


# Channeling in the Use of Nonprescription Paracetamol and Ibuprofen in an Electronic Medical Records Database: Evidence and Implications

Rachel B. Weinstein<sup>1</sup>  · Patrick Ryan<sup>1</sup> · Jesse A. Berlin<sup>2</sup> · Amy Matcho<sup>3</sup> · Martijn Schuemie<sup>1</sup> · Joel Swerdel<sup>1</sup> · Kayur Patel<sup>4</sup> · Daniel Fife<sup>1</sup>

Published online: 5 August 2017

© The Author(s) 2017. This article is an open access publication

## Abstract

**Introduction** Over-the-counter analgesics such as paracetamol and ibuprofen are among the most widely used, and having a good understanding of their safety profile is important to public health. Prior observational studies estimating the risks associated with paracetamol use acknowledge the inherent limitations of these studies. One threat to the validity of observational studies is channeling bias, i.e. the notion that patients are systematically exposed to one drug or the other, based on current and past comorbidities, in a manner that affects estimated relative risk.

**Objectives** The aim of this study was to examine whether evidence of channeling bias exists in observational studies that compare paracetamol with ibuprofen, and, if so, the extent to which confounding adjustment can mitigate this bias.

**Study Design and Setting** In a cohort of 140,770 patients, we examined whether those who received any paracetamol (including concomitant users) were more likely to have prior diagnoses of gastrointestinal (GI) bleeding, myocardial infarction (MI), stroke, or renal disease than those who received ibuprofen alone. We compared propensity score

distributions between drugs, and examined the degree to which channeling bias could be controlled using a combination of negative control disease outcome models and large-scale propensity score matching. Analyses were conducted using the Clinical Practice Research Datalink.

**Results** The proportions of prior MI, GI bleeding, renal disease, and stroke were significantly higher in those prescribed any paracetamol versus ibuprofen alone, after adjusting for sex and age. We were not able to adequately remove selection bias using a selected set of covariates for propensity score adjustment; however, when we fit the propensity score model using a substantially larger number of covariates, evidence of residual bias was attenuated.

**Conclusions** Although using *selected* covariates for propensity score adjustment may not sufficiently reduce bias, large-scale propensity score matching offers a novel approach to consider to mitigate the effects of channeling bias.

**Electronic supplementary material** The online version of this article (doi:10.1007/s40264-017-0581-7) contains supplementary material, which is available to authorized users.

✉ Rachel B. Weinstein  
Rweinst1@its.jnj.com

<sup>1</sup> Janssen Research and Development, LLC, 1125 Harborton-Trenton Rd, Titusville, NJ 08560, USA

<sup>2</sup> Johnson and Johnson, 1125 Harborton-Trenton Rd, Titusville, NJ 08560, USA

<sup>3</sup> Janssen Research and Development, LLC, 920 US Route 202, Raritan, NJ 08869, USA

<sup>4</sup> Jan-Cil UK, High Wycombe, UK

## Key Points

Channeling bias exists in the dispensing of single-ingredient paracetamol versus ibuprofen. In previously published papers, models attempting to control for such channeling may not have adequately controlled for this bias.

Propensity score model diagnostics and negative control outcomes can be used to check the adequacy of models to control for channeling bias.

Large-scale propensity score matching can reduce channeling bias and should be considered for bias reduction in future observational studies in which the possibility of channeling bias is a major concern.

## 1 Introduction

Over-the-counter analgesics such as paracetamol and ibuprofen are among the most widely used drugs in the world. Therefore, having a good understanding of their safety profile is an important public health consideration. A key challenge to studying the safety of paracetamol and ibuprofen is that they are used primarily without prescription, therefore exposure status can be difficult to ascertain accurately in both prospective and retrospective epidemiology studies. Several observational studies have been conducted that attempt to estimate the risks associated with the use of paracetamol [1–8], but all acknowledge the inherent limitations of observational studies in this context. One approach that has been taken [7] is to use electronic medical records that capture prescriptions, and attempt to estimate the risks using only prescription exposure to paracetamol compared with prescription exposure to other pain medication, such as ibuprofen. While this approach may have promise, a threat to the validity of this design is channeling bias, also called selection by contraindication, i.e. the notion that patients may be systematically exposed to one drug or the other, based on current and past comorbidities, in a manner that could affect the estimates of relative risk.

This study aims to examine whether channeling bias exists in the context of studies that compare paracetamol with ibuprofen, and, if so, the extent to which various confounding adjustment strategies can mitigate this bias when estimating average treatment effects. We do so using the Clinical Practice Research Datalink (CPRD) database, which was the basis of a study [7] that found an association between paracetamol and increased risk of several adverse events, including upper gastrointestinal (GI) bleeding, myocardial infarction (MI), stroke, and acute renal failure. That study adjusted its estimates for several potential confounders and clearly acknowledged the possibility of confounding by indication (actually, contraindication in this instance), but did not document the existence of this bias or attempt to assess its impact.

The primary research question addressed in the current study is whether evidence of channeling bias in the prescription of single-ingredient paracetamol versus single-ingredient ibuprofen can be detected in an electronic medical records database. If so, a second research question assessed the prospects for studies confronted with that bias to reliably control it through multivariable methods or propensity score adjustment methods that assume the groups being compared are drawn from a single population.

## 2 Methods

We used two approaches to determine if there was evidence of bias. In the first, we looked for evidence that patients who received paracetamol were more likely to have had specific prior diagnoses of interest, including GI bleeding, MI, ischemic or hemorrhagic stroke, or acute or chronic renal disease, than those who received ibuprofen. These are known risks associated with ibuprofen, therefore the question is whether doctors selectively prescribe paracetamol to individuals for whom ibuprofen would be less appropriate. Second, we conducted a more global assessment of bias through the use of propensity score distributions to demonstrate the extent of differences between the populations exposed to these two drugs in terms of their propensity to be prescribed paracetamol versus ibuprofen. Finally, we examined the degree to which channeling bias could be controlled via propensity score matching. To do this, we fitted logistic regression models of negative control outcomes to propensity score matched, paracetamol-exposed and ibuprofen-exposed groups. For the comparison of these groups, we identified negative control outcomes, conditions chosen specifically because they were evaluated a priori to not have a differential risk associated with either OTC analgesic, i.e. conditions for which the relative risk associated with paracetamol versus ibuprofen was therefore believed to be approximately one. In this analysis, successful control of channeling by the propensity score matching would be evidenced by the lack of association of the negative control outcomes with the exposure group, i.e. by observing no more negative controls associated with the exposure group than would be expected to be observed by chance. The criteria for defining these negative controls are described below. The protocol for this study is available at ClinicalTrials.gov (NCT identifier: NCT02830178).

