We greatly appreciate the opportunity to revise our manuscript and to respond to comments and questions from the reviewers and editors. We feel that the manuscript is now greatly improved in conveying the clinical relevance of our results as well as the innovative features of our methodologic approach.

We thank the Editor for guidance on submitting a revised version and have responded to each of the points outlined in their letter.

- Our statistician quite liked this as some aspects are very good, but there were statistical issues that are detailed in his report (see Reviewer 2's comments)

We thank the statistical reviewer for his insightful comments and have carefully addressed them below. In particular, please see our reply to Reviewer 2 and the revised manuscript.

- One editor felt that this was a clinically relevant topic, but he didn't think this added much.

We thank the editor for giving us the opportunity to emphasise the novel aspects of our work.
This IPD MA contributes the following 5 new findings:

1. **Naproxen MI risk**
   Naproxen was associated with increases in risk of acute myocardial infarction. Previous work has generally been underpowered to fully assess the risk of naproxen and while point estimates have been < 1, confidence intervals have not been able to exclude clinically meaningful risks. Our study has the required power to address this shortcoming.
2. **Celecoxib MI risk**
   With celecoxib, associated risk of acute MI were not greater than that of diclofenac, ibuprofen, or naproxen, were and lower than those of rofecoxib. Again the power of our study enables these additional insights.
3. **Dose-MI risk relationship**
   Acute MI risks were dose-related and use for 1 to 4 weeks at high-dose seemed particularly harmful with rofecoxib (> 50 mg/day), ibuprofen (> 1200 mg/day), and with naproxen (> 750 mg/day). Previous studies could not adequately assess the dose-MI risk response with NSAIDs.
4. **Early onset of risk**
   For all studied NSAIDs, the onset of associated MI risk occurred within the first week of use. Virtually no previous studies have extensively examined very short term risk, again perhaps due to a lack of events in this earlier time window. Documenting a rapid onset of risk is important information for both clinicians and patients and may help improve decision making about the risks and benefits trade-off with these drugs.
5. **Patterns of MI risk**
   NSAIDs exhibited distinct patterns of MI risk, which suggest that time dependency of risk may differ between NSAIDs. No previous research has employed the methodology required to examine this issue.

The finding that naproxen, along with other NSAIDs, was associated with an increased risk of acute MI clearly is of interest to a broad audience of clinicians. Indeed, the present general impression in the scientific community is that naproxen is the least cardiotoxic of NSAIDs [1] and some consider high-dose (500 mg twice daily) naproxen as being cardiac neutral. [2] Reporting of meta-analysis results may contribute to the perception that naproxen is safer. For example, the latest network meta-analysis (NMA) of NSAIDs RCTs, the Coxib and traditional NSAID Trialists' (CNT) Collaboration (Lancet 2013) [3], states that "high-dose naproxen was not associated with any significant excess risk of major vascular events (0·93, 0·69–1·27; figure 4), and nor was there an increase in major coronary events (0·84, 0·52–1·35)." We submit that the latter adjusted rate ratio of naproxen vs placebo lacks sufficient precision to rule out clinically meaningful differences. This is because the confidence interval is compatible with a decrease in risk of major coronary events of 48% but also with an increase in risk of MI or CHD death of 35%. The lack of precision in NMA of RCTs is not surprising, as it can be estimated that NSAIDs RCTs provide fewer than 1000 acute MI events – 532 over all exposures in the NMA of Trelle 2011 [4] – whereas our IPD MA comprised 61 460 acute MI events.

Of note, in the CNT meta-analysis, MI risk was not considered separately for the various coxibs. The findings of our IPD MA suggest there are notable differences between rofecoxib and celecoxib for this outcome – and this is also of interest to clinicians.

Certainly, our findings are not immune to bias such as misclassification due to partial non-adherence or to residual confounding. However, the two latest and much cited works (the Trelle 2011 and CNT 2013 NMAs) are direct or indirect comparisons of RCT data via a network meta-analysis, a technique that also presents issues of potential bias, indirectness, inconsistency, and imprecision (see Salanti G. et al. Evaluating the quality of evidence from a network meta-analysis. PLoS One. 2014;9(7):e99682).

In particular, various biases that may occur post-randomization in a RCT [5, 6] may potentially have affected the validity of trials included in these NMAs. These may include: 1) violation of treatment groups exchangeability due to non-persistence with assigned NSAID treatment, or 2) violation of treatment consistency due to group contamination (e.g. OTC

---

[1] Schmidt M, Lamberts M, Olsen AS, et al. Cardiovascular safety of non-aspirin non-steroidal anti-inflammatory drugs: review and position paper by the working group for Cardiovascular Pharmacotherapy of the European Society of Cardiology. Eur Heart J. 2016 Mar 16.

[2] Patrono C, Baigent C. Nonsteroidal anti-inflammatory drugs and the heart. Circulation. 2014;129(8):907-16.

[3] Bhala N, for the Coxib and traditional NSAID Trialists' (CNT) Collaboration, Emberson J, et al. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials. Lancet. 2013;382(9894):769-79.

[4] Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ. 2011;342:c7086.

[5] Hernán MA, Hernandez-Diaz S, Robins JM. Randomized trials analyzed as observational studies. Ann Intern Med. 2013 Oct 15;159(8):560-2.

[6] Manson JE, Shufelt CL, Robins JM. The potential for postrandomization confounding in randomized clinical trials. JAMA. 2016 Jun 7;315(21):2273-4.

NSAIDs) or to differential concomitant treatments potentially affecting the risk of the MI outcome during follow-up (e.g. low-dose aspirin) or sub-optimal adherence to the NSAID. For example, placebo-controlled RCTs of naproxen were mainly performed in populations with or at risk of Alzheimer's disease. [7] Non-adherence bias has been documented in the main naproxen study (ADAPT) [8] and this may translate into an underestimation of the risk of acute MI with naproxen in NMAs of RCTs.

One would hope that the PRECISION trial in patients with cardiovascular (CV) disease or at high CV risk will give more definitive inferences on MI with naproxen, ibuprofen, and celecoxib. However, only 580 Anti-Platelet Trialists' Collaboration composite events (cardiovascular death, non-fatal stroke, or non-fatal MI) are expected. [9]

The increased statistical power of this collaborative project has allowed answers to clinically pertinent research questions, which were to better characterize the effects of individual NSAID dose, duration of continuous use, and the time course of any associated risk for acute MI.

Therefore we submit that this IPD MA adds to– and, in certain instances, challenges conventional aspects of RCT and observational evidence on NSAIDs and acute MI risk.

- Another editor felt that the manuscript was clearly written and that most of the difficulty is in understanding the methods used.

We thank the editor and have amended the Methods section in manuscript to improve overall clarity. Please also see our reply to Reviewer 2.

- Another editor felt that this was an Important research question and believed that this data substantially added to the existing evidence. The methods used look very robust and the used data sources are very reasonable, particularly for a "real life setting". He added that you address confounding by indication (i.e., the biggest threat for the causal interpretation). He did not identify a fatal flaw.

One issue that the authors may want to discuss is a potential interaction between aspirin and non-aspirin NSAIDS and the fact the some studies also suggest increased risk of MI for users of

---

[7] Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ. 2011;342:c7086.

[8] Steering Committee of the Alzheimer's Disease Anti-inflammatory Prevention Trial (ADAPT). Statement for communication to the FDA Arthritis Advisory Committee and the Drug Safety and Risk Management Advisory Committee. 18 February 2005. Available at: https://jhuccs1.us/adapt/pdf%20documents/FDA%20ADAPT%20STATEMENT_web%20posting.pdf. Last accessed August 31, 2016.

[9] Gaffney M. Statistical Issues in the Design, Conduct and Analysis of Large Safety Studies. Available at: www.med.upenn.edu/cceb/biostat/.../MichaelGaffney_2015.pptx. Last accessed February 28, 2016.

acetaminophen. He believed that this study would be of great interest for the readership of the journal.

We thank the editor for his comments and suggestions. We have revised the manuscript to add the following paragraph:

> "We might have taken advantage of this one-stage IPD MA to investigate clinical heterogeneity via a formal analysis of interactions between NSAIDs and major CV risk profiles. [10, 11, 12] However, interaction studies involving a categorical NSAID exposure, such as in this IPD MA, would have less statistical power [13] and would overly depend on classifying exposure without error. [14]"

Of note, using the RAMQ cohort, we investigated a potential interaction between aspirin and non-aspirin NSAIDs. In this completed study to be submitted shortly for publication we found that concurrent cardioprotective aspirin does not appear to mitigate the risk of MI associated with NSAIDs.

We would rather not comment here or in the manuscript about possibly increased risk of MI for users of acetaminophen as we have not reviewed this literature nor studied this exposure in the IPD MA since the over-the-counter prescribing of acetaminophen excludes good exposure data.

- Another editor made the following comments:
1. We need more details about the search strategies. Some are in the appendix but ideally they should be placed in the main article. What databases were searched? When?

We thank the editor for many useful comments and questions.

In the revised Table 1 of Web Extra 1, we provide all search strategies. In the revised manuscript we added the following paragraph:

> "We searched MEDLINE and EMBASE via Ovid and PubMed by applying filters for retrieval of observational studies and by combining these with the appropriate search terms for NSAIDs and for myocardial infarction (Web Appendix 1,Table 1). We also retrieved systematic reviews of non-randomized

---

[10] Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Stat Med. 2007;26(15):2982-99.

[11] Fisher DJ, Copas AJ, Tierney JF, et al. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. J Clin Epidemiol. 2011;64(9):949-67.

[12] Debray TP, Moons KG, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods. 2015;6(4):293-309.

[13] Greenland S. Basic problems in interaction assessment. Environ Health Perspect. 1993;101(Suppl 4):59-66.

[14] Greenland S. Interactions in epidemiology: relevance, identification, and estimation. Epidemiology. 2009;20(1):14-7.

studies of CV adverse events associated with NSAIDs and manually searched their bibliography. To increase sensitivity for the MI outcome, we included broader terms for cardiac morbidity. After an initial search (2007) we set-up weekly updates (June 2010 to November 2013)."

You mention the criteria you used for identifying studies but some are somewhat vague and open to interpretation. We need more details. The I/E criteria are not easy to replicate, in my opinion.

For this systematic review and IPD meta-analysis, we elaborated *a priori* study selection criteria by considering all main sources of biases and how they specifically applied to our research questions. This allowed controlling for confounding (including that due to calendar time) and for misclassification at the stage of study selection. While we acknowledge that this could make it more difficult to replicate inclusion/exclusion of studies, we submit that the enhanced control of confounding and of information bias that ensues from meta-analysing highly internally valid observational studies is a major quality feature of work.

Two of us (MB and JMB) independently assessed the articles for inclusion once, based on the title and abstract, and again, based on a full-text review. We resolved our few disagreements through discussion. We checked the conformity of the identified studies against the a priori inclusion criteria (these are reported in the manuscript). Non-conformity with one or more of the inclusion criteria resulted in the exclusion of that study (this is reported in Web Extra 1, Figure 1).

In the table below we present the inclusion criteria and the main reasons for excluding studies side-by-side.

Table – Study inclusion criteria and main reason for excluding studies

Study selection criteria	**Number of studies excluded, with main reasons (N=76)**
1) Sourced from validated computerized drug prescription or medical databases	• Availability of larger source cohort (n=1, previous smaller RAMQ study)
2) Conducted in the general or an elderly population	• Limited external validity due to selection of population (n=6)
3) Appropriately allowed for time-dependent analyses	• Time dependency of drug exposure not adequately established (n=5) • Potential for important misclassification of exposure (n=7) • Survival run-in period when analysing outcome (n=5)
4) Documented acute MI separately from other CV outcomes	• Outcome is a composite or another cardiovascular endpoint (n=28)
5) Considered both selective COX-2 inhibitors and traditional NSAIDs	• Did not consider both traditional and COX-2 selective NSAIDs (n=11)
6) Compared acute MI risk in NSAID users with	• Risk in users not compared with risk in non-

Study selection criteria	Number of studies excluded, with main reasons (N=76)
non-users	users (n=5)
7) Minimized the effects of confounding by indication or by contraindication and of selective prescribing. This was achieved by limiting the inclusion of studies to those which were conducted before the withdrawal of rofecoxib, matched cases and controls for calendar time, and documented comorbidities and concomitant treatments.	• Not limited to period before rofecoxib withdrawal (n=8)

Please note that although tools exist for assessing the quality of observational studies, [15, 16] these critical appraisal instrument have a generic purpose and we considered them as being unsuitable in the context of our research question. For example, whereas a study defining "current" use of an NSAID as a having a prescription in the last 180 days qualified for inclusion in a standard meta-analysis of NSAIDs and acute MI, such as that by the SOS group, [17] due to potential for important misclassification of exposure we excluded such a study from our IPD MA, which aimed at precisely characterizing the time course of MI risk associated with NSAIDs.

2. I do not understand why "each study had to show this increased MI risk [as an internal quality control measure]. Would this bias the results?

We understand the editor's concern for bias and indeed limiting the selection of studies to those that showed increased risk of MI with rofecoxib would induce selection bias if the IPD MA was aimed at determining whether or not rofecoxib is associated with an increased risk of MI. This however was not our intent.

Because the RCT evidence of MI risk with rofecoxib is strong, consistent, and well established, [18, 19, 20, 21] our purpose in limiting the inclusion of studies to those showing an

[15] Margulis AV, Pladevall M, Riera-Guardia N, et al. Quality assessment of observational studies in a drug-safety systematic review, comparison of two tools: the Newcastle-Ottawa Scale and the RTI item bank. Clin Epidemiol. 2014;6:359-68.

[16] Cochrane Methods Bias. The ROBINS-I tool (Risk Of Bias In Non-randomized Studies - of Interventions). Available at: https://sites.google.com/site/riskofbiastool/. Last accessed August 13, 2016.

[17] Varas-Lorenzo C, Riera-Guardia N, Calingaert B, et al. Myocardial infarction and individual nonsteroidal anti-inflammatory drugs meta-analysis of observational studies. Pharmacoepidemiol Drug Saf. 2013;22(6):559-70.

[18] Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. N Engl J Med. 2000;343(21):1520-8.

[19] Bresalier RS, Sandler RS, Quan H, et al. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. N Engl J Med. 2005;352(11):1092-102.

[20] Baron JA, Sandler RS, Bresalier RS, et al. Cardiovascular events associated with rofecoxib: final analysis of the APPROVe trial. Lancet. 2008;372(9651):1756-64.

increased MI risk for rofecoxib was:
1) to ensure that only highly internally valid observational studies were included in the IPD MA (studies not showing an increased MI risk for rofecoxib are suspect for a bias to the null for other NSAIDs);
2) to use the results for rofecoxib in our IPD MA as an internal quality check for the measures of association reported for the other studied NSAIDs.

In the APPROVe trial of rofecoxib 25 mg vs placebo the hazard ratio (95% CI) of acute MI up to 14 days after end of treatment was 2.65 (1.21 to 5.75). [22] The Trelle NMA of RCTs reported a rate ratio (95%CrI) for rofecoxib (any daily dose) vs placebo of 2.12 (1.26 to 3.56). [23] A weighted average of the IPD MA findings across the five categories of current rofecoxib use (see Table 2 and Figure 6, Panel b of the manuscript) would yield results that are slightly lower than those of the APPROVe trial and Trelle NMA of RCTs. Assuming that the distribution of residual confounding and misclassification is similar across exposures this suggests that the results of the IPD MA might be somewhat biased to the null and thus that the 'true' but unknown risk of acute MI with associated NSAIDs might be possibly be greater than what we found.

We have specified the above in the manuscript to help with comparison with previous literature.

3. Some of the data discussed in the methods section under "data sources and outcome" should be placed in the results section (8 databases, four had to be excluded...)

We agree that these should be reported in the Results section and have moved the paragraph.

4. What outcomes were selected? were they the primary outcomes for the original studies.

The outcome of the IPD MA is acute myocardial infarction and this was also the primary outcome in original studies contributing IPD. We have specified this in the manuscript, which now reads:

> "The similarity of definition and ascertainment of cases of acute MI and controls in each study confirmed the appropriateness of pooling in a meta-analysis (Web Table 3)."

---

[21] Ross JS, Madigan D, Hill KP, et al. Pooled analysis of rofecoxib placebo-controlled clinical trial data: lessons for postmarket pharmaceutical safety surveillance. Arch Intern Med. 2009;169(21):1976-85.
[22] Baron JA, Sandler RS, Bresalier RS, et al. Cardiovascular events associated with rofecoxib: final analysis of the APPROVe trial. Lancet. 2008;372(9651):1756-64.
[23] Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ. 2011;342:c7086

Why was the Quebec database used as the reference standard?

The large RAMQ IPD dataset was created specifically for the purpose of this IPD MA, which required retrospective harmonization.

Griffith et al.[24] write "To achieve IPD harmonization, investigators of any retrospective harmonization initiative will need to follow a series of practical steps…. Secondly, based on documentation obtained and the scientific aims of the harmonization initiative, variables targeted to serve as reference for data harmonization across studies are selected. A priori selection of variables targeted for harmonization is generally guided by a balance between enabling integration of a significant number of studies to provide the benefits of large sample sizes, while restricting integration to studies providing the lowest level of heterogeneity possible. Finally, following the identification of reference variables and the selection of studies collecting the valid information required to construct these variables, various methodologies can be applied to transform study-specific data items under the target variable format."

This above description fits our approach to retrospective harmonization. The reference variables in the IPD MA were those of the RAMQ dataset. We have removed the word 'standard' in the revised manuscript as we realize that 'reference standard' may be confusing. The revised text reads as follows:

> "From the universal, computerized public insurance databases of Quebec, Canada, we created a nested case-control dataset (RAMQ), which served as reference for harmonizing the IPD from other studies."

5. I do not undertand how you used time. How did they account for the fact that some patients took NSAIDS for a while, then stopped, then resumed?

We strived to faithfully account for time-varying use of NSAIDs, including stopping (thus being unexposed to NSAIDs for some time), resuming the previous NSAID, starting a new one, or switching between NSAIDs.

The key for doing was creating episodes of NSAID use (or non-use) by using computer-recorded variables in each healthcare database. For each continuous episode of NSAID use, we documented the daily dose and the duration of treatment.

---

[24] Griffith L, van den Heuvel E, Fortier I, et al. Harmonization of Cognitive Measures in Individual Participant Data and Aggregate Data Meta-Analysis [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Mar. Introduction. Available from: http://www.ncbi.nlm.nih.gov/books/NBK132539/. Last accessed August 13, 2016.

