

Drug combination-wide association studies of cancer

Corresponding Author: Dr Rachel Melamed

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This is a very laudable effort and kudos to the authors. The appeal I believe is chiefly methodological. The biggest strength of the study area a) Using a hierarchical approach is very sound and the methods are very well narrated; b) The large sample size of observations. The biggest limitation, which the authors acknowledge as their second limitation in the Discussion, is that the follow-up period is pretty short (around 3-4 years). For most patients / tumours much longer follow-up will be needed, certainly for true novel tumour development / oncogenesis. Therefore I think the findings are pretty limited in their value in terms of pharmacovigilance / monitoring of the safety of drug combinations from the point of view of inducing / increasing cancer risk. I appreciate the authors cite an aspirin trial with 4-year period follow-up, but the effect there is protective, not cancer inducing. I think the authors will have much better use of their method in terms of safety monitoring of drug combinations for illnesses with much shorter 'incubation' /latency period, such as acute psychosis, or cardiovascular events. In short, I think it reads very well - I just think there is not truly long enough follow-up for this to be more than methodological interest when it comes to monitoring excess cancer risk potentially induced by drug combinations. There may be a limited value though in terms of discovering putative cancer risk reduction associated with certain drug combinations that can be then tested in other real-world datasets or trials.

Reviewer #2

(Remarks to the Author)

Many thanks for this submission. Topic and concept are relevant addressing an unmet need - assessing the effect of drug combinations on incident cancer. The paper is very well written.

The huge IBM dataset goes back to 2015 only.
9460 drug combinations were tested, but not entirely clear to me why this number.

The potential issues with DrugWAS are nicely explored.
The rationale for the methodology developed by the authors is quite clear and explanation is easy to follow.

The main issue is lack of validation and biological plausibility for most of the findings> Validation of findings using Mendelian Randomization and a detailed discussing of biological plausibility for the observed associations would be of interest.

How do the authors explain that amlodipine + pravastatin is associated with reduced risk of cancer, and amlodipine + other stronger statins is not also a hit?

Similarly, for most of the detected associations, it seems hard to understand and believe that there is a protective effect based on my large clinical experience.

Reviewer #3

(Remarks to the Author)

This paper focused on the "Drug combination-wide association studies of cancer".
The aim of this study sounds very promising, however the authors have ignored recent developments in epidemiology. Over the last decade, the optimal way to perform comparative effectiveness research (especially in pharmacoepidemiology) is to utilise the target trial emulation framework
<https://jamanetwork.com/journals/jama/article-abstract/2799678>
<https://pubmed.ncbi.nlm.nih.gov/26994063/>

(even if the authors use 3-4 references of papers that use the TTE framework, but they cite them only on how they estimated IPW)

Specifically, to estimate the effect of medication A vs dual medication A and B, the authors should specify 7 key components of both the target trial and the emulated trial protocol: the eligibility criteria, treatment strategies being compared (including their start and end times), assignment procedures, follow-up period, outcome of interest, causal contrast(s) of interest, and analysis plan.

Their analysis is based on the definition of causality of the Bradford-Hill criteria, which had revolutionized the way of thinking for epidemiologists back in 1965, however this definition cannot take into consideration the use of modern Big Data and the challenges that come with that. One a minor point, the language is outdated in causal inference in some cases, as for example the term “consistency” for Bradford-Hill is translated as “consistent findings observed by different persons with different samples strengthens the likelihood of an effect”, while in modern causal inference theory, which has adopted the potential outcome framework, consistency is about well-defined interventions (or more formally that the observed outcome for every treated individual equals her outcome if she had received treatment, and that the observed outcome for every untreated individual equals her outcome if she had remained untreated).

The focus of the authors is only around confounding – of course this is a very important problem that is responsible for bias in medical research. However, an equally important problem is the clarification of the protocol of this comparison of medication A vs dual medication A and B. And then the protocol of the target trial will dictate the analysis plan, e.g. for the estimation of the observational analogue of the intention-to-treat effect, one should not censor individuals that discontinued treatment (and should not use adherence weights), while for the observational analogue of the per-protocol effect (i.e. effect of the intervention, had everybody adhered), one should not censor individuals that discontinued treatment (and should not use adherence weights).

Moreover, there is confusion about the use of negative controls. A negative control outcome is a tool for detecting confounding and bias in observational studies (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3053408/>). Specifically, a negative control outcome is selected in such a way so that it should not be related with the treatment. If the treatment-outcome relationship is not null, then there is an indication that there is unmeasured confounding in the data, so we cannot really trust the results of our main analysis. The authors seem to have selected their negative control on a different basis, that is not completely clear.

There is also confusion about the use of IPW. The weights described in the paper are in fact the adherence weights that should have been used only in the per-protocol analysis, to account for those individuals who discontinue their treatment (<https://www.nature.com/articles/s41591-019-0597-x>). The description of these weights is not clear from the authors – it seems that they use them to account for confounders without providing many clarifications.

Overall, I believe that any comparison between single and dual medication should be based on a certain protocol, that has to be clearly stated. I don't think that a “drug combination-wide association study” will give more insight about the benefits of the dual medication, even in the absence of unmeasured confounding. The analyses of observational data can result in substantial selection bias, if the design of these studies are flawed, especially in pharmacoepidemiology. There are many examples in the literature that highlight this problem, e.g. the effect of statin on cancer has been measured up to 65% beneficial in some studies (that accounted for confounding), however the effect was way more modest (almost null), when the TTE was implemented to study this relationship <https://pubmed.ncbi.nlm.nih.gov/31591592/>

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I raised the question about whether 3-4 years of follow-up are adequate for inducing new cancers - I do not think they are. The authors provide a lot of references, which fall into two categories: 1. Papers with breast or other, hormone-sensitive tumours. 2. Few other papers by Stoer <https://aacrjournals.org/cebp/article/30/4/682/670683/Drug-Use-and-Cancer-Risk-A-Drug-Wide-Association>, or Pottegard where the authors claim that the follow-up was only 2 years, but from looking at these papers, I could not find this information (I admit I may have missed it).