### 2.1 Database Used

The CPRD, a UK primary care database containing de-identified data from 1 January 1988 through 31 December 2014 and covering 11.65 million individuals, was used for these analyses. The database includes data on demographics, conditions diagnosed, observations, measurements, and procedures that the general practitioner (GP) is made aware of, in addition to any made by the GP. A key strength of the data is the long-term follow-up. Overall, median follow-up time for individual patients is approximately 5 years (interquartile range 1.8–11.1 years) [9]. Data for this study were from practices classified as ‘up to standard’ (UTS) by the CPRD during the period of interest.

The UTS designation reflects a minimum, practice-level measure of quality based on continuity of recording and number of deaths recorded. The protocol for this study (reference number 15\_173R) was approved by the Independent Scientific Advisory Committee (ISAC).

## 2.2 Determination of Exposures and Outcomes

The CPRD data in this analysis were converted to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) version 5 [10]. The OMOP CDM allows for standardization of clinical structure and content across databases, and also facilitates the development of covariates in the large-scale propensity score models. Native source codes for outcomes and comorbid conditions are mapped to the OMOP standardized vocabularies for each domain (conditions, procedures, etc.). For instance, all Read condition codes (diagnoses, procedures, and clinical observations) are mapped to SNOMED-CT concepts, and Gemscript drug codes are mapped to RxNorm codes [10].

## 2.3 Cohort Definition

For subjects included in the study, the index date was defined as the day when they received a first prescription for single-ingredient paracetamol and/or single-ingredient ibuprofen. Subjects were included if they were aged 18 years or older on the index date and enrolled in an UTS practice for the 2 years prior to and 1 year following the index date. To reduce the risk of observing prevalent prescription use, we required at least 6 months of continuous observation without prescriptions prior to the first prescriptions of paracetamol or ibuprofen in 2012. Only index dates in 2012 were considered. In addition, subjects with a prescription for paracetamol- or ibuprofen- containing combination medication on the index date were excluded.

We classified analgesic use into two cohorts: (1) ‘any paracetamol’, i.e. patients with new paracetamol exposure alone or in combination with a new ibuprofen exposure; and (2) ‘ibuprofen alone’, i.e. patients with new, single-ingredient ibuprofen alone. Where multiple exposures on the index date qualified a patient for inclusion, cohort and index date assignment were based on the qualifying exposure(s) occurring on the index date only. Approximately 6% of the cohort had concomitant use and were classified as ‘any paracetamol’, to better reflect prescribing in clinical practice. A sensitivity analysis excluding concomitant users, to assess the impact of these users on the evidence of channeling, was performed.

## 2.4 Statistical Methods Used

### 2.4.1 Channeling Bias Models

To gain an understanding of how the study populations differ from each other, a descriptive comparison by 10-year age group, sex, and baseline health-related characteristics was performed, stratified by prescriptions for any paracetamol versus ibuprofen, and age 18 years or older in 2012 (see the ‘Before Match’ columns of Table 1).

We examined potential channeling bias in the prescription of any paracetamol versus ibuprofen. We used multivariable logistic regression with any paracetamol versus ibuprofen as the outcome, and diagnoses of GI bleeding, MI, ischemic or hemorrhagic stroke, or acute or chronic renal disease as the ‘exposure’ (predictor) variable, controlling for age group in 5-year increments and sex. The presence or absence of a given diagnosis was determined by diagnoses in the database for the 2 years prior to the index date (date of the first paracetamol/ibuprofen prescription).

### 2.4.2 New-User Cohort Analysis Using Propensity Scores

A comparative cohort analysis was performed, comparing new users of paracetamol with new users of ibuprofen. Two propensity score models were created, both with any paracetamol use (yes/no) as the ‘outcome’ variable. For the first propensity score model, the independent variables were baseline patient characteristics defined as described in previous studies described in the literature [1–5, 7]. These covariates, which we will refer to as ‘publication covariates’, were not consistently included in each study, but represent a super-set of the covariates measured, and include the following conditions and drug exposures (the number of studies that included the variable is given in parentheses): obese (6), morbidly obese (6), smoker (5), alcohol abuser (5), upper GI events (1), osteoarthritis (1), rheumatoid arthritis (1), ischemic heart disease (3), heart failure (2), hypertension (7), cerebrovascular disease (1), diabetes mellitus (4), hyperthyroidism (1), stroke or transient ischemic attack (1), cancer [excluding non-melanoma skin cancer] (1), inflammatory bowel (1), autoimmune disease (1), depression (1), drug abuse (1), anticoagulants (1), oral glucocorticoids (1), diuretics (2), cardiac glycosides (1), statins (2), angiotensin receptor blockers (3), hypnotics and anxiolytics (1), antipsychotics (1), antibacterials (1), aminosaliculates (1), antidepressants (1), aspirin (2), oral corticosteroids (1), proton-pump inhibitors (1), histamine-2 receptor antagonists (1), hyperlipidemia (2), non-ibuprofen non-steroidal anti-inflammatory drugs (2), and other analgesics (5).

**Table 1** Distribution of selected characteristics 1 year prior to first use of paracetamol and ibuprofen in 2012 among users in the study populations before and after matching on the publication variables PS model and the large-scale PS model