So far in single observational studies or in their meta-analyses, NSAID exposure has typically been modelled using simple binary time-varying indicators of either current use (exposed or unexposed) or current dose (low or high dose). [25, 26]

Some of these studies have also separately explored the effect of categorized treatment duration and recency of use. [27, 28, 29] However, to more accurately relate acute MI to time-varying NSAID exposure and to characterize the time course of risk, we felt it was best to combine information about dose, duration, and timing of past treatments – which is what we did for measurement and modelling of NSAID exposure in this IPD MA.

Briefly, we represented time-varying use of NSAIDs by creating several multidimensional exposure categories with different indicator variables corresponding to the contribution of (i) recency of use, (ii) dose level, and (iii) treatment duration. The 8 indicator (dummy) NSAID exposure variables were: non-use, past use, recent use, current use for 1-7 days, current use at low daily dose for 8-30 days, current use at high daily dose for 8-30 days, current use at low daily dose for greater than 30 days, and current use at high daily dose for greater than 30 days.

Please see also reply 5) to Reviewer 2 for additional information and illustration of how episodes of use were built, NSAID categorical exposure variables were created, and how these indicator variables were included in the IPD MA model.

I think the authors should justify their use of some frequentist and some Bayesian analyses. They did both. As an informed but not sophisticated readers, I would like to know why they used both approaches.

The IPD MA was preceded by a frequentist analysis of each study. This helped with setting up the IPD MA because it informed the between-study heterogeneity. Indeed, by doing a frequentist analysis, we could observe how the ORs of acute MI varied across the four studies for each NSAID exposure category. We however realize that the interest in frequentist analysis is essentially technical and that presenting frequentist and Bayesian

---

[25] McGettigan P, Henry D. Cardiovascular risk with non-steroidal anti-inflammatory drugs: systematic review of population-based controlled observational studies. PLoS Med. 2011;8(9):e1001098.

[26] Varas-Lorenzo C, Riera-Guardia N, Calingaert B, et al. Myocardial infarction and individual nonsteroidal anti-inflammatory drugs meta-analysis of observational studies. Pharmacoepidemiol Drug Saf. 2013;22(6):559-70.

[27] Hammad TA, Graham DJ, Staffa JA, et al. Onset of acute myocardial infarction after use of non-steroidal anti-inflammatory drugs. Pharmacoepidemiol Drug Saf. 2008;17(4):315-21.

[28] Schjerning Olsen AM, Fosbol EL, Lindhardsen J, et al. Duration of treatment with nonsteroidal anti-inflammatory drugs and impact on risk of death and recurrent myocardial infarction in patients with prior myocardial infarction: a nationwide cohort study. Circulation. 2011;123(20):2226-35.

[29] Varas-Lorenzo C, Riera-Guardia N, Calingaert B, et al. Myocardial infarction and individual nonsteroidal anti-inflammatory drugs meta-analysis of observational studies. Pharmacoepidemiol Drug Saf. 2013;22(6):559-70.

results are not helpful for the reader. We have therefore removed mentions of frequentist analysis.

Why is the Bayesian approach superior in this setting?

In Bayesian analysis, combining prior information (as probability distributions) with the data on hand i.e. the current experience (via the likelihood) yields a posterior probability distribution for each parameter of interest, as follows:

$$\underset{\text{prior probability}}{p(\text{hypothesis})} \times \underset{\text{likelihood (frequentist p-value)}}{p(\text{data}|\text{hypothesis})} \propto \underset{\text{Bayesian posterior probability}}{p(\text{hypothesis}|\text{data})}$$

1) For this research, one major reason for a Bayesian framework was the opportunity to make clinically useful inferences by using the posterior distributions above and beyond obtaining pooled estimates of acute MI risk. Indeed, one huge advantage of the Bayesian approach is that it allows making direct probability statements. Such probability statements, based on the posterior distribution of an outcome parameter (here the pooled ORs of acute MI), may allow comparing interventions or assessing whether certain thresholds of risk are exceeded. This is straightforward as it can be done by writing the appropriate corresponding lines of code in the model and monitoring the results as output of the Bayesian analysis. Panels (a) of Figures 2 to 6, Figure 7, and Web Extra 1 Figure 3 illustrate how we took advantage of these features of Bayesian analysis. Note also that these probability statements have a correspondence with the credible interval (CrI) in a Bayesian analysis (usually represented by the values of 2.5th percentile up to the 97.5th percentile of the posterior distribution).
   Take for example the summary OR of acute MI of 2.65 (1.46, 4.67) for rofecoxib >25 mg/day used for greater than 30 days vs non-use. With a frequentist confidence interval (CI), which represents uncertainty through repetition of the experience, all OR values from 1.46 to 4.67 might seem equally likely in a repeated sampling of similar studies. In contrast, the Bayesian approach though it results in a numerically similar 95% credible interval also allows us to calculate that there is an 83% probability that this OR is greater than 2.00. This illustrates how useful the Bayesian approach is for decision making.

In addition to allowing direct probability statements to be made, a one-stage Bayesian model had the following technical advantages: [30, 31] for this IPD MA:

2) The available data for the IPD MA are gathered at multiple levels and therefore have a hierarchical structure. Within each of the four database studies (higher level of analysis), are the individual patients (lower level of analysis). We had the ability to fit a

[30] Debray TP, Moons KG, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods. 2015;6(4):293-309.

[31] Spiegelhalter DJ, Myles JP, Jones DR, et al. Bayesian methods in health technology assessment: a review. Health Technol Assess. 2000;4(38):1-130.

hierarchical logistic regression model to as little as four studies, including for some very sparse or systematically missing exposure categories;
3) We could explicitly model the heterogeneity between the four studies without having to choose a priori between a fixed-effect or a random-effects meta-analysis (as we would have had to do if using a frequentist framework);
4) Unlike what we experienced in pilot work using a frequentist approach, the Bayesian framework allowed overcoming systematic missingness and retaining the NSAID exposures and the 'IPD MA confounders' that were available in each study. Hence we did not lose information on exposures and we adjusted for confounding as well as we could within the limitations of each dataset
Furthermore, for NSAID exposure categories that were systematically missing in a given study, the Bayesian analysis permitted the estimation of an odds ratio and credible interval, despite the fact that values were missing for that study.
The OR and CrI reported in Table 2 of the main manuscript for systematically missing exposures reflect the similarity (or exchangeability) of the data across the four studies.

In our reply to Reviewer 2, we provide further technical details on the advantage of using a one-stage IPD MA for this research.

The methodology is complex. I think it should be better described for average BMJ readers. The results of the Bayesian analysis have to be explained to a frequentist audience.

Producing research that can easily translate into practice is one of the motivations for using a Bayesian approach and for expressing results as posterior probabilities of acute MI. This is not new to the BMJ readership and, in fact, the Trelle NMA of NSAIDs RCTs previously presented various adverse cardiovascular outcomes as Bayesian posterior probabilities (BMJ 2011;342:c7086, Figure 4).

However, we do agree that the methodology is complex and we have tried to better describe it for the BMJ readers. Please see the revised manuscript.

- Another editor was very interested in this from a clinical point of view and was supportive.

We thank the editor for being supportive. Musculoskeletal disorders, which are a major indication for NSAIDs, are among the fastest-growing cause of worldwide disability. [32] NSAIDs continue to be among the most prescribed [33] and advocated [34] therapies for the

[32] Vos T, Barber RM, Bell B, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015;386(9995):743-800.

[33] Kingsbury SR, Gross HJ, Isherwood G, et al. Osteoarthritis in Europe: impact on health status, work productivity and use of pharmacotherapies in five European countries. Rheumatology (Oxford). 2014;53(5):937-47.

treatment of OA; in fact, use of NSAIDs is ubiquitous.[35, 36] Consequently, although acute MI is a rare adverse event, given the high prevalence of NSAID use, we feel it is paramount from a population perspective to better understand the association of this adverse CV outcome with real-world NSAID use.

- Another editor was not sure what the clinical implications are here. The problem is already well known and this offers no extra advice/evidence.

Indeed, the problem is already well known, although perhaps not with a full recognition of the residual uncertainty (cf. the conventional view that naproxen is less cardiotoxic or cardioneutral). However current treatment guidelines promoting the lowest effective NSAID dose for the shortest possible duration[37, 38, 39] are not sufficiently informative. This work specifically addressed questions stemming from routine patient-centred clinical decisions such as initiating NSAID treatment, increasing the dose or renewing a prescription to extend continuous use, for which NSAID-specific guidance is currently lacking. We believe that this research contributes actionable extra advice/evidence.

For example, we have found that the onset of NSAIDs-associated MI risk increase occurs within the first week of use and that use at high dose for 1 to 4 weeks is particularly harmful with naproxen and with ibuprofen. A clinician might consider this new information and decide not to use NSAIDs in a given patient. He or she might also opt to maintain a patient on a lower NSAID dose and seek alternative interventions for treating pain and inflammation.

It is noteworthy that a recent systematic review of Clinical Practice Research Datalink (formerly GPRD) research in NICE guidance documents concludes there seems to be an increasing trend in the use of healthcare system data to inform clinical practice, especially as the real-world validity of clinical trials is being questioned.[40]

---

[34] da Costa BR, Reichenbach S, Keller N, et al. Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis. Lancet. 2016 March 17.

[35] Conaghan PG, Porcheret M, Kingsbury SR, et al. Impact and therapy of osteoarthritis: the Arthritis Care OA Nation 2012 survey. Clin Rheumatol. 2015;34(9):1581-8.

[36] McGettigan P, Henry D. Use of non-steroidal anti-inflammatory drugs that elevate cardiovascular risk: an examination of sales and essential medicines lists in low-, middle-, and high-income countries. PLoS Med. 2013;10(2):e1001388.

[37] American Geriatrics Society Panel on Pharmacological Management of Persistent Pain in Older Persons. Pharmacological management of persistent pain in older persons. J Am Geriatr Soc. 2009;57(8):1331-46.

[38] Abdulla A, Adams N, Bone M, et al. Guidance on the management of pain in older people. Age Ageing. 2013;42 Suppl 1:i1-57.

[39] Scarpignato C, Lanas A, Blandizzi C, et al. Safe prescribing of non-steroidal anti-inflammatory drugs in patients with osteoarthritis--an expert consensus addressing benefits as well as gastrointestinal and cardiovascular risks. BMC Med. 2015;13:55.

[40] Oyinlola JO, Campbell J, Kousoulis AA. Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances. BMC Health Serv Res. 2016;16(1):299.

Comments from Reviewers

Reviewer: 1
Please enter your name: Rahul Mukherjee
Job Title: NIHR Academic Clinical Fellow in Cardiology
Institution: King's College London

Comments:
Overall, the study addresses an important clinical question. A large cohort of patients studied with high numbers of patients with acute myocardial infarction - although the 4 databases did have heterogenous populations, this does help understand 'real world' data as the authors assert. The methodology used for the systematic review appears to be robust. However, as the authors point out, the number of different studies used were small (4) - with 4 studies excluded as data were not available due to restrictions placed - the addition of this data would have increased the strength of the authors' conclusions which is unfortunate.

The authors have identified a number of limitations to the study but overall I feel that the conclusions stated adequately reflect the data that has been analysed.

We are thankful for the above comments.
This research is, to our knowledge, the first one-stage Bayesian IPD MA of drug safety data sourced from healthcare databases. As underlined in the Discussion, it is unfortunate that accessibility of administrative healthcare data remains very limited.

Reviewer: 2
Please enter your name: Richard Riley
Job Title: Professor of Biostatistics
Institution: Keele University

Comments:

This is a very interesting and generally well conducted Bayesian meta-analysis. Some of the limitations of the study are well noted, and I think in general the authors have done as well as they could. They did a systematic review to identify relevant databases, followed by IPD meta-analysis of 4 of these. Only 4 out of 8 databases were obtained. But they argue this was for independent reasons to the research question, and I am sympathetic to this. It is hard to get all the data desired, and they have done the best they can in this situation, whilst noting the limitation. I like, very much, the probabilistic inferences from the analysis (e.g. Fig 2),

However, I do have some areas for clarification as the reporting of the methods needs to be more detailed in places.Indeed, the longer I spent reading it, the more confused I was becoming to be honest.

We sincerely thank the reviewer for his time and his thoughtful comments, which gave us insights into areas that needed clarifying and better reporting. We do hope that our comments and explanations have addressed all questions raised and that reporting in the Methods and other sections of the manuscript has improved and is now satisfactory.

1) Looking at 'real-world' evidence is important. But these are observational studies. It is debatable how real world this is, though it could be strongly argued that the population is more real world than an RCT. Yet, the issue of confounding then comes into play. I think there should be more objective discussion about whether this is indeed 'real world'.

This is an interesting issue, which led us to re-examine the literature. A prevalent definition for 'real-world data' is that of the ISPOR Real-World Data Task Force, which is "data used for decision-making that are not collected in conventional RCTs." [41] The current GetReal initiative (a project of the Innovative Medicines Initiative – https://www.imi-getreal.eu/) adopts the above definition and specifies that "Instead, real world data (RWD) is collected both prospectively and retrospectively from observations of routine clinical practice. RWD can be obtained from many sources including patient registries, electronic medical records, and observational studies." [42] One such source for RWD in the UK, is the Clinical Practice Research Datalink (CPRD) formerly known as the General Practice Research Datalink

[41] Garrison LP, Jr., Neumann PJ, Erickson P, et al. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Health. 2007;10(5):326-35.

[42] Makady A, Goettsch W, Willemsen A. Review of Policies and Perspectives on Real-World Data (IMI-GetReal Deliverable). Draft Report dated 30.01.2015 available at: https://www.imi-getreal.eu/Portals/1/Documents/Publications/D1.3%20GetReal%20Glossary.pdf. Last accessed August 10, 2016.

(GPRD).[43] GPRD along with the RAMQ, Finland, and Saskatchewan's health services databases were the sources of real-world data for this IPD MA.

However, although the IPD MA is based on real-world data, a key consideration is whether it also provides real-world evidence, as " 'data' conjures the idea of simple factual information, whereas 'evidence' connotes the organization of the information to inform a conclusion or judgment."[44] We believe that our research contributes 'real-world evidence' – and evidence that is new – because of the way NSAIDs exposures were measured and modelled in this IPD MA.

One of the reasons for poor external validity of NSAID RCTs (and their inability to represent real-life behaviour) is that the NSAIDs were to be taken continuously over long follow-up periods (see for example Trelle NMA RCTs, Table 1). However, in real-life patients take NSAIDs intermittently (with periods of non-exposure), change the daily dose they use over time, and switch one NSAID for another.[45, 46]

For example, the table below taken from RAMQ shows that NSAIDs are used intermittently, which argues that the population is indeed more real world than is the study sample in an RCT.

Table – Duration of continuous episodes of use for NSAIDs of interest in the RAMQ nested case-control dataset

Over cohort time, prescribed	Number of episodes	Duration of continuous episodes of use (days)						
		Median	Mean	SD	P5	IQR	P95	Maximum
Celecoxib	186 012	30	43.1	73.5	7	15-30	148	1659
Diclofenac	216 479	15	22.6	33.3	6	10-30	60	2089
Ibuprofen	56 770	12	18.1	26.3	4	8-21	40	1468
Naproxen	161 554	14	19.6	27.7	6	10-24	48	1187
Rofecoxib	125 421	30	40.1	63.7	7	15-30	129	1469

IQR= interquartile range; P5= 5th percentile; P95= 95th percentile; SD= standard deviation

Moreover, in this RAMQ dataset, we documented a high proportion of NSAID switching, particularly during the short period of 1999-2000.

[43] Bate A, Juniper J, Lawton AM, et al. Designing and incorporating a real world data approach to international drug development and use: what the UK offers. Drug Discovery Today. 2016 3//;21(3):400-5.

[44] Garrison LP, Jr., Neumann PJ, Erickson P, et al. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Health. 2007;10(5):326-35.

[45] Gore M, Sadosky A, Leslie D, et al. Patterns of therapy switching, augmentation, and discontinuation after initiation of treatment with select medications in patients with osteoarthritis. Clin Ther. 2011;33(12):1914-31.

[46] Langman M, Kahler KH, Kong SX, et al. Drug switching patterns among patients taking non-steroidal anti-inflammatory drugs: a retrospective cohort study of a general practitioners database in the United Kingdom. Pharmacoepidemiol Drug Saf. 2001;10(6):517-24.

Consequently, we believe that real-life use of NSAIDs has to be described in terms of the duration of each episode of use, the daily dose, and must also reflect the fact that patients may have stopped taking NSAIDs recently or more remotely. This can be achieved by simultaneously modelling multiple components of treatment history. In the IPD MA this was done by creating several multidimensional exposure categories with indicator variables corresponding to the contribution of (i) recency of use, (ii) dose level, and (iii) treatment duration. With this approach, acute MI risk for 'High' versus 'Low' dose should inform on the NSAID dose-MI response relationship, risk with various durations of 'Current' use should represent cumulative effects of the NSAID, and risk with 'Recent' or 'Past' use should correspond to its residual effects (persistence of effects after dosing).

Recency was defined as follows: 'Past' use (supply ended 31 to 365 days prior to index date), 'Recent' use (supply ended 1 to 30 days prior to index date), and 'Current' use (duration of NSAID supply overlapped with index date). We sought to produce real-world evidence by selecting cut points for daily dose and durations that correspond to available dosage forms, usual dosage range, and typical or recommended duration of use. Daily dose was dichotomized as 'Low' or 'High' based on clinical consensus for each drug with cut-offs also corresponding to median daily dose in the RAMQ study population. Continuous duration of current use was categorized as 'Use for 1 to 7 days, 'Short' (use for 8 to 30 days), and 'Long' (use for more than 30 days). These cut-offs were chosen also because of their clinical relevance. Use of an NSAID for 1 to 7 days corresponds to seeking relief for an acute condition. Duration of 8 to 30 days represents short-term use, with receipt of one or two consecutive prescriptions, while duration longer than 30 days reflects more chronic indications. Finally, 'Non-use' corresponds to individuals not exposed to any NSAIDs in the year preceding the index date. Note that this conservative definition of non-use reduces the possibility of confounding that might arise if patients who are sicker recently stopped taking NSAIDs. It also takes into account the possibility that a cumulative effect from distant past exposure might exist (this was investigated in separate work, to be published).

2) I urge the authors to tone down, indeed remove, causal language. E.g.: "In real-world use, all NSAIDs – including naproxen – appear to increase the risk of acute MI". It is associated with an increase, but this suggests it does increase directly.

They are associated with an increased risk, but do they actually increase it?
Another example "rofecoxib is systematically more cardiotoxic …"

We certainly agree that we should be careful about making direct causal statement. What we meant to do is report our observed associations. We have modified the manuscript throughout to specify that we are reporting associations.

For example the Abstract now reads:

> "In real-world use, all NSAIDs – including naproxen – were found to be associated with an increase in risk of acute MI"

and the phrase on rofecoxib says:

"… rofecoxib is systematically associated with more cardiotoxicity"

3) Confounders were included, and were pre-defined based on a detailed process. I like that (avoid data dredging in the main analysis). But the list of confounders is not documented in the methods.

