgroups. We compare the performance of these ontology-based hierarchical models, in terms of their recovery of comorbidity-treatment interaction effects for individual drugs, with that of standard, single-drug meta-analyses (see **Figure 1** for an overview). In addition, we highlight specific properties of these models that emphasize their potential utility in IPD meta-analyses of comorbidity-based sub-group effects within clinical trial data.

Methods

Identification and classification of existing trials as basis for simulation

We opted to base our simulations on the characteristics of real trials of an exemplar drug grouping: non-insulin glucose lowering drugs for diabetes. We identified all relevant existing trials on the US clinical trials register (clinicaltrials.gov) that met a set of pre-specified selection criteria [Prospero protocol CRD42018048202 (31)]. Briefly, these included a minimum enrolment of 300 participants, a start date of 1 January 1990 or later, being a phase 2/3, 3, or 4 trial, and having an upper age limit of 60 years or more. We used trial-level descriptive information that is publicly available on their clinicaltrials.gov record trials to define the structure of the simulated data to reflect, as closely as possible, the characteristics of real IPD that is (theoretically) available from trial sponsors. Specifically, we obtained information about the number of trials available per drug and class, and the number of participants enrolled in each trial, and used these characteristics as the basis of our simulations of subgroup effects for each trial. For simplicity, all trials were treated as single-arm versus placebo/usual care in the simulation.

After classification according to the WHO-ATC ontology, we included 161 trials involving 210,046 participants, of 24 separate drugs from 7 different WHO-ATC 5-level classes (e.g., DPP-4 inhibitors, SGLT-2 inhibitors). Full details of the classifications are provided in **sAppendix 1** and **sTable 1**.

Data generation procedure

We simulated data to generate trial-level subgroup effects for each of the 161 trials. This was to reflect a situation where individual patient level data for these trials had been shared, and the effect of an interaction between a particular comorbidity and the drug treatment under investigation had been estimated for each trial in preparation for meta-analysis – which is a common analytical approach in IPD meta-analysis. (17)

Data were simulated based on an overall comorbidity-treatment interaction of -0.1 standard deviations at the level of the wider drug grouping (i.e., the top level of the hierarchy, reflecting the average interaction effect across all drugs). This was chosen as a minimum difference which might plausibly be important for decision-making, recognising that sub-group interactions are likely to be modest in real applications. This effect size would mean that, for a treatment minus control arm difference in efficacy of 0.2 standard deviations, the treatment efficacy in patients with multimorbidity would be 0.1 standard deviations.

Trial-level effects were simulated by adding random variation around the overall comorbidity-treatment interaction effect at each level of the hierarchy (i.e., at the level of drug class, drug, and trial). We simulated 1000 datasets for each of a range of scenarios, reflecting different degrees of between-trial, between-drug, and between-class variability:

All levels: low variation

In this scenario, we simulated 1000 datasets with trial-level interaction effects by adding random variation of 0.05 SDs at the levels of drug class, drug, and trial to the fixed overall effect of -0.10 SDs. Datasets in this scenario represent situations where all trials of a given drug, all drugs in a given class, and all classes of drugs in the hierarchy have highly similar estimates for a given comorbidity-treatment interaction effect.

All levels: medium variation

In this scenario, we simulated 1000 datasets with trial-level interaction effects by adding random variation of 0.15 SDs at the level of drug class, drug, and trial to the fixed overall effect of -0.1 standard deviations. Datasets in this scenario represent situations where all trials of a given drug, all drugs in a given class, and all classes of drugs in the hierarchy have moderately similar estimates for a given comorbidity-treatment interaction effect.

All levels: high variation

In this scenario, we simulated 1000 datasets with trial-level interaction effects by adding random variation of 0.25 SDs at the level of drug class, drug, and trial to the fixed overall effect of -0.1 standard deviations. Datasets in this scenario represent situations where all trials of a given drug, all drugs in a given class, and all classes of drugs in the hierarchy have relatively dissimilar estimates for a given comorbidity-treatment interaction effect.

Other scenarios: variation manipulated at a specific level of the hierarchy

We additionally simulated sets of 1000 datasets in scenarios where, during the data generation procedure, we manipulated variation at each level of the hierarchy in turn, while keeping variation at the other levels constant at 0.05 SDs. So, for example, this allowed us to represent situations where trial-level estimates of comorbidity-treatment interactions for a given drug were highly dissimilar, but drugs and drug classes behaved more consistently (i.e., a "Trial-level: high variation" scenario, where trial-level interaction effects were simulated by adding random variation of 0.05 SDs at the level of drug class and drug, but 0.25 SDs at the level of trial, to the fixed overall effect of -0.1 standard deviations.)

In the main analyses, we assumed the prevalence of the comorbidity that defines the subgroup to be 20%. This value is used in determining the precision of the simulated trial-level interaction estimate, which is also based on the number of individuals enrolled in the trial and is the same across datasets and scenarios. Further details of the simulation procedure are given in **sAppendix 2** and an abbreviated example of a simulated dataset is presented in **sTable 2**.

Modelling

To each simulated dataset, we fitted: i) a hierarchical generalized linear model with all trials nested within drugs, nested within ATC-5 drug classes (henceforth "the full model"); and ii) hierarchical generalized linear models for all trials of each of the 24 drugs (henceforth "single-drug models"). We fit these models using the R-INLA package.(32) Although integrated nested Laplacian approximation (INLA) performs approximate Bayesian inference, and offers less flexibility than software which uses Markov-chain Monte Carlo (MCMC) methods to fit models (specifically, hyperpriors must be Gaussian – an acceptable restriction in this case), model-fitting using R-INLA is very rapid, and gave good agreement to MCMC. Using R-INLA meant we could run the models on a larger number of simulated iterations and scenarios.

Full model description

Interactions at the various levels of the hierarchy were specified as follows:

Trial-specific comorbidity-treatment interactions

$$y_{z,d,c} \sim N(\mu_{z,d,c}, s^2_{z,d,c})$$

Between-trial variation in comorbidity-treatment interactions

$$\mu_{z,d,c} \sim N(\beta_{d,c}, \tau^2_{d,c})$$

Between-drug variation in comorbidity-treatment interaction

$$\beta_{d,c} \sim N(\gamma_c, \sigma^2_c)$$

Between-class comorbidity-treatment interaction

$$\gamma_c \sim N(\alpha, \zeta^2)$$

The observed quantities y and s represent the comorbidity-treatment interaction and standard error at the level of the individual trial. Normal distributions are parameterised as mean and variance. The z subscript indicates the trial, the d subscript indicates the drug, and the c subscript indicates the drug class. The prior for the overall mean comorbidity-treatment interaction α was a Normal distribution $[N(0, 2^2)]$. This was chosen to correspond to the assumption that covariate treatment interactions are uncommon (when trial data are analysed on an appropriate scale). For the between drug class variation ζ , the between-drug (i.e., within-class) variation (σ_c) and the between-trial (i.e., within-drug) variation $(\tau_{d,c})$ we used half-Normal priors on the standard deviations: $halfN(0, 1^2)$. The priors were selected to be relatively non-informative in relation to the values for variance at these levels used during the data generation (0.05^2, 0.15^2, 0.25^2), in order to ensure that the performance of the full model was not artificially aided by our knowledge of these values.

Single-drug models were specified using the lowest two levels of the full model (i.e., trial-specific and between-trial) as outlined above, and with the prior for the mean comorbidity-treatment interaction for a given drug $\beta_{d,c}$ parameterized as a Normal distribution $[N(0, 2^2)]$.

Performance evaluation and sensitivity testing

We evaluated the performance of the full model against that of the single drug models on their recovery of the drug-level interaction effect. In accordance with the framework outlined by Morris et al (33), we compare the two approaches on several established performance measures: bias (the extent to which the effect is systematically over-/underestimated); mean squared error (MSE; the average extent to which the effect is over or underestimated) and root mean squared error (RMSE; equivalent to MSE but interpretable on the same scale as the data); change in precision relative to the single drug model; and coverage (the proportion of credible intervals containing the true value). We used the R package *rsimsum* (34)to derive estimates and Monte Carlo standard errors for each of the measures (RMSE was derived manually and standard errors approximated using the delta method).

To evaluate the sensitivity of the models to the prevalence of the subgroup-defining comorbidity, we re-ran all analyses with this value set at 10 and 50% respectively.

The R code for the simulation and modelling are available at https://github.com/dmcalli2/simlt_interactions/blob/master/scripts/.

Results

The relative performance of the full and single drug models is summarized in **Table 1**. Performance measures are aggregated across datasets and scenarios according to amount of variability around the overall average interaction effect of -0.1 that was introduced at each level of the hierarchy during data generation, as described in the Data generation procedure section above. As such, the top section of the table shows results for all datasets in three main scenarios: “All levels: low variation”, “All levels: medium variation”, “All levels: high variation”. In the lower section of the table, results are summarized for scenarios reflecting the effects of increasing variation at specific levels of the hierarchy.

The full model estimated drug level comorbidity-treatment interaction effects without bias to the same extent as drug only models. MSE/RMSE values were similar in the full models and drug models in all cases, indicating that the degree of accuracy of the point estimates was at least equivalent. The largest difference in accuracy occurred in the “Trial level: high variation” scenario, where simulated trial-level effects were highly variable but drugs and drug classes relatively similar, when the full model was more precise by approximately 0.05 SDs ($RMSE_{drug} = 0.13$; $RMSE_{full} = 0.08$). The models differed more markedly on the other two measures of performance, precision and coverage, for related reasons. The full model estimated drug level comorbidity-treatment interactions, on average, more precisely in all scenarios, and substantially so in most cases. This is the expected result of information sharing at the level of drug class. The relative precision of estimates from the two approaches is illustrated, as a function of drug class, in **Figure 2**. Precision gains related to use of the full model are most substantial for drugs with a limited number of trials (or only small trials; see **sTable 1** for drug-specific details), and when drugs and drug classes are more similar and trial-level estimates more varied (e.g., “Trial level: high variation” scenario, middle panel). Precision in the full model was similar to or worse than in the drug model in all classes only when drug classes in the hierarchy were relatively dissimilar (e.g., “Class level: high variation” scenario, bottom-right panel).

Coverage – the proportion of credible intervals including the “true” effect – was reduced in most instances in the full model, but this too is an expected feature of these models. It results from the combination of increased precision and shrinkage of extreme-for-class drug level effect estimates toward the class average. These features are illustrated in **Figure 3**, which shows the posterior distributions of effects of drugs in a specific class, as estimated in the full model (middle panel) and single drug models (lower panel), as they relate to the effect at the class level effect (top panel). Drug level effect distributions are shrunk (drawn towards the class level mean) and estimated more precisely in the full model when drugs in the hierarchy are sufficiently similar (e.g., “All levels: low variation” scenario, left-hand panels of **Figure 3**). This means that the simulated effect for a given drug has more chance of falling outside the 95% credible intervals – but this is clearly desirable if the exchangeability assumption is met, as information from similar drugs has been used alongside the evidence available from trials of that drug to improve the estimate. In higher variation scenarios, the extent to which drug-level estimates are influenced by class-level information is flexible and proportionate to the homogeneity of effects within the class. In the example shown in **Figure 3** in the “All levels: high variation” scenario (right-hand panels), shrinkage is minimal and only effects for gemigliptin and linagliptin are estimated more precisely in the full model, reflecting the fact that drugs and classes in this scenario are much less similar in terms of their subgroup effects.