	Before match ( <i>N</i> = 143,126)			After matched on variables in the publication variables PS model ( <i>N</i> = 98,048)			After matched on variables in the large-scale PS model ( <i>N</i> = 66,070)		
	Paracetamol, %	Ibuprofen, %	Std diff <sup>a</sup>	Paracetamol, %	Ibuprofen, %	Std diff <sup>a</sup>	Paracetamol, %	Ibuprofen, %	Std diff <sup>a</sup>
Female	60.6	59.1	0.029	59.8	61.1	-0.026	60.0	60.4%	-0.008
Mean age, years (SD)	62.2 (19.6)	52.6 (18.3)	-	56.1 (19.0)	55.8 (18.7)	-	56.4 (19.2)	57.3 (19.2)	-
Age, years									
15-19	1.6	3.1	-0.098	2.4	2.4	0.001	2.6	2.3	0.016
20-24	3.2	5.5	-0.115	4.6	4.6	0.001	4.6	4.5	0.005
25-29	3.5	5.4	-0.091	4.8	4.8	0.002	4.6	4.3	0.011
30-34	4.0	5.9	-0.090	5.4	5.5	-0.003	5.2	5.0	0.009
35-39	4.0	6.3	-0.105	5.3	5.4	-0.002	5.1	4.9	0.008
40-44	4.8	8.0	-0.134	6.4	6.4	0.001	6.4	5.8	0.023
45-49	5.4	9.1	-0.146	7.2	7.2	0.002	7.1	6.6	0.018
50-54	5.8	8.8	-0.118	7.5	7.3	0.005	7.2	6.9	0.011
55-59	6.2	8.0	-0.072	7.5	7.5	0.000	7.2	7.1	0.005
60-64	8.8	9.9	-0.038	10.1	10.4	-0.010	10.0	9.9	0.003
65-69	10.7	10.0	0.021	11.2	11.6	-0.011	11.4	11.6	-0.005
70-74	10.3	7.4	0.102	9.5	9.9	-0.014	9.5	10.1	-0.020
75-79	11.0	5.9	0.184	8.0	8.1	-0.004	8.4	9.2	-0.027
80-84	9.8	3.8	0.241	5.5	5.2	0.015	6.0	6.5	-0.023
85-89	7.0	1.9	0.249	3.0	2.6	0.023	3.1	3.5	-0.019
90-94	3.3	0.7	0.185	1.2	1.0	0.015	1.3	1.4	-0.012
95-99	0.8	0.1	0.101	0.2	0.1	0.012	0.2	0.2	0.007
Conditions									
Obese	4.0	3.3	0.040	3.8	3.6	0.007	3.7	3.7	-0.001
Smoker	8.3	7.7	0.022	8.7	7.9	0.028	8.3	8.3	0.000
Osteoarthritis of the knee	15.1	9.0	0.188	11.6	11.3	0.012	11.9	12.5	-0.021
Neoplasm of the prostate	1.4	0.8	0.054	0.9	1.1	-0.021	1.2	1.2	-0.003
Hyperthyroidism	0.6	0.5	0.015	0.5	0.5	-0.001	0.5	0.5	-0.005
Inflammatory bowel disease	0.2	0.2	0.004	0.1	0.2	-0.009	0.1	0.2	-0.009
Rheumatoid arthritis	1.8	0.7	0.091	1.3	0.9	0.033	1.1	1.2	-0.005
Heart failure	1.2	0.3	0.106	0.5	0.4	0.020	0.5	0.5	0.003
Ischemic heart disease	4.9	2.0	0.156	2.9	2.8	0.006	2.9	3.1	-0.011
Essential hypertension	29.0	17.5	0.275	21.1	21.6	-0.010	21.9	23.2	-0.033
Hyperlipidemia	5.3	3.4	0.092	4.1	4.2	-0.006	4.1	4.3	-0.008
Cerebrovascular disease	1.9	0.5	0.124	0.9	0.7	0.019	0.7	0.9	-0.022
Drugs									
Cardiac glycosides	2.3	0.5	0.151	0.8	0.7	0.015	0.9	0.9	-0.005
Corticosteroids, systemic	23.8	19.1	0.115	20.5	21.5	-0.026	21.1	21.5	-0.009
Antibacterials, systemic	53.1	46.9	0.123	50.1	51.2	-0.022	49.9	50.8	-0.018
Proton-pump inhibitors	36.9	28.8	0.175	32.4	32.4	0.000	32.0	32.9	-0.019
Angiotensin system	29.5	17.9	0.277	21.8	22.1	-0.007	22.5	23.8	-0.031

**Table 1** continued

	Before match ( <i>N</i> = 143,126)			After matched on variables in the publication variables PS model ( <i>N</i> = 98,048)			After matched on variables in the large-scale PS model ( <i>N</i> = 66,070)		
	Paracetamol, %	Ibuprofen, %	Std diff <sup>a</sup>	Paracetamol, %	Ibuprofen, %	Std diff <sup>a</sup>	Paracetamol, %	Ibuprofen, %	Std diff <sup>a</sup>
Non-ibuprofen NSAIDs	29.8	20.0	0.228	26.0	26.9	-0.021	23.7	24.5	-0.018
Other analgesics	39.9	20.8	0.424	28.0	28.2	-0.005	27.8	29.7	-0.042
Anxiolytics	8.2	6.1	0.083	7.3	7.3	0.000	6.8	7.1	-0.014
Hypnotics and sedatives	9.8	6.9	0.105	8.1	8.4	-0.011	8.1	8.4	-0.010
Statins	32.5	19.5	0.300	24.2	24.6	-0.010	24.4	26.0	-0.038
Vitamin K antagonists	5.4	1.5	0.217	2.6	2.0	0.037	2.4	2.5	-0.010

PS propensity score, *Std diff* standardized difference, *NSAIDs* non-steroidal anti-inflammatory drugs

<sup>a</sup> The difference in prevalence of the characteristics in the two cohorts divided by the standard deviation

For the second propensity score model, a much larger set of baseline covariates were defined, which we will refer to as the ‘full set of covariates available’ or ‘large-scale propensity score covariates’:

Age in 5-year increments  
Sex  
Race  
Year of index date  
Month of index date  
Charlson Comorbidity Index  
Diabetes Complications Severity Index (DCSI) score  
CHADS2 score

#### Conditions<sup>1</sup>

Presence/absence of a condition in a 365-day window prior to or on the index date  
Presence/absence of a condition in a 30-day window prior to or on the index date  
Presence/absence of a condition diagnosed in an inpatient setting in a 180-day window prior to or on the index date  
Presence/absence of an aggregation of episodes of care over time for a condition (‘condition era’) any time prior to or on the index date  
Presence/absence of a condition era that overlaps the index date  
Condition classes (based on the SNOMED hierarchy of conditions)  
Presence/absence of a drug in a 365-day window prior to or on the index date

<sup>1</sup> Examples of conditions include essential hypertension, dyspnea and osteoarthritis.

Presence/absence of a drug in a 30-day window prior to or on the index date

#### Drugs<sup>2</sup>

Presence/absence of an inference of length of time of exposure to a drug product (‘drug era’) for all in a 365-day window prior to or on the index date  
Presence/absence of a drug era in a 30-day window prior to or on the index date  
Presence/absence of a drug era that overlaps the index date  
Presence/absence of a drug era that overlaps the index date  
Presence/absence of a drug era any time prior to or on the index date  
Drug classes based on the Anatomical Therapeutic Chemical hierarchy

#### Procedures, observations and measurements<sup>3</sup>

Presence/absence of a procedure in a 365-day window prior to or on the index date  
Presence/absence of a procedure in a 30-day window prior to or on the index date  
Procedures classes (based on the SNOMED hierarchy of procedures)  
Presence/absence of an observation in a 365-day window prior to or on the index date  
Presence/absence of an observation in a 30-day window prior to or on the index date

<sup>2</sup> Examples of drugs include analgesics, antithrombotic agents, and codeine.

<sup>3</sup> Examples of a procedure, measurement and observation include heart valve replacement, creatinine measurement and CHADS2 stroke risk score, respectively.