The author state: "The authors of the breast cancer studies cited above hypothesized that the shorter term effect is likely due to effect on existing neoplastic but precancerous tissue." This is totally true, here we are not talking about genuine cancer development de novo / oncogenesis, but an 'acceleration' of the growth of existing tumours, such as the case for as-yet-asymptomatic / very early breast cancer being accelerated into a clinical phase by hormone-active therapies. I think most of the effects that are being described will be due to such mechanisms, not oncogenesis.

This needs to be declared as such in the paper.

Reviewer #2

(Remarks to the Author)

Thanks for addressing the previously raised concerns. I am happy with the answers and believe the manuscript now reads better.

Reviewer #3

(Remarks to the Author)

In principle, I strongly disagree with the performance of exposure-wide (or combination-wide) epidemiology. Such exposure-wide studies are prone to numerous biases because of the inability to make simultaneous confounding control. Certain variables can be considered as confounders for a certain exposure-outcome relationships, however the same variables can be mediators for other exposure-outcome relationships. For more details, see

<https://pmc.ncbi.nlm.nih.gov/articles/PMC5378648/> which presents certain limitations of

<https://pubmed.ncbi.nlm.nih.gov/26646432/>

In target trial emulation (TTE) studies, the situation is even more complicated. When we perform a certain TTE study, we try to think of the hypothetical (target) trial and then emulate it. We try to do this by emulating each of the 7 components (that now the authors did add in their revised version, however, the description is very vague). The problem is that each of these trials can be very different in terms of the eligibility criteria, the definition of the treatment strategy and the causal contract. Let's take a certain example. Imagine a certain target trial which would enrol hypertensive individuals. The 2 different arms would be

a) Thiazide diuretics and b) Thiazide diuretics & ACE inhibitors

In the observational data, we can find individuals (say individuals A, B and C) that initiated thiazide diuretics for 6 months and then they additionally initiated ACE inhibitors

The authors have allocated all these individuals in the arm b) Thiazide diuretics & ACE inhibitors, by considering as time zero the 1st time that individuals initiated ACE inhibitors. This is wrong, as in this way it is introduced a severe amount of bias. The correct approach is to break the follow-up time by month and clone individuals. The first 6 months, all 3 individuals A, B and C are contributing in arm a) Thiazide diuretics and they are censored, as they deviated from their strategy. Then they are eligible to b) Thiazide diuretics & ACE inhibitors (i.e. the same individuals are eligible to both arms with different follow-up time, outcome etc). Or alternatively they contribute at random to only one of the 2 arms (e.g. individual A to a) Thiazide diuretics and B & C to b) Thiazide diuretics & ACE inhibitors). For more details, see

<https://pubmed.ncbi.nlm.nih.gov/27237061/>

Another issue is that each trial is different and for example some trials utilise a "grace period" before treatment initiation. This issue is tackled using cloning, censoring and weighing. In brief, TTE studies are not like GWAS studies, they cannot be conducted all at once. In a TTE study, we cannot say over simplistically that we compare drug X vs drugs X & Y. We need to describe the 7 components of the trial explicitly (e.g. eligibility criteria: individuals aged >50years old without cancer). This has not been done by the authors and cannot be done in general as these components (and especially the eligibility criteria) will vary from trial to trial. In summary, it is not correct to perform treatment-wide (or combination-wide) TTE studies, because each of these trials is different.

I have other comments as well, e.g. the causal contrast is vague, we usually utilise the observational analogue of the intention-to-treat (ITT) or the observational analogue of the per-protocol effect... In any case, the main problem of the paper is that it is wrong to perform blindly treatment-wide TTE studies, as I mentioned before.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Thank you for addressing my comments regarding the phenomena relating to acceleration of latent (pre-existing) cancers as opposed to true / de novo oncogenesis. I have no other comments.

Reviewer #3

(Remarks to the Author)

There is a fundamental issue with this paper that I attempted to describe in my previous revision as well: The TTE studies are not GWAS studies. We cannot "estimate the effects of over 9,000 drug combinations on the incidence of all common cancer types" all at once, simply because we would need to perform 9,000 different TTE studies, each with different eligibility criteria, treatment strategies, follow-up time, outcomes, estimands, and analysis plans. This paper has not described in detail these 9,000 different studies, and this would not even be possible. I strongly disagree with this type of papers, as the results would be highly uninterpretable

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

We thank the reviewers for their time and thoughtful critiques. Below, we have included the full text of the reviews, and for each concern, we include a numbered response indented below, with selected quoted changes to the text indented further and marked with a vertical line.

We hope that these responses satisfy all concerns, and we believe the changes have greatly strengthened the manuscript.

Sincerely,
The authors

Referee expertise:

Referee #1: Oncology

Referee #2: Cardiology, phenome-wide scans

Referee #3: Statistics, trial emulation

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

This is a very laudable effort and kudos to the authors. The appeal I believe is chiefly methodological. The biggest strength of the study area a) Using a hierarchical approach is very sound and the methods are very well narrated; b) The large sample size of observations. The biggest limitation, which the authors acknowledge as their second limitation in the Discussion, is that the follow-up period is pretty short (around 3-4 years). For most patients / tumours much longer follow-up will be needed, certainly for true novel tumour development / oncogenesis. Therefore I think the findings are pretty limited in their value in terms of pharmacovigilance / monitoring of the safety of drug combinations from the point of view of inducing / increasing cancer risk. I appreciate the authors cite an aspirin trial with 4-year period follow-up, but the effect there is protective, not cancer inducing. I think the authors will have much better use of their method in terms of safety monitoring of drug combinations for illnesses with much shorter 'incubation' /latency period, such as acute psychosis, or cardiovascular events. In short, I think it reads very well - I just think there is not truly long enough follow-up for this to be more than methodological interest when it comes to monitoring excess cancer risk potentially induced by drug combinations. There may be a limited value though in terms of discovering putative cancer risk reduction associated with certain drug combinations that can be then tested in other real-world datasets or trials.