Figure 4 illustrates the potentially clinically-meaningful impact of the increases in precision afforded by using the full model to estimate treatment interactions for related drugs in a hierarchy. It summarizes the proportion of all datasets in three main scenarios, in which a true drug-level subgroup effect of -0.10 or larger was able to be detected (i.e., with credible intervals not including zero) in i) both models; ii) the single drug model only; and iii) the full model only. For context, the panel on the right-hand side of the figure shows enrollment (i.e., *N*) for the largest trial per drug and aggregated across all trials of a drug. Effects in drugs with large trials and/or high aggregated enrollment were generally well-detected in both models, though a substantial proportion were only detected in the full model, especially in the "All levels: low variation" scenario. The drug-class level information-sharing in the full model was most beneficial for drugs with smaller/fewer trials (e.g., taspoglutide, and all drugs in classes A10BB and A10BX), where true effects were only detected in the full model, regardless of the extent of variation in the hierarchy.

The results of sensitivity analyses for different rates of subgroup-defining comorbidity prevalence are presented in the supplement (**sTables 3-4 and sFigures 1-4**).

Discussion

In this paper, we have demonstrated the feasibility of improving the estimation of treatment effects in sub-groups by using Bayesian hierarchical meta-analytic models that share information across related trials based on established classification ontologies. Our simulations, based on characteristics of real trials of non-insulin glucose lowering drugs for diabetes, show that partial-pooling of subgroup effects across classes of drugs is: a) feasible, given the amount of data that is theoretically available from trial sponsors; and b) effective at increasing the potential of sub-group effect estimation in the context of multimorbidity to influence clinical decision-making.

In our simulations, Bayesian hierarchical models structured around the ATC ontology were unbiased, and compared favourably to standard meta-analytic approaches in terms of both their precision (estimates are more precise) and conservatism (extreme estimates at drug and trial levels are shrunk towards class means) for estimating subgroup effects. Both of these represent non-trivial improvements in the estimation of drug level subgroup effects, as lack of data from individual trials/standard meta-analyses has typically meant that estimates are often too imprecise to be clinically useful and concerns about 'false positives' are commonly expressed in the literature around subgroups (12,35). More precisely and reliably estimated subgroup effects have greater potential to be incorporated into guidelines and influence clinical decision-making. This is particularly important in the context of multimorbid patients – who represent more than 50% of individuals with any chronic condition – since current guidelines lack specific trial-based recommendations for the treatment of these individuals. (2)

The core features of the model we have outlined will not be novel to anyone familiar with Bayesian hierarchical modelling, and with concepts such as shrinkage and exchangeability. Such readers will also likely be aware of the difficulties inherent in formalizing prior knowledge for use in such models. We propose that the use of existing ontologies – specifically, though not exclusively, of drugs – such as the ATC system, to structure hierarchical models for meta-analyses of trial data is a widely applicable solution to this problem. In particular, we believe it to be immediately applicable to the challenge of estimating treatment effects in sub-groups of patients likely to be poorly represented in clinical trials, such as those with specific comorbidities. To illustrate the portability of this approach and make it easier for others to use the WHO-ATC classification system, we have developed an online application (**Figure 5**) that can be used to visualise hierarchical classifications for a large set of trials registered in the clinicaltrials.gov database with relevant metadata. The tool is available at https://ihwph-hehta.shinyapps.io/duk_example_app/. Users can select trial types, wider drug groupings, and conditions of interest to create a hierarchy that can then be visualised in different ways, and for which the constituent trials (complete with NCT ID numbers) can be exported as a table. The tool can also be used to visualize networks of trials including drug-drug comparisons, and the principles of the model evaluated in this paper can be straightforwardly extended to perform network meta-analyses including data from such trials. The R code for the diagram is also available https://github.com/dmcalli2/ctg_network_diagram. We anticipate continuing to update this tool using more recent data from clinicaltrials.gov.

An advantage of using existing ontologies such as the ATC system,(23,24) is that they have already codified a considerable body of expert knowledge about similarities/differences between different drugs. However, such ontologies might be used with modifications in real settings, where a decision may be made to exclude a drug from a given class or exclude a class from the modelling. If, for example, a new class of drug was developed to address a perceived loss of efficacy in a particular subgroup (e.g., there is some evidence that certain classes of antiplatelet have a lower relative efficacy in women than in men (36)) it would not be appropriate to include other drug-classes with the same physiological action within the modelling. Future work could also explore the use of more complex relationships between drugs, by incorporating multiple ontologies.

Limitations and assumptions

The core assumption of this approach is that partially pooling interaction estimates across different drugs and, especially, across drug classes, is reasonable. When considering the validity of this assumption, it is worth taking into account the context for examining treatment effects in multimorbidity. In current practice, imprecise covariate-treatment interactions are typically either interpreted as evidence that no difference exists, or as evidence that the treatment is not efficacious in patients with the multimorbidity, often according to some unstated prior belief. More precise estimates can be obtained from large observational datasets; however, such analyses are subject to confounding by indication, which has been called an “intractable” problem of epidemiology.(9) Secondly, it is worth reiterating that the flexibility of these models means that hierarchical structures can be defined (and subsequently refined) based on expert opinion and empirical evidence regarding the validity of the core assumption for specific drug groupings. We anticipate that sensitivity analyses involving dropping specific classes and drugs from a hierarchy and comparing model fit will become a standard facet of this approach, but acknowledge that further work is needed to develop formal assumption testing measures for use with real data. In particular, it will be important to develop contingencies to ascertain when a drug level estimate is extreme because that drug truly behaves differently from others in its class, and hence when the shrinkage afforded by a Bayesian hierarchical model is undesirable. Nonetheless, it should be borne in mind that an implicit assumption of single-drug meta-analyses is that drugs with similar mechanisms of action are no more likely to have similar sub-group effects associated with them than those operating via entirely different biological pathways. This assumption, were it to be made explicitly each time a single drug meta-analysis is performed, would likely be at least as debatable – if not more so – than the notion that related drugs may behave similarly to one another.

The scale of IPD sharing that is required for network meta-analyses is clearly greater than that which is needed for individual-drug meta-analyses. However, an important facet of the models we propose is that their benefits can be propagated to future work. Once a large network meta-analysis has been run, the posterior distributions of effects at drug class and drug levels can be used as priors in subsequent analyses. Indeed, as more trial sponsors provide access to individual-level participant data for increasing numbers of trials (e.g., via ClinicalStudyDataRequest.com (37)) it is possible to envisage the eventual compiling of a database of 'off-the-shelf' priors for treatment-comorbidity interactions, which will enable health economists and others to more easily model the effect of treatments in people with multimorbidity.

The simulations in our study are subject to certain limitations. First, although we restricted our simulation to the trial-level (rather than simulating IPD at the patient level) for computational reasons, we have only considered a situation where IPD are available for all trials. That is, IPD would almost certainly be needed from all trials to get results stratified by particular comorbidities. To more pragmatically reflect the likely availability of data from clinical trials, it would be useful to explore models designed to accommodate aggregate data alongside IPD. However, one issue that this would exacerbate is inconsistency of reporting of covariates. Given that covariate reporting is likely to be missing not-at-random, such models would need to account for bias or rely on specific covariate results being obtainable from sponsors (at an aggregate level) on request. Even within IPD, trials may not consistently record or define specific covariates, and the impact of these potential inconsistencies are not considered here. However, in the case of multimorbidity at least, we have recently demonstrated using IPD for over 100 trials shared by commercial sponsors, that it is possible to use generally well-recorded concomitant medication use data to facilitate the investigation of comorbidities.(8) Second, for simplicity we considered only a single comorbidity-treatment interaction. It would be useful in future studies to consider multiple comorbidities. This would mean simulating the impact of between-trial information sharing in models where there is also within-trial sharing via, for example, the Dixon-Simon model, where a common prior is placed on all treatment-covariate interactions.(20) Third, there are a range of important possible scenarios that we do not address in the current simulation. These include scenarios (i) with a smaller overall interaction effect (including no interaction), (ii) where an interaction effect is differs in magnitude or direction across classes within a hierarchy, and scenarios where comorbidities are absent in some trials. These (and many others) are relevant and realistic considerations for the challenges that real data may pose. However, the multiplicities created by so many possible scenarios are a limitation for all simulation studies, and the drawing up of bounds on the simulated universe(s) to be investigated is an inherent part of study design. In future, potentially informed by the characteristics of real IPD where it is obtained, explorations of the capability of this approach to be informative in different scenarios would undoubtedly be beneficial. We have published our code which we or others could modify to examine these, and many other scenarios in future.

Summary and conclusions

Determining treatment effectiveness in multimorbidity is a challenging problem. If we are willing to assume - informed by existing ontologies – a level of similarity between drugs, hierarchical models can be used to estimate comorbidity-treatment interactions with improved precision. This has the potential to support trial-based decision making for patients with multimorbidity.

Acknowledgements

LJH was supported during the completion of this work by a grant from the Norwegian South-Eastern Regional Health Authority (2018059). DMP was supported by MRC grants MR/P015298/1 and MR/R025223/1. NJW was part-supported by the NIHR Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care. DM is supported by an Intermediate Clinical Fellowship and Beit Fellowship from the Wellcome Trust [201492/Z/16/Z].

Declaration of conflicting interests

DMP reports grants from the Medical Research Council during the conduct of the study, honoraria from the Association of the British Pharmaceutical Industry and personal fees from UCB and Bristol Myers-Squibb outside of the submitted work.