Count of each observation concept in a 365-day window prior to or on the index date

Presence/absence of a measurement in a 365-day window prior to or on the index date

Presence/absence of a measurement in a 30-day window prior to or on the index date

Count of each measurement concept in a 365-day window prior to or on the index date

Presence/absence of a measurement with a numeric value below the normal range for the latest value within 180 days of the cohort index

Presence/absence of a measurement with a numeric value above the normal range for the latest value within 180 days of the cohort index

Counts of the number of concepts that a person has within each domain (i.e. condition, drug, procedure, clinical observation, laboratory measurement)

All propensity models were fitted using regularized regression with an  $L_1$  (LASSO) prior [11]. The optimal regularization hyperparameters were estimated using ten-fold cross-validation. Based on the work of Walker et al. [12], we estimated preference scores, which are propensity scores normalized for the prevalence of the drugs being compared. The state of clinical equipoise between different treatments exists when the treatments are considered equivalent and there is no clinical reason for choosing one treatment over another. Walker et al. [12] suggested the use of a minimum threshold of 50% of subjects from each population with preference scores between 0.3 and 0.7 to provide evidence of clinical equipoise. We compared the preference scores for the probability of exposure to a first prescription for any paracetamol versus ibuprofen alone using these guidelines.

We examined the standardized mean differences (SMDs) for both the publication covariates and the full set of available covariates between paracetamol and ibuprofen in the two propensity score models. SMDs are a measure of covariate balance between the two treatment groups and are the difference in prevalence in each cohort divided by the standard deviation of the difference. A large absolute value SMD on a covariate is an indication of a significant disparity in the proportion of subjects with the covariate between the two groups. An SMD > 0.1 has been used as an ad hoc heuristic for what constitutes ‘large’ [13].

## 2.5 Negative Control Outcomes Analyses for Residual Confounding

Negative controls are outcomes determined a priori to have no association with the exposure of interest [14]. In particular, there must be no mention of these outcomes in the US FDA-structured product drug label for either

paracetamol or ibuprofen, a manual review of the literature must find no studies showing the drug causing the outcome, and the drugs cannot be listed as a ‘causative agent’ for the outcome according to data provided in *Drug-Induced Diseases: Prevention, Detection and Management* [15]. The resulting list contained 31 outcomes. Schuemie et al. [16] determined that a sample size of at least 25 negative controls is sufficient for constructing an empirical null distribution. Any measurement error in the negative controls is intended to shape, at least in part, the empirical distribution.

Models with these negative controls as outcome variables (outcome models), which adequately control for systematic bias, should produce relative risk estimates close to the null value of 1.0. Negative control models allow for the examination of the extent of bias in the database, study design and analysis, to the degree that they produce significant odds ratios (ORs) different from 1.0. We fit a series of logistic regression models to examine the association between treatment with paracetamol or ibuprofen for each of the 31 negative control incident outcomes within 1 year after the index date for treatment. Review of all available time prior to the outcome was the constraint in determining whether an event was incident or not. The negative controls chosen for this study were:

1. Melanocytic nevus of skin
2. Impotence
3. Lipoma
4. Labyrinthitis
5. Panic attack
6. Acute stress disorder
7. Chronic sinusitis
8. Seborrheic keratosis
9. Infected ulcer of skin
10. Bronchopneumonia
11. Restless legs
12. Perforation of tympanic membrane
13. Type 1 diabetes mellitus
14. Hemangioma
15. Lichen planus
16. Ocular hypertension
17. Bronchiectasis
18. Hyperthyroidism
19. Lymphedema
20. Obsessive-compulsive disorder
21. Vitiligo
22. Strabismus
23. Open-angle glaucoma
24. Biliary calculus
25. Lactose intolerance
26. Ptosis of eyelid
27. Urethritis

28. Keratoacanthoma
29. Vasomotor rhinitis
30. Granuloma annulare
31. Phobic disorder

For each of the negative control outcomes, we fit six models:

1. *Demographic Adjustment in Outcome Model*: Performed a multivariable logistic regression to estimate the effect of exposure (any paracetamol versus ibuprofen) on the incidence of each outcome, adjusting for sex and age in 5-year increments. No matching or restriction (via propensity score or otherwise) was performed prior to final outcome model fitting.
2. *Publication Variable Outcome Model Adjustment*: Performed a multivariable logistic regression to estimate the effect of exposure (any paracetamol versus ibuprofen) on the incidence of each outcome, adjusting for sex, age in 5-year increments, and the set of 38 publication covariates. No matching or restriction (via propensity score or otherwise) was performed prior to final outcome model fitting.
3. *Large-Scale Adjusted Outcome Model* Performed a regularized logistic regression using  $L_1$  (LASSO) shrinkage [11] to estimate the effect of exposure (any paracetamol versus ibuprofen) on the incidence of each outcome, adjusting for the full set of data covariates available. No shrinkage was applied to the effect of exposure, and no matching or restriction (via propensity score or otherwise) was performed prior to final outcome model fitting.
4. *Publication Variable Propensity Score Adjustment* Performed a multivariable logistic regression to estimate a propensity score that predicts treatment assignment (any paracetamol versus ibuprofen) using baseline covariates for sex, age in 5-year increments, and publication variables. One-on-one matching on the propensity score was performed using a standardized caliper of  $0.25 \times$  propensity score standard deviation. The matched sets were used within a univariate conditional logistic regression, which estimated the effect of exposure (any paracetamol versus ibuprofen) on the incidence of each outcome, without further adjustment.
5. *Large-Scale Propensity Score Adjustment* Performed a regularized logistic regression using  $L_1$  shrinkage [11] to estimate a propensity score that predicts treatment assignment (any paracetamol versus ibuprofen) using the full set of data covariates available. The propensity score was used to perform 1:1 matching (using a standardized caliper of  $0.25 \times$  propensity score standard deviation). The matched sets were used within a

univariate conditional logistic regression, which estimated the effect of exposure (any paracetamol versus ibuprofen) on the incidence of each outcome, without further adjustment.

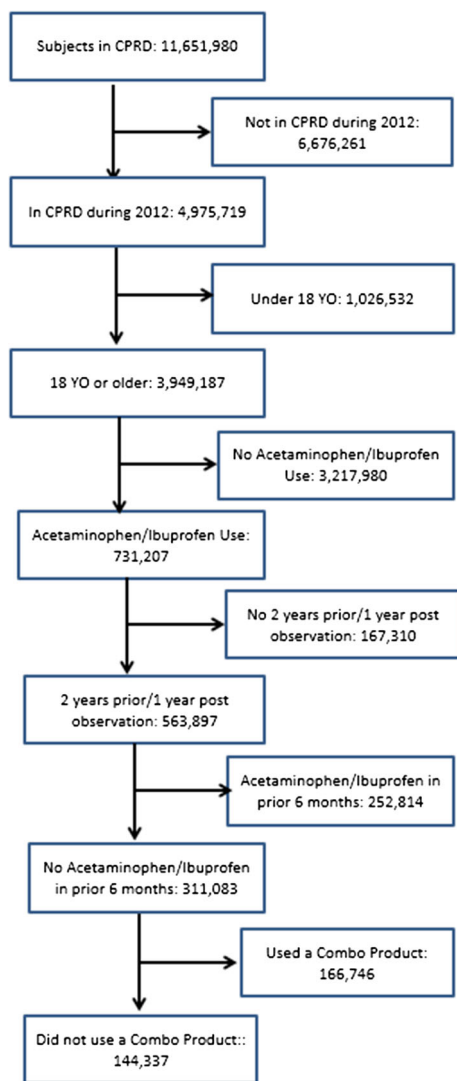
6. *Large-Scale Propensity Score Adjustment with Large-Scale Outcome Modeling* Performed a regularized logistic regression using  $L_1$  shrinkage which estimated a propensity score that predicts treatment assignment (any paracetamol versus ibuprofen) using the full set of data covariates. The propensity score was used to perform 1:1 matching (using a standardized caliper of  $0.25 \times$  propensity score standard deviation). The matched sets were then used to perform a regularized conditional logistic regression using  $L_1$  shrinkage to estimate the effect of exposure (any paracetamol versus ibuprofen) on the incidence of each outcome, adjusted for the full set of data covariates available. No shrinkage was applied to the effect of exposure.