Response 1.1: We appreciate the reviewer's concern will represent an important point for many readers, but we maintain that the 3.25 year median follow-up time does not preclude our study having value in the context of cancer. In order to explain our reasoning, we have included here a short summary of the points and evidence that is presented in the revised Discussion.

- *Drug-cancer effects described in randomized trials have similar follow-up time.* [Women's Health Initiative](#) uncovered effects of menopausal hormonal therapies on breast cancer in their original 2.75 year study period. This follow-up time was sufficient to identify [beneficial effect of estrogen](#) reducing risk of breast cancer, and detrimental effect of estrogen plus progesterone increasing cancer risk. Similarly, the study by [Burn](#) of protective effects of aspirin followed participants for a maximum of 5 years.
- *Epidemiological observational studies often have a similar follow-up time.* While many observational examples are available, we described two well-cited examples. A [study](#) investigating beneficial effects of tricyclic antidepressants on glioma incidence found the effect at 6 months of follow-up (over 60 citations). Similarly, the effect of hormonal contraception on breast cancer was observable for patients with 1-4 years of use and only three months of follow-up time ([Fitzpatrick 2023](#), cited 20 times)
- *Previous comprehensive drug-wide association studies have had a similar follow-up time.* Two major drug-wide studies, similar to our own, each had a long-term usage follow-up period of two years [[Støer, 2020](#), [Pottegård, 2016](#)].
- *Drug effect on cancer can actually attenuate over a longer follow-up time.* In the Women's Health Initiative studies described above, the protective effects of estrogen alone on breast cancer decreased to non-significance in the post-intervention period after 2.75 years, while the increased risk in the combined estrogen and progesterone group decreased in the post-intervention period. Similarly, harmful effects of contraceptives on breast cancer disappeared in the late follow-up period (by 4-5 years [Burchhardt, 2022](#), [Fitzpatrick, 2023](#)) .
- *While we agree that true oncogenesis takes many more years, a biomedically relevant drug effect need not be restricted to oncogenesis.* The authors of the breast cancer studies cited above hypothesized that the shorter term effect is likely due to effect on existing neoplastic but precancerous tissue. Their [observation](#) that abnormal mammograms increased within one year of hormone use supports this proposition.
- *A lack of longer follow-up may bias findings toward the null, but this does not invalidate non-null findings.* While we cannot exclude an effect of drug combinations on cancer that could occur on a longer time scale, such as an impact on oncogenesis, this lack of longer follow-up does not in any way invalidate or bias findings that we do discover.

Reviewer #2 (Remarks to the Author):

Many thanks for this submission. Topic and concept are relevant addressing an unmet need - assessing the effect of drug combinations on incident cancer. The paper is very well written.

The huge IBM dataset goes back to 2015 only.

Response 2.1: We clarify that the data set ranges from 2003 to 2015 (first section of Results).

9460 drug combinations were tested, but not entirely clear to me why this number.

Response 2.2: We have edited the text to clarify that we test all combinations where the resulting cohort study included at least 100 cases of some type of common cancer (first section of Results)

The potential issues with DrugWAS are nicely explored.

The rationale for the methodology developed by the authors is quite clear and explanation is easy to follow.

The main issue is lack of validation and biological plausibility for most of the findings> Validation of findings using Mendelian Randomization and a detailed discussing of biological plausibility for the observed associations would be of interest.

Response 2.3: Although validation would require another similarly sized data set, which, to our knowledge, does not exist, and we do not know of any Mendelian randomization methods appropriate for testing a possible drug combination effect, we have taken a number of steps to perform evaluation. Previously we performed rigorous sensitivity analyses, including varying the definition of drug exposure; a dose-specific analysis; and an analysis that assesses drugs per active ingredient.

To address this concern and to strengthen confidence in our findings we have added a new section to the Results, called "*Per-combination results concord with prior drug-wide studies of cancer*"; including a new Table 1; a new Supplementary Figure S2; and we have included all per-combination results in Supplementary Data S2 and S3. In this new Results section, we investigate whether our findings align with a number of systematic drug-wide association studies of cancer. A key part of the new section is excerpted below:

In order to evaluate our results, we endeavor to compare them with previous studies. However, it is important to note first, that there are no previous drug combination-wide association studies of cancer, making a direct comparison with our results difficult. Instead, we compare to previous drug-wide association studies, under the assumption that drugs with a previously reported single-drug effect may also have a drug combination effect. Second, we have explicitly stated our focus is on discovering effects that are consistent across all cancers due to obvious confounding; in contrast, previous drug-wide association studies investigated cancer site-specific associations.

We focus on two recent drug wide association studies, McDowell¹ and Støer¹ (Table 1). Of these, Støer made an effort to evaluate their results against earlier reports by Patel¹ and Pottegård¹, along with a less formal comparison to the much older work of Selby¹. Out of 21 drug-cancer signals they highlight, including 18 replicated from earlier studies and three novel signals discussed in the study, we could test 15 of these drug-cancer associations, finding 10 out of 15 concordant signals. By comparing the results to McDowell, who studied the top 250 most common medications in their primary care data set, we found that 14 of the medication associations from Støer could be tested, and 10 of 14 showed a similar signal. Combining the two results, of nine signals present in both McDowell and Støer that were testable in our data set, seven of the nine replicated. This replication rate in published studies, then, is similar to the replication rate we achieved, despite the noted different goals and design of our studies.

We agree on discussion of biological plausibility, and we have added follow-up literature review. Alongside our previous discussion of possible mechanisms of the combination effect of omega 3 fish oil and fenofibrate, we added references regarding the combination of tiotropium and diclofenac; niacin and oseltamivir; and amlodipine and pravastatin.

How do the authors explain that amlodipine + pravastatin is associated with reduced risk of cancer, and amlodipine + other stronger statins is not also a hit?

