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 and some consider high-dose (500 mg twice daily) naproxen as being cardiac neutral. 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) 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 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 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. Non-adherence bias has been documented in the main naproxen study (ADAPT) 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.

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

---

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. However, interaction studies involving a categorical NSAID exposure, such as in this IPD MA, would have less statistical power and would overly depend on classifying exposure without error.”

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

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

<table>
<thead>
<tr>
<th>Study selection criteria</th>
<th>Number of studies excluded, with main reasons (N=76)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Sourced from validated computerized drug prescription or medical databases</td>
<td>• Availability of larger source cohort (n=1, previous smaller RAMQ study)</td>
</tr>
<tr>
<td>2) Conducted in the general or an elderly population</td>
<td>• Limited external validity due to selection of population (n=6)</td>
</tr>
<tr>
<td>3) Appropriately allowed for time-dependent analyses</td>
<td>• 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)</td>
</tr>
<tr>
<td>4) Documented acute MI separately from other CV outcomes</td>
<td>• Outcome is a composite or another cardiovascular endpoint (n=28)</td>
</tr>
<tr>
<td>5) Considered both selective COX-2 inhibitors and traditional NSAIDs</td>
<td>• Did not consider both traditional and COX-2 selective NSAIDs (n=11)</td>
</tr>
<tr>
<td>6) Compared acute MI risk in NSAID users with</td>
<td>• Risk in users not compared with risk in non-</td>
</tr>
<tr>
<td>Study selection criteria</td>
<td>Number of studies excluded, with main reasons (N=76)</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>non-users</td>
<td>users (n=5)</td>
</tr>
<tr>
<td>7) Minimized the effects of confounding by indication or by contraindication and of</td>
<td>• Not limited to period before rofecoxib withdrawal (n=8)</td>
</tr>
<tr>
<td>selective prescribing. This was achieved by limiting the inclusion of studies to those</td>
<td></td>
</tr>
<tr>
<td>which were conducted before the withdrawal of rofecoxib, matched cases and controls for</td>
<td></td>
</tr>
<tr>
<td>calendar time, and documented comorbidities and concomitant treatments.</td>
<td></td>
</tr>
</tbody>
</table>

Please note that although tools exist for assessing the quality of observational studies, these critical appraisal instruments 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 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, 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, our purpose in limiting the inclusion of studies to those showing an

---

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).”

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.\textsuperscript{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 understand 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.

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).\textsuperscript{25, 26}

Some of these studies have also separately explored the effect of categorized treatment duration and recency of use.\textsuperscript{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.

\textit{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

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:

\[ p(\text{hypothesis}) \times p(\text{data}|\text{hypothesis}) \propto p(\text{hypothesis}|\text{data}) \]

prior probability likelihood (frequentist p-value) Bayesian posterior probability

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 2.5\text{th} percentile up to the 97.5\text{th} 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:

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

---


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


treatment of OA; in fact, use of NSAIDs is ubiquitous. \cite{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 \cite{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. \cite{40}

---


\cite{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.