We analyzed each of the negative controls to determine the level of bias remaining in the six different models. If there was no residual bias in a particular model, we would expect no more than one or two risk ratios to be significantly greater or less than unity based on a  $p$  value of  $<0.05$  (i.e. 5% of  $31 \approx 1.5$ , the number of significant parameter estimates found by 5% chance alone).

### 3 Results

The number of participants in the CPRD dataset eligible for inclusion in the study was 3,949,187 (Fig. 1), of whom 144,337 received prescriptions for paracetamol or ibuprofen (Fig. 1). Prior to any matching, more females received either prescription than males, and those receiving ibuprofen prescriptions were somewhat younger than those prescribed paracetamol (Table 1). The paracetamol cohort had higher rates of cardiovascular and musculoskeletal morbidity. Smoking and obesity were also more prevalent in the paracetamol group (Table 1).

Regarding evidence of channeling, approximately 13% of patients who received a prescription for any paracetamol had a prior condition of GI bleed, MI, renal disease and/or stroke in the 2 years prior to the prescription (Table 2). In those patients prescribed ibuprofen, approximately 6% had a prior condition of GI bleed, MI, renal disease and/or stroke in the 2 years prior to the prescription. The ORs comparing prior conditions in those prescribed any paracetamol versus ibuprofen were all significantly greater than unity for all of the conditions, both in the unadjusted models and the models adjusted for sex and age (Table 2). For those patients prescribed any paracetamol, the odds of



**Fig. 1** Flow of patients from Clinical Practice Research Datalink to analytic study population. *CPRD* Clinical Practice Research Datalink, *yo* years old, *combo* combination

prior MI were approximately 2.6 times the odds of prior MI in those prescribed ibuprofen, after adjusting for sex and age (adjusted OR 2.6, 95% confidence interval [CI] 2.3–3.0). Similarly, the OR of a GI bleed prior to paracetamol prescription versus an ibuprofen prescription was 1.4 (95% CI 1.3–1.5), OR of renal disease was 1.8 (95% CI 1.7–1.9), and OR of a stroke was 2.1 (95% CI 1.9–2.3). For MI, renal disease and stroke, the adjustment for sex and age moved the OR toward unity, reflecting the preferential prescription of paracetamol in patients over 70 years of age (Table 1). If the assumption of clinical equipoise were valid, the ORs for each of these conditions would not differ from unity.

Table 2 also presents the corresponding information for paracetamol only versus ibuprofen only, omitting those patients who were prescribed both ibuprofen and

paracetamol on their index date. The impact of excluding those with both was small but consistent. Both the crude and age- and sex-adjusted ORs were higher without the concomitant group than including it with those receiving paracetamol.

### 3.1 Propensity Score Models

We developed two propensity score models to assess the differences between using the covariates described in prior literature (the ‘publication covariates’) versus the full set of data covariates available. The area under the receiver operating characteristics curve (AUC) for the propensity scores based on ‘publication covariates’ was 0.71, while the AUC for the full set of data covariates available was 0.85. The full set of covariates provided better discrimination between patients receiving paracetamol and those receiving ibuprofen compared with the publication covariates. The larger AUC found using the full set of covariates is an indication that discrimination between subjects receiving prescriptions for paracetamol and ibuprofen is achievable and is further evidence that the two treatments do not follow the assumption of clinical equipoise.

The distributions of the preference scores for paracetamol and ibuprofen utilizing the publication covariates are shown in Fig. 2a. The overlap in the preference scores in the 0.3–0.7 range, as described by Walker et al. [12], was >50% for the publication covariates (Table 3). Using this set of covariates would fulfill the criteria for necessary overlap to assume clinical equipoise for the two treatment options. However, when utilizing the full set of covariates (Fig. 2b), the degree of overlap was below 50% and is evidence against the assumption of clinical equipoise for paracetamol and ibuprofen prescriptions. While matching by preference scores using the published covariates reduced the number of eligible matches by approximately 32% (See Supplement for sample attrition table), matching on the full set of available covariates reduced the eligible matches by over 50%.

The change in covariate balance for the publication covariates between paracetamol and ibuprofen following propensity score matching using publication covariates for a representative negative outcome—melanocytic nevus of skin—is shown in Fig. 3a. Each point represents the (absolute value) SMD for a single covariate prior to matching (on the horizontal axis) and after matching (on the vertical axis). Prior to matching, the SMDs ranged from –0.15 to 0.40 in the publication covariates, indicating appreciable differences in the distributions of these covariates between the paracetamol and ibuprofen populations. Following matching, among the matched cohorts, the SMDs ranged from –0.02 to 0.01, indicating more homogenous populations (Fig. 3a).



**Table 2** Unadjusted and adjusted OR of prior conditions for patients prescribed any paracetamol versus ibuprofen alone and paracetamol only versus ibuprofen only ( $N = 144,337$ )

Condition	Any paracetamol <sup>a</sup> [% (N)]	Ibuprofen [% (N)]	Unadjusted OR (95% CI)	Adjusted OR <sup>b</sup> (95% CI)
GI bleed	3.1 (2372)	2.2 (1480)	1.5 (1.4–1.6)	1.4 (1.3–1.5)
MI	1.6 (1175)	0.4 (288)	3.7 (3.3–4.2)	2.6 (2.3–3.0)
Renal disease	7.4 (5643)	2.8 (1887)	2.8 (2.7–3.0)	1.8 (1.7–1.9)
Stroke	2.5 (1874)	0.7 (494)	3.5 (3.2–3.8)	2.1 (1.9–2.3)
	Paracetamol only	Ibuprofen only	Unadjusted OR (95% CI)	Adjusted OR <sup>b</sup> (95% CI)
GI bleed	3.3 (2201)	2.2 (1480)	1.5 (1.4–1.6)	1.5 (1.4–1.6)
MI	1.7 (1149)	0.4 (288)	4.1 (3.6–4.7)	2.9 (2.5–3.3)
Renal disease	8.0 (5382)	2.8 (1887)	3.1 (2.9–3.2)	1.9 (1.8–2.1)
Stroke	2.7 (1791)	0.7 (494)	3.8 (3.4–4.2)	2.3 (2.0–2.5)