References

1. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. Lancet. 2012;380(9836):37–43.
2. Guiding Principles for the Care of Older Adults with Multimorbidity: An Approach for Clinicians. J Am Geriatr Soc. 2012;
3. Bursi F, Vassallo R, Weston SA, Killian JM, Roger VL. Chronic obstructive pulmonary disease after myocardial infarction in the community. Am Heart J. 2010 Jul;160(1):95–101.
4. Wang PS, Avorn J, Brookhart MA, Mogun H, Schneeweiss S, Fischer MA, et al. Effects of noncardiovascular comorbidities on antihypertensive use in elderly hypertensives. Hypertension. 2005;
5. Smith DJ, Martin D, McLean G, Langan J, Guthrie B, Mercer SW. Multimorbidity in bipolar disorder and undertreatment of cardiovascular disease: a cross sectional study. BMC Med. 2013 Dec 23;11(1):263.
6. Quipourt V, Jooste V, Cottet V, Faivre J, Bouvier AM. Comorbidities alone do not explain the undertreatment of colorectal cancer in older adults: A French population-based study. J Am Geriatr Soc. 2011;
7. Johnston MC, Crilly M, Black C, Prescott GJ, Mercer SW. Defining and measuring multimorbidity: A systematic review of systematic reviews. Eur J Public Health. 2019;
8. Hanlon P, Hannigan L, Rodriguez-Perez J, Fischbacher C, Welton NJ, Dias S, 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 Dec 12;17(1):201.
9. Langan SM, Schmidt SA, Wing K, Ehrenstein V, Nicholls SG, Filion KB, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ. 2018;363:k3532.
10. Nørgaard M, Ehrenstein V, Vandenbroucke JP. Confounding in observational studies based on large health care databases: Problems and potential solutions – a primer for the clinician. Clin Epidemiol. 2017;9:185–93.
11. Izem R, Liao J, Hu M, Wei Y, Akhtar S, Wernecke M, et al. Comparison of propensity

score methods for pre-specified subgroup analysis with survival data. J Biopharm Stat. 2020;30(4):734–51.

12. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; J Clin Epidemiol. 2004 Mar;57(3):229–36.
13. Alosh M, Huque MF, Bretz F, D'Agostino RB. Tutorial on statistical considerations on subgroup analysis in confirmatory clinical trials. Stat Med. 2017 Apr 15;36(8):1334–60.
14. Wang R, Drazen JM, Lagakos SW, Ware JH, Hunter DJ. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. N Engl J Med. 2007;357(21):2189–94.
15. Wallac JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JPA. Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. JAMA Intern Med. 2017;177(4):554–60.
16. Yusuf S. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA J Am Med Assoc. 1991;266(1):93–8.
17. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration [Internet]. The Cochrane Collaboration. 2011. Available from: www.cochrane-handbook.org.
18. Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. J Clin Epidemiol. 2004;57(7):683–97.
19. Jones HE, Ohlssen DI, Neuenschwander B, Racine A, Branson M. Bayesian models for subgroup analysis in clinical trials. Clin Trials. 2011 Apr;8(2):129–43.
20. Dixon DO, Simon R. Bayesian subset analysis. Biometrics. 1991 Sep;47(3):871–81.
21. Berry SM, Berry DA. Accounting for Multiplicities in Assessing Drug Safety: A Three-Level Hierarchical Mixture Model. Biometrics. 2004 Jun;60(2):418–26.
22. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A (Statistics Soc. 2009;172(1):137–59.
23. WHO. WHO ATC/DDD Index 2019 [Internet]. 2019. Available from: https://www.whocc.no/atc_ddd_index/
24. U.S. Department of Veterans Affairs. Medication Reference Terminology (MED- RT ™) Documentation. 2018;(June):1–16.
25. Hoehndorf R, Schofield PN, Gkoutos G V. The role of ontologies in biological and biomedical research: A functional perspective. Brief Bioinform. 2015;16(6):1069–80.
26. Sim I, Tu SW, Carini S, Lehmann HP, Pollock BH, Peleg M, et al. The Ontology of Clinical Research (OCRe): An informatics foundation for the science of clinical research. J Biomed Inform. 2014;52:78–91.
27. Kondylakis H, Claerhout B, Keyur M, Koumakis L, van Leeuwen J, Marias K, et al. The INTEGRATE project: Delivering solutions for efficient multi-centric clinical research and trials. J Biomed Inform. 2016;
28. Patel C, Cimino J, Dolby J, Fokoue A, Kalyanpur A, Kershenbaum A, et al. Matching patient records to clinical trials using ontologies. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2007.
29. Huang Z, Ten Teije A, Van Harmelen F. SemanticCT: A semantically-enabled system for clinical trials. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2013.
30. Davies S. The importance of PROSPERO to the National Institute for Health

Research. Syst Rev. 2012;1(1):5.

31. McAllister DA, Rodriguez-Perez J, Hannigan LJ. Assessing heterogeneity in treatment efficacy by age, sex and comorbidity. PROSPERO. 2018;CRD4201804.
32. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B Stat Methodol. 2009;
33. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019;38(11):2074–102.
34. Gasparini A. rsimsum: Summarise results from Monte Carlo simulation studies. J Open Source Softw. 2018;
35. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials. 2010 Dec 12;11(1):85.
36. Lee KK, Welton N, Shah AS, Adamson PD, Dias S, Anand A, et al. Differences in relative and absolute effectiveness of oral P2Y 12 inhibition in men and women: A meta-analysis and modelling study. Heart. 2018;104(8):657–64.
37. Navar AM, Pencina MJ, Rymer JA, Louzao DM, Peterson ED. Use of open access platforms for clinical trial data. JAMA - J Am Med Assoc. 2016;315(12):1283–4.

Tables

Scenario			**Single-drug model**		**Full model**	
Level(s)	**Variation**	**Performance measure**	**Estimate**	**MCSE**	**Estimate**	**MCSE**
All	Low	Bias	0.013	0.000	0.001	0.000
All	Low	MSE	0.003	0.000	0.003	0.000
All	Low	RMSE	0.058	0.000	0.056	0.000
All	Low	Rel. prec.	-	-	89.004	1.415
All	Low	Coverage	0.968	0.001	0.852	0.002
All	Medium	Bias	0.012	0.001	0.000	0.001
All	Medium	MSE	0.025	0.000	0.028	0.000
All	Medium	RMSE	0.159	0.001	0.166	0.001
All	Medium	Rel. prec.	-	-	13.212	0.435
All	Medium	Coverage	0.812	0.003	0.674	0.003
All	High	Bias	0.010	0.002	-0.003	0.002
All	High	MSE	0.069	0.001	0.076	0.001
All	High	RMSE	0.263	0.001	0.276	0.001
All	High	Rel. prec.	-	-	3.952	0.287
All	High	Coverage	0.759	0.003	0.624	0.003
Trial	Medium	Bias	0.012	0.001	0.000	0.000
Trial	Medium	MSE	0.008	0.000	0.005	0.000
Trial	Medium	RMSE	0.091	0.000	0.068	0.000
Trial	Medium	Rel. prec.	-	-	127.704	1.767
Trial	Medium	Coverage	0.958	0.001	0.897	0.002
Trial	High	Bias	0.014	0.001	0.001	0.000
Trial	High	MSE	0.017	0.000	0.006	0.000
Trial	High	RMSE	0.132	0.001	0.077	0.000
Trial	High	Rel. prec.	-	-	251.074	2.848
Trial	High	Coverage	0.959	0.001	0.944	0.001
Drug	Medium	Bias	0.013	0.001	0.001	0.001
Drug	Medium	MSE	0.004	0.000	0.007	0.000
Drug	Medium	RMSE	0.064	0.000	0.082	0.001
Drug	Medium	Rel. prec.	-	-	34.080	0.558
Drug	Medium	Coverage	0.968	0.001	0.902	0.002
Drug	High	Bias	0.013	0.001	0.000	0.001
Drug	High	MSE	0.006	0.000	0.008	0.000
Drug	High	RMSE	0.077	0.000	0.090	0.001
Drug	High	Rel. prec.	-	-	10.278	0.274
Drug	High	Coverage	0.968	0.001	0.911	0.002
Class	Medium	Bias	0.013	0.001	0.001	0.001

Class	Medium	MSE	0.019	0.000	0.020	0.000
Class	Medium	RMSE	0.138	0.001	0.142	0.001
Class	Medium	Rel. prec.	-	-	3.648	0.443
Class	Medium	Coverage	0.785	0.003	0.542	0.003
Class	High	Bias	0.010	0.002	-0.003	0.002
Class	High	MSE	0.053	0.001	0.061	0.001
Class	High	RMSE	0.230	0.001	0.246	0.001
Class	High	Rel. prec.	-	-	-10.158	0.283
Class	High	Coverage	0.676	0.003	0.380	0.003

Table 1. Summary of performance measures for full and single drug only models across all simulated datasets for different scenarios

Note – See Data generation procedure in Methods for full definition of scenarios; MSE =mean squared error; RMSE=root mean squared error; Rel. precision = % change in precision for full vs. drug model; Coverage = proportion of 95% credible intervals containing true effect; MCSE = Monte Carlo standard errors; RMSE estimates and corresponding MCSEs are not calculated by default in the rsimsum package and so are instead derived, with the MCSE approximated using the delta method, i.e.:

$$SE_{RMSE} = \sqrt{\frac{Var(\sqrt{MSE})}{n}} \approx \sqrt{\frac{n(SE_{MSE})^2}{4n \times \widehat{MSE}}} = \frac{SE_{MSE}}{2 \times \sqrt{\widehat{MSE}}}.$$

Figure legends

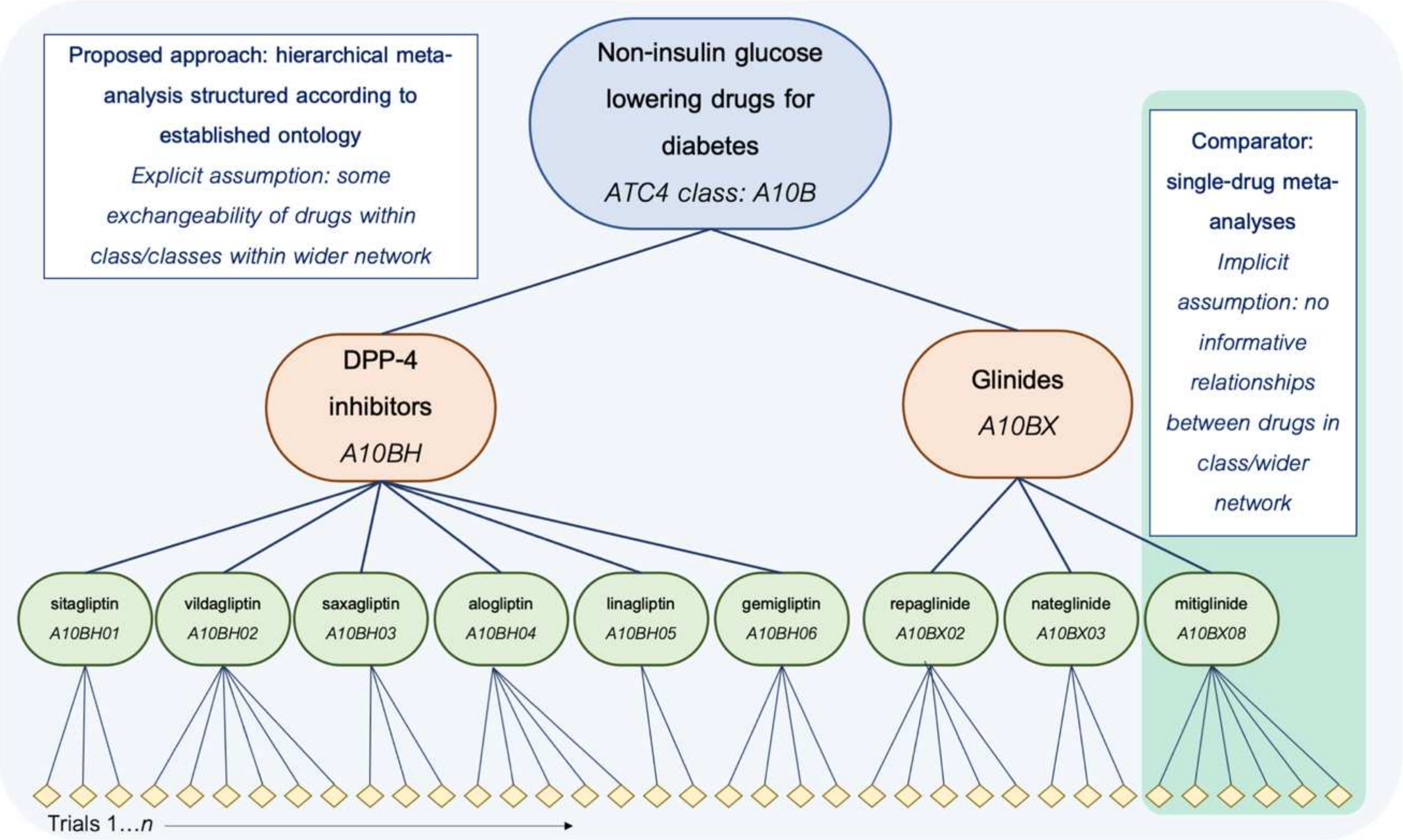
Figure 1. Schematic overview of the proposed and comparator approaches in the current simulation study – shown for a subset of drugs within two classes the A10B ATC4 class used as the basis for the simulation