Response 2.4: To address this concern we have undertaken an analysis of whether drug combination effects with pravastatin are also present in other statins. For the amlodipine and pravastatin combination specifically, we find a similar effect for amlodipine and rosuvastatin (log HR = -0.2, posterior interval range [-.4, 0.01], versus for amlodipine + pravastatin, log HR = -26, [-0.48,-0.09]); the rosuvastatin combination has only 75% of the number of patients as the pravastatin combination. Overall, the correlation between the posterior bounds (which is the quantity we use to identify drug effect) is 0.46 for the 1% bound, and 0.33 for the 99% bound. To strengthen confidence in our findings, we have added a *Supplementary Figure S2C* summarizing the correlation in combination drug effects for all statins.

Similarly, for most of the detected associations, it seems hard to understand and believe that there is a protective effect based on my large clinical experience.

Response 2.5: We are grateful for the experienced feedback of this reviewer, but we respectfully suggest that it is not possible for any health provider, no matter how experienced, to notice these patterns via experience. We believe that the exciting potential of large data is in the ability to analyze cohorts that cannot be collected in any one medical center: for example, in our large data set of over 100 million individuals, only around 3000 are recorded to be treated with omega-3 fatty acids and then fenofibrate. We hope that the edits above address this concern.

Reviewer #3 (Remarks to the Author):

This paper focused on the “Drug combination-wide association studies of cancer”.

The aim of this study sounds very promising, however the authors have ignored recent developments in epidemiology.

Over the last decade, the optimal way to perform comparative effectiveness research (especially in pharmacoepidemiology) is to utilise the target trial emulation framework

<https://jamanetwork.com/journals/jama/article-abstract/2799678>

<https://pubmed.ncbi.nlm.nih.gov/26994063/>

(even if the authors use 3-4 references of papers that use the TTE framework, but they cite them only on how they estimated IPW)

Specifically, to estimate the effect of medication A vs dual medication A and B, the authors should specify 7 key components of both the target trial and the emulated trial protocol: the eligibility criteria, treatment strategies being compared (including their start and end times), assignment procedures, follow-up period, outcome of interest, causal contrast(s) of interest, and analysis plan.

Response 3.1: We agree with this criticism and we have adjusted our presentation to clarify that all of our analysis is heavily inspired off of the target trial framework and the work of Hernán and colleagues. This includes the overall design of the hypothetical randomized trial we are emulating; the marginal structural model approach to adjust for time-varying confounding; and the framework for creating a sequence of nested trials. We appreciate the feedback that the clarity would be improved by explicitly comparing our design to the seven key components of a target trial. To address this issue we have added Supplementary Figure S3, which includes a flow chart containing the seven key components and the protocol to

emulate them. As well, matching this figure, we have revised the first section of the Methods which is now named "Overview of approach using observational data to emulate a target trial of drug combinations", to clarify how the analysis follows this framework.

Their analysis is based on the definition of causality of the Bradford-Hill criteria, which had revolutionized the way of thinking for epidemiologists back in 1965, however this definition cannot take into consideration the use of modern Big Data and the challenges that come with that. One a minor point, the language is outdated in causal inference in some cases, as for example the term "consistency" for Bradford-Hill is translated as "consistent findings observed by different persons with different samples strengthens the likelihood of an effect", while in modern causal inference theory, which has adopted the potential outcome framework, consistency is about well-defined interventions (or more formally that the observed outcome for every treated individual equals her outcome if she had received treatment, and that the observed outcome for every untreated individual equals her outcome if she had remained untreated).

Response 3.2: We hope that the previous comment addresses some of this concern, but to address the conflicting meanings of the term "consistency" we have revised our text to refer to "Bradford-Hill consistency" throughout. Although we agree that the potential outcomes use of consistency is important, this is an assumption in most observational studies. Our revision includes the following:

The Bradford Hill consistency criterion posits that a true effect should result in consistently observed signals, such as in repeated studies or in similar contexts. (Note that this should not be confused with the identifiability condition called consistency, which is required for any observational study to emulate a target trial and is an independent concept.)

The focus of the authors is only around confounding – of course this is a very important problem that is responsible for bias in medical research. However, an equally important problem is the clarification of the protocol of this comparison of medication A vs dual medication A and B. And then the protocol of the target trial will dictate the analysis plan, e.g. for the estimation of the observational analogue of the intention-to-treat effect, one should not censor individuals that discontinued treatment (and should not use adherence weights), while for the observational analogue of the per-protocol effect (i.e. effect of the intervention, had everybody adhered), one should not censor individuals that discontinued treatment (and should not use adherence weights).

Response 3.3: We hope the revision described in Response 3.1 clarifying the target trial addresses this concern. In our main analysis we compare those who sustain drug A alone against those who sustain drug A and additionally take drug B; this entails censoring those who discontinue drug A. In a sensitivity analysis, we ignore whether or not they sustained drug A.

Moreover, there is confusion about the use of negative controls. A negative control outcome is a tool for detecting confounding and bias in observational studies (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3053408/>). Specifically, a negative control outcome is selected in such a way so that it should not be related with the treatment. If the treatment-outcome relationship is not null, then there is an indication that there is unmeasured confounding in the data, so we cannot really trust the results of our main analysis. The authors seem to have selected their negative control on a different basis, that is not completely clear.

Response 3.4: The test we previously labeled as a negative control test was intended to operate on a similar principle to the idea the reviewer describes, but with the unmeasured confounding in this case being overall predisposition to disease that could cause people to both take a drug and to develop many health conditions including cancer. However, we appreciate that this was a confusing way to describe the concept, and we have removed all references to "negative control" and replaced it with the more precise terminology, describing these as a way to exclude the hypothesis that general sickness explains the association.

There is also confusion about the use of IPW. The weights described in the paper are in fact the adherence weights that should have been used only in the per-protocol analysis, to account for those individuals who discontinue their treatment (<https://www.nature.com/articles/s41591-019-0597-x>). The description of these weights is not clear from the authors – it seems that they use them to account for confounders without providing many clarifications.

Response 3.5: We appreciate this feedback and to address this issue we have edited the text to better describe the weights, in the Methods subsection revised with the title "Censoring and adherence weights". We did not previously use the term "adherence weights" but referred simply to the probability of not being censored. This unclear terminology is fixed, and in that section we describe that the adherence weights are the result of modeling the probability of discontinuing drug A. The adherence weights are not the same as the treatment weights, which are described in "Implementing the marginal structural model for time to cancer incidence" and in the section "High-dimensional propensity score model". The treatment weights are the result of modeling the probability of initiating the combination A+B, among the population eligible to initiate.