OR odds ratio, CI confidence interval, GI gastrointestinal, MI myocardial infarction

<sup>a</sup> Any paracetamol includes patients who received both paracetamol and ibuprofen on the index date

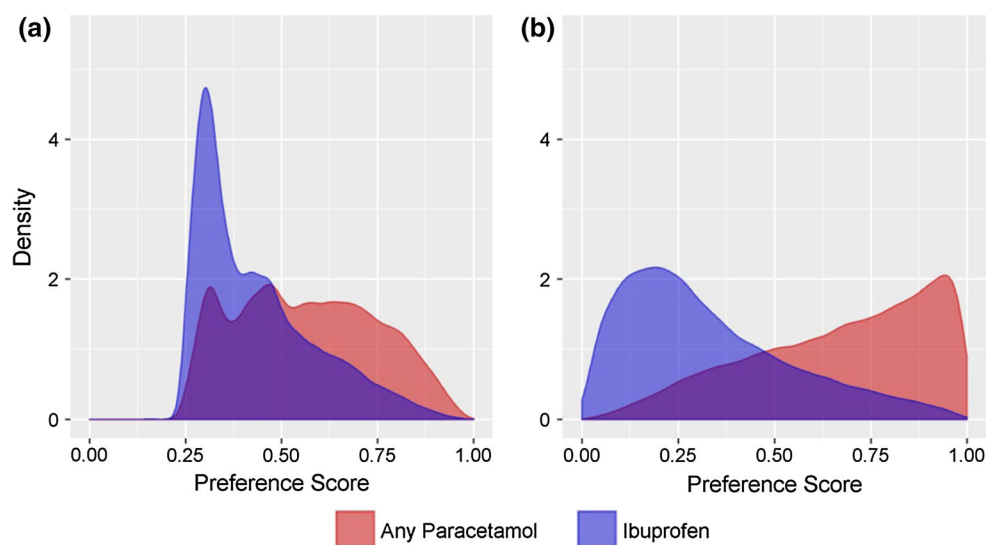
<sup>b</sup> Adjusted by age group and sex

The purpose of propensity score matching is to achieve a balanced sample on all sample characteristics. Figure 3b shows the covariate balance before and after matching on the publication variables propensity score model, but showing the covariate balance among the full set of available covariates. While all SMDs for the publication covariates were reduced to below 2% after matching by the publication covariates, there were a large number (nearly 200) of other covariates for which the SMD between paracetamol and ibuprofen subjects remained  $>0.10$  after matching. This indicates that there were still many covariates that were unbalanced between the two treatment groups, and brings into question the assumption of clinical equipoise for the treatments. Figure 3c, in

which subjects were matched by propensity scores developed from the full set of covariates available, demonstrates that the full set of covariates achieved adequate balance after matching.

### 3.2 Residual Confounding in Negative Controls

As shown in Table 4, in our examination of residual bias in the six models for our negative controls, we found that only those models matched on all available covariates (models 5 and 6) removed confounding bias sufficiently to ensure that any significant finding would likely be due to chance as opposed to being due to residual bias. All models without matching performed poorly,



**Fig. 2** Distribution of propensity scores from **a** publication covariates and **b** the full set of covariates for any paracetamol compared with ibuprofen

**Table 3** Propensity score models distribution between 0.3 and 0.7

Analgesic	All scores <i>N</i>	$0.3 \leq \text{preference} \leq 0.7$	
		<i>N</i>	%
<b>Publication variables model</b>			
Any paracetamol	75,919	51,249	67.5
Ibuprofen alone	68,418	49,728	72.7
<b>Large-scale model</b>			
Any paracetamol	75,919	30,219	39.8
Ibuprofen alone	68,418	26,064	38.1

presumably due to the lack of balance in the clinical characteristics of the sample. In model 5, where we performed univariable logistic regression (with exposure as the single variable) after 1:1 matching by a propensity score developed using all available covariates, no risk ratios in the 31 negative controls differed significantly from unity. By comparison, in model 4, where we performed univariate logistic regression after 1:1 matching on propensity scores developed using the publication covariates, we found seven risk ratios out of 31 negative controls (approximately 23%) that differed significantly from unity. We also found no risk ratios different from unity in model 6, where we matched by propensity score using all available covariates and used a logistic regression model that also adjusted for all available covariates.

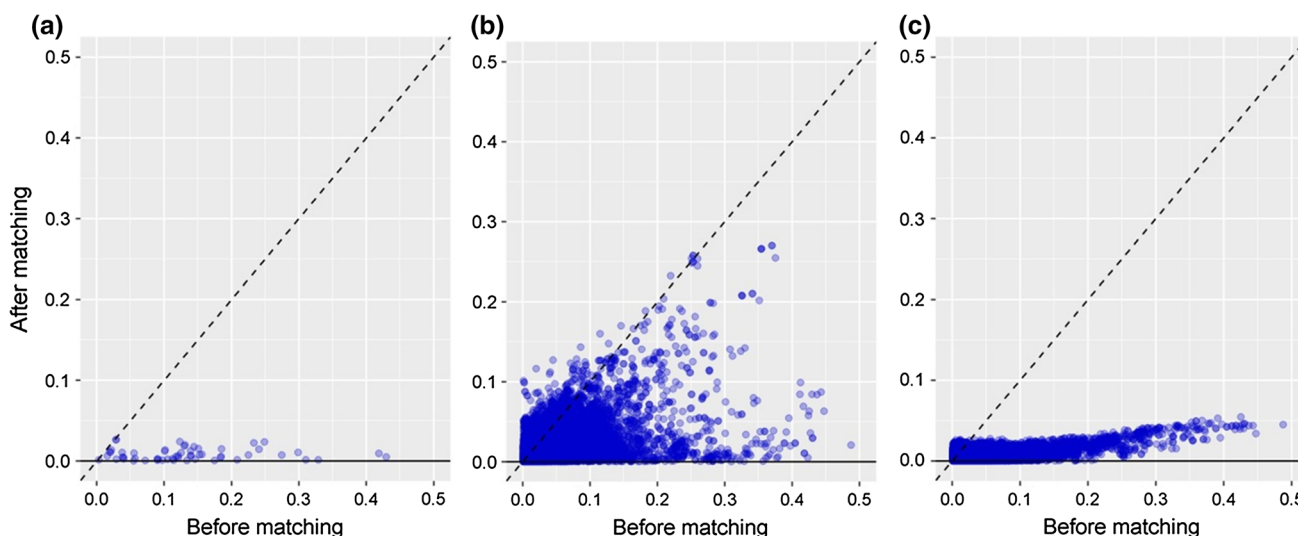
It should be noted that in model 6, using both propensity score matching and a fully adjusted outcome model, we were unable to model 10 of the negative controls. This was

due to the methodological constraint that to fit the full outcome model, we needed to perform regularization. To perform regularization, we needed to know the appropriate hyperparameter (variance of the regularization prior) to use, and for that we used cross-validation. We required a minimum amount of data for the cross-validation (100 ‘informative strata’, which is basically the same as the number of outcome events in the two treatment cohorts taken together), and ten of the negative controls did not meet this criterion.