Figure 2. Summary of relative precision of drug level comorbidity-treatment interaction effects in full *vs.* single drug model as a function of drug class

Figure 3. Posterior densities estimated for interaction effects at the drug class (top panel) and drug level (middle panel) from the full model and at the drug level (bottom panel) from single drug models for drugs in the A10BH class in a single randomly-selected dataset in the *All levels: low variation* and *All levels: high variation* scenarios, illustrating properties of shrinkage at the drug level in the full model

Figure 4. Illustration of the impact of increased precision in the full model: summarising the proportion of all datasets with "true" effects in the three main scenarios in which credible intervals for the interaction effect estimate for each drug excluded zero (i.e., no interaction) in i) both models; ii) the single drug model only; and iii) the full model only, alongside enrollment information

Figure 5. Introduction to an online tool for drawing network hierarchies of trials nested within drugs and drug WHO-ATC drug classes ascertained based on clinical trials with relevant metadata on clinicaltrials.gov



Proposed approach: hierarchical meta-analysis structured according to established ontology
Explicit assumption: some exchangeability of drugs within class/classes within wider network
Non-insulin glucose lowering drugs for diabetes
ATC4 class: A10B
Comparator: single-drug meta-analyses
Implicit assumption: no informative relationships between drugs in class/wider network
DPP-4 inhibitors
A10BH
Glinides
A10BX
sitagliptin
A10BH01
vildagliptin
A10BH02
saxagliptin
A10BH03
alogliptin
A10BH04
linagliptin
A10BH05
gemigliptin
A10BH06
repaglinide
A10BX02
nateglinide
A10BX03
mitiglinide
A10BX08
Trials 1…n

Figure 1. Schematic overview of the proposed and comparator approaches in the current simulation study – shown for a subset of drugs within two classes the A10B ATC4 class used as the basis for the simulation

Note – only a subset of the hierarchy is shown in the interests of managing space constraints; in the study the full hierarchical meta-analytic model is applied to a network incorporating all A10B drugs, and single drug meta-analyses are similarly run for all drugs in the network

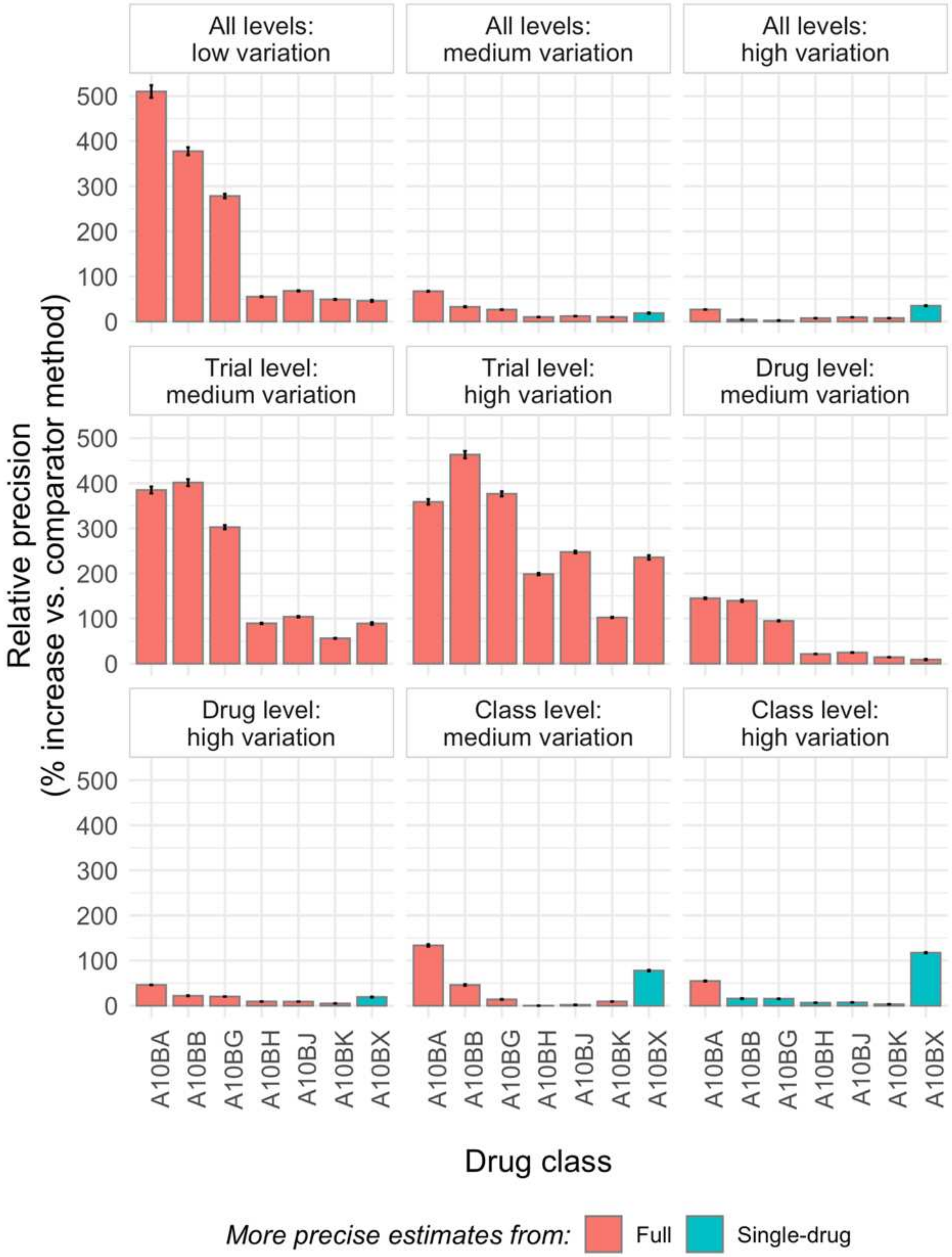


Figure 2. Summary of relative precision of drug level comorbidity-treatment interaction effects in full *vs.* single drug model as a function of drug class

Note – error bars show Monte-Carlo standard errors; information on the number and size of trials for each drug class is found in sTable 1; in contrast to the values in Table 1 where relative precision is always displayed as % change in precision for the full model relative to the drug model, here the "comparator method" is selected as whichever of the full or drug model are less precise in order to facilitate visual comparisons

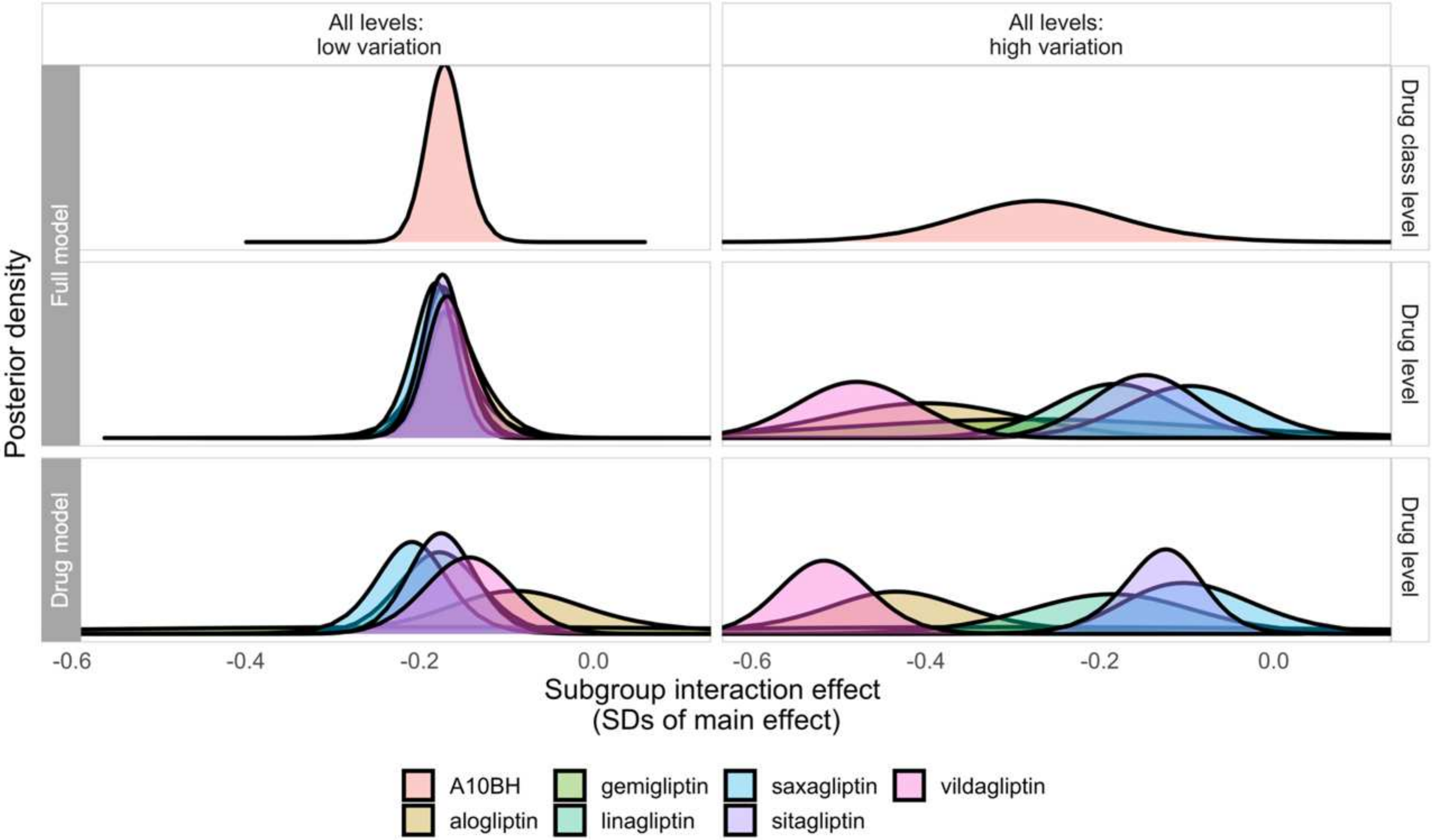


Figure 3. Posterior densities estimated for interaction effects at the drug class (top panel) and drug level (middle panel) from the full model and at the drug level (bottom panel) from single drug models for drugs in the A10BH class in a single randomly-selected dataset in the *All levels: low variation* and *All levels: high variation* scenarios, illustrating properties of shrinkage at the drug level in the full model