Indeed detailed adjustment was only reported in footnote to Table 2 in our original manuscript, which we have amended to add the following paragraph:

"All studies were adjusted for past or recent use of the other NSAIDs, age at index date, diabetes, hyperlipidemia, hypertension, coronary heart disease, and rheumatoid arthritis. Individual studies were also adjusted for additional 'IPD MA confounders' based of their availability (see footnote to Table 2)"

4) I like aspects of the Bayesian analysis. Clustering is accounted for, and the prior distributions are informed by prior evidence, which is sensible and I think the authors should be applauded for taking this viewpoint.

But some of the modelling is somewhat black box. In particular, for each confounder, was a different adjustment term included per study? Or was a common one assumed? Are there random effects in the analysis? On which variables? I would like to see the 1-stage model given in full (in equation form) and all parameter estimates provided (e.g. in supplementary material).

We had previously included the `jags` program and all parameter estimates under Supplementary Material. We provide (as an attachment) the program and all parameter estimates. Here are the model equations:

One-stage IPD MA

Level-1 (Within-study level)

In each study s, a regression model was used to relate the outcome of interest (Y_s) to a vector of NSAID exposure variables (X_s) and a vector of covariates (Z_s), as follows

$$g(Y_s) = a_s + b_s X_s + c_s Z_s,$$

where $g(.)$ is a link function. As explained in the Methods, the link function used differed slightly between studies in accordance with the study design. Systematically missing variables in X_s and Z_s were set to 0.

The a_s and c_s parameters were treated as fixed effects and assigned low information Normal (Mean=0, SD=1000) priors.

Level-2 (Between-study level)

The parameters $b_s = (b_{s1}, ..., b_{sJ})$ were treated as (independent) random effects as follows:

$$b_{sj} = \mu_j + \xi_j \eta_{sj}$$
$$\eta_{sj} \sim \text{Normal(Mean=0, SD=1)}$$

The following hyper-prior distributions were used

$$\mu_j \sim \text{Normal(Mean=0, SD=1000)}$$
$$\xi_j \sim \text{Normal(Mean=0, SD=0.181)}$$

This results in a half-normal hyperprior distribution for the between-study standard deviation in the parameters b_s.[47]

[47] Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis. 2006;1(3):515-34.

5) Most importantly, the analysis uses logistic regression, but why is a survival model not used? I assume that patients are dropping out over time, and there is censoring. Logistic regression would only be sensible if follow-up was complete up to 1 particular time-point … but that is not clear here.

The choice of analysis in the IPD MA respected the design (dynamic population-based cohort with nested case-control analysis or population-based case-control study) and matching strategies (individual- or frequency matching) and matching variables (demographics and time) of the original studies. For convenience an excerpt of previously filed Web Extra Table 2 is given hereafter.

Table –Overview and design of the four healthcare database studies included in the one-stage Bayesian IPD MA of NSAIDs and acute MI – Dataset supplied by Investigator

	RAMQ (N=233 816)	**Finland** (N=172 219)	**GPRD** (N=17 561)	**Saskatchewan** (N=23 167)
Design	Cohort with nested case-control analysis	Population-based case-control study	Cohort with nested case-control analysis	Cohort with nested case-control analysis
Matching of controls to cases on time	Individual matching on: 1) year and month of cohort entry 2) duration of follow-up in cohort	Individual matching on index date	Individual matching on: 1) year of cohort entry 2) duration of follow-up in cohort	Frequency matching on distribution of cases on index date (± 3 months) of non-cases within their person-time of follow-up
Matching of controls to cases on demographic or healthcare variables	Age (± 1 year) and sex	Age at end of calendar year, sex, and hospital catchment area	Age (± 2 year), sex, medical practice	None

As seen from the above, the RAMQ and GPRD cohort design involved matching on calendar time. Control subjects were sampled from the risk sets that are used in the corresponding Cox proportional hazards model.[48] On a practical note, Essebag et al.[49] compared a nested case-control approach for the study of time-dependent exposure with cohort analysis using Cox

[48] Vandenbroucke JP, Pearce N. Case–control studies: basic concepts. Int J Epidemiol. 2012;41(5):1480-9.

[49] Essebag V, Platt RW, Abrahamowicz M, et al. Comparison of nested case-control and survival analysis methodologies for analysis of time-dependent exposure. BMC Med Res Methodol. 2005;5(1):5.

regression including time-dependent covariates and found similar risk estimates are obtainable with superior computational efficiency.

For clarity, we have added the following sentence in the Statistical analysis sub-section:

> "As appropriate for each study design (RAMQ and GPRD: individually-matched nested case-control studies; Saskatchewan: frequency-matched nested case-control study; Finland: population-based matched case-control study), we implemented conditional (RAMQ, Finland, and GPRD) or unconditional logistic regression (Saskatchewan) to analyze the IPD."

Follow-up was indeed complete to a particular time point, and this was specified in Web Extra 1 Table 2 as excerpted hereafter:

Table –Overview and design of the four healthcare database studies included in the one-stage Bayesian IPD MA of NSAIDs and acute MI – Dataset supplied by Investigator

	RAMQ (N=233 816)	**Finland** (N=172 219)	**GPRD** (N=17 561)	**Saskatchewan** (N=23 167)
Cohort entry	Date of first dispensing of single NSAID prescription after 1 January 1993 in individuals without NSAID prescriptions for ≥ 1 year	Not applicable	Date of NSAID first prescription after 1 June 2000	15 November 1999
Cohort exit	Earliest of the following dates: first hospital admission for acute MI, death, termination of health system coverage, or end of the study (30 Sep 2004)	Not applicable	Earliest of the following dates: date of acute MI, death, end of registration with the practice, or end of the study (31 Oct 2004)	Earliest of the following dates: first hospital admission for acute MI, diagnosis of one of the exclusion criteria, 85th birthday, death, termination of health system coverage, or end of the study (31 Dec 2001)

6) How was the use of treatment over time accounted for? What if they stopped then re-started?

Time-varying use of NSAIDs was accounted for in the IPD MA, including stopping and starting NSAID (being unexposed) and switching between NSAIDs.

For example, in the RAMQ cohort, NSAID exposure was determined for each day of follow-up to correspond to expected real-life variation. We assessed, separately, exposure to each of the following drugs: celecoxib, diclofenac, ibuprofen, naproxen, and rofecoxib, and to all other NSAIDs, grouped together. Computer-recorded variables allowed the direct calculation of daily dose of NSAIDs of interest as pill strength times number of pills divided by number of days supplied. Days supplied and consecutive prescriptions dates confirmed the duration of each dispensing and allowed identifying gaps between the end of a preceding prescription and the start of a next one. A priori rules were specified to capture usage behaviours such as intermittent use, dose changes, and drug switches. Continuous episodes of exposure were built from consecutive prescriptions only if the drug and daily dose were the same and any gap did not exceed 14 days. Additional rules (see Table, further below) addressed situations where dispensing started on the same day or overlapped, such that patients could not be concurrently exposed to more than one NSAID.

The effect of NSAID exposure on risk of acute MI was analysed by simultaneously modelling these multiple components of treatment history, i.e. (i) recency, (ii) dose, and (iii) duration, as shown for celecoxib in the Table below (see also Table 2 of manuscript for complete data for the other NSAIDs of interest and Figure 1 of the manuscript for an illustration of how exposure was defined).

Table – NSAID multidimensional exposure categories defined by recency of use, daily dose, and duration in the one-stage Bayesian IPD MA

NSAID use [*†]	**Exposure category**
Non-use	1: No NSAIDs in year prior to index date
Celecoxib	
Past	2: Ended 31-365 days prior to index date
Recent	3: Ended 1-30 days prior to index date
Current	4: Use for 1-7 days, any dose
Current	5: ≤ 200 mg/day for 8 to ≤30 days
Current	6: > 200 mg/day for 8 to ≤30 days
Current	7: ≤ 200 mg/day for >30 days
Current	8: >200 mg/day for >30 days

[*] *IPD MA*
All studies
For a given NSAID, 'current', 'recent', and 'past' use categories are mutually exclusive
Among NSAIDs, 'current' use categories are mutually exclusive (subjects could be 'current' users of only one NSAID)
RAMQ, GPRD, and Saskatchewan
Among NSAIDs, 'current', 'recent', and 'past' use categories are not mutually exclusive

('current' users of one NSAID could be 'recent' users of another NSAID, and 'past' user of yet another NSAID)
Finland
Among NSAIDs, 'current' use is exclusive of 'recent', and 'past' use
('current' users of an NSAID could not be 'recent' users or 'past' users of another NSAID)

† *Original RAMQ and Finland studies:* current= use of this NSAID overlapped with the index date; recent= use of this NSAID ended 1-30 days prior to index date;
past= use of this NSAID ended 31-365 days prior to index date
Original GPRD study: current= use of this NSAID overlapped with the index date or lasted into the 14-day period before the index date; recent use category systematically missing;
past= use of this NSAID ended 184-365 days prior to index date
Original Saskatchewan study: current= use of this NSAID overlapped with the index date or lasted into the 7-day period before the index date; recent= use of this NSAID ended 8-60 days prior to index date; past= use of this NSAID ended 61-365 days prior to index date

The sample patient depicted in the Figure below was included in the IPD MA model as a current user of rofecoxib at index date (exposure category 4, Use for 1-7 days, any dose), a recent user of celecoxib and diclofenac (exposure category 3, Use ended 1-30 days prior to index date), and a past user of rofecoxib and naproxen (exposure category 2, Use ended 31-365 days prior to index date)

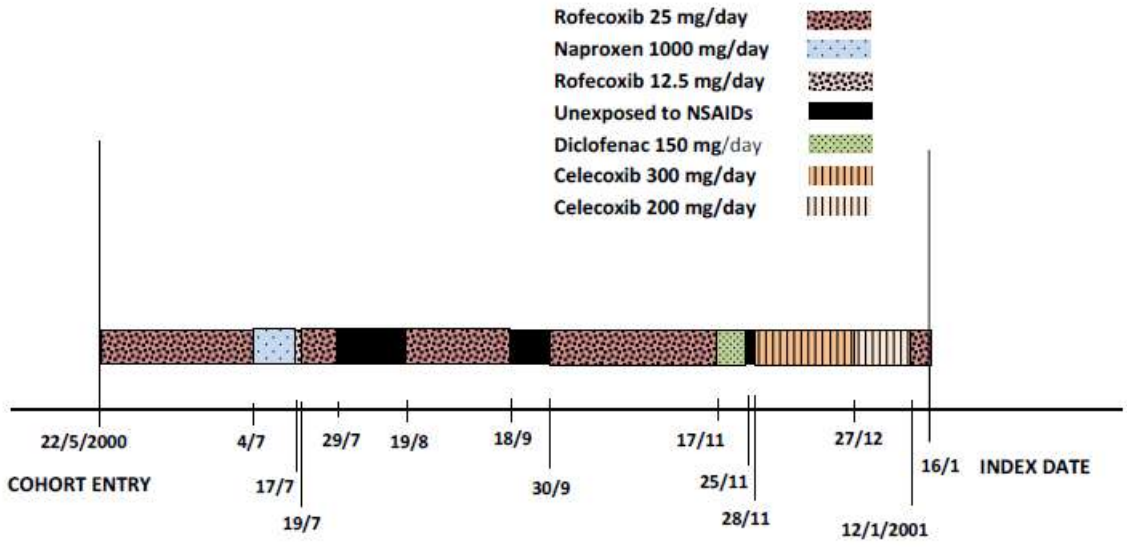


Figure – Sample RAMQ study patient depicting time-varying NSAID exposure during follow-up

The Figure below illustrates how an episode of continuous use was tallied from consecutive dispensing of the same NSAID and daily dose, in case of overlapping prescriptions or of fill gaps in the RAMQ dataset.

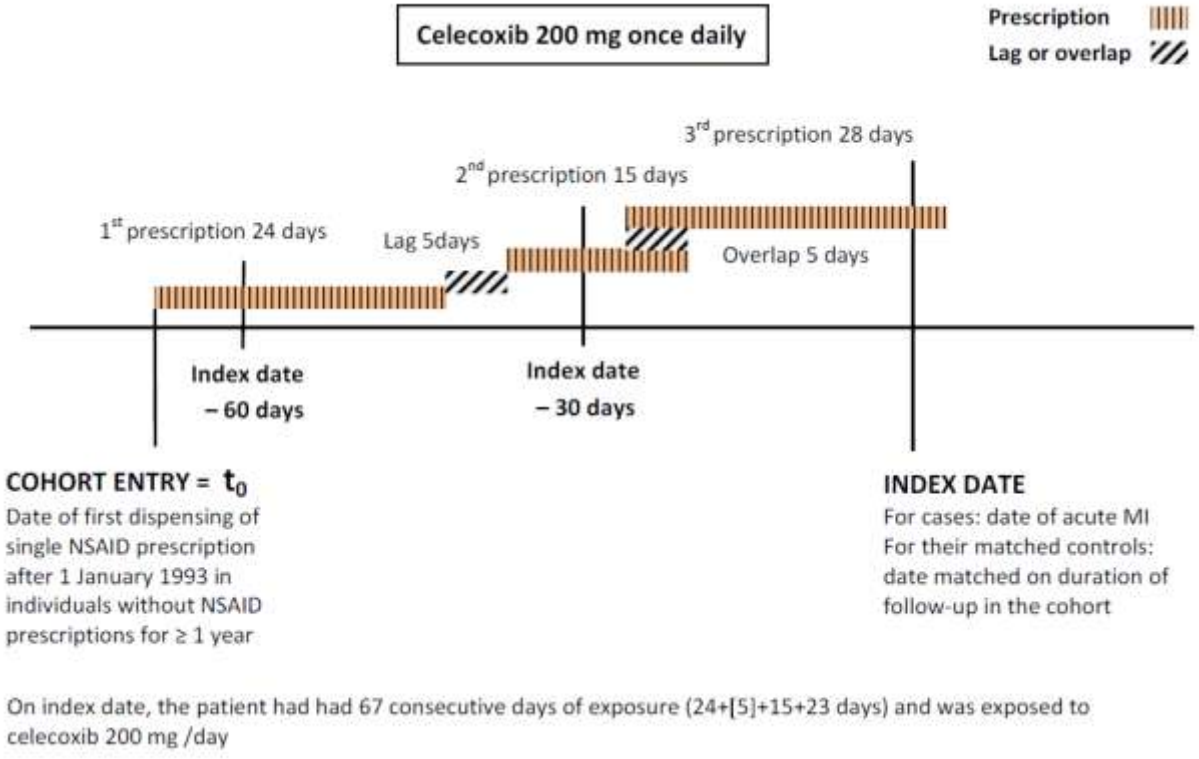
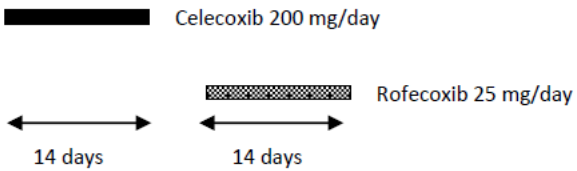
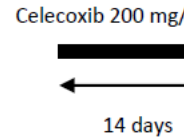
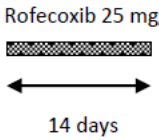
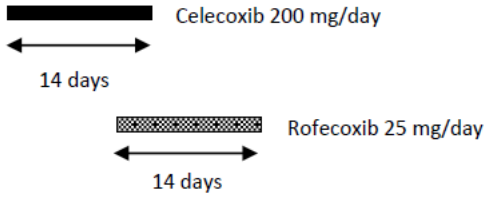
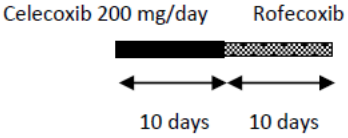
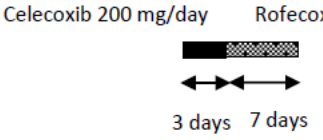


Figure – Example of an episode of continuous use from consecutive prescriptions of celecoxib in the RAMQ nested case-control dataset used for the one-stage Bayesian IPD MA

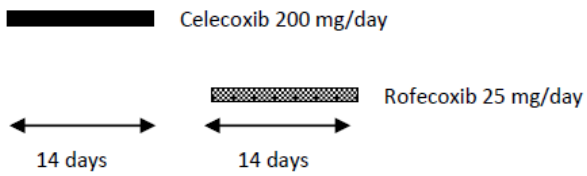
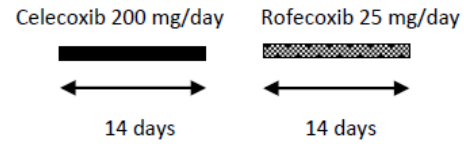
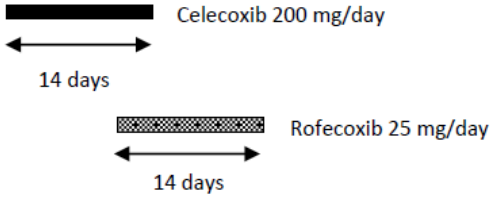
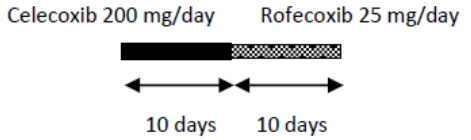
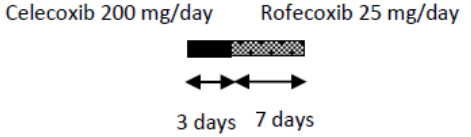
Finally, the Tables in the next pages illustrate the creation of episodes of use given various scenarios for consecutive prescriptions of the same NSAID and of two different NSAIDs in the RAMQ dataset.

Table – Creation of episodes of use given various scenarios for consecutive prescriptions of the same NSAID in the RAMQ nested case-control dataset used for the one-stage Bayesian IPD MA

NSAID prescriptions scenario *	Illustration using celecoxib and rofecoxib as examples	Decision for creating episodes of use
Second prescription starts and ends after the first one No overlap of coverage dates Different NSAIDs		Start a new episode of use  
Second prescription starts and ends after the first one Overlap of coverage dates Different NSAIDs		Start a new episode of use 
Second prescription within coverage dates of the first one Different NSAIDs		Start a new episode of use 

* Usage on an 'as needed' basis was estimated by reconciling usual daily doses and dispensing data for each drug fill, allowing to account for irregular use consistently when building episodes of use

Table – Creation of episodes of use given various scenarios for consecutive prescriptions of two different NSAIDs in the RAMQ nested case-control dataset used for the one-stage Bayesian IPD MA

NSAID prescriptions scenario *	Illustration using celecoxib and rofecoxib as examples	Decision for creating episodes of use
Second prescription starts and ends after the first one No overlap of coverage dates Different NSAIDs		Start a new episode of use 
Second prescription starts and ends after the first one Overlap of coverage dates Different NSAIDs		Start a new episode of use 
Second prescription within coverage dates of the first one Different NSAIDs		Start a new episode of use 

* Usage on an 'as needed' basis was estimated by reconciling usual daily doses and dispensing data for each drug fill, allowing to account for irregular use consistently when building episodes of use

7) Could important confounders been missed?