Overall, I believe that any comparison between single and dual medication should be based on a certain protocol, that has to be clearly stated. I don't think that a "drug combination-wide association study" will give more insight about the benefits of the dual medication, even in the absence of unmeasured confounding.

Response 3.6: We hope that the edits above clarifying the target trial and its implementation address this concern.

The analyses of observational data can result in substantial selection bias, if the design of these studies are flawed, especially in pharmacoepidemiology. There are many examples in the literature that highlight this problem, e.g. the effect of statin on cancer has been measured up to 65% beneficial in some studies (that accounted for confounding), however the effect was way more modest (almost null), when the TTE was implemented to study this relationship
<https://pubmed.ncbi.nlm.nih.gov/31591592/>

Response 3.7: We hope the edits above address this concern. We have carefully avoided issues such as prevalent user bias or selection into the study based on any subject data collected in the follow-up period, and we emphasize that we closely follow the procedure of emulating a randomized trial even though our previous manuscript did not use the term "target trial". When post-baseline information is used, only for censoring, we carefully adjust for the resulting possible selection bias with the adherence weights. We appreciate the rigor of this critique and we hope we have satisfied all concerns about our methods.

Dear Reviewers,

We appreciate your time spent on our manuscript and we believe we have clarified all remaining issues with edits to the text and to Supplementary Figure S3. Below, we have included the full text of the reviews in italics, and for each concern, we include a numbered response indented below, with all changes to the text copied in and indented further and marked with a vertical line.

Sincerely,
The Authors

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

I raised the question about whether 3-4 years of follow-up are adequate for inducing new cancers - I do not think they are. The authors provide a lot of references, which fall into two categories: 1. Papers with breast or other, hormone-sensitive tumours. 2. Few other papers by Stoer

<https://aacrjournals.org/cebp/article/30/4/682/670683/Drug-Use-and-Cancer-Risk-A-Drug-Wide-Association> , or Pottegard where the authors claim that the follow-up was only 2 years, but from looking at these papers, I could not find this information (I admit I may have missed it).

The author state: "The authors of the breast cancer studies cited above hypothesized that the shorter term effect is likely due to effect on existing neoplastic but precancerous tissue." This is totally true, here we are not talking about genuine cancer development de novo / oncogenesis, but an 'acceleration' of the growth of existing tumours, such as the case for as-yet-asymptomatic / very early breast cancer being accelerated into a clinical phase by hormone-active therapies. I think most of the effects that are being described will be due to such mechanisms, not oncogenesis.

This needs to be declared as such in the paper.

Response 1.1: We agree with the Reviewer that detecting true oncogenic effects of drugs is challenging with the current available data. We appreciate that the Reviewer agrees it is possible to discover effects of drugs on early stage tumors, but we clarify that such effects **need not be restricted to hormone therapies**. For example, epidemiological work on a similar time scale found antidepressants were associated with incidence of colorectal cancers and gliomas ([Walker 2011](#) *British Journal of Cancer*). The effect of these drugs on glioma growth was experimentally validated in mouse models of glioma ([Shchors 2015](#) *Cancer Cell*). Discovering drug combination effects that impact growth of latent early stage tumors is still biomedically impactful.

In addition to underlining this issue in the Discussion, given this Reviewer's emphasis we have added this point to the Introduction so readers will not be misled into

thinking we claim to discover true oncogenesis. We hope that these declarations satisfy these concerns:

Edits to Introduction:

While true oncogenesis may take decades, longer than the follow-up time in current health data, drug combinations that impact the growth of latent tumors may result in observable change in cancer risk in exposed individuals.

Edits to Fourth paragraph of Discussion:

A second limitation of our work is the limited follow-up time available for observing drug combination effects. In our analysis with no censoring, we have 3.25 years of follow-up time available. This is short, but it is in line with some single drug^{41,42} and combination drug effects¹⁰ discovered in randomized trials. The follow-up time is also similar to both highly cited observational studies of specific drug effects on cancer^{43,44}, and recent landmark drug-wide association studies of cancer^{3,4}. It is not obvious that a longer follow-up time is more meaningful: in the Women's Health Initiative study, beneficial effects of estrogen decreased to non-significance in the post-intervention period after 2.75 years, while detrimental effect of combined estrogen and progesterone on breast cancer decreased⁴², a pattern also seen in effects of contraceptives on breast cancer^{43,45}. While true effects on oncogenesis can only be discovered over many years, drugs may also impact cancer by an effect on precancerous neoplastic tissues: abnormal mammograms were found after only one year of hormone use⁴⁶. Therefore, the available follow-up time may have limited the discoveries possible in our study, but future studies will have longer follow-up time available, and our method will be directly applicable in that work. Though our analysis cannot exclude a possible longer-term drug effect of any drug combinations, this issue does not invalidate findings made with shorter available follow-up data.

Edits to Last paragraph of Discussion:

Future health data will contain longer follow-up time. Applying our methods to these data will then allow discovery not only of the effects of drug combinations on tumor growth, as was possible here, but also for discovery of impacts of drug combinations on oncogenesis.

Reviewer #2 (Remarks to the Author):

Thanks for addressing the previously raised concerns. I am happy with the answers and believe the manuscript now reads better.

Reviewer #3 (Remarks to the Author):

In principle, I strongly disagree with the performance of exposure-wide (or combination-wide) epidemiology. Such exposure-wide studies are prone to numerous biases because of the inability to make simultaneous confounding control.

Response 3.1: Overall, the Reviewer points out possible pitfalls of an exposure-wide approach, but we emphasize that we agree with these points and have carefully designed our study to consider and avoid them. We respond regarding **collider bias** in Response 3.2, **immortal time bias** in Response 3.4, and other details of **emulating a target trial** in an exposure-wide setting in Response 3.3 and 3.6. Therefore, we maintain that an exposure-wide study that scrupulously emulates the steps of the target trial does not necessarily introduce bias, and these sources of bias have been adequately addressed by the design of our particular exposure-wide study.