#### 4 Discussion

This is the first study that we know of to examine evidence directly and quantitatively for channeling in the use of paracetamol and ibuprofen, two widely used analgesics. The purpose of this study was to answer two questions. First, was there evidence of channeling in the prescription of paracetamol versus ibuprofen, and, if so, could we adequately adjust for channeling in studies examining outcomes after exposure?

We found that patients with a prescription of single-ingredient paracetamol were more likely to have a prior history of GI bleeding, MI, renal disease, and stroke compared with ibuprofen after controlling for age and sex. Given this evidence of preferential prescribing, we explored whether various strategies for confounding adjustment in observational studies were adequate to allow for unbiased assessments of outcomes from these exposures. We used propensity score adjustment as it is



**Fig. 3** Absolute value standardized difference of the mean of the **a** publication covariates and **b** full set of data covariates available prior to and after matching on publication variable propensity scores,

and **c** full set of data covariates available prior to matching and after matching on propensity scores using the full set of data covariates available for melanocytic nevus of skin

**Table 4** Negative control model point estimates and confidence intervals

	Model 1 Demographic adjustment	Model 2 Publication variables adjustment	Model 3 Large-scale adjustment	Model 4 Publication variables propensity score matched	Model 5 Large-scale propensity score matched	Model 6 Large-scale propensity score matched and adjustment
No. of negative controls significant	7	3	7	7	0	0
Negative association	3	1	3	5	0	0
Positive association	4	2	4	2	0	0
No. not significant	24	28	24	24	31	21
Fraction significant, %	22.6	9.7	22.6	22.6	0.0	0.0
Outcome	RR (95% CI)	RR (95% CI)	RR (95% CI)	RR (95% CI)	RR (95% CI)	RR (95% CI)
Acute stress disorder	1.13 (0.8–1.61)	0.98 (0.52–2.10)	1 (0.71–1.41)	0.88 (0.60–1.28)	1.13 (0.73–1.74)	0.96 (0.54–1.68)
Biliary calculus	1.34 (0.8–2.58)	1.19 (0.71–2.35)	1.14 (0.97–1.34)	<b>1.20 (1.01–1.42)</b>	0.95 (0.77–1.17)	0.95 (0.74–1.21)
Bronchiectasis	1.31 (0.9–2.02)	1.09 (0.78–1.56)	1.18 (0.88–1.61)	1.21 (0.86–1.70)	1.1 (0.75–1.6)	0.75 (0.27–2.18)
Bronchopneumonia	<b>2.71 (1.41–5.61)</b>	<b>2.71 (1.41–5.61)</b>	<b>2.25 (1.15–4.76)</b>	1.90 (0.90–4.25)	0.75 (0.25–2.16)	NA
Chronic sinusitis	0.96 (0.55–1.68)	0.98 (0.65–1.64)	1.1 (0.87–1.39)	1.07 (0.82–1.39)	1.2 (0.89–1.62)	1.42 (0.95–2.16)
Granuloma annulare	1.65 (0.63–4.8)	1.65 (0.63–4.80)	0.93 (0.72–1.2)	1.40 (0.45–4.73)	1.26 (0.88–1.81)	1.19 (0.73–1.93)
Hemangioma	0.65 (0.42–2.22)	0.63 (0.34–2.82)	0.96 (0.69–1.33)	1.00 (0.69–1.45)	0.94 (0.58–1.53)	0.97 (0.59–1.61)
Hypothyroidism	1.06 (0.58–1.82)	0.93 (0.45–2.17)	0.96 (0.72–1.3)	0.88 (0.62–1.25)	0.82 (0.54–1.24)	0.74 (0.43–1.23)
Impotence	1.07 (0.91–1.26)	0.89 (0.79–1.01)	0.93 (0.52–1.83)	0.89 (0.79–1.00)	1.09 (0.94–1.27)	1.06 (0.78–1.43)
Infected ulcer of skin	<b>5.41 (1.47–34.7)</b>	<b>5.41 (1.47–34.74)</b>	<b>5.41 (1.47–34.74)</b>	4.00 (1.00–26.50)	4.00 (0.59–78.2)	NA <sup>a</sup>
Keratoacanthoma	1.42 (0.73–3.02)	1.39 (0.60–4.74)	<b>2.04 (1.11–3.96)</b>	1.25 (0.59–2.72)	1.78 (0.8–4.2)	NA <sup>a</sup>
Labyrinthitis	0.72 (0.43–1.58)	0.67 (0.31–1.65)	<b>0.74 (0.61–0.89)</b>	<b>0.70 (0.57–0.85)</b>	0.82 (0.64–1.05)	0.8 (0.61–1.05)
Lactose intolerance	1.08 (0.33–3.75)	1.08 (0.33–3.75)	1.08 (0.33–3.75)	1.33 (0.29–6.77)	2.00 (0.53–9.48)	NA <sup>a</sup>
Lichen planus	0.85 (0.58–1.28)	0.76 (0.34–1.80)	0.82 (0.56–1.19)	0.83 (0.56–1.24)	1.13 (0.7–1.84)	1.34 (0.63–2.76)
Lipoma	<b>0.76 (0.59–0.99)</b>	0.72 (0.42–1.19)	<b>0.75 (0.61–0.93)</b>	<b>0.74 (0.59–0.92)</b>	0.81 (0.62–1.06)	0.86 (0.65–1.15)
Lymphedema	<b>1.93 (1.4–2.82)</b>	1.57 (0.76–3.23)	<b>1.47 (1.05–2.1)</b>	<b>1.48 (1.02–2.16)</b>	1.33 (0.85–2.11)	1.78 (0.76–4.45)
Melanocytic nevus of skin	<b>0.82 (0.73–0.92)</b>	0.76 (0.49–1.18)	<b>0.87 (0.76–0.98)</b>	<b>0.79 (0.69–0.90)</b>	0.98 (0.83–1.15)	0.97 (0.81–1.16)
Obsessive-compulsive disorder	0.79 (0.47–1.31)	0.80 (0.27–2.79)	0.92 (0.54–1.57)	1.05 (0.57–1.92)	1.00 (0.5–2.01)	NA <sup>a</sup>
Ocular hypertension	0.88 (0.55–1.45)	0.86 (0.57–1.33)	0.86 (0.61–1.2)	0.94 (0.64–1.39)	0.73 (0.44–1.19)	0.71 (0.36–1.36)
Open-angle glaucoma	0.81 (0.56–1.17)	0.81 (0.45–1.58)	0.92 (0.65–1.3)	0.85 (0.57–1.26)	0.87 (0.54–1.38)	0.6 (0.01–43.54)
Panic attack	1.01 (0.84–1.24)	0.84 (0.65–1.11)	0.81 (0.31–2.33)	0.84 (0.69–1.03)	0.81 (0.64–1.04)	0.83 (0.53–1.31)
Perforation of tympanic membrane	0.68 (0.35–2.67)	0.75 (0.27–2.97)	0.9 (0.61–1.32)	0.75 (0.50–1.13)	1.13 (0.7–1.82)	1.7 (0.66–4.55)
Phobic disorder	1.01 (0.79–1.32)	0.85 (0.42–1.90)	0.93 (0.72–1.2)	0.94 (0.71–1.24)	1.08 (0.77–1.53)	1.05 (0.6–1.87)
Ptosis of eyelid	1.18 (0.66–2.16)	0.94 (0.46–4.02)	1.15 (0.65–2.09)	0.89 (0.46–1.72)	1.64 (0.78–3.58)	NA <sup>a</sup>
Restless legs	1.34 (1–1.92)	1.11 (0.51–3.01)	1.18 (0.92–1.53)	1.03 (0.78–1.37)	1.13 (0.8–1.58)	0.94 (0.54–1.6)
Seborrheic keratosis	<b>0.83 (0.73–0.96)</b>	<b>0.84 (0.76–0.95)</b>	1.02 (0.67–1.71)	<b>0.87 (0.79–0.97)</b>	1.04 (0.91–1.18)	1.03 (0.88–1.2)