According to current clinical knowledge we do not believe that we failed to identify important confounders. As illustrated in Figure 2 in Web Extra 1 and explained in the legend to that figure, there exist unmeasured (smoking, obesity, socio-economic status, exercise, OTC NSAIDs, OTC low-dose aspirin), incompletely measured (blood pressure increase, renal dysfunction, heart failure [confounding by contraindication]), and unmeasurable risk factors for myocardial infarction (genes) that we could not adjust for.

We did our best to adjust for confounding by measured confounders. In addition to using multivariable regression, confounding control in the IPD MA included matching on time including for index date, year of cohort entry, and duration of follow-up in cohort (see answer to previous Question 5 or Web Extra 1 –Table 2). This matching on time is an important quality feature, given the potential confounding effect of calendar time around the launch of celecoxib and rofecoxib (1999-2000).

For refining confounding adjustment, we entertained a few options including propensity score (PS) methods and instrumental variable (IV) analysis (prescriber's preference or calendar time) but considered them unsuitable. Briefly, a sound propensity score approach [50] would have to consider time-varying use of each individual NSAIDs and change in measured confounder values over time. Ultimately, it might improve confounding adjustment only marginally because the effect of mediating variables between NSAID exposure and the MI outcome cannot be factored in the determination of PS. As for instrumental variable analysis, the literature [51, 52, 53] suggests that confounding of the IV and the acute MI outcome association is likely to occur such that IV analysis is a poor method for dealing with unmeasured confounders in our research.

However, the literature provides useful insight on the risk of bias due to confounders that are unobserved in database studies. Graham and coll. [54] assessed the potential for confounding from low-dose aspirin use, OTC NSAID use, smoking history, and family history of acute MI via a standardized telephone survey of a random sample of the controls who were currently or remotely exposed to NSAIDs in their study. They found that these

---

[50] Mack CD, Glynn RJ, Brookhart MA, et al. Calendar time-specific propensity scores and comparative effectiveness research for stage III colon cancer chemotherapy. Pharmacoepidemiol Drug Saf. 2013;22(8):810-8.

[51] Tamblyn R, McLeod P, Hanley JA, et al. Physician and practice characteristics associated with the early utilization of new prescription drugs. Med Care. 2003;41(8):895-908.

[52] Solomon DH, Schneeweiss S, Glynn RJ, et al. Determinants of selective cyclooxygenase-2 inhibitor prescribing: are patient or physician characteristics more important? Am J Med. 2003;115(9):715-20.

[53] Franklin JM, Schneeweiss S, Huybrechts KF, et al. Evaluating possible confounding by prescriber in comparative effectiveness research. Epidemiology. 2015;26(2):238-41.

[54] Graham DJ, Campen D, Hui R, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet. 2005;365(9458):475-81.

factors were not differentially distributed with respect to NSAID exposure. Schneeweiss [55] used data collected in the Medicare Current Beneficiary Survey and calculated the confounding bias of not having adjusted an NSAID study in Medicare beneficiaries [56] for BMI, OTC aspirin use, current smoking, income, and educational attainment. Failure to adjust for these unmeasured confounders led to a small underestimation of the association between coxibs and MI.[57]

We believe that the main source of confounding bias is the inability to adjust for unmeasured mediators (NSAID-induced blood pressure increases, renal deterioration) between NSAID exposure and the acute MI outcome. In an attempt to estimate confounding bias due to unmeasured mediators, we considered the prevalence and the strength of the association of each mediator [58] with the acute MI outcome documented in the literature [59, 60, 61] We also reviewed the strength of association between NSAIDs and blood pressure increases [62, 63, 64, 65, 66] and renal failure. [67] In the end, the overall extent and the direction of any residual confounding due to mediators are very hard to predict as these may be causally ordered and may co-exist, and because some biasing effects might cancel each other. In theory, the likelihood of residual confounding due to mediating blood pressure

---

[55] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005;16(1):17-24.

[56] Solomon DH, Schneeweiss S, Glynn RJ, et al. Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults. Circulation. 2004;109(17):2068-73.

[57] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005;16(1):17-24.

[58] Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology. 2011;22(1):42-52.

[59] Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990;335(8693):827-38.

[60] Meisinger C, Doring A, Lowel H. Chronic kidney disease and risk of incident myocardial infarction and all-cause and cardiovascular disease mortality in middle-aged men and women from the general population. Eur Heart J. 2006;27(10):1245-50.

[61] Coca SG, Yusuf B, Shlipak MG, et al. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;53(6):961-73.

[62] Pope JE, Anderson JJ, Felson DT. A meta-analysis of the effects of nonsteroidal anti-inflammatory drugs on blood pressure. Arch Intern Med. 1993;153(4):477-84.

[63] Johnson AG, Nguyen TV, Day RO. Do nonsteroidal anti-inflammatory drugs affect blood pressure? A meta-analysis. Ann Intern Med. 1994;121(4):289-300.

[64] Farkouh ME, Verheugt FW, Ruland S, et al. A comparison of the blood pressure changes of lumiracoxib with those of ibuprofen and naproxen. J Clin Hypertens (Greenwich). 2008;10(8):592-602.

[65] Chan CC, Reid CM, Aw TJ, et al. Do COX-2 inhibitors raise blood pressure more than nonselective NSAIDs and placebo? An updated meta-analysis. J Hypertens. 2009;27(12):2332-41.

[66] Snowden S, Nelson R. The effects of nonsteroidal anti-inflammatory drugs on blood pressure in hypertensive patients. Cardiol Rev. 2011;19(4):184-91.

[67] Ungprasert P, Cheungpasitporn W, Crowson CS, et al. Individual non-steroidal anti-inflammatory drugs and risk of acute kidney injury: A systematic review and meta-analysis of observational studies. Eur J Intern Med. 2015;26(4):285-91.

increase or renal deterioration may increase with longer durations of NSAID use over follow-up time – such that the estimates of acute MI risk in the 'Long' duration categories might be biased to the null.

In summary, on the basis of our assessment of the literature and anticipated direction of bias to the null, we think that unmeasured and unmeasurable confounders are unlikely to affect the substantive conclusions of this IPD MA, which found associations between current NSAID exposure and increased risk of acute MI.

We have added the following text in the revised manuscript:

> "We suspect that residual confounding exists because substantive knowledge [68, 69, 70] ascertains there are mediating intermediate variables on the causal pathway between NSAID exposures and acute MI (Web Figure 2). The likelihood of residual confounding due to mediating blood pressure increase or renal deterioration may increase with longer durations of NSAID use over follow-up time – such that the ORs of acute MI in the 'Long' duration categories might be biased to the null (underestimated). Previous work [71, 72] provide useful insight on the risk of bias due to confounders unobserved in database studies (obesity, OTC aspirin or NSAID use, smoking, income, or educational attainment), which suggests that failure to adjust for these confounders might slightly underestimate MI risk. [73] On the basis of our assessment of the literature [74, 75, 76] and the anticipated direction of bias to the null, we believe that unmeasured and incompletely measured confounders are unlikely to affect the substantive

---

[68] Johnson AG, Nguyen TV, Day RO. Do nonsteroidal anti-inflammatory drugs affect blood pressure? A meta-analysis. Ann Intern Med. 1994;121(4):289-300.

[69] Chan CC, Reid CM, Aw TJ, et al. Do COX-2 inhibitors raise blood pressure more than nonselective NSAIDs and placebo? An updated meta-analysis. J Hypertens. 2009;27(12):2332-41.

[70] Ungprasert P, Cheungpasitporn W, Crowson CS, et al. Individual non-steroidal anti-inflammatory drugs and risk of acute kidney injury: A systematic review and meta-analysis of observational studies. Eur J Intern Med. 2015;26(4):285-91.

[71] Graham DJ, Campen D, Hui R, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet. 2005;365(9458):475-81.

[72] Solomon DH, Schneeweiss S, Glynn RJ, et al. Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults. Circulation. 2004;109(17):2068-73.

[73] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005;16(1):17-24.

[74] Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990;335(8693):827-38.

[75] Meisinger C, Doring A, Lowel H. Chronic kidney disease and risk of incident myocardial infarction and all-cause and cardiovascular disease mortality in middle-aged men and women from the general population. Eur Heart J. 2006;27(10):1245-50.

[76] Coca SG, Yusuf B, Shlipak MG, et al. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;53(6):961-73.

conclusions of this IPD MA, which found associations between current NSAID exposure and increased risk of acute MI."

8) The databases possibly also contain clustering, for example by hospital, region, practice etc, but it is not clear if this was handled.

We did not include any further random effects for clustering, for example by hospital, region, practice etc, within the four studies. However, the Saskatchewan and GPRD studies matched cases and controls for demographics while the Finland study additionally matched cases and controls by hospital catchment area and the GPRD study matched by medical practice (see Web Extra 1 –Table 2). The original matching was preserved when analysing IPD from each study (see attachment for program and all parameter estimates or alternatively, Web Extra 2 – Model for one-stage Bayesian IPD MA of NSAIDs and risk of acute MI, p. 7)

9) Many of the issues above may be addressed by reporting better, by drawing from PRISMA-IPD.
Stewart LA, Clarke M, Rovers M, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. Jama 2015;313(16):1657-65.

Thank you. We hope that the revised version allows for a better understanding of our methods and results.

10) There were systematically missing predictors (confounders not recorded by 1 or more studies) and these were assumed to be completely missing at random. But that does not explain how they could be included in the 1-stage analysis when they were missing. It depends on the modelling structure, and we need more details about how other confounders were included.

In addition to the JAGS model code (previously submitted in Web Extra 2), we hope that the model equation, which we provided above, answers this question.

In practice, the Bayesian framework allowed overcoming systematic missingness and retaining the 'IPD MA confounders' that were available in each study (we specify this in the revised manuscript). Hence we adjusted for confounding as well as we could within the limitations of each dataset. All studies were adjusted for age at index date, diabetes, hyperlipidemia, hypertension, coronary heart disease, and rheumatoid arthritis. Individual studies were also adjusted for additional 'IPD MA confounders' based of their availability, as follows:

a) RAMQ: Also adjusted for previous myocardial infarction, congestive heart failure, cerebrovascular disease, peripheral vascular disease, chronic obstructive pulmonary

disease, gastrointestinal ulcer disease, gastrointestinal bleed, acute or chronic renal failure, and concomitant use of oral corticosteroids, clopidogrel, and cardioprotective aspirin

b) Finland: Also adjusted for clopidogrel use
c) GPRD: Also adjusted for cerebrovascular disease
d) Saskatchewan: Also adjusted for previous myocardial infarction, congestive heart failure, cerebrovascular disease, peripheral vascular disease, chronic obstructive pulmonary disease, gastrointestinal ulcer disease, gastrointestinal bleed, and acute or chronic renal failure

We have revised the manuscript as follows:

> "On the basis of substantive knowledge and confirmation by a search of the literature [77,78,79,80] we identified risk factors for the outcome and potential confounders, which formed a set of candidate covariates to include in multivariable regression analysis. Via a simplified causal graph, [81] we mapped relationships between variables,[82] (Web Figure 2), including the special case of time-dependent confounders that are mediating intermediates on the causal pathway between NSAID exposures and acute MI outcome. [83] Such mediator variables must be identified at the analysis planning stage as they require appropriate measurement and analytical strategies. Using the reference RAMQ study, we then substantiated the confounder status for each candidate covariate, by calculating the odds ratio (OR) of association between the covariate and exposure to NSAIDs among controls and the OR of association between the covariate and MI outcome in the unexposed, leading to the final set of target confounders (referred to as 'IPD MA confounders') selected for adjustment in the meta-analysis (Web Table 5). The final set of target 'IPD MA confounders' comprised: age at index date, male sex, diabetes, hyperlipidemia, hypertension, previous myocardial infarction, coronary heart disease (excluding prior MI), congestive heart failure, cerebrovascular disease, peripheral vascular disease, chronic obstructive pulmonary disease, gastrointestinal ulcer disease, gastrointestinal bleed, acute or chronic renal failure, rheumatoid arthritis, and

---

[77] Coca SG, Yusuf B, Shlipak MG, et al. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;53(6):961-73.

[78] Hall AJ, Stubbs B, Mamas MA, et al. Association between osteoarthritis and cardiovascular disease: systematic review and meta-analysis. Eur J Prev Cardiol. 2015 Oct 13.

[79] Han MK, McLaughlin VV, Criner GJ, et al. Pulmonary diseases and the heart. Circulation. 2007;116(25):2992-3005.

[80] Meisinger C, Doring A, Lowel H. Chronic kidney disease and risk of incident myocardial infarction and all-cause and cardiovascular disease mortality in middle-aged men and women from the general population. Eur Heart J. 2006;27(10):1245-50.

[81] Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10(1):37-48.

[82] Sauer BC, Brookhart MA, Roy J, et al. A review of covariate selection for non-experimental comparative effectiveness research. Pharmacoepidemiol Drug Saf. 2013.

[83] Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550-60.

concomitant treatment with oral corticosteroids, clopidogrel, or cardioprotective aspirin. The definitions and determination strategies for 'IPD MA confounders' available from each study can be found in Web Tables 6 to 9. All studies were adjusted for the following 'common IPD MA confounders': age at index date, diabetes, hyperlipidemia, hypertension, coronary heart disease, and rheumatoid arthritis. Each study was also adjusted for additional 'IPD MA confounders', based on measurement in the original dataset, as described in the footnote to Table 2. When an 'IPD MA confounder' was unavailable, it was missing systematically for all subjects in that study."

11 But I am pleased to see a sensitivity analysis on this, though it is hidden in the discussion: "To examine how possibly imperfect adjustment resulting from systematically missing confounders might affect estimates of acute MI, we compared results of the RAMQ study obtained with two adjustment strategies – full set of confounders available in this study versus only those confounders that are common to the four studies comprised in the IPD MA. Only adjusting for common confounders in the RAMQ study had a limited effect on the estimates obtained from fully adjusted models (Web Table 12)."
I suggest this is actually moved to the results, and the methods of this sensitivity analysis actually detailed in the Methods section

We have detailed the methods of this sensitivity analysis in the Methods section and presented the findings in the Results section.

12) I wonder, could the authors use the approach outline here to deal with systematically missing predictors? It is a simply approach, and just slight extension of their sensitivity analysis anyway:
1. Fibrinogen Studies Collaboration, Jackson D, White I, et al. Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. Stat Med 2009;28(8):1218-37.

Jackson et al. (2009) write: "We propose a joint model for the fully and partially adjusted estimates. We allow for statistical heterogeneity between studies by using the standard bivariate random-effects model for meta-analysis [4–6] for the partially and fully adjusted estimates of effect. This enables inferences concerning both partially and fully adjusted effects to 'borrow strength' [6] from the other type of estimate. Most importantly, the fully adjusted estimate borrows strength from the studies that only provide partially adjusted estimates. …"
"With IPD, a one-stage approach to analysis is possible: in this a single model for all studies, typically incorporating study level random effects, is fitted directly, thus avoiding the within-study quadratic approximation. …
In this paper, we do not adopt a one-stage approach because of its computational complexity with time-to-event outcomes [14], because of the large size of our motivating data set, and because it is not clear how to encompass both fully adjusted and partially adjusted models."

The one-stage Bayesian model written for this IPD MA did allow encompassing both fully adjusted studies (RAMQ) and partially adjusted studies (Finland, GRPD, and Saskatchewan) (see attachment for the program and all parameter estimates or alternatively Web Extra 2 – Model for one-stage Bayesian IPD MA of NSAIDs and risk of acute MI, p. 7).

13) I do not see what the word 'one-stage' adds to the title.

We have removed 'one-stage' from the title but believe it is important that this analytical framework be clearly specified elsewhere in the manuscript.

One of the advantages of using a one-step approach is that the uncertainty in parameter estimates derived at the lower level of analysis is accounted for. [84] We feel this is particularly important in the context of this IPD MA that includes only four studies.

It seems that the respective advantages and disadvantages of one- and two-stage modelling is a topic of current interest and debate. See for example:

Debray TP, Moons KG, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods. 2015;6(4):293-309.

14) "adjusted median pooled odds ratios": this is not very easy for the BMJ reader to interpret. I would call them summary adjusted odds ratios, and in the methods simply explain that these are the median values from the posterior distribution.

We modified the manuscript based on the above suggestion.

15) It is not always clear (e.g. in the abstract) who the comparison group is. Those not taking NSAIDs? It is not stated.

The comparison group was non-use of any NSAID in the year preceding the index date. Indeed this was missing from the Abstract and this has been corrected.

16) It says there is at least an 80% probability of increasing MI risk by X-Y%. This does not make sense to me. Why restrict it to a category, and not simply say by at least X%.?

In our submitted manuscript we provided the confidence in magnitude of risk increase associated with short-term use of NSAIDs (use for 1-7 days, Low dose for 1-4 weeks, High dose for 1-4 weeks) and this is why we reported 'at least an 80% probability of associated

[84] Debray TP, Moons KG, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods. 2015;6(4):293-309.

increased MI risk as a range of X-Y%. We realize that this may be confusing. Therefore, we have modified the text as follows:

> "Confidence in magnitude of risk increase associated with short-term use
> We examined the plots of Bayesian posterior probability that ORs of acute MI were greater than a series of pre-specified thresholds. Selecting a benchmark of 80% (horizontal black line), and considering use for 1-7 days (intersecting gold line), there was at least 80% probability that the OR of acute MI is greater than 1.07 for celecoxib (Figure 2a), 1.30 for diclofenac and naproxen (Figures 3a and 5a), 1.25 for ibuprofen (Figure 4a), and 1.35 for rofecoxib (Figure 6a). Clinicians may wish to consider different exposure categories, cut points for posterior probability of risk, and OR thresholds. For example, with naproxen ≤ 750 mg/day for 1-4 weeks, there is a 58% probability that the OR of acute MI is greater than 1.20 (green line in Figure 5a)."

17) The abstract does not give us the summary meta-analysis results (ORs), despite detailing them in the short methods.

We struggled with what to present in the Abstract for results as this meta-analysis involves some 25 summary ORs of interest since there are 5 NSAIDs and 5 categories of current exposure. While readers are used to seeing a point estimate and its confidence or credible interval, we felt that presenting direct probability statements of risk were more informative than ORs and CrIs.