As well, we are not the first nor only to implement exposure-wide analyses (also known as high-throughput target trial emulation). Many high impact recent works demonstrate that this strategy is of growing interest for discovery in chronic diseases. Beyond the publications we cited for cancer (Pottegard 2016 *EBioMedicine*; Stør 2021 *Cancer Epidemiology Biomarkers Prevention* (AACR); Patel 2016 *Scientific Reports*; McDowell 2021 *BMC Medicine*), there are analogous studies for other chronic diseases including Coronary Artery Disease ([Liu 2021](#), *Nature Machine Intelligence*), Alzheimer's (Zang 2023, *Nature Communications*, particularly see [Table 1](#)), as well as ongoing efforts by the Observational Health Data Sciences and Informatics consortium (for example, studying hypertension in [Schuemie 2020](#), *J American Medical Informatics Assoc*). Although the Reviewer has a strong objection on principle, and we appreciate and agree with the rigor of these concerns, we hope that the Reviewer will consider that our careful design provides a valuable contribution to the literature and is meritorious of publication in this revised form.

Certain variables can be considered as confounders for a certain exposure-outcome relationships, however the same variables can be mediators for other exposure-outcome relationships. For more details, see <https://pmc.ncbi.nlm.nih.gov/articles/PMC5378648/> which presents certain limitations of <https://pubmed.ncbi.nlm.nih.gov/26646432/>

Response 3.2: The Reviewer and the reference they provide point out that it is crucial to avoid **collider bias**, which can occur when a **mediator** (defined here as a downstream effect of an exposure, that also affects the outcome) is adjusted for as a confounder. We fully agree that this can be a source of bias in cross-sectional outcome-wide studies, but we believe our design carefully avoids this risk by exploiting the **temporal ordering of events in the health record**. The reference that the reviewer cites (VanderWeele 2017 *Epidemiology*) describes examples such as the influence of race on education mediating the effect of race on health. In that type of exposure-wide analysis, there is no temporal ordering of variables, and therefore it is impossible to disentangle mediators (with their risk of introducing collider bias) from confounders that influence whether someone is exposed. The reference the Reviewer cites also suggests that collider bias can be avoided by using the temporal ordering of events.

Our analysis is not cross-sectional but uses temporal data, carefully adjusting only for medical history occurring before the exposure. To clarify this point, we have added edits to the Methods to clarify, in the section "Overview of approach using observational data to emulate a target trial of drug combinations", as copied below:

In order to perform a drug-wide scan, we must consider that many confounders, including unexpected confounders, could affect prescription. Therefore, our set of possible confounders is drawn from medical history before prescription. Because of the time-specific nature of medical data, colliders (VanderWeele 2009 *Epidemiology*) are unlikely: this would require both the treatment and the exposure to affect some variable that occurs *before* either of them.

As well as edits in the Methods section "High-dimensional propensity score model":

It is crucial that this model uses the temporal information to adjust only for data occurring before exposure, in order to mitigate risk of collider bias that can particularly flaw exposure-wide studies (VanderWeele 2017 *Epidemiology*).

As well, we have included this issue as a point in the Discussion:

While investigating drug combinations one at a time could enable experts to craft a smaller set of confounding variables for each combination under study, by using the temporal ordering of events we are able to mitigate collider bias that some exposure-wide studies may suffer (VanderWeele 2017 *Epidemiology*). A drug combination-wide study is able to generate findings that may never have been investigated. As well, since the study is performed in a consistent and systematic way for all drug pairs, investigator bias is minimized (Ionnidis 2016 *Statistics in Medicine*).

In target trial emulation (TTE) studies, the situation is even more complicated.

When we perform a certain TTE study, we try to think of the hypothetical (target) trial and then emulate it. We try to do this by emulating each of the 7 components (that now the authors did add in their revised version, however, the description is very vague). The problem is that each of these trials can be very different in terms of the eligibility criteria, the definition of the treatment strategy and the causal contract.

Let's take a certain example. Imagine a certain target trial which would enrol hypertensive individuals. The 2 different arms would be