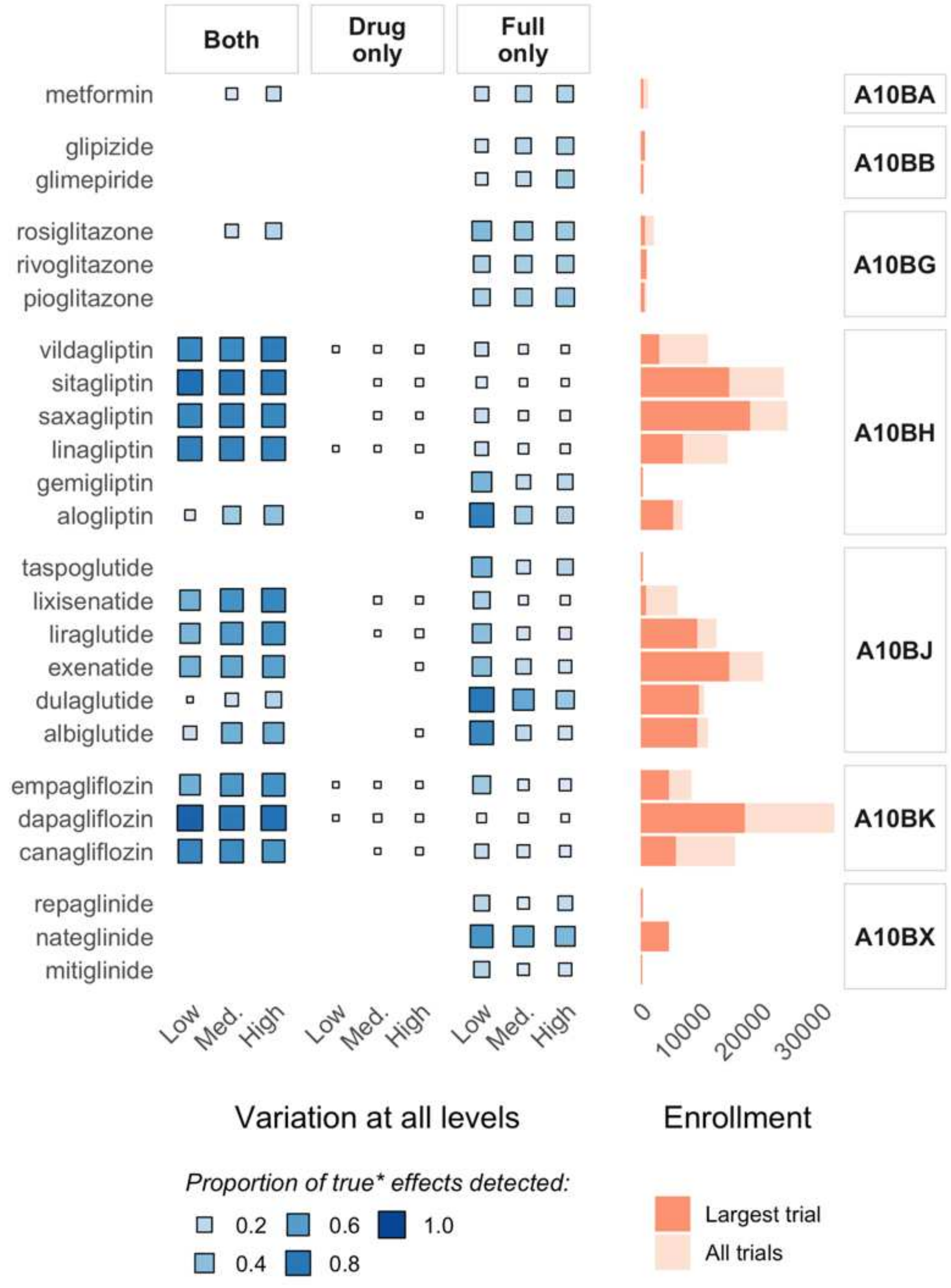
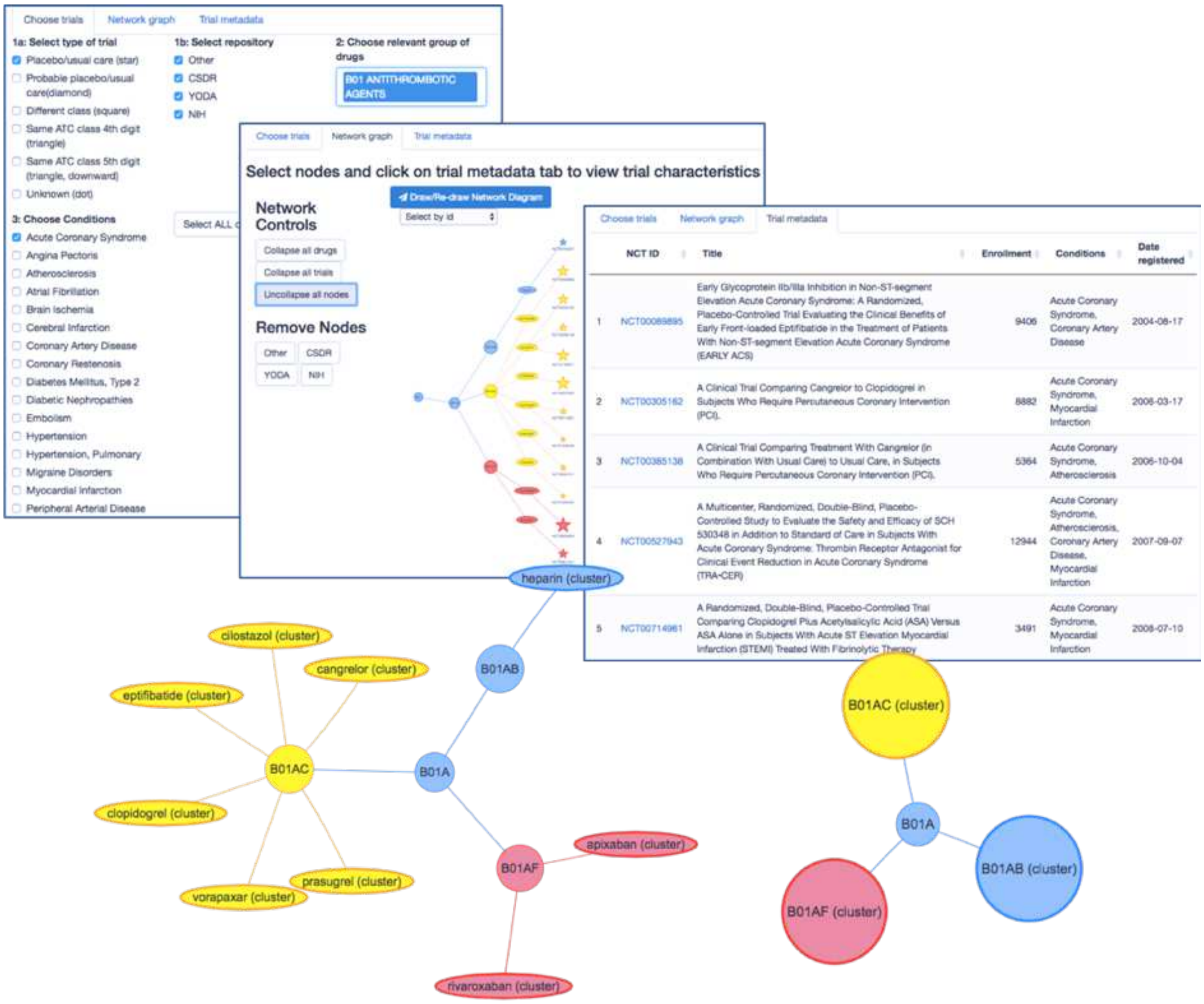


Figure 4. Illustration of the impact of increased precision in the full model: summarising the proportion of all datasets with "true" effects in the three main scenarios in which credible intervals for the interaction effect estimate for each drug excluded zero (i.e., no interaction) in i) both models; ii) the single drug model only; and iii) the full model only, alongside enrollment information



Choose trials | Network graph | Trial metadata

1a: Select type of trial

☑ Placebo/usual care (star)
☐ Probable placebo/usual care(diamond)
☐ Different class (square)
☐ Same ATC class 4th digit (triangle)
☐ Same ATC class 5th digit (triangle, downward)
☐ Unknown (dot)

1b: Select repository

☑ Other
☑ CSDR
☑ YODA
☑ NIH

2: Choose relevant group of drugs

B01 ANTITHROMBOTIC AGENTS

3: Choose Conditions

☑ Acute Coronary Syndrome
☐ Angina Pectoris
☐ Atherosclerosis
☐ Atrial Fibrillation
☐ Brain Ischemia
☐ Cerebral Infarction
☐ Coronary Artery Disease
☐ Coronary Restenosis
☐ Diabetes Mellitus, Type 2
☐ Diabetic Nephropathies
☐ Embolism
☐ Hypertension
☐ Hypertension, Pulmonary
☐ Migraine Disorders
☐ Myocardial Infarction
☐ Peripheral Arterial Disease