We have modified the Abstract (Methods and Results), which now reads as follows:

> **"Methods**
> We performed a systematic review followed by a one-stage Bayesian individual patient data meta-analysis of internally valid healthcare database studies reporting MI risks with traditional NSAIDs (t-NSAIDs), celecoxib, and rofecoxib. To capture the complexities of real-world patient behaviour, we modelled drug exposure as an indicator variable incorporating the specific NSAID, its recency, duration of use, and dose. We calculated summary adjusted odds ratios (OR) of acute MI for each NSAID use versus non-use in the preceding year. The prior distribution of MI heterogeneity was informed by previous observational research
> **Results**
> We had access to four healthcare databases (Provinces of Quebec and of Saskatchewan [Canada], Finland, and the UK) and identified a cohort of 446 763 individuals including 61 460 with acute myocardial infarction. All dose-duration categories of current NSAID use were associated with an increased risk of MI. With use for 1-7 days (any dose) the posterior probability of the OR of acute MI being greater than 1.0 was 92% for celecoxib, 97% for ibuprofen, and 99% for naproxen, diclofenac, and rofecoxib. The corresponding ORs (95% credible

interval) with use for 1-7 days were: for celecoxib: 1.24 (0.91, 1.82), ibuprofen: 1.48 (1.00, 2.26), diclofenac: 1.50 (1.06, 2.04), naproxen: 1.53 (1.07, 2.33), and rofecoxib: 1.58(1.07, 2.17). High NSAID doses were associated with greater MI risk. With use for >1 month risks did not appear to exceed those associated with shorter durations."

18) The ORs relate to what time-point?

The ORs related to the index date: date of hospitalization with acute MI for cases and matched date for controls.

19) I was extremely confused when the authors suddenly mentioned cases and controls, as I had assumed that databases essential represent a cohort study. This needs to be clarified. ". Differences in local prescribing habits during the study period – including uptake of coxibs in the various study jurisdictions, 54-56 and in matching strategies – may explain the observed contrasting prevalences of exposure on index date (MI for cases, matched date for controls – See Web Table 2 for details on matching) within- and between studies".

As was reported in the Methods – Data source and outcome in the submitted manuscript "From the universal, computerized public insurance databases of Quebec, Canada, we created a nested case-control dataset (RAMQ), which served as reference for harmonizing the IPD from other studies. The other included datasets were a population-based case-control study from Finland [85] and two nested case-control studies – one from the UK General Practice Research Database (GPRD) [86] and the other from Saskatchewan [87]"

Web Extra 1 Table 2 described the design of the original four healthcare database studies included in the IPD MA (dataset supplied by Investigator).

This issue relates to point 15 and the comparison group I sense. If there is no comparison group, then how do we know we have increased risks?

The comparison group was non-use of any NSAID in the year preceding the index date.

---

[85] Helin-Salmivaara A, Virtanen A, Vesalainen R, et al. NSAID use and the risk of hospitalization for first myocardial infarction in the general population: a nationwide case-control study from Finland. Eur Heart J. 2006;27(14):1657-63.

[86] Andersohn F, Suissa S, Garbe E. Use of first- and second-generation cyclooxygenase-2-selective nonsteroidal antiinflammatory drugs and risk of acute myocardial infarction. Circulation. 2006;113(16):1950-7.

[87] Varas-Lorenzo C, Castellsague J, Stang MR, et al. The use of selective cyclooxygenase-2 inhibitors and the risk of acute myocardial infarction in Saskatchewan, Canada. Pharmacoepidemiol Drug Saf. 2009;18(11):1016-25.

20) How were covariates such as 'use for 1 to 7 days' included in the model? And again, what is the comparison group? This analysis suggests that the model is looking at interactions between effects and covariates like duration, but how was this modelled. See for example below on the need to carefully model interactions in a 1-stage analysis
Fisher DJ, Copas AJ, Tierney JF, et al. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. J Clin Epidemiol 2011;64(9):949-67.
Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. Journal of Reseach Synthesis Methods 2010;1:2-9.

Please see our answers to Questions 6), 10), and 15).

We did not model the effects of NSAID, such as duration, as an interaction with the main effect.

Note that for the purpose of answering a question from an Editor regarding interactions with cardiovascular risk profiles, we have revised the manuscript to add the following paragraph:

> "We might have taken advantage of this one-stage IPD MA to investigate clinical heterogeneity via a formal analysis of interactions between NSAIDs and major CV risk profiles. [88, 89, 90] However, interaction studies involving a categorical NSAID exposure, such as in this IPD MA, would have less statistical power [91] and would overly depend on classifying exposure without error. [92]"

21) I struggle with the main conclusion that "For celecoxib, risks of acute MI do not appear to be greater than those of t-NSAIDs and are lower than those of rofecoxib"

We believe that this conclusion is reasonable, as we are not stating that celecoxib is safer than the other NSAID or 'best'. Instead we are simply summarizing what we have consistently observed from our data – See Table 2 and Figures 2 to 6. Furthermore, in additional analyses we have explored the possibility but found no evidence that residual confounding or exposure misclassification might affect the various NSAIDs in a differential way.

---

[88] Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Stat Med. 2007;26(15):2982-99.

[89] Fisher DJ, Copas AJ, Tierney JF, et al. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. J Clin Epidemiol. 2011;64(9):949-67.

[90] Debray TP, Moons KG, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods. 2015;6(4):293-309.

[91] Greenland S. Basic problems in interaction assessment. Environ Health Perspect. 1993;101(Suppl 4):59-66.

[92] Greenland S. Interactions in epidemiology: relevance, identification, and estimation. Epidemiology. 2009;20(1):14-7.

Figure 7 nicely shows the probabilities that one is the best. But can we believe these? These are not randomised comparisons, so the authors should be very careful. If the authors are to make inferences about which drug is best/worst, then they need to be extremely cautious here and justify why. It relates to the causality issue, how many confounders were adjusted for, whether there are missing confounders etc.
Though the findings are interesting, surely the findings need to be tempered considerably?

We propose replacing Figure 7 by the following Table, which by providing direct probability statements, allows the reader to better appreciate the uncertainty of increased risk of acute MI associated with current NSAID use.

Table – Probability of increased risk of acute MI (OR > 1.0) associated with current NSAID use compared with non-use of any NSAID in the year preceding index in the one-stage Bayesian IPD MA

Current use category [*]	**Bayesian posterior probability of summary adjusted OR** [†] **of acute MI >1.0**
Celecoxib	
Any dose for ≤ 7 days	92.4
≤ 200 mg/day for 8 to ≤30 days	97.4
> 200 mg/day for 8 to ≤30 days	85.8
≤ 200 mg/day for >30 days	97.9
>200 mg/day for >30 days	95.0
Diclofenac	
Any dose for ≤ 7 days	98.6
≤ 100 mg/day for 8 to ≤30 days	93.3
> 100 mg/day for 8 to ≤30 days	92.7
≤ 100 mg/day for >30 days	99.9
> 100 mg/day for >30 days	98.8
Ibuprofen	
Any dose for ≤ 7 days	97.3
≤ 1200 mg/day for 8 to ≤30 days	59.3 [‡]
> 1200 mg/day for 8 to ≤30 days	97.5
≤ 1200 mg/day for >30 days	98.0
> 1200 mg/day for >30 days	98.3
Naproxen	
Any dose for ≤ 7 days	98.8
≤ 750 mg/day for 8 to ≤30 days	92.1
> 750 mg/day for 8 to ≤30 days	99.2
≤ 750 mg/day for >30 days	94.9
> 750 mg/day for >30 days	91.6
Rofecoxib	
Any dose for ≤ 7 days	98.8
≤ 25 mg/day for 8 to ≤30 days	88.6
> 25 mg/day for 8 to ≤30 days	99.8
≤ 25 mg/day for >30 days	99.9

Current use category [*]	**Bayesian posterior probability of summary adjusted OR** [†] **of acute MI >1.0**
> 25 mg/day for >30 days	99.0

[*†] Refer to Table 2 of the revised manuscript
[‡] Sparse data

Comparisons between the various NSAIDs are left as a supplementary analysis (given in Web Extra 1, Figure 3)

Anyway, I wil stop there. In summary, there is much that I like, and the authors have done a considerable piece of work. But there is lack of clarity and over-stating of strength of findings, in my opinion. Thus it is too preliminary for the BMJ to make a decision on this article, until these issues are resolved and clarified.

I hope my comments help the authors improve the article going forward.
Richard Riley, BMJ Statistics Editor.

Reviewer: 3
Please enter your name: Jacob Udell
Job Title: Cardiologist and Clinician Scientist
Institution: Women's College Hospital, University of Toronto

Comments:
The paper by Bally et al is a Bayesian, individual patient level data, meta-analysis of observational cohort studies investigating the risk of acute MI associated with NSAID use. The authors used state-of-the-art advanced health services research methodology to perform a sophisticated analysis. They utilized a systematic review though really a convenience sample of datasets that permitted access to and pooling of IPD. They performed multidimensional analysis to capture time-varying NSAID use and claim to have identified all potential confounders and risk factors for the outcome based on substantive knowledge and literature search.

However, my key issue with their methods is the critical consideration of a control group, and the major confounder of the entire analysis for which both are not at all dwelled on for any period of time in the methods, discussion, or limitation section, yet it seems to me the most pivotal decision for which the entire analysis hinges on. First, take the major confounder not discussed which is the indication for taking a pain relief drug for any period of time, which is the underlying exacerbation of pain symptoms, or typically an acute viral systemic illness, which is not at all captured using simple administrative health terms/coding. No where is this indication discussed, or the potential for either to act as a major confounder for triggering use of NSAIDs and an acute MI, such as by increases in SNS/HTN/tachycardia/fever/and other potential mechanisms discussed at length in the literature. Moreover, none of the methods currently account or come close to triangulate this bias by indication. Yes the investigators account to the extent possible for propensity to be treated with one over another NSAID, but merely by showing the acute nature of the associated risk, typically consistently elevated across all drugs in the class, without much of an impressive dose-response association with risk, suggests an acute intercedent indication to take these drugs is unaccounted for.

How best to mitigate the risk of this competing bias? I suggest, as has been performed numerously throughout the literature of high quality health services research when an issue like this arises, is to compare the risk associated with these NSAIDs to an active control. Although this may seem challenging, in fact there is a very good active control therapy which tracks with either indication (pain or fever) sufficiently well enough that results, which presumably will be attenuated, can be interpreted robustly if a significant bump in risk remains - that is acetaminophen use. I imagine the reason this active control was not selected was because of the challenge of capturing the use of an OTC medication, however that remains the threshold for the authors to overcome in ultimately proceeding with analyzing only that data which provides an active control, otherwise there is no method of mitigation that can avoid this potential bias.

The other, not alternative but complimentary, methodologic approach to consider such a factor is to select an alternative outcome in a sensitivity analysis that also would be expected

to be associated with pain/fever/viral illness but not considered an outcome associated with NSAID use, to determine if such an unmeasured confounder is truly acting within these analyses. I'm not sure off hand whether I have a solution to provide here, seeing that acute kidney injury or GI outcomes would also not help here, but some capture of dehydration might. Such a 'negative control/tracer' outcome would be handy here to at least to demonstrate and quantify the impact of my proposed unmeasured confounder.

Without either major revision to address these potential limitations, I'm very concerned an alarmist analysis without appropriate consideration of the intrinsic indication for use (I emphasize again that pain or fever are neither at all discussed throughout the paper) and its potential impact on MI risk will result in the avoidance of pain/fever medication, some of which may in fact not increase risk beyond that of acetaminophen, translating into millions of ill patients avoiding these drugs (including acetaminophen by association due to misinterpretation) and prolonged exposure to my suggested mediating factors that ironically could paradoxically lead to an increase in adverse MI events.

We appreciate the time and reflection that Reviewer 3 has accorded our manuscript. The reviewer's main concern about our manuscript is the potential for bias by indication, a recognized threat to most observational research where the totality of indication factors, even beyond the principal diagnosis, are generally not available. We would strongly argue that this is not a fatal flaw.

The main use of NSAIDs in an elderly population is osteoarthritis.[93, 94] Osteoarthritis was not a confounder for acute MI in our data (properly substantiating the confounder status of a variable requires calculating the odds ratio (OR) of association between the variable and exposure to NSAIDs among controls and the OR of association between the variable and MI outcome in the unexposed). The next most common NSAID indications in this population are for systemic inflammatory states such as rheumatoid arthritis, which is associated with increased risk of acute MI.[95] However, this is captured in our datasets and has been controlled for in multivariable analyses. Regarding a diagnosis of a viral illness, in our clinical experience it is distinctly unusual for *prescription* NSAIDs to be used in this indication. Moreover, the confounding triad in this situation is further undermined by a lack of any consistent relationship between viral illness and CV outcomes. In addition to the lack of good observational evidence to support a viral illness–adverse CV outcomes association a recent Cochrane review of 8 RCTs looking at the effect of vaccination has

[93] Kingsbury SR, Gross HJ, Isherwood G, et al. Osteoarthritis in Europe: impact on health status, work productivity and use of pharmacotherapies in five European countries. Rheumatology (Oxford). 2014;53(5):937-47.

[94] da Costa BR, Reichenbach S, Keller N, et al. Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis. Lancet. 2016 March 17.

[95] Baghdadi LR, Woodman RJ, Shanahan EM, et al. The impact of traditional cardiovascular risk factors on cardiovascular outcomes in patients with rheumatoid arthritis: a systematic review and meta-analysis. PLoS One. 2015;10(2):e0117952.

concluded “Not enough evidence was available to establish whether influenza vaccination has a role to play in the primary prevention of cardiovascular disease.” [96]

The Reviewer suggested 1) using an active control i.e. acetaminophen or 2) using a 'negative control/tracer' outcome. Approach 1 is impractical (acetaminophen is as suggested by the reviewer an OTC drug and not available in the databases). Moreover, there is little to suggest that prescription indications, comorbidities and physician preferences for acetaminophen and NSAIDs are identical. For Approach 2, although this has been used in the pharmacoepidemiology literature, there is no evidence of its validity for this research question.

Reviewer 3 is also concerned about “an alarmist analysis”. Here again we must respectfully disagree with any attempt to characterize our analysis as “alarmist”. What our analysis does show is the current uncertainty surrounding the small but non-negligent MI risks associated with NSAIDs. We believe that acknowledging the uncertainty of NSAID CV safety is scientifically appropriate and justified.

We have amended the manuscript to discuss residual confounding as follows:

> “We suspect that residual confounding exists because substantive knowledge [97, 98, 99] ascertains there are mediating intermediate variables on the causal pathway between NSAID exposures and acute MI (Web Figure 2). The likelihood of residual confounding due to mediating blood pressure increase or renal deterioration may increase with longer durations of NSAID use over follow-up time – such that the ORs of acute MI in the ‘Long’ duration categories might be biased to the null (underestimated). Previous work [100, 101] provide useful insight on the risk of bias due to confounders unobserved in database studies (obesity, OTC aspirin or NSAID use, smoking, income, or educational attainment), which suggests that failure to adjust for these confounders might slightly underestimate MI risk. [102] On the basis

---

[96] Clar C, Oseni Z, Flowers N, Keshtkar-Jahromi M, Rees K. Influenza vaccines for preventing cardiovascular disease. Cochrane Database Syst Rev. 2015 May 5;(5):CD005050. doi: 10.1002/14651858.CD005050.pub3.

[97] Johnson AG, Nguyen TV, Day RO. Do nonsteroidal anti-inflammatory drugs affect blood pressure? A meta-analysis. Ann Intern Med. 1994;121(4):289-300.

[98] Chan CC, Reid CM, Aw TJ, et al. Do COX-2 inhibitors raise blood pressure more than nonselective NSAIDs and placebo? An updated meta-analysis. J Hypertens. 2009;27(12):2332-41.

[99] Ungprasert P, Cheungpasitporn W, Crowson CS, et al. Individual non-steroidal anti-inflammatory drugs and risk of acute kidney injury: A systematic review and meta-analysis of observational studies. Eur J Intern Med. 2015;26(4):285-91.

[100] Graham DJ, Campen D, Hui R, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet. 2005;365(9458):475-81.

[101] Solomon DH, Schneeweiss S, Glynn RJ, et al. Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults. Circulation. 2004;109(17):2068-73.

[102] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005;16(1):17-24.

of our assessment of the literature [103, 104, 105] and the anticipated direction of bias to the null, we believe that unmeasured and incompletely measured confounders are unlikely to affect the substantive conclusions of this IPD MA, which found associations between current NSAID exposure and increased risk of acute MI."

[103] Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990;335(8693):827-38.

[104] Meisinger C, Doring A, Lowel H. Chronic kidney disease and risk of incident myocardial infarction and all-cause and cardiovascular disease mortality in middle-aged men and women from the general population. Eur Heart J. 2006;27(10):1245-50.

[105] Coca SG, Yusuf B, Shlipak MG, et al. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;53(6):961-73.

Reviewer: 4
Additional Questions:
Please enter your name: Stefanos Bonovas
Job Title: Researcher
Institution: Humanitas Clinical and Research Institute, Milan, Italy

Comments:
The authors report the results of a Bayesian meta-analysis of individual patient data sourced from health care databases. They aimed to examine the effects of individual NSAID dose and duration of use on the risk of acute myocardial infarction. The study is well-done and interesting.
Please find below a few comments:

We thank the Reviewer for his comments, questions, and for his time.

(i) Is there a registered protocol for this IPD meta-analysis? Please clarify.

This work was part of a doctoral research program and stemmed from a funded grant. There is no registered protocol for this IPD MA.

(ii) The authors miss to discuss, in the Limitations Section, that because of the observational nature of the data included in this meta-analysis, associations may have been confounded by unknown or unmeasured factors associated with both exposure and outcome.

Certainly we agree that the issue of confounding must be carefully considered and we have revised the manuscript accordingly (see below for text added to the manuscript). Please let us put this new paragraph in context, by reporting the following additional information:

According to current clinical knowledge we do not believe that we failed to identify important confounders. As illustrated in Figure 2 in Web Extra 1 and explained in the legend to that figure, there exist unmeasured (smoking, obesity, socio-economic status, exercise, OTC NSAIDs, OTC low-dose aspirin), incompletely measured (blood pressure increase, renal dysfunction, heart failure [confounding by contraindication]), and unmeasurable risk factors for myocardial infarction (genes) that we could not adjust for.

We did our best to adjust for confounding by measured confounders. In addition to using multivariable regression, confounding control in the IPD MA included matching on time including for index date, year of cohort entry, and duration of follow-up in cohort (see answer to previous Question 5 or Web Extra 1 –Table 2). This matching on time is an important quality feature, given the potential confounding effect of calendar time around the launch of celecoxib and rofecoxib (1999-2000).