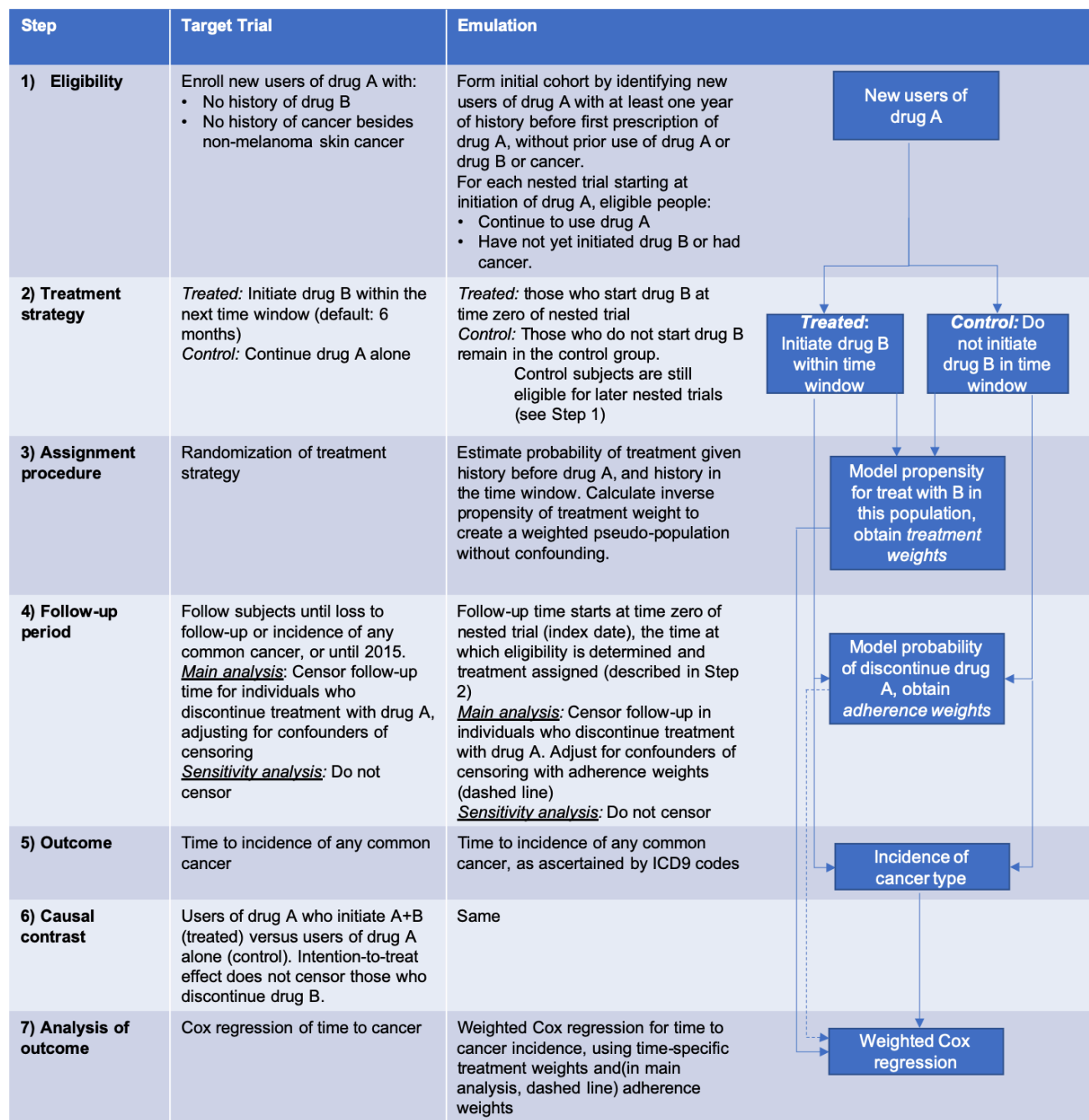
a) Thiazide diuretics and b) Thiazide diuretics & ACE inhibitors

Response 3.3: We agree with the Reviewer about the principles of target trial emulation, and we have improved the text to clarify our approach for rigorously emulating the seven steps of the target trial in the exposure-wide context, including description of the causal contrast (see also *Response 3.6*). The Target Trial Emulation is described in the

Methods section "Overview of approach using observational data to emulate a target trial of drug combinations", and it is also presented in Supplementary Fig S3 (updated version copied below). Below, we have created a table demonstrating three of the steps of target trial emulation mentioned above. This table shows how the example drug combination described by the Reviewer can indeed be presented in generic terms, where "drug A" is Thiazides and "drug B" is ACE inhibitors.

Target Trial Step	Single Target Trial Emulation comparing Thiazide diuretics against combination of Thiazide diuretics & ACE inhibitors	High-throughput version (Supplementary Fig S3)
1: Eligibility Criteria	Form initial cohort by identifying new users of Thiazides with at least one year of history before first prescription of Thiazides, without prior use of Thiazides or ACE inhibitors or cancer. For each nested trial starting at initiation of Thiazides, eligible people: - Continue to use Thiazides - Have not yet initiated ACE inhibitors or had cancer.	Form initial cohort by identifying new users of drug A with at least one year of history before first prescription of drug A, without prior use of drug A or drug B or cancer. For each nested trial starting at initiation of drug A, eligible people: - Continue to use drug A - Have not yet initiated drug B or had cancer.
2: Treatment Strategy	<i>Treated:</i> those who start ACE inhibitors at time zero of nested trial <i>Control:</i> Those who do not start ACE inhibitors are in the control group.	<i>Treated:</i> those who start drug B at time zero of nested trial <i>Control:</i> Those who do not start drug B are in the control group.
6: Causal Contrast	Users of Thiazides who initiate Thiazides + ACE inhibitors (treated) versus users of Thiazides alone (control). Intention-to-treat effect does not censor those who discontinue ACE inhibitors.	Users of drug A who initiate A+B (treated) versus users of drug A alone (control). Intention-to-treat effect does not censor those who discontinue drug B.

All 7 components of target trial emulation are presented in Supplementary Fig S3, below:



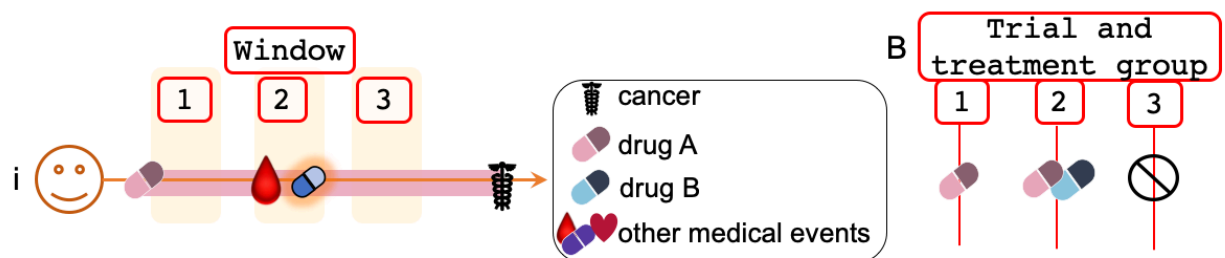
Supplementary Fig S3 Summary of emulation of a target trial for testing drug combination effect on cancer incidence.

In the observational data, we can find individuals (say individuals A, B and C) that initiated thiazide diuretics for 6 months and then they additionally initiated ACE inhibitors. The authors have allocated all these individuals in the arm b) Thiazide diuretics & ACE inhibitors, by considering as time zero the 1st time that individuals initiated ACE inhibitors. This

is wrong, as in this way it is introduced a severe amount of bias. The correct approach is to break the follow-up time by month and clone individuals. The first 6 months, all 3 individuals A, B and C are contributing in arm a) Thiazide diuretics and they are censored, as they deviated from their strategy. Then they are eligible to b) Thiazide diuretics & ACE inhibitors (i.e. the same individuals are eligible to both arms with different follow-up time, outcome etc). Or alternatively they contribute at random to only one of the 2 arms (e.g. individual A to a) Thiazide diuretics and B & C to b) Thiazide diuretics & ACE inhibitors). For more details, see <https://pubmed.ncbi.nlm.nih.gov/27237061/>

Another issue is that each trial is different and for example some trials utilise a “grace period” before treatment initiation. This issue is tackled using cloning, censoring and weighing.