Select ALL c

Choose trials | Network graph | Trial metadata

Select nodes and click on trial metadata tab to view trial characteristics

Network Controls

Draw/Re-draw Network Diagram

Select by id

Collapse all drugs

Collapse all trials

Uncollapse all nodes

Remove Nodes

Other | CSDR | YODA | NIH

Choose trials | Network graph | Trial metadata

	NCT ID	Title	Enrollment	Conditions	Date registered
1	NCT00089895	Early Glycoprotein IIb/IIIa Inhibition in Non-ST-segment Elevation Acute Coronary Syndrome: A Randomized, Placebo-Controlled Trial Evaluating the Clinical Benefits of Early Front-loaded Eptifibatide in the Treatment of Patients With Non-ST-segment Elevation Acute Coronary Syndrome (EARLY ACS)	9406	Acute Coronary Syndrome, Coronary Artery Disease	2004-08-17
2	NCT00305162	A Clinical Trial Comparing Cangrelor to Clopidogrel in Subjects Who Require Percutaneous Coronary Intervention (PCI).	8882	Acute Coronary Syndrome, Myocardial Infarction	2006-03-17
3	NCT00385138	A Clinical Trial Comparing Treatment With Cangrelor (in Combination With Usual Care) to Usual Care, in Subjects Who Require Percutaneous Coronary Intervention (PCI).	5364	Acute Coronary Syndrome, Atherosclerosis	2006-10-04
4	NCT00527943	A Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of SCH 530348 in Addition to Standard of Care in Subjects With Acute Coronary Syndrome: Thrombin Receptor Antagonist for Clinical Event Reduction in Acute Coronary Syndrome (TRA-CER)	12944	Acute Coronary Syndrome, Atherosclerosis, Coronary Artery Disease, Myocardial Infarction	2007-09-07
5	NCT00714961	A Randomized, Double-Blind, Placebo-Controlled Trial Comparing Clopidogrel Plus Acetylsalicylic Acid (ASA) Versus ASA Alone in Subjects With Acute ST Elevation Myocardial Infarction (STEMI) Treated With Fibrinolytic Therapy	3491	Acute Coronary Syndrome, Myocardial Infarction	2008-07-10

Figure 5. Introduction to an online tool for drawing network hierarchies of trials nested within drugs and drug WHO-ATC drug classes ascertained based on clinical trials with relevant metadata on clinicaltrials.gov

Improving the estimation of subgroup effects for clinical trial participants with multimorbidity by incorporating drug class-level information in Bayesian hierarchical models: a simulation study (Supplementary material)

sAppendix 1

Each set of trials was organised into a hierarchy based on the biological mechanism of action of the drugs under study. For the diabetes trial-set, the WHO-ATC-5 level grouping of included drugs was closely reflective of the mechanism of action, so this was used as a node in the hierarchy. As a result, diabetes trials were nested within drugs, drugs were nested within WHO-ATC-5 drug classes, and drug classes were nested within the wider drug grouping (the WHO-ATC-4 level code A10B). The full network diagram for the diabetes trials can be viewed online here: https://ihwph-hehta.shinyapps.io/duk_example_app/. The classifications for the diabetes drugs are shown in sTable 1.

sAppendix 2

In order to simulate interaction effects and their precision at the trial-level (but without simulating IPD in full), four groups per trial were defined overall: two for each arm (those with and without the comorbidity). We assumed one-to-one randomisation within a two-group parallel design, and that the comorbidity was equally common in the intervention and control arms. The precision of the comorbidity-treatment interaction estimate was approximated as the inverse of the sum of the standard errors-squared for each of the four groups, where the standard error for each group was estimated as the standard deviation for the standardised outcome (by definition one), divided by the square root of the number of participants in that group. The N-per-group was calculated by combining the trial-specific enrolment (as recorded in clinicaltrials.gov) and the specified prevalence for the comorbidity (see below).

An example dataset from a single iteration of the simulation for the diabetes trial-set is shown in sTable 2.

Supplementary references

1. Pathak J, Chute CG. Analyzing categorical information in two publicly available drug terminologies: RxNorm and NDF-RT. J Am Med Informatics Assoc. 2010;

2. Iglay K, Hannachi H, Joseph Howie P, Xu J, Li X, Engel SS, et al. Prevalence and co-prevalence of comorbidities among patients with type 2 diabetes mellitus. Curr Med Res Opin. 2016;32(7):1243–52.

3. Matcham F, Rayner L, Steer S, Hotopf M. The prevalence of depression in rheumatoid arthritis: a systematic review and meta-analysis. Rheumatology. 2013;52(12):2136–48.

4. Mikocka-Walus AA, Turnbull DA, Moulding NT, Wilson IG, Andrews JM, Holtmann GJ. Controversies surrounding the comorbidity of depression and anxiety in inflammatory bowel disease patients: A literature review. Inflamm Bowel Dis. 2007;13(2):225–34.

5. Hanlon P, Hannigan L, Rodriguez-Perez J, Fischbacher C, Welton NJ, Dias S, 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 Dec 12;17(1):201.

Supplementary tables

Drug	WHO-ATC level 5 code	ATC-5 class description	ATC-7 code	N trials	Total N participants
metformin	A10BA	biguanide	A10BA02	4	1220
glipizide	A10BB	A10BB Sulfonylureas	A10BB07	1	700
glimepiride	A10BB	A10BB Sulfonylureas	A10BB12	2	618
rivoglitazone	A10BG	thiazolidinedione	A10BG_2	1	910
rosiglitazone	A10BG	thiazolidinedione	A10BG02	4	2134
pioglitazone	A10BG	thiazolidinedione	A10BG03	2	948
sitagliptin	A10BH	dpp-4 inhibitor	A10BH01	20	23766
vildagliptin	A10BH	dpp-4 inhibitor	A10BH02	16	11158
saxagliptin	A10BH	dpp-4 inhibitor	A10BH03	11	24404
alogliptin	A10BH	dpp-4 inhibitor	A10BH04	5	6944
linagliptin	A10BH	dpp-4 inhibitor	A10BH05	13	14452
gemigliptin	A10BH	dpp-4 inhibitor	A10BH06	1	288
taspoglutide	A10BJ	glp-1 receptor agonist	A10BJ_3	1	332
exenatide	A10BJ	glp-1 receptor agonist	A10BJ01	7	20402
liraglutide	A10BJ	glp-1 receptor agonist	A10BJ02	8	12534
lixisenatide	A10BJ	glp-1 receptor agonist	A10BJ03	13	6058
albiglutide	A10BJ	glp-1 receptor agonist	A10BJ04	6	11142
dulaglutide	A10BJ	glp-1 receptor agonist	A10BJ05	4	10504
dapagliflozin	A10BK	A10BK Sodium-glucose co-transporter 2 (SGLT2) inhibitors	A10BK01	22	32198
canagliflozin	A10BK	A10BK Sodium-glucose co-transporter 2 (SGLT2) inhibitors	A10BK02	9	15724
empagliflozin	A10BK	A10BK Sodium-glucose co-transporter 2 (SGLT2) inhibitors	A10BK03	8	8390
repaglinide	A10BX	glinide	A10BX02	1	324
nateglinide	A10BX	glinide	A10BX03	1	4652
mitiglinide	A10BX	glinide	A10BX08	1	244

sTable 1. Classification of drugs from the diabetes trial-set into classes based on WHO-ATC-5

Trial ID	WHO-ATC level 5 code (class)	Drug	Number of participants per experimental group in trial	Simulated covariate-treatment interaction effect estimate	Standard error of interaction effect estimate (derived[1])
NCT00819741	A10BA	metformin	216	0.08	0.19
NCT01512979	A10BA	metformin	158	0.07	0.23
NCT01545388	A10BA	metformin	112	0.07	0.27
NCT02068443	A10BA	metformin	124	0.13	0.25
NCT00131664	A10BB	glimepiride	130	-0.14	0.25
NCT01459809	A10BB	glimepiride	179	-0.11	0.21
NCT00086515	A10BB	glipizide	350	-0.19	0.15
NCT00094757	A10BG	pioglitazone	173	-0.26	0.22
NCT00220961	A10BG	pioglitazone	301	-0.23	0.16
NCT00484198	A10BG	rivoglitazone	455	-0.10	0.13
NCT00241605	A10BG	rosiglitazone	300	0.05	0.16
NCT00359112	A10BG	rosiglitazone	272	0.03	0.17
NCT00386100	A10BG	rosiglitazone	344	-0.03	0.15
NCT00499707	A10BG	rosiglitazone	151	-0.05	0.23
NCT00286494	A10BH	alogliptin	164	-0.19	0.22
NCT00432276	A10BH	alogliptin	401	-0.34	0.14
NCT00968708	A10BH	alogliptin	2690	-0.20	0.05
NCT01318070	A10BH	alogliptin	113	-0.24	0.27
NCT01318083	A10BH	alogliptin	104	-0.26	0.28
NCT01787396	A10BH	gemigliptin	144	-0.10	0.24
…					

sTable 2. Example dataset from simulation (abbreviated to 20 rows)

Notes: [1] The standard error of the interaction effect estimate for each trial was derived by assuming a SD of 1 for the main effect in a trial, and using the number of participants per group in each trial to calculate standard errors for subgroups with and without a hypothetical comorbidity (prevalence set at 0.2), which were combined to give the overall variance of the interaction estimate

Scenario			Single-drug model		Full model	
Level(s)	Variation	Performance measure	Estimate	MCSE	Estimate	MCSE
All	Low	Bias	0.013	0.000	0.001	0.000
All	Low	MSE	0.003	0.000	0.003	0.000
All	Low	RMSE	0.058	0.000	0.056	0.000
All	Low	Rel. prec.	-	-	122.727	1.868
All	Low	Coverage	0.989	0.001	0.889	0.002
All	Medium	Bias	0.013	0.001	0.000	0.001
All	Medium	MSE	0.025	0.000	0.028	0.000
All	Medium	RMSE	0.160	0.001	0.167	0.001
All	Medium	Rel. prec.	-	-	20.311	0.543
All	Medium	Coverage	0.855	0.002	0.719	0.003
All	High	Bias	0.010	0.002	-0.003	0.002
All	High	MSE	0.069	0.001	0.077	0.001
All	High	RMSE	0.263	0.001	0.277	0.001
All	High	Rel. prec.	-	-	7.028	0.345
All	High	Coverage	0.783	0.003	0.651	0.003
Trial	Medium	Bias	0.012	0.001	0.000	0.000
Trial	Medium	MSE	0.009	0.000	0.005	0.000
Trial	Medium	RMSE	0.093	0.000	0.069	0.000
Trial	Medium	Rel. prec.	-	-	134.132	1.922
Trial	Medium	Coverage	0.977	0.001	0.918	0.002
Trial	High	Bias	0.014	0.001	0.002	0.001
Trial	High	MSE	0.018	0.000	0.007	0.000
Trial	High	RMSE	0.134	0.001	0.081	0.000
Trial	High	Rel. prec.	-	-	221.363	2.696
Trial	High	Coverage	0.968	0.001	0.950	0.001
Drug	Medium	Bias	0.013	0.001	0.001	0.001
Drug	Medium	MSE	0.004	0.000	0.008	0.000
Drug	Medium	RMSE	0.065	0.000	0.090	0.001
Drug	Medium	Rel. prec.	-	-	56.857	0.797
Drug	Medium	Coverage	0.989	0.001	0.938	0.002
Drug	High	Bias	0.013	0.001	0.000	0.001
Drug	High	MSE	0.006	0.000	0.011	0.000
Drug	High	RMSE	0.077	0.000	0.107	0.001
Drug	High	Rel. prec.	-	-	22.996	0.399
Drug	High	Coverage	0.989	0.001	0.953	0.001
Class	Medium	Bias	0.014	0.001	0.001	0.001
Class	Medium	MSE	0.019	0.000	0.019	0.000
Class	Medium	RMSE	0.138	0.001	0.138	0.001
Class	Medium	Rel. prec.	-	-	12.471	0.544
Class	Medium	Coverage	0.846	0.002	0.606	0.003