For refining confounding adjustment, we entertained a few options including propensity score (PS) methods and instrumental variable (IV) analysis (prescriber's preference or calendar time) but considered them unsuitable. Briefly, a sound propensity score approach

[106] would have to consider time-varying use of each individual NSAIDs and change in measured confounder values over time. Ultimately, it might improve confounding adjustment only marginally because the effect of mediating variables between NSAID exposure and the MI outcome cannot be factored in the determination of PS. As for instrumental variable analysis, the literature [107, 108, 109] suggests that confounding of the IV and the acute MI outcome association is likely to occur such that IV analysis is a poor method for dealing with unmeasured confounders in our research.

However, the literature provides useful insight on the risk of bias due to confounders that are unobserved in database studies. Graham and coll. [110] assessed the potential for confounding from low-dose aspirin use, OTC NSAID use, smoking history, and family history of acute MI via a standardized telephone survey of a random sample of the controls who were currently or remotely exposed to NSAIDs in their study. They found that these factors were not differentially distributed with respect to NSAID exposure. Schneeweiss [111] used data collected in the Medicare Current Beneficiary Survey and calculated the confounding bias of not having adjusted an NSAID study in Medicare beneficiaries [112] for BMI, OTC aspirin use, current smoking, income, and educational attainment. Failure to adjust for these unmeasured confounders led to a small underestimation of the association between coxibs and MI.[113]

We believe that the main source of confounding bias is the inability to adjust for unmeasured mediators (NSAID-induced blood pressure increases, renal deterioration) between NSAID exposure and the acute MI outcome. In an attempt to estimate confounding bias due to unmeasured mediators, we considered the prevalence and the strength of the association of each mediator [114] with the acute MI outcome documented in the literature [115,

---

[106] Mack CD, Glynn RJ, Brookhart MA, et al. Calendar time-specific propensity scores and comparative effectiveness research for stage III colon cancer chemotherapy. Pharmacoepidemiol Drug Saf. 2013;22(8):810-8.

[107] Tamblyn R, McLeod P, Hanley JA, et al. Physician and practice characteristics associated with the early utilization of new prescription drugs. Med Care. 2003;41(8):895-908.

[108] Solomon DH, Schneeweiss S, Glynn RJ, et al. Determinants of selective cyclooxygenase-2 inhibitor prescribing: are patient or physician characteristics more important? Am J Med. 2003;115(9):715-20.

[109] Franklin JM, Schneeweiss S, Huybrechts KF, et al. Evaluating possible confounding by prescriber in comparative effectiveness research. Epidemiology. 2015;26(2):238-41.

[110] Graham DJ, Campen D, Hui R, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet. 2005;365(9458):475-81.

[111] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005;16(1):17-24.

[112] Solomon DH, Schneeweiss S, Glynn RJ, et al. Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults. Circulation. 2004;109(17):2068-73.

[113] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005;16(1):17-24.

[114] Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology. 2011;22(1):42-52.

[116, 117] We also reviewed the strength of association between NSAIDs and blood pressure increases [118, 119, 120, 121, 122] and renal failure. [123] In the end, the overall extent and the direction of any residual confounding due to mediators are very hard to predict as these may be causally ordered and may co-exist, and because some biasing effects might cancel each other. In theory, the likelihood of residual confounding due to mediating blood pressure increase or renal deterioration may increase with longer durations of NSAID use over follow-up time – such that the estimates of acute MI risk in the 'Long' duration categories might be biased to the null.

In summary, on the basis of our assessment of the literature and anticipated direction of bias to the null, we think that unmeasured and unmeasurable confounders are unlikely to affect the substantive conclusions of this IPD MA, which found associations between current NSAID exposure and increased risk of acute MI.

We have added the following text in the revised manuscript:

> "We suspect that residual confounding exists because substantive knowledge [124, 125, 126] ascertains there are mediating intermediate variables on the causal pathway between NSAID exposures and acute MI (Web Figure 2). The likelihood of residual confounding due to mediating blood pressure increase or renal

---

[115] Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990;335(8693):827-38.

[116] Meisinger C, Doring A, Lowel H. Chronic kidney disease and risk of incident myocardial infarction and all-cause and cardiovascular disease mortality in middle-aged men and women from the general population. Eur Heart J. 2006;27(10):1245-50.

[117] Coca SG, Yusuf B, Shlipak MG, et al. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;53(6):961-73.

[118] Pope JE, Anderson JJ, Felson DT. A meta-analysis of the effects of nonsteroidal anti-inflammatory drugs on blood pressure. Arch Intern Med. 1993;153(4):477-84.

[119] Johnson AG, Nguyen TV, Day RO. Do nonsteroidal anti-inflammatory drugs affect blood pressure? A meta-analysis. Ann Intern Med. 1994;121(4):289-300.

[120] Farkouh ME, Verheugt FW, Ruland S, et al. A comparison of the blood pressure changes of lumiracoxib with those of ibuprofen and naproxen. J Clin Hypertens (Greenwich). 2008;10(8):592-602.

[121] Chan CC, Reid CM, Aw TJ, et al. Do COX-2 inhibitors raise blood pressure more than nonselective NSAIDs and placebo? An updated meta-analysis. J Hypertens. 2009;27(12):2332-41.

[122] Snowden S, Nelson R. The effects of nonsteroidal anti-inflammatory drugs on blood pressure in hypertensive patients. Cardiol Rev. 2011;19(4):184-91.

[123] Ungprasert P, Cheungpasitporn W, Crowson CS, et al. Individual non-steroidal anti-inflammatory drugs and risk of acute kidney injury: A systematic review and meta-analysis of observational studies. Eur J Intern Med. 2015;26(4):285-91.

[124] Johnson AG, Nguyen TV, Day RO. Do nonsteroidal anti-inflammatory drugs affect blood pressure? A meta-analysis. Ann Intern Med. 1994;121(4):289-300.

[125] Chan CC, Reid CM, Aw TJ, et al. Do COX-2 inhibitors raise blood pressure more than nonselective NSAIDs and placebo? An updated meta-analysis. J Hypertens. 2009;27(12):2332-41.

[126] Ungprasert P, Cheungpasitporn W, Crowson CS, et al. Individual non-steroidal anti-inflammatory drugs and risk of acute kidney injury: A systematic review and meta-analysis of observational studies. Eur J Intern Med. 2015;26(4):285-91.

deterioration may increase with longer durations of NSAID use over follow-up time – such that the ORs of acute MI in the 'Long' duration categories might be biased to the null (underestimated). Previous work [127, 128] provide useful insight on the risk of bias due to confounders unobserved in database studies (obesity, OTC aspirin or NSAID use, smoking, income, or educational attainment), which suggests that failure to adjust for these confounders might slightly underestimate MI risk. [129] On the basis of our assessment of the literature [130, 131, 132] and the anticipated direction of bias to the null, we believe that unmeasured and incompletely measured confounders are unlikely to affect the substantive conclusions of this IPD MA, which found associations between current NSAID exposure and increased risk of acute MI."

(iii) Post-marketing observational data lack the experimental random allocation which is necessary to optimally test exposure-outcome hypotheses. Such data often leave open to discussion whether events are associated with the exposure, or perhaps with the underlying diseases (indications of NSAIDs).

We agree that experimental random allocation is a highly desirable feature for causal inference. Unfortunately, except for rofecoxib, for which the single RCT evidence [133, 134] of increased MI risk is strong and consistent, meta-analyses (MA) have been necessary for assessing the MI risk of other NSAIDs.

However, the two latest and much cited works (the Trelle 2011[135] and CNT 2013[136] NMAs) are direct or indirect comparisons of RCT data via a network meta-analysis, a technique

---

[127] Graham DJ, Campen D, Hui R, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet. 2005;365(9458):475-81.

[128] Solomon DH, Schneeweiss S, Glynn RJ, et al. Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults. Circulation. 2004;109(17):2068-73.

[129] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005;16(1):17-24.

[130] Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990;335(8693):827-38.

[131] Meisinger C, Doring A, Lowel H. Chronic kidney disease and risk of incident myocardial infarction and all-cause and cardiovascular disease mortality in middle-aged men and women from the general population. Eur Heart J. 2006;27(10):1245-50.

[132] Coca SG, Yusuf B, Shlipak MG, et al. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;53(6):961-73.

[133] Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. N Engl J Med. 2000;343(21):1520-8.

[134] Bresalier RS, Sandler RS, Quan H, et al. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. N Engl J Med. 2005;352(11):1092-102.

[135] Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ. 2011;342:c7086.

that also presents issues of potential bias, indirectness, inconsistency, and imprecision (see Salanti G. et al. Evaluating the quality of evidence from a network meta-analysis. PLoS One. 2014;9(7):e99682).

With respect to the individual RCT, random allocation should ensure that measured, unmeasured, and unmeasurable confounders are balanced among intervention groups at the very beginning of the trial. However, the issue of post-randomization confounding should not be overlooked, especially with modern RCTs, which are pragmatic and tend to run over long periods of time. Of note, various biases that may occur post-randomization in a RCT [137, 138] may potentially have affected the validity of trials included in these NMAs. These may include: 1) violation of treatment groups exchangeability due to non-persistence with assigned NSAID treatment, or 2) violation of treatment consistency due to group contamination (e.g. OTC NSAIDs) or to differential concomitant treatments potentially affecting the risk of the MI outcome during follow-up (e.g. low-dose aspirin) or sub-optimal adherence to the NSAID. For example, placebo-controlled RCTs of naproxen were mainly performed in populations with or at risk of Alzheimer's disease. [139] Non-adherence bias has been documented in the main naproxen study (ADAPT) [140] and this may translate into an underestimation of the risk of acute MI with naproxen in NMAs of RCTs.

---

[136] Bhala N, for the Coxib and traditional NSAID Trialists' (CNT) Collaboration, Emberson J, et al. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials. Lancet. 2013;382(9894):769-79.

[137] Hernán MA, Hernandez-Diaz S, Robins JM. Randomized trials analyzed as observational studies. Ann Intern Med. 2013 Oct 15;159(8):560-2.

[138] Manson JE, Shufelt CL, Robins JM. The potential for postrandomization confounding in randomized clinical trials. JAMA. 2016 Jun 7;315(21):2273-4.

[139] Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ. 2011;342:c7086.

[140] Steering Committee of the Alzheimer's Disease Anti-inflammatory Prevention Trial (ADAPT). Statement for communication to the FDA Arthritis Advisory Committee and the Drug Safety and Risk Management Advisory Committee. 18 February 2005. Available at: https://jhuccs1.us/adapt/pdf%20documents/FDA%20ADAPT%20STATEMENT_web%20posting.pdf. Last accessed August 31, 2016.

Reviewer: 5
Please enter your name: Fumiaki Imamura
Job Title: Senior Investigator Scientist
Institution: MRC Epidemiology Unit, University of Cambridge

Comments:
Dear Editors and Authors,
Thank you very much for the opportunity of reviewing the rigorous meta-analysis examining NSAID use and a risk of myocardial infarction (MI). I am sorry my comments became late. The topic is of interest in clinical epidemiology over more than a decade and remains to be important, given that musculoskeletal pain is one of global burdens around the world.

The first question coming up is if this study is really necessary given the evidence from available meta-analyses based on trials and based on observational studies in which effects of different dosage were evaluated. The authors are saying yes, of course, but I wish to see a stronger argument than the current one. With this concern, the main comments are provided below. Overall, I was impressed by this manuscript, but am not confident to say this paper merits BMJ, BMJ's readers and clinicians. Minor comments are reserved at this moment, I am afraid. I wish my comments are helpful and informative.

Sincerely,
Fumiaki Imamura
MRC Epidemiology Unit
University of Cambridge

We thank Reviewer 5 for his time and for allowing us to better explain our approach to IPD meta-analysis, outline its advantages, and stress the quality and originality of our work.

1. The strength of this study is not so convincing. The authors argue strengths of the estimations for 'real world use' and of dose and duration, above and beyond what meta-analysis of RCTs can achieve.

In this IPD MA, we provided important evidence above and beyond what meta-analysis (MA) of RCTs could achieve. Let us explain why.

<u>1- Real-word data are needed</u>
First, estimating the effects of 'real-world use' requires real-world data. Real-world data is "data used for decision-making that are not collected in conventional RCTs." [141, 142] Our IPD MA comprised 446 763 individuals from the general population.

---

[141] Garrison LP, Jr., Neumann PJ, Erickson P, et al. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Health. 2007;10(5):326-35.
[142] Makady A, Goettsch W, Willemsen A. Review of Policies and Perspectives on Real-World Data (IMI-GetReal Deliverable). Draft Report dated 30.01.2015 available at: https://www.imi-getreal.eu/Portals/1/Documents/Publications/D1.3%20GetReal%20Glossary.pdf. Last accessed August 10, 2016.

2- Safety meta-analyses of NSAIDs RCTs are not immune to bias

Certainly, our findings are not immune to bias such as misclassification due to partial non-adherence or to residual confounding, and we addressed this further down and in the revised manuscript. However, the two latest and much cited works (the Trelle 2011 and CNT 2013 NMAs) are direct or indirect comparisons of RCT data via a network meta-analysis, a technique that also presents issues of potential bias, indirectness, inconsistency, and imprecision (see Salanti G. et al. Evaluating the quality of evidence from a network meta-analysis. PLoS One. 2014;9(7):e99682).

The reporting of meta-analysis results may need to be challenged. For example, The CNT NMA [143] states that "high-dose naproxen was not associated with any significant excess risk of major vascular events (0·93, 0·69–1·27; figure 4), and nor was there an increase in major coronary events (0·84, 0·52–1·35)." We submit that the latter adjusted rate ratio of naproxen vs placebo lacks sufficient precision to rule out clinically meaningful differences. This is because the confidence interval is compatible with a decrease in risk of major coronary events of 48% but also with an increase in risk of MI or CHD death of 35%. In the CNT meta-analysis, MI risk was not considered separately for the various coxibs. The findings of our IPD MA suggest there are notable differences between rofecoxib and celecoxib for this outcome.

Regarding the validity of individual NSAID RCTs, various biases that may occur post-randomization in a RCT [144, 145] may potentially have affected the validity of trials included in these NMAs. These may include: 1) violation of treatment groups exchangeability due to non-persistence with assigned NSAID treatment, or 2) violation of treatment consistency due to group contamination (e.g. OTC NSAIDs) or to differential concomitant treatments potentially affecting the risk of the MI outcome during follow-up (e.g. low-dose aspirin) or sub-optimal adherence to the NSAID. For example, placebo-controlled RCTs of naproxen were mainly performed in populations with or at risk of Alzheimer's disease. [146] Non-adherence bias has been documented in the main naproxen study (ADAPT) [147] and this may translate into an underestimation of the risk of acute MI with naproxen in NMAs of RCTs.

---

[143] Bhala N, for the Coxib and traditional NSAID Trialists' (CNT) Collaboration, Emberson J, et al. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials. Lancet. 2013;382(9894):769-79.

[144] Hernán MA, Hernandez-Diaz S, Robins JM. Randomized trials analyzed as observational studies. Ann Intern Med. 2013 Oct 15;159(8):560-2.

[145] Manson JE, Shufelt CL, Robins JM. The potential for postrandomization confounding in randomized clinical trials. JAMA. 2016 Jun 7;315(21):2273-4.

[146] Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ. 2011;342:c7086.

[147] Steering Committee of the Alzheimer's Disease Anti-inflammatory Prevention Trial (ADAPT). Statement for communication to the FDA Arthritis Advisory Committee and the Drug Safety and Risk Management Advisory Committee. 18 February 2005. Available at: https://jhuccs1.us/adapt/pdf%20documents/FDA%20ADAPT%20STATEMENT_web%20posting.pdf. Last accessed August 31, 2016.

3- Dose effects cannot be estimated from meta-analyses of NSAID RCTs

The Trelle 2011 and CNT 2013 network meta-analyses of RCTs do not allow concluding on the relationship between dose of the NSAID and acute MI.

In the Trelle NMA (Web appendix 2: Additional results, p.6,Table 9: Results of sensitivity analyses on the influence of dose and outliers), the rate ratio (95% credibility interval) for myocardial infarction for naproxen vs. placebo was 0.82 (0.37-1.67) for the Main analysis 0.55 (0.20-1.56) and High-dose trials only.

The CNT NMA main paper reports that: "Almost all (roughly 99%) of primary outcomes occurred in trials involving a coxib or high-dose tNSAID (diclofenac 150 mg daily, ibuprofen 2400 mg daily, or naproxen 1000 mg daily)". But what is the risk with lower daily doses of traditional NSAIDs that many patients use in real-life? This is answered in our IPD MA. As for celecoxib, although the CNT NMA reported an analysis of events by dose compared with placebo, this was for the *composite endpoint of major vascular events* (Webfigure 15: Effect of different coxib regimens on major vascular events, by dose: trials of a coxib vs placebo, p. 18). But is there a dose-related increased risk of *acute MI* with celecoxib? Again our work provides an answer.

4- Duration effects cannot be estimated from meta-analyses of NSAID RCTs

The Trelle 2011 and CNT 2013 network meta-analyses of RCTs do not allow making a conclusion on the relationship between duration of NSAID use and acute MI.

In real-life patients take NSAIDs intermittently (with periods of non-exposure), change the daily dose they use over time, and switch one NSAID for another. [148, 149] For example in the RAMQ study (data available on request), the median duration of treatment episodes is 4 weeks with coxibs and 2 weeks with traditional NSAIDs (t-NSAIDs). In contrast, RCTs typically mandate taking NSAIDs continuously for lengthy periods of time. See for example the Trelle NMA, Table 1, which shows that follow-up ranged 12 to 176 weeks in arthritis indications and from 52 to 235 weeks in Alzheimer's disease or cancer prevention indications). The CNT NMA specified the inclusion of RCTs of at least 4 weeks duration (see the main paper – Identification of trials and eligibility assessment). But what is the acute MI risk associated with duration of use of less than 4 weeks? This is not available from meta-analyses of NSAIDs RCTs and is given in our results.

In summary, our IPD MA characterized the effects of individual NSAID dose and duration of use, and the time course of any associated risk for acute MI and this does provide new clinical guidance.

---

[148] Gore M, Sadosky A, Leslie D, et al. Patterns of therapy switching, augmentation, and discontinuation after initiation of treatment with select medications in patients with osteoarthritis. Clin Ther. 2011;33(12):1914-31.

[149] Langman M, Kahler KH, Kong SX, et al. Drug switching patterns among patients taking non-steroidal anti-inflammatory drugs: a retrospective cohort study of a general practitioners database in the United Kingdom. Pharmacoepidemiol Drug Saf. 2001;10(6):517-24.

However, the authors did not assess the degree of compliance with prescriptions of NSAIDS and other medications. The true effects of different dose and duration are not reliable. The approach is similar to intention-to-treat analysis in RCT.