Response 3.4: Regarding the issue of correct assignment of **time zero** and risk of **immortal time bias**, we appreciate the Reviewer's concern for this bias when an analysis fails to align time zero (index date) as 1) trial eligibility date, 2) treatment assignment, and 3) follow-up time zero. That is, we also agree with the Reviewer's description of the correct implementation of the **Treatment Strategy** (Step 2) in the Target Trial. We have clarified the text to show that we closely adhere to the design cited by the Reviewer, as outlined in the Methods, with an example in Supplementary Figure S4 and Supplementary Table 1. Our example *Person i* in Supplementary Figure S4A (excerpted below) follows the Reviewer's example of starting Thiazides and only later starting ACE inhibitors:



As shown in Supplementary Figure S4B, *Person i* is cloned in the way the reviewer recommends, to participate in both Trial 1 (which initiates with time zero in the first time window, at the time of drug A initiation) and Trial 2 (which initiates at the with time zero being the time interval 2). In Trial 1 they are assigned to the Thiazide-only group, and followed for four windows from Time 1 to Time 4, where they incur the cancer outcome. In the later Trial 2, which initiates in the time window where *Person i* starts ACE inhibitors, *Person i* is assigned to the combination arm. Therefore, follow-up time is only assigned to the time under exposure, and time where an individual is only exposed to the single drug is allocated to that group. As this is an intention-to-treat analysis (see also *Response 3.6*), follow-up time is not censored for non-adherence. An excerpt of Supplementary Table 1 shows the expansion into person-time for the two trials *Person i* participates in:

Supplementary Table				
Person	Trial (time zero)	Follow-up index (calendar time = time zero + follow-up - 1)	Treatment	Outcome
i	1	1	A	no
i	1	2	A	no
i	1	3	A	no
i	1	4	A	yes
i	2	1	A+B	no
i	2	2	A+B	no
i	2	3	A+B	yes

To address this issue, we have edited the text to clarify this point, removing the impression that our analysis is vulnerable to immortal time bias. The edits clarify that **time zero** (the start of follow-up time) is the start of each nested trial, and eligibility and treatment assignment are determined at that time. In the section "Overview of approach using observational data to emulate a target trial of drug combinations" we added:

Importantly, each trial is defined by initiation at a time zero, which is the time at which eligibility and treatment assignment is determined. Since follow-up commences at this time zero, only person-time after exposure to the drug combination is included in the combination exposure arm, avoiding immortal time bias (Hernán 2016 *J Clin Epidemiol*).

As well, in the section "Expansion of medical record into person time" we added:

In this approach, individuals are cloned and they are only assigned to the combination arm after they are exposed to the combination. Because follow-up time under combination exposure only begins after the joint exposure, this strategy is able to avoid immortal time bias. In order to account for time-varying confounding, we follow the approach of Danaei, et. al., in which participant time after enrollment (initiation of drug A) is used to form a sequence of randomized trials, each defined by trial initiation date, known as time zero (Danaei 2013 *J Clinical Epidemiology*, Danaei 2018 *Stat Methods Med Res*). In this approach, individuals are cloned and each individual's eligibility and treatment group is determined for each clone at each trial time zero. Therefore, person-time is only assigned to the combination arm in the time after that person is exposed

to the combination. Because follow-up time under combination exposure only begins after the joint exposure, the design avoids immortal time bias.

The reference cited by the Reviewer also alludes to the possibility of **selection bias** (bias for presence of data). When censoring those who are not eligible to participate in further trials due to discontinuation of the first drug, we model the influence of history before censoring. As we are emulating the intention-to-treat effect, censoring follow-up time for nonadherence is not needed.

In brief, TTE studies are not like GWAS studies, they cannot be conducted all at once. In a TTE study, we cannot say over simplistically that we compare drug X vs drugs X & Y. We need to describe the 7 components of the trial explicitly (e.g. eligibility criteria: individuals aged >50years old without cancer). This has not been done by the authors and cannot be done in general as these components (and especially the eligibility criteria) will vary from trial to trial. In summary, it is not correct to perform treatment-wide (or combination-wide) TTE studies, because each of these trials is different.

Response 3.5: We respond to this critique as well in *Response 3.1*. We emphasize that we have explicitly described the seven components of the target trial in the Methods section "Overview of approach using observational data to emulate a target trial of drug combinations" and in Supplementary Figure S3. Our approach to describing the high-throughput implementation of the seven components of a target trial is similar to that taken by [Zang 2023, Nature Communications](#).

I have other comments as well, e.g. the causal contrast is vague, we usually utilise the observational analogue of the intention-to-treat (ITT) or the observational analogue of the per-protocol effect... In any case, the main problem of the paper is that it is wrong to perform blindly treatment-wide TTE studies, as I mentioned before.

Response 3.6: We have clarified the **Causal Contrast** in the Methods section "Overview of approach using observational data to emulate a target trial of drug combinations" and in Supplementary Figure S3. The causal contrast compares the initiation of the drug combination as opposed to sustained use of the single drug alone, so we have clarified that this can be classified as an Intention-to-treat effect:

This can be interpreted as the intention-to-treat effect of the drug combination, as we do not additionally censor those who discontinue use of drug B.

While it is only possible for us to respond to comments included by the reviewer, we hope that these comments have clarified the listed concerns, and that the reviewer will appreciate the efforts we have made to address the comments written. We believe that we have concisely described all steps taken to emulate a target trial, for each combination of drugs. We disagree that it is not possible to do so, because we have carefully considered factors raised by the reviewer including immortal time bias, collider bias, and meticulous specification and emulation of a target trial.