Class	High	Bias	0.010	0.001	-0.003	0.002
Class	High	MSE	0.052	0.001	0.058	0.001
Class	High	RMSE	0.229	0.001	0.241	0.001
Class	High	Rel. prec.	-	-	-6.145	0.322
Class	High	Coverage	0.737	0.003	0.447	0.003

sTable 3. Summary of performance measures for full and single drug only models across all simulated datasets for different scenarios – **LOW SIMULATED COMORBIDITY PREVALENCE (10%)**

Note – See Data generation procedure in Methods for full definition of scenarios; MSE =mean squared error; RMSE=root mean squared error; Rel. precision = % change in precision for full vs. drug model; Coverage = proportion of 95% credible intervals containing true effect; MCSE = Monte Carlo standard errors; RMSE estimates and corresponding MCSEs are not calculated by default in the rsimsum package and so are instead derived, with the MCSE approximated using the delta method, i.e.:

$$SE_{RMSE} = \sqrt{\frac{Var(\sqrt{MSE})}{n}} \approx \sqrt{\frac{n(SE_{MSE})^2}{4n \times \widehat{MSE}}} = \frac{SE_{MSE}}{2 \times \sqrt{\widehat{MSE}}}.$$

Scenario			**Single-drug model**		**Full model**	
Level(s)	**Variation**	**Performance measure**	**Estimate**	**MCSE**	**Estimate**	**MCSE**
All	Low	Bias	0.013	0.000	0.001	0.000
All	Low	MSE	0.003	0.000	0.003	0.000
All	Low	RMSE	0.057	0.000	0.056	0.000
All	Low	Rel. prec.	-	-	65.898	1.127
All	Low	Coverage	0.943	0.001	0.821	0.002
All	Medium	Bias	0.012	0.001	0.000	0.001
All	Medium	MSE	0.025	0.000	0.027	0.000
All	Medium	RMSE	0.159	0.001	0.165	0.001
All	Medium	Rel. prec.	-	-	9.408	0.370
All	Medium	Coverage	0.783	0.003	0.647	0.003
All	High	Bias	0.010	0.002	-0.003	0.002
All	High	MSE	0.069	0.001	0.076	0.001
All	High	RMSE	0.263	0.001	0.276	0.001
All	High	Rel. prec.	-	-	2.398	0.261
All	High	Coverage	0.748	0.003	0.609	0.003
Trial	Medium	Bias	0.012	0.001	0.000	0.000
Trial	Medium	MSE	0.008	0.000	0.004	0.000
Trial	Medium	RMSE	0.090	0.000	0.067	0.000
Trial	Medium	Rel. prec.	-	-	129.466	1.703
Trial	Medium	Coverage	0.944	0.001	0.884	0.002
Trial	High	Bias	0.013	0.001	0.001	0.000
Trial	High	MSE	0.017	0.000	0.006	0.000
Trial	High	RMSE	0.131	0.001	0.075	0.000
Trial	High	Rel. prec.	-	-	263.698	2.887
Trial	High	Coverage	0.955	0.001	0.942	0.002
Drug	Medium	Bias	0.013	0.001	0.001	0.001
Drug	Medium	MSE	0.004	0.000	0.006	0.000
Drug	Medium	RMSE	0.064	0.000	0.076	0.000
Drug	Medium	Rel. prec.	-	-	21.332	0.425
Drug	Medium	Coverage	0.943	0.001	0.860	0.002
Drug	High	Bias	0.012	0.002	0.000	0.001
Drug	High	MSE	0.006	0.000	0.006	0.000
Drug	High	RMSE	0.076	0.000	0.080	0.001
Drug	High	Rel. prec.	-	-	2.542	0.215
Drug	High	Coverage	0.943	0.001	0.867	0.002
Class	Medium	Bias	0.013	0.001	0.001	0.001
Class	Medium	MSE	0.019	0.000	0.021	0.000
Class	Medium	RMSE	0.139	0.001	0.145	0.001
Class	Medium	Rel. prec.	-	-	-1.838	0.384
Class	Medium	Coverage	0.739	0.003	0.494	0.003

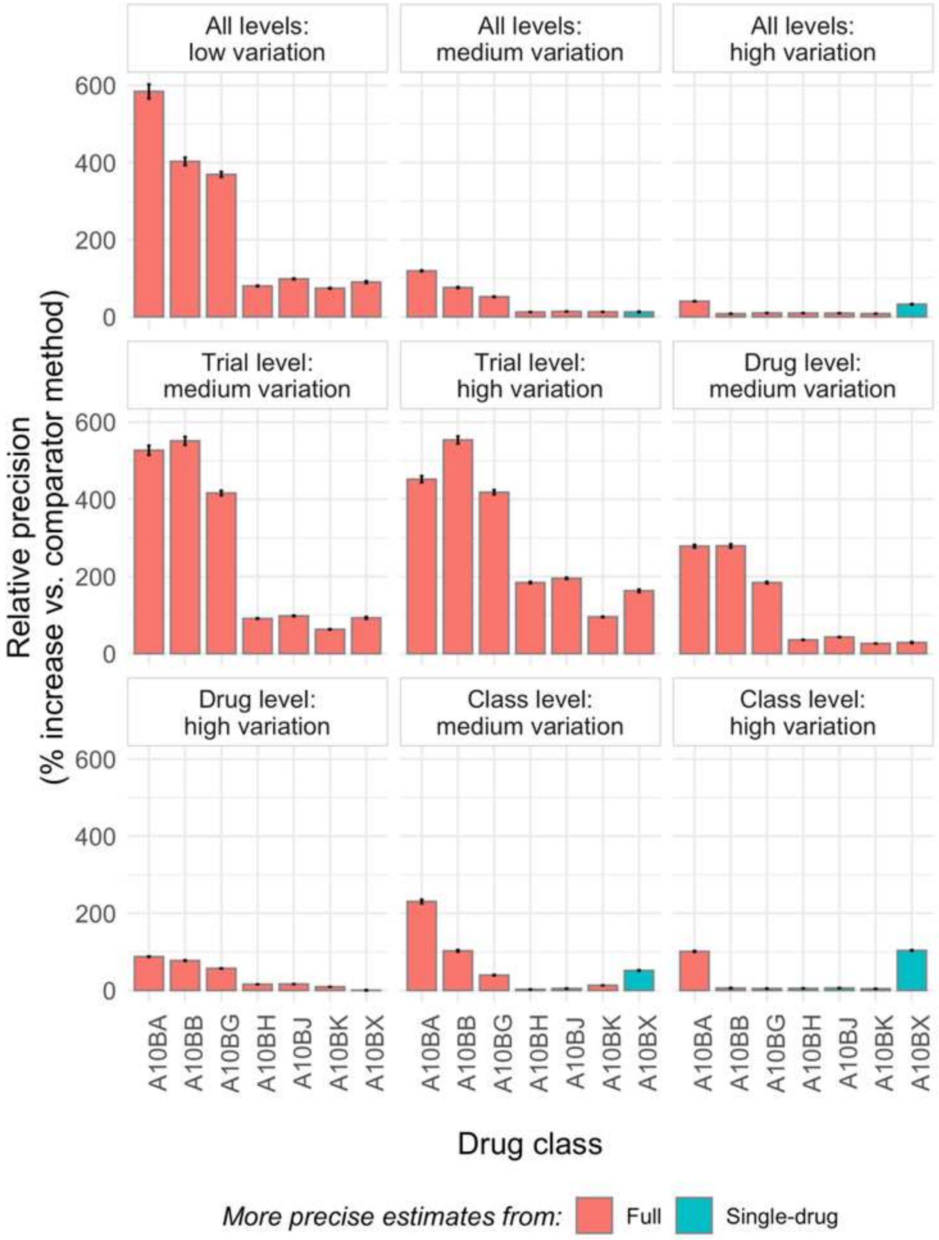
Class	High	Bias	0.010	0.002	-0.003	0.002
Class	High	MSE	0.053	0.001	0.062	0.001
Class	High	RMSE	0.230	0.001	0.249	0.001
Class	High	Rel. prec.	-	-	-12.440	0.263
Class	High	Coverage	0.635	0.003	0.333	0.003

sTable 4 Summary of performance measures for full and single drug only models across all simulated datasets for different scenarios – **HIGH SIMULATED COMORBIDITY PREVALENCE (50%)**

Note – See Data generation procedure in Methods for full definition of scenarios; MSE =mean squared error; RMSE=root mean squared error; Rel. precision = % change in precision for full vs. drug model; Coverage = proportion of 95% credible intervals containing true effect; MCSE = Monte Carlo standard errors; RMSE estimates and corresponding MCSEs are not calculated by default in the rsimsum package and so are instead derived, with the MCSE approximated using the delta method, i.e.:

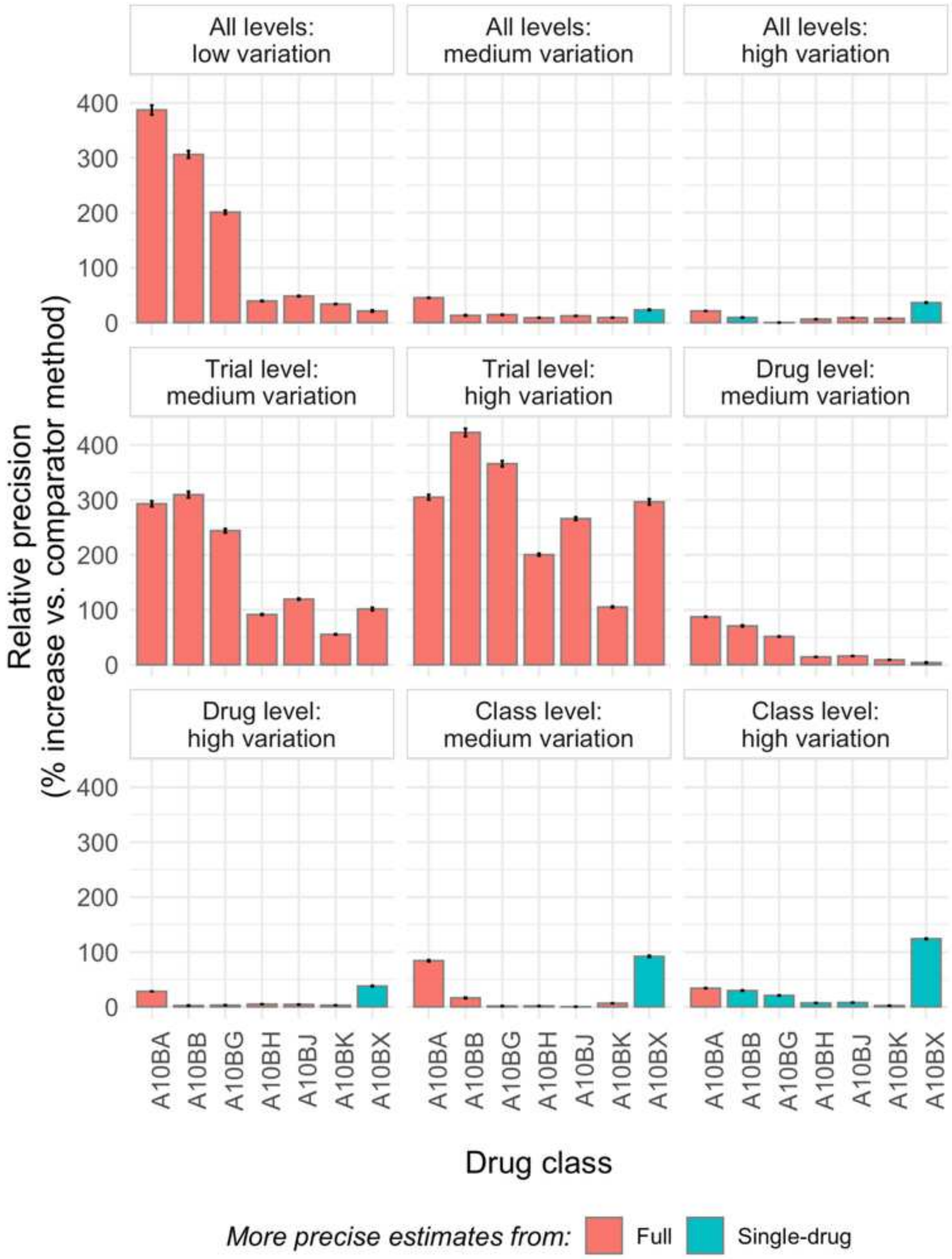
$$SE_{RMSE} = \sqrt{\frac{Var(\sqrt{MSE})}{n}} \approx \sqrt{\frac{n(SE_{MSE})^2}{4n \times \widehat{MSE}}} = \frac{SE_{MSE}}{2 \times \sqrt{\widehat{MSE}}}.$$

Supplementary figures



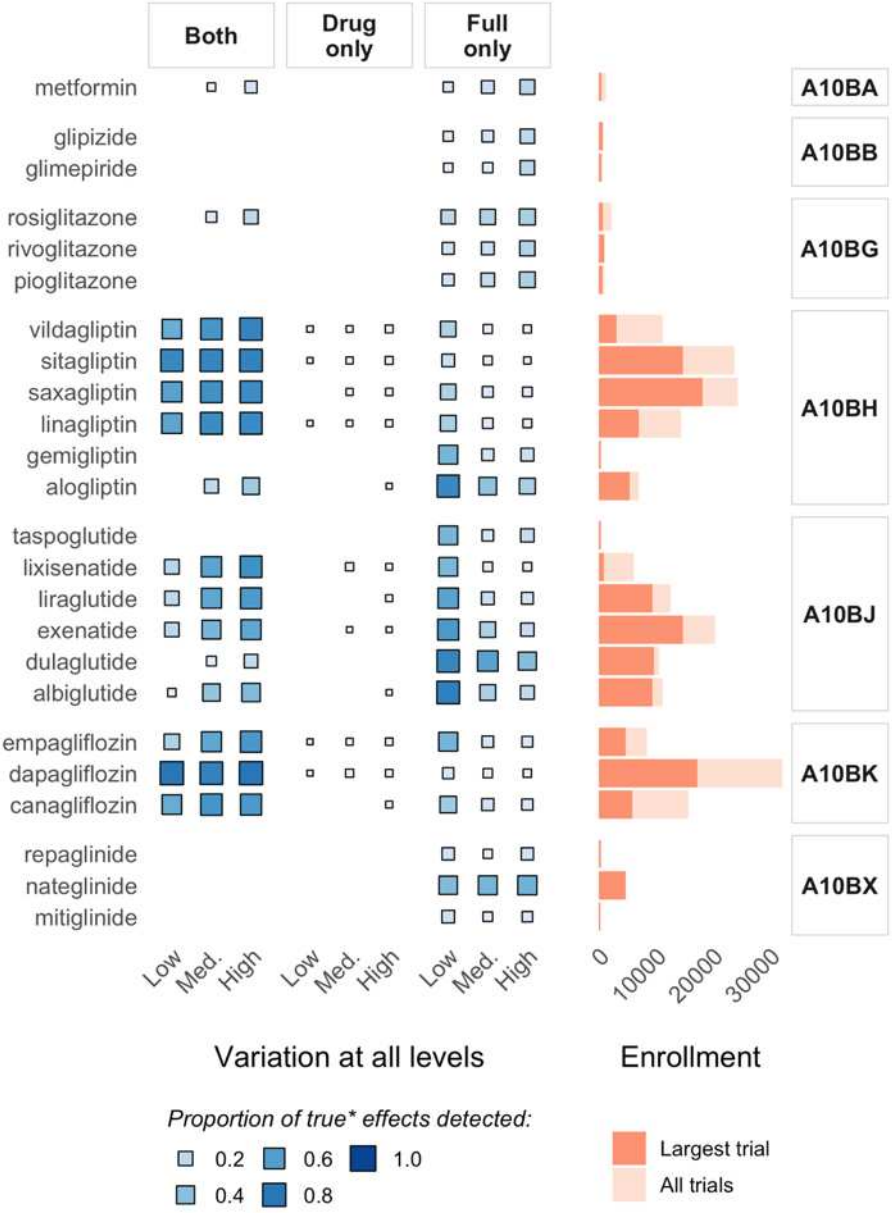
sFigure 1 Summary of relative precision of drug level comorbidity-treatment interaction effects in full *vs.* single drug model as a function of drug class– **LOW SIMULATED COMORBIDITY PREVALENCE (10%)**

Note – error bars show Monte-Carlo standard errors; information on the number and size of trials for each drug class is found in sTable 1

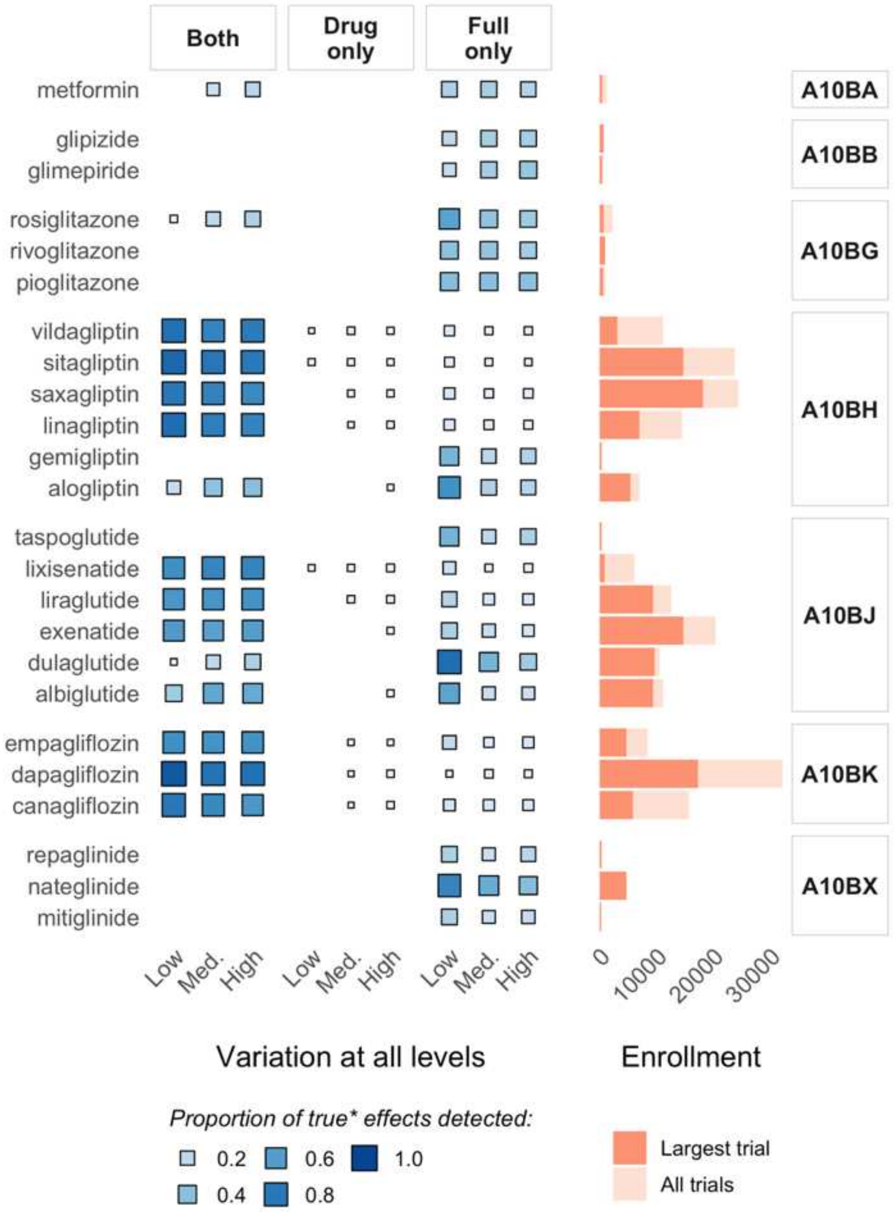


sFigure 2 Summary of relative precision of drug level comorbidity-treatment interaction effects in full *vs.* single drug model as a function of drug class– **HIGH SIMULATED COMORBIDITY PREVALENCE (50%)**

Note – error bars show Monte-Carlo standard errors; information on the number and size of trials for each drug class is found in sTable 1



sFigure 3 Illustration of the impact of increased precision in the full model: summarising the proportion of all datasets with "true" effects in the three main scenarios in which credible intervals for the interaction effect estimate for each drug excluded zero (i.e., no interaction) in i) both models; ii) the single drug model only; and iii) the full model only, alongside enrollment information – **LOW SIMULATED COMORBIDITY PREVALENCE (10%)**



sFigure 4 Illustration of the impact of increased precision in the full model: summarising the proportion of all datasets with “true” effects in the three main scenarios in which credible intervals for the interaction effect estimate for each drug excluded zero (i.e., no interaction) in i) both models; ii) the single drug model only; and iii) the full model only, alongside enrollment information – **HIGH SIMULATED COMORBIDITY PREVALENCE (50%)**