A- The IPD MA assessment of compliance is not similar to intention-to-treat analysis in RCT.
In a RCT, intention-to-treat analysis estimates the effect of *being assigned* to an intervention, regardless of intervention received. [150]

In the IPD MA, we estimated the effect of *using NSAID*s, as estimated by the computer prescription fills at pharmacy level (RAMQ, Finland, Saskatchewan) or writing a prescription by doctors (GPRD). This is similar to estimating a per-protocol effect and by analogy requires adjustment for time-varying treatment and time-varying confounding. [151]

B-The extent of misclassification of NSAID exposure on the estimates of dose and duration effects can be estimated
As we stated in the Limitations section of the manuscript, this work shares the limitations of all database studies, which measure drug dispensing and not actual drug intake. Does that mean that pharmacoepidemiological studies are not reliable *to estimate* the true effects of different dose and duration? We argue against this viewpoint.

We vigorously attempted to minimize misclassification of time-varying NSAID use in our work and we investigated its extent. For example, in the RAMQ cohort, NSAID exposure was determined for each day of follow-up to correspond to expected real-life variation. We assessed, separately, exposure to each of the following drugs: celecoxib, diclofenac, ibuprofen, naproxen, and rofecoxib, and to all other NSAIDs, grouped together. Computer-recorded variables allowed the direct calculation of daily dose of NSAIDs of interest as pill strength times number of pills divided by number of days supplied. Days supplied and consecutive prescriptions dates confirmed the duration of each dispensing and allowed identifying gaps between the end of a preceding prescription and the start of a next one. A priori rules were specified to capture usage behaviours such as intermittent use, dose changes, and drug switches.

Using this methods, we documented that use could be overestimated when a prescription was dispensed on an 'as needed' basis. 'As needed' prescription was estimated to represent,

---

[150] Hernán MA. Estimating the Effects of Time-varying Treatments on Cancer Risk in Randomized and Nonrandomized Studies. Methodological Considerations in Evaluation of Cancer as an Adverse Outcome Associated With Use of NonOncological Drugs and Biological Products in the Postapproval Setting FDA/NCI, September 10-11, 2014. p. 11. Available at: http://www.fda.gov/downloads/Drugs/NewsEvents/UCM414812.pdf. Last accessed August 16, 2016.
[151] Hernán MA. Estimating the Effects of Time-varying Treatments on Cancer Risk in Randomized and Nonrandomized Studies. Methodological Considerations in Evaluation of Cancer as an Adverse Outcome Associated With Use of NonOncological Drugs and Biological Products in the Postapproval Setting FDA/NCI, September 10-11, 2014. p.12. Available at: http://www.fda.gov/downloads/Drugs/NewsEvents/UCM414812.pdf. Last accessed August 16, 2016.

the following proportion of single prescriptions: ibuprofen: 21.3%, diclofenac: 6.3%, naproxen: 4.9%, rofecoxib: 3.6%, and celecoxib: 2.3%. This is mentioned in our manuscript as we state that "In particular, the ibuprofen estimates may be globally more fragile due to the universal availability of over-the-counter purchase and because use 'as needed' in our data was documented to be more prevalent for ibuprofen than for the other NSAIDs."

The authors also argue importance of this study by pointing out a lack of precision in prior RCTs. However, the authors' estimates are not so precise, either. The uncertainty measures of estimates of different dose levels and durations are largely overlapping.

It can be estimated that NSAIDs RCTs provide fewer than 1000 acute MI events (532 over all exposures in Trelle 2011), whereas our IPD MA comprised 61 460 acute MI events. Had we analysed MI risk overall 'by NSAID' we would have found extremely precise results.

So far in single observational studies or in their meta-analyses, NSAID exposure has typically been modelled using simple binary time-varying indicators of either current use (exposed or unexposed) or current dose (low or high dose). [152, 153] Some of these studies have also separately explored the effect of categorized treatment duration and recency of use. [154, 155, 156] However, to more accurately relate acute MI to time-varying NSAID exposure and to characterize the time course of risk, we felt it was best to combine information about dose, duration, and timing of past treatments – which is what we did for measurement and modelling of NSAID exposure in this IPD MA.

Loss of precision is inevitable when characterizing the effects of dose and duration of treatment, as NSAID exposure becomes more granular. The clinical inferences that are nonetheless possible from this IPD MA offset lower precision in some of the current dose-duration categories. Consider also that is IPD MA was done in the Bayesian framework and that precision read from the credible intervals around each median summary OR of acute MI should be interpreted as the (95%) range of posterior probabilities of ORs of acute MI (see also answer to Comment 2 below).

---

[152] McGettigan P, Henry D. Cardiovascular risk with non-steroidal anti-inflammatory drugs: systematic review of population-based controlled observational studies. PLoS Med. 2011;8(9):e1001098.

[153] Varas-Lorenzo C, Riera-Guardia N, Calingaert B, et al. Myocardial infarction and individual nonsteroidal anti-inflammatory drugs meta-analysis of observational studies. Pharmacoepidemiol Drug Saf. 2013;22(6):559-70.

[154] Hammad TA, Graham DJ, Staffa JA, et al. Onset of acute myocardial infarction after use of non-steroidal anti-inflammatory drugs. Pharmacoepidemiol Drug Saf. 2008;17(4):315-21.

[155] Schjerning Olsen AM, Fosbol EL, Lindhardsen J, et al. Duration of treatment with nonsteroidal anti-inflammatory drugs and impact on risk of death and recurrent myocardial infarction in patients with prior myocardial infarction: a nationwide cohort study. Circulation. 2011;123(20):2226-35.

[156] Varas-Lorenzo C, Riera-Guardia N, Calingaert B, et al. Myocardial infarction and individual nonsteroidal anti-inflammatory drugs meta-analysis of observational studies. Pharmacoepidemiol Drug Saf. 2013;22(6):559-70.

In addition, precision changed easily by altering a prior uncertainty fed into Bayesian analysis according to one of supplementary tables.

Of course precision changed when altering the prior for the between-study standard deviation as the 4 analyses were designed specifically for the purpose of testing the robustness of the results. However, even with an unrealistic scenario of ORs of acute MI varying by a factor of 8-fold across studies – which is twice the range documented by historical data – results of the IPD MA (last column of Web Extra 1, Table 10) allow making a conclusion on dose effects and on duration effects.

Therefore, it is not convincing that this study is strong enough to achieve the authors' aims beyond available evidence from RCTs (Trelle et al., BMJ, 2011; NSAID Trialists' Collaboration, Lancet, 2013).

This IPD MA contributes the following 5 new findings:

1. **Naproxen MI risk**
   Naproxen was associated with increases in risk of acute myocardial infarction. Previous work has generally been underpowered to fully assess the risk of naproxen and while point estimates have been < 1, confidence intervals have not been able to exclude clinically meaningful risks. Our study has the required power to address this shortcoming.
2. **Celecoxib MI risk**
   With celecoxib, associated risk of acute MI were not greater than that of diclofenac, ibuprofen, or naproxen, were and lower than those of rofecoxib. Again the power of our study enables these additional insights.
3. **Dose-MI risk relationship**
   Acute MI risks were dose-related and use for 1 to 4 weeks at high-dose seemed particularly harmful with rofecoxib (> 50 mg/day), ibuprofen (> 1200 mg/day), and with naproxen (> 750 mg/day). Previous studies could not adequately assess the dose-MI risk response with NSAIDs.
4. **Early onset of risk**
   For all studied NSAIDs, the onset of associated MI risk occurred within the first week of use. Virtually no previous studies have extensively examined very short term risk, again perhaps due to a lack of events in this earlier time window. Documenting a rapid onset of risk is important information for both clinicians and patients and may help improve decision making about the risks and benefits trade-off with these drugs.
5. **Patterns of MI risk**
   NSAIDs exhibited distinct patterns of MI risk, which suggest that time dependency of risk may differ between NSAIDs. No previous research has employed the methodology required to examine this issue.

This IPD MA specifically addressed questions stemming from routine patient-centred clinical decisions such as initiating NSAID treatment, increasing the dose or renewing a prescription to extend continuous use, for which NSAID-specific guidance is currently

lacking. Current treatment guidelines promoting the lowest effective NSAID dose for the shortest possible duration [157, 158, 159] are not sufficiently informative.

We have found that the onset of NSAIDs-associated MI risk increase occurs within the first week of use and that use at high dose for 1 to 4 weeks is particularly harmful with naproxen and with ibuprofen. A clinician might consider this new information and decide not to use NSAIDs in a given patient. He or she might also opt to maintain a patient on a lower NSAID dose and seek alternative interventions for treating pain and inflammation.

Moreover, the prior meta-analysis of NSAIDs in observational studies is also available, including evaluation of dose levels (Varas-Lorenzo et al., Pharmacoepidemiol Drug Saf, 2013), as cited by the authors. The authors' approach of Bayesian analysis and IPD was fabulous, but there is no clear difference between this IPD meta-analysis and the prior one in terms of strength of evidence, information for audience, and influence on clinical medicine. Also, at least, this study does not provide strong information that alters available knowledge on NSAIDs.

Whereas the SOS project (safety of NSAIDs project) is a relevant real-world data initiative, [160], there are important methodological differences with this IPD MA.

<u>1- The SOS meta-analysis of NSAIDs and acute MI did not investigate how the internal validity of component studies might affect their findings</u>
In accordance with most aggregate meta-analyses, no restriction based on methodological quality was imposed in the SOS MA such that it included field case-control studies possibly affected by recall bias, [161,162] a cohort study with immortal time bias, [163] studies that compared risk in current users with users in the recent past [164, 165] and studies defining as

---

[157] American Geriatrics Society Panel on Pharmacological Management of Persistent Pain in Older Persons. Pharmacological management of persistent pain in older persons. J Am Geriatr Soc. 2009;57(8):1331-46.
[158] Abdulla A, Adams N, Bone M, et al. Guidance on the management of pain in older people. Age Ageing. 2013;42 Suppl 1:i1-57.
[159] Scarpignato C, Lanas A, Blandizzi C, et al. Safe prescribing of non-steroidal anti-inflammatory drugs in patients with osteoarthritis--an expert consensus addressing benefits as well as gastrointestinal and cardiovascular risks. BMC Med. 2015;13:55.
[160] Makady A, Goettsch W, Willemsen A. Review of Policies and Perspectives on Real-World Data (IMI-GetReal Deliverable). Draft Report dated 30.01.2015 available at: https://www.imi-getreal.eu/Portals/1/Documents/Publications/D1.3%20GetReal%20Glossary.pdf. Last accessed August 10, 2016.
[161] Bueno H, Bardaji A, Patrignani P, et al. Use of non-steroidal antiinflammatory drugs and type-specific risk of acute coronary syndrome. Am J Cardiol. 2010;105(8):1102-6.
[162] Kimmel SE, Berlin JA, Reilly M, et al. Patients exposed to rofecoxib and celecoxib have different odds of nonfatal myocardial infarction. Ann Intern Med. 2005;142(3):157-64.
[163] Mamdani M, Rochon P, Juurlink DN, et al. Effect of selective cyclooxygenase 2 inhibitors and naproxen on short-term risk of acute myocardial infarction in the elderly. Arch Intern Med. 2003;163(4):481-6.
[164] Graham DJ, Campen D, Hui R, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet. 2005;365(9458):475-81.

current use a time window possibly remote from the event date. [166, 167] These dissimilarities and potential biases, if occurring in a non-differential manner, would be expected to move pooled relative risks of acute MI towards the null. Estimates of acute MI risks found by the SOS MA of observational studies [168] are globally lower than those of our IPD MA. Although the SOS Investigators identified various potential biases, they did not report any sensitivity analyses for methodological quality.

For this systematic review and IPD meta-analysis, we elaborated *a priori* study selection criteria by considering all main sources of biases and how they specifically applied to our research questions. This allowed controlling for confounding (including that due to calendar time) and for misclassification at the stage of study selection. We submit that a major quality feature of our work is the enhanced control of confounding and of information bias that ensues from meta-analyzing highly internally valid observational studies.

2- The SOS meta-analysis does not accurately inform on NSAID dose or duration effects

The SOS Investigators analysed the effects of NSAIDs on acute MI risk stratified by daily dose. They found a significant dose-response relationship for rofecoxib, with higher dose indicating greater risk. However, cut-off values defining low-medium and high daily doses of NSAIDs were inconsistent across studies, and this was particularly true for naproxen with 'high dose' definition ranging from > 500 mg to > 1000 mg depending on the study (see Web Supplement to Varas-Lorenzo 2013 – p. 8). Possible misclassification, together with insufficient statistical power, hindered ability to conclude about the impact of NSAID dose other than for rofecoxib in the SOS MA

In the SOS MA (see Web Supplement to Varas-Lorenzo 2013 – p.9), definitions of duration of treatment varied across studies to an extent that precluded pooling of the effect estimates across different exposure windows.

Our IPD MA allows for improved coherence of the definitions of current exposure to NSAIDs. The fact that we were able to draw conclusions about the effect of dose and duration for dose levels and treatment durations corresponding to various scenarios of real-world use reveals a key methodological advantage of patient-level MA (our IPD MA) compared with aggregate-data MA (the SOS MA) as IPD MA improves the coherence of

---

[165] van der Linden MW, van der Bij S, Welsing P, et al. The balance between severe cardiovascular and gastrointestinal events among users of selective and non-selective non-steroidal anti-inflammatory drugs. Ann Rheum Dis. 2009;68(5):668-73.

[166] Abraham NS, El-Serag HB, Hartman C, et al. Cyclooxygenase-2 selectivity of non-steroidal anti-inflammatory drugs and the risk of myocardial infarction and cerebrovascular accident. Aliment Pharmacol Ther. 2007;25(8):913-24.

[167] Hippisley-Cox J, Coupland C. Risk of myocardial infarction in patients taking cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. BMJ. 2005;330(7504):1366.

[168] Varas-Lorenzo C, Riera-Guardia N, Calingaert B, et al. Myocardial infarction and individual nonsteroidal anti-inflammatory drugs meta-analysis of observational studies. Pharmacoepidemiol Drug Saf. 2013;22(6):559-70.

exposure and outcome definitions, [169] through data harmonization before pooling in a meta-analysis.

2. The inference from results and model fitting would not stop readers getting reluctant to interpret quantitative information.

The main results documented in the abstract and the main text are based on 80% probability or more of adverse cardiovascular effects of NSAID. That indicates 20% probability of no harm or benefit; or indicates that the authors argue findings based on Bayes p-values < 0.20, not p < 0.05.

We do not argue findings on the basis of Bayes p-values < 0.20, we argue on the basis of posterior probability statements.

In Bayesian analysis, combining prior information (as probability distributions) with the data on hand i.e. the current experience (via the likelihood) yields a posterior probability distribution for each parameter of interest, as follows:

$$\underset{\text{prior probability}}{p(\text{hypothesis})} \times \underset{\text{likelihood (frequentist p-value)}}{p(\text{data}|\text{hypothesis})} \propto \underset{\text{Bayesian posterior probability}}{p(\text{hypothesis}|\text{data})}$$

One can see from the above that the concept of hypothesis testing i. e the probability of the "data given a 'true' null hypothesis" (or the frequentist p-value) is absent from Bayesian analysis. In fact, through the Bayes theorem, the respective places of the 'hypothesis' and the 'data' are inverted in Bayesian vs frequentist analysis. Probabilities of the "hypothesis given the data" have a distribution, the posterior probability distribution in Bayesian analysis.

The "80% probability" we referred to in the manuscript is a statement that refers to posterior probabilities and was obtained in the following manner:
The point estimate (median) of summary ORs of acute MI is accompanied by its probability distribution. Typically one is interested in the 95% credible interval, represented by the values of 2.5th percentile up to the 97.5th percentile of the posterior distribution. This is often how the uncertainty in the current experience is presented to the reader. However the Bayesian framework also allows making direct probability statements, based on the posterior distribution of a parameter of interest, here the summary OR of acute MI. For exposures corresponding to current use for each NSAID, we obtained probabilities that the summary adjusted ORs of acute MI were greater than a series of pre-specified threshold of harms. These probabilities were compiled from the posterior distribution of summary adjusted ORs of acute MI by repeated sampling via a function that took the value of '1' each time the summary adjusted OR exceeded a given OR threshold and took the value of '0' otherwise. This is straightforward, as it can be done by writing the appropriate lines of

[169] van Walraven C. Individual patient meta-analysis--rewards and challenges. J Clin Epidemiol. 2010;63(3):235-7.

code in the model and monitoring the results as output of the Bayesian analysis. These results (the sum of '1's) are plotted in Figures 2 to 6 – (a) panels, which present the posterior probability of exceeding certain threshold ORs of MI for each NSAID in each current dose-duration category. The plots displayed in the manuscript help the reader making a conclusion about increasingly important degrees of harms associated with current use of each NSAID.

As we now report in the revised manuscript,

> **"Confidence in magnitude of risk increase associated with short-term use**
> We examined the plots of Bayesian posterior probability that ORs of acute MI were greater than a series of pre-specified thresholds. Selecting a benchmark of 80% (horizontal black line), and considering use for 1-7 days (intersecting gold line), there was at least 80% probability that the OR of acute MI is greater than 1.07 for celecoxib (Figure 2a), 1.30 for diclofenac and naproxen (Figures 3a and 5a), 1.25 for ibuprofen (Figure 4a), and 1.35 for rofecoxib (Figure 6a). Clinicians may wish to consider different exposure categories, cut points for posterior probability of risk, and OR thresholds. For example, with naproxen ≤ 750 mg/day for 1-4 weeks, there is a 58% probability that the OR of acute MI is greater than 1.20 (green line in Figure 5a)."

Considering what readers are used to reading, all may need to be replaced with medians and 95% credible intervals from posterior distributions of estimates.

We agree that readers are more used to seeing point estimates and 95% confidence or credible intervals and have amended the Abstract accordingly.

As the authors did this study to provide information useful for clinicians, the authors may want to provide types of information clinicians are used to reviewing. In other words, the authors are providing information on what they estimated, but not highlighting common measures of strengths of evidence.

We have revised the manuscript to help the reader interpret Bayesian posterior probabilities and probability statement that may be clinically relevant.

3. It is not clear why the authors selected categorical variables to define dose levels and duration of treatment. If the authors' interest was in dose levels and duration, the authors should not lose continuous information by creating crude categories. The rationale of the categorisation is missing.

We sought to produce real-world evidence by selecting cut points for daily dose and durations that correspond to available dosage forms, usual dosage range, and typical or recommended duration of use. Daily dose was dichotomized as 'Low' or 'High' based on clinical consensus for each drug with cut-offs also corresponding to median daily dose in the RAMQ study population. Continuous duration of current use was categorized as 'Use for

1 to 7 days, 'Short' (use for 8 to 30 days), and 'Long' (use for more than 30 days). These cut-offs were chosen also because of their clinical relevance. Use of an NSAID for 1 to 7 days corresponds to seeking relief for an acute condition. Duration of 8 to 30 days represents short-term use, with receipt of one or two consecutive prescriptions, while duration longer than 30 days reflects more chronic indications. Finally, 'Non-use' corresponds to individuals not exposed to any NSAIDs in the year preceding the index date. Note that this conservative definition of non-use reduces the possibility of confounding that might arise if patients who are sicker recently stopped taking NSAIDs.