Table 4 continued

Outcome	RR (95% CI)	RR (95% CI)	RR (95% CI)	RR (95% CI)	RR (95% CI)	RR (95% CI)
Strabismus	0.99 (0.55–1.8)	0.99 (0.55–1.80)	0.99 (0.55–1.8)	0.80 (0.37–1.71)	0.73 (0.28–1.8)	NA <sup>a</sup>
Type 1 diabetes mellitus	<b>1.46 (1.1–1.95)</b>	0.98 (0.64–1.60)	1.01 (0.73–1.4)	0.95 (0.67–1.35)	0.76 (0.5–1.16)	0.84 (0.05–10.69)
Urethritis	0.53 (0.2–4.19)	0.90 (0.39–2.10)	0.9 (0.39–2.1)	0.80 (0.31–2.03)	0.6 (0.12–2.45)	NA <sup>a</sup>
Vasomotor rhinitis	0.54 (0.18–1.95)	0.64 (0.32–1.25)	0.68 (0.34–1.34)	<b>0.24 (0.07–0.64)</b>	0.67 (0.22–1.85)	NA <sup>a</sup>
Vitiligo	0.82 (0.43–1.52)	0.82 (0.43–1.52)	0.83 (0.43–1.6)	1.00 (0.49–2.06)	1.00 (0.41–2.44)	NA <sup>a</sup>

Bolded values indicate RR significantly different from 1.0

RR relative risk, CI confidence interval, NA not available

<sup>a</sup> Outcome model was attempted but had insufficient counts to make an estimate

designed to control for systematic differences in patient populations between those receiving a treatment and a comparator.

When we constructed a propensity score model using baseline characteristics previously used in prior published studies, we found that propensity score matching did adequately balance these covariates but failed to balance other observable covariates. This suggests that current publications may still be subject to channeling bias that could influence reported effect estimates. Moreover, while the publication-inspired propensity score distributions may suggest that the two populations may be sufficiently near clinical equipoise, we showed that a more complete confounding adjustment would result in much greater discrimination between the two populations that calls clinical equipoise into question. When we applied a large-scale regularized model to estimate the propensity score using all available covariates, we observed that matching achieved consistently greater balance across all covariates, inclusive of the baseline characteristics used in prior publications. These findings suggest that the magnitude and complexity of channeling bias is substantial, and while it may not be adequately mitigated with traditional approaches, could be better managed through more complete multivariable models combined with matching.

Negative control outcomes were selected a priori, meeting criteria for having no known association with exposure to paracetamol or ibuprofen. The distribution of relative risk estimates from negative control outcomes provides a measure of the extent of residual bias after controlling for possible confounders since we know these should produce relative risk estimates near unity. We selected 31 negative control outcomes to empirically evaluate the expectation that (under the null hypothesis), 5% of the relative risks would be statistically significantly different from 1.0 (at  $\alpha = 0.05$ ), by chance alone. When we only used publication variables in the propensity score model, we observed there was substantial residual bias, resulting in the model failing to have nominal operating characteristics; namely, 23% (7 of 31) of the negative control outcomes yielded significant findings, with five of the null outcomes showing significant decreased risk associated with paracetamol, and two outcomes showing an increased risk. For many of the negative controls, paracetamol looks numerically protective, indicating that we cannot make assumptions about direction of the bias when we do not adjust properly. These results call into question whether statistical significance under this adjustment strategy can be confidently inferred without calibration [16]. In contrast, when we fit the propensity score model using a larger set of covariates, all significant associations observed using literature covariates were

attenuated; no negative control outcomes produced significant findings after full propensity score matching. The negative control results corroborate the covariate balance findings, which showed that modeling limited to publication covariates left residual bias, but more complete adjustment with matching appears to be effective at mitigating the threats of channeling, although at the cost of reducing the degree of overlap in propensity score distributions between the cohorts.

This assessment of channeling bias has important limitations that require consideration. The bias evaluation using negative controls requires an exchangeability assumption that may not hold for future unknown outcomes of interest. Specifically, we assume the magnitude (but not specific source) of residual bias for the unknown outcome could have been sampled from the distribution of residual bias that was observed for the negative controls. As such, it is possible there is additional residual bias that we did not adequately observe based on the measured covariate balance and empirical null distribution diagnostics. This analysis was restricted to prescription use of paracetamol and ibuprofen, and it is unknown whether these results would generalize to non-prescription exposures. There are several reasons for a GP to prescribe these medications in the CPRD, including record keeping and giving the patient access to the medication at a lower cost because the patient qualifies for free filling of prescriptions. In addition, those using these medications chronically may need larger quantities than typically available over the counter. Thus, it is likely that, by relying on prescriptions, we skewed our study population toward elderly subjects with chronic conditions who may also be at the low end of the economic spectrum. While being selected into the study population limits the generalizability of the results, it does not limit their validity, i.e. we expect that the selection mechanism should be the same for both drug cohorts.

## 5 Conclusions

While it is widely understood that there are substantial differences in the baseline characteristics of patients treated with acetaminophen versus ibuprofen, the adjustment for these differences must be confirmed with comparisons before and after matching or with appropriate diagnostics if other methods are used. We were not able to adequately remove selection bias using a selected set of covariates for propensity score adjustment. Large-scale propensity score matching offers a novel approach that should be considered when attempting to mitigate the risk of channeling bias.

## Compliance with Ethical Standards

**Funding** No sources of funding were used to