In other words, the authors assumed that 100 mg/day of celecoxib has the same effect as 200 mg/day of celecoxib, by creating a category of 'low' dose <=200 mg/day. By doing that, the authors cannot argue the strength of this study on effects of different dose levels.

We do not assume that 100 mg/day of celecoxib has the same effect as 200 mg/day of celecoxib, this is an inference that the reader may or may not want to make. We provided OR of acute MI for celecoxib <=200 mg/day. We believe that our decisions for data categorization, as explained above, achieve a very good trade-off between loss of precision and clinical relevance.

4. The authors estimated odds ratios to association of NSAID use with MI risk. When the authors state 'risk', this would let readers interpret that the results reflect the risk of developing MI among those free from heart disease. This is because, epidemiologically, 'risk' is probability of getting the event over a certain period of time among those without it.

We refer to dictionary definition of 'risk' in epidemiology (A Dictionary of Epidemiology. Edited by Miquel Porta. Fifth edition. Published in print January 2008 | ISBN: 9780195314496. Published online January 2014. p. 240 http://www.irea.ir/files/site1/pages/dictionary.pdf), which is:
"Risk – The probability of an adverse or beneficial event in a defined population over a specified time interval. In epidemiology and in clinical research it is commonly measured through the cumulative incidence and the incidence proportion.1-3,5,6,8,270 Other concepts and measures of risk are also relevant to study and control threats to human health.14,38,106-108,212,248,279,303,304,332-336,361,539,603,712-71."

We do not understand why or how when we "state 'risk', this would let readers to interpret that the results reflect the risk of developing MI among those free from heart disease". Nothing indicates that the acute MI outcome is a first-time event.

In this study, the authors included also adults who had a prior history of ischaemic heart disease (IHD) which included non-fatal MI. Thus, the authors examined the first occurrence of MI among adults without history of IHD and also MI recurrence among adults with MI history.

The authors should clearly state that this study evaluated both types of outcomes. Clinicians would anticipate much higher risk of MI when seeing adults with MI history than adults

without MI history. Therefore, this study must let readers clearly and quickly recognise the inclusion of those with MI history and those without MI history.

Table 1, which can be copyedited to accompany the second paragraph of Results (describing study populations) provides for the benefit of readers the prevalence of cardiac comorbidities, cardiovascular risk factors and cardiovascular diseases, in each study.

Moreover, the authors should do stratified analysis by MI history, at least.

We reported results for our overall study population in this IPD MA. Investigation in patients with certain cardiovascular risk profile was the object of a separate study.

5. Considering the inclusion of adults with MI history or adults at very high risk of developing MI, developing other diseases, or dying, some assumption becomes concerning.

The authors did adjustment for potential confounders that might have caused confounding by indication. Although the adjustment cannot get rid of the problem, that is a fair approach. But, the model assumes that there is no heterogeneity in effects of NSAIDs. Given the diversity of the clinical conditions and heterogeneity between the populations, how did the authors have confidence about the assumption?

The authors estimated random effects of NSAIDs in addition to the main effects, assuming random heterogeneity in effects. Unlike a report from a meta-analysis in general, no measures of heterogeneity are presented in this manuscript. The authors should present them, explore sources of heterogeneity, and discuss implications from potential heterogeneity of effects.

As specified in the revised manuscript,

> "The available data for the IPD MA are gathered at multiple levels. At the higher level of analysis we have the characteristics of the four studies. At the lower level we have individual patient characteristics. For pooling individual patient data, we opted for a one-stage model in the Bayesian framework. Such a model has a hierarchical structure allowing the IPD to be combined in a single step while accounting for clustering of data within each database study. [170, 171]"

The model <u>does assume that there may be heterogeneity</u> in effects of NSAIDs. Not only can a hierarchical Bayesian model assume that there is effect heterogeneity but it allows the extent of that heterogeneity to be revealed. Indeed, as with other parameters this heterogeneity can be monitored as output of the posterior distribution. As a side note, we

---

[170] Debray TP, Moons KG, Abo-Zaid GM, et al. Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? PLoS One. 2013;8(4):e60650.

[171] Stewart GB, Altman DG, Askie LM, et al. Statistical analysis of individual participant data meta-analyses: a comparison of methods and recommendations for practice. PLoS One. 2012;7(10):e46042.

could explicitly model the heterogeneity between the four studies without having to choose a priori between a fixed-effect or a random-effects meta-analysis (as we would have had to do if using a frequentist framework);

While we provided this result in Web Extra 1, the revised manuscript reads:

> **"Heterogeneity of effects**
> Values of the between-study standard deviations on the log OR summarized across exposure categories had a mean =0.100 and a median= 0.108 in the primary analysis…."

As we stated in the previously submitted manuscript, we go on by saying, "For a given NSAID exposure category the ORs are similar in magnitude across studies. This indicates good robustness despite populations being heterogeneous and exposure duration and dose being constructed based on differing sources of raw data."
….

"Because of the underlying low between-study heterogeneity, information was effectively 'shared' by studies (Table 2) allowing more definitive conclusions to be reached. "

6. As commented above, the authors examined a population including people with prior history of major clinical events, i.e. at a high risk of dying. The authors examined effects of interest in populations derived from mainly nested case-control studies. In a risk-set sampling, participants evaluated ought to be survivors until the time of an outcome (MI), thereby ignoring anyone who died due to non-MI causes. Non-MI causes could include congestive heart failure (CHF) which appeared to be related to NSAID use (Lancet, 2013).

One of the limitation in this study is missing approach to handling competing risk. Sensitivity analysis to examine robustness against possible competing risk is necessary. Overall conclusion in the Abstract and in the Discussion should also account for the possibility. Discussion on this issue would have to be discussed.

As mentioned in the manuscript, the IPD MA pooled healthcare databases and is representative of general populations using NSAIDs in real-life. Subjects with history of major clinical events at high risk of dying are indeed represented in these databases.

The Finland study was a case-control study of the Finnish population, with MI case identified from the Finnish Hospital Discharge Register and, subsequently, controls (matched on age at end of calendar year, sex, and hospital catchment area) randomly selected from the Population Register.

The RAMQ, GPRD, and Saskatchewan studies were case-control studies nested within a dynamic populations, the healthcare or medical databases (see Web Extra Table 2 for a description of the data sources). 'Dynamic population' refers to populations in which the

members vary over time; the membership is not fixed. [172] The study population may be regarded as dynamic and open in terms of calendar time, with database participants 'joining' at different times. In such healthcare or medical databases studies, it is reasonable to acknowledge that individuals may die without the outcome of interest, and therefore that inferences are made 'in the presence of the competing risk of dying'. [173] Therefore, a competing risk sensitivity analysis to examine robustness is not necessary.

7. The tone of language should be revised throughout. For example, the authors argued that "We identified all potential confounders and risk factors for outcome based on substantive knowledge and …". The authors did not capture any continuous characteristics of blood pressure measures, blood cholesterol levels, and any lifestyle characteristics except smoking. With possibility of residual confounding, which can be argued limitlessly, some degrees of uncertainty should be apparent in sentences.

Maybe the following revised paragraph in our Methods is clearer:

> "On the basis of substantive knowledge and confirmation by a search of the literature [174,175,176,177] we identified risk factors for the outcome and potential confounders, which formed a set of candidate covariates to include in multivariable regression analysis. Via a simplified causal graph, [178] we mapped relationships between variables,[179] (Web Figure 2), including the special case of time-dependent confounders that are mediating intermediates on the causal pathway between NSAID exposures and acute MI outcome. [180] Such mediator variables must be identified at the analysis planning stage as they require appropriate measurement and analytical strategies. Using the reference RAMQ study, we then substantiated the confounder status for each candidate covariate, by calculating the odds ratio (OR) of association between the covariate and exposure to NSAIDs among controls and the OR of association

---

[172] Vandenbroucke JP, Pearce N. Incidence rates in dynamic populations. Int J Epidemiol. 2012 Oct;41(5):1472-9

[173] Andersen PK, Geskus RB, de Witte T, et al. Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol. 2012 Jun;41(3):861-70.

[174] Coca SG, Yusuf B, Shlipak MG, et al. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;53(6):961-73.

[175] Hall AJ, Stubbs B, Mamas MA, et al. Association between osteoarthritis and cardiovascular disease: systematic review and meta-analysis. Eur J Prev Cardiol. 2015 Oct 13.

[176] Han MK, McLaughlin VV, Criner GJ, et al. Pulmonary diseases and the heart. Circulation. 2007;116(25):2992-3005.

[177] Meisinger C, Doring A, Lowel H. Chronic kidney disease and risk of incident myocardial infarction and all-cause and cardiovascular disease mortality in middle-aged men and women from the general population. Eur Heart J. 2006;27(10):1245-50.

[178] Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10(1):37-48.

[179] Sauer BC, Brookhart MA, Roy J, et al. A review of covariate selection for non-experimental comparative effectiveness research. Pharmacoepidemiol Drug Saf. 2013.

[180] Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550-60.

between the covariate and MI outcome in the unexposed, leading to the final set of target confounders (referred to as 'IPD MA confounders') selected for adjustment in the meta-analysis (Web Table 5). The final set of target 'IPD MA confounders' comprised: age at index date, male sex, diabetes, hyperlipidemia, hypertension, previous myocardial infarction, coronary heart disease (excluding prior MI), congestive heart failure, cerebrovascular disease, peripheral vascular disease, chronic obstructive pulmonary disease, gastrointestinal ulcer disease, gastrointestinal bleed, acute or chronic renal failure, rheumatoid arthritis, and concomitant treatment with oral corticosteroids, clopidogrel, or cardioprotective aspirin. The definitions and determination strategies for 'IPD MA confounders' available from each study can be found in Web Tables 6 to 9. All studies were adjusted for the following 'common IPD MA confounders': age at index date, diabetes, hyperlipidemia, hypertension, coronary heart disease, and rheumatoid arthritis. Each study was also adjusted for additional 'IPD MA confounders', based on measurement in the original dataset, as described in the footnote to Table 2. When an 'IPD MA confounder' was unavailable, it was missing systematically for all subjects in that study."

According to current clinical knowledge we do not believe that we failed to identify important confounders. As illustrated in Figure 2 in Web Extra 1 and explained in the legend to that figure, there exist unmeasured (smoking, obesity, socio-economic status, exercise, OTC NSAIDs, OTC low-dose aspirin), incompletely measured (blood pressure increase, renal dysfunction, heart failure [confounding by contraindication]), and unmeasurable risk factors for myocardial infarction (genes) that we could not adjust for. However whether these risk factors are associated with NSAID exposure, a required element to complete the triad of confounding is not established as shown below.

We did our best to adjust for confounding by measured confounders. In addition to using multivariable regression, confounding control in the IPD MA included matching on time including for index date, year of cohort entry, and duration of follow-up in cohort (see answer to previous Question 5 or Web Extra 1 –Table 2). This matching on time is an important quality feature, given the potential confounding effect of calendar time around the launch of celecoxib and rofecoxib (1999-2000).

For refining confounding adjustment, we entertained a few options including propensity score (PS) methods and instrumental variable (IV) analysis (prescriber's preference or calendar time) but considered them unsuitable. Briefly, a sound propensity score approach [181] would have to consider time-varying use of each individual NSAIDs and change in measured confounder values over time. Ultimately, it might improve confounding adjustment only marginally because the effect of mediating variables between NSAID exposure and the MI outcome cannot be factored in the determination of PS. As for

[181] Mack CD, Glynn RJ, Brookhart MA, et al. Calendar time-specific propensity scores and comparative effectiveness research for stage III colon cancer chemotherapy. Pharmacoepidemiol Drug Saf. 2013;22(8):810-8.

instrumental variable analysis, the literature [182, 183, 184] suggests that confounding of the IV and the acute MI outcome association is likely to occur such that IV analysis is a poor method for dealing with unmeasured confounders in our research.

However, the literature provides useful insight on the risk of bias due to confounders that are unobserved in database studies. Graham and coll. [185] assessed the potential for confounding from low-dose aspirin use, OTC NSAID use, smoking history, and family history of acute MI via a standardized telephone survey of a random sample of the controls who were currently or remotely exposed to NSAIDs in their study. They found that these factors were not differentially distributed with respect to NSAID exposure. Schneeweiss [186] used data collected in the Medicare Current Beneficiary Survey and calculated the confounding bias of not having adjusted an NSAID study in Medicare beneficiaries [187] for BMI, OTC aspirin use, current smoking, income, and educational attainment. Failure to adjust for these unmeasured confounders led to a small underestimation of the association between coxibs and MI. [188]

We believe that the main source of confounding bias is the inability to adjust for unmeasured mediators (NSAID-induced blood pressure increases, renal deterioration) between NSAID exposure and the acute MI outcome. In an attempt to estimate confounding bias due to unmeasured mediators, we considered the prevalence and the strength of the association of each mediator [189] with the acute MI outcome documented in the literature [190, 191, 192] We also reviewed the strength of association between NSAIDs and blood pressure

---

[182] Tamblyn R, McLeod P, Hanley JA, et al. Physician and practice characteristics associated with the early utilization of new prescription drugs. Med Care. 2003;41(8):895-908.

[183] Solomon DH, Schneeweiss S, Glynn RJ, et al. Determinants of selective cyclooxygenase-2 inhibitor prescribing: are patient or physician characteristics more important? Am J Med. 2003;115(9):715-20.

[184] Franklin JM, Schneeweiss S, Huybrechts KF, et al. Evaluating possible confounding by prescriber in comparative effectiveness research. Epidemiology. 2015;26(2):238-41.

[185] Graham DJ, Campen D, Hui R, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet. 2005;365(9458):475-81.

[186] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005;16(1):17-24.

[187] Solomon DH, Schneeweiss S, Glynn RJ, et al. Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults. Circulation. 2004;109(17):2068-73.

[188] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005;16(1):17-24.

[189] Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology. 2011;22(1):42-52.

[190] Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990;335(8693):827-38.

[191] Meisinger C, Doring A, Lowel H. Chronic kidney disease and risk of incident myocardial infarction and all-cause and cardiovascular disease mortality in middle-aged men and women from the general population. Eur Heart J. 2006;27(10):1245-50.

increases [193, 194, 195, 196, 197] and renal failure. [198] In the end, the overall extent and the direction of any residual confounding due to mediators are very hard to predict as these may be causally ordered and may co-exist, and because some biasing effects might cancel each other. In theory, the likelihood of residual confounding due to mediating blood pressure increase or renal deterioration may increase with longer durations of NSAID use over follow-up time – such that the estimates of acute MI risk in the 'Long' duration categories might be biased to the null.

In summary, on the basis of our assessment of the literature and anticipated direction of bias to the null, we think that unmeasured and unmeasurable confounders are unlikely to affect the substantive conclusions of this IPD MA, which found associations between current NSAID exposure and increased risk of acute MI.

We have added the following text in the revised manuscript:

> "We suspect that residual confounding exists because substantive knowledge [199, 200, 201] ascertains there are mediating intermediate variables on the causal pathway between NSAID exposures and acute MI (Web Figure 2). The likelihood of residual confounding due to mediating blood pressure increase or renal deterioration may increase with longer durations of NSAID use over follow-up time – such that the ORs of acute MI in the 'Long' duration categories might be biased to the null (underestimated). Previous work [202, 203] provide useful insight

---

[192] Coca SG, Yusuf B, Shlipak MG, et al. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;53(6):961-73.

[193] Pope JE, Anderson JJ, Felson DT. A meta-analysis of the effects of nonsteroidal anti-inflammatory drugs on blood pressure. Arch Intern Med. 1993;153(4):477-84.

[194] Johnson AG, Nguyen TV, Day RO. Do nonsteroidal anti-inflammatory drugs affect blood pressure? A meta-analysis. Ann Intern Med. 1994;121(4):289-300.

[195] Farkouh ME, Verheugt FW, Ruland S, et al. A comparison of the blood pressure changes of lumiracoxib with those of ibuprofen and naproxen. J Clin Hypertens (Greenwich). 2008;10(8):592-602.

[196] Chan CC, Reid CM, Aw TJ, et al. Do COX-2 inhibitors raise blood pressure more than nonselective NSAIDs and placebo? An updated meta-analysis. J Hypertens. 2009;27(12):2332-41.

[197] Snowden S, Nelson R. The effects of nonsteroidal anti-inflammatory drugs on blood pressure in hypertensive patients. Cardiol Rev. 2011;19(4):184-91.

[198] Ungprasert P, Cheungpasitporn W, Crowson CS, et al. Individual non-steroidal anti-inflammatory drugs and risk of acute kidney injury: A systematic review and meta-analysis of observational studies. Eur J Intern Med. 2015;26(4):285-91.

[199] Johnson AG, Nguyen TV, Day RO. Do nonsteroidal anti-inflammatory drugs affect blood pressure? A meta-analysis. Ann Intern Med. 1994;121(4):289-300.

[200] Chan CC, Reid CM, Aw TJ, et al. Do COX-2 inhibitors raise blood pressure more than nonselective NSAIDs and placebo? An updated meta-analysis. J Hypertens. 2009;27(12):2332-41.

[201] Ungprasert P, Cheungpasitporn W, Crowson CS, et al. Individual non-steroidal anti-inflammatory drugs and risk of acute kidney injury: A systematic review and meta-analysis of observational studies. Eur J Intern Med. 2015;26(4):285-91.

[202] Graham DJ, Campen D, Hui R, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet. 2005;365(9458):475-81.

on the risk of bias due to confounders unobserved in database studies (obesity, OTC aspirin or NSAID use, smoking, income, or educational attainment), which suggests that failure to adjust for these confounders might slightly underestimate MI risk.[204] On the basis of our assessment of the literature[205, 206, 207] and the anticipated direction of bias to the null, we believe that unmeasured and incompletely measured confounders are unlikely to affect the substantive conclusions of this IPD MA, which found associations between current NSAID exposure and increased risk of acute MI."

---

[203] Solomon DH, Schneeweiss S, Glynn RJ, et al. Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults. Circulation. 2004;109(17):2068-73.

[204] Schneeweiss S, Glynn RJ, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. Epidemiology. 2005;16(1):17-24.

[205] Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990;335(8693):827-38.

[206] Meisinger C, Doring A, Lowel H. Chronic kidney disease and risk of incident myocardial infarction and all-cause and cardiovascular disease mortality in middle-aged men and women from the general population. Eur Heart J. 2006;27(10):1245-50.

[207] Coca SG, Yusuf B, Shlipak MG, et al. Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;53(6):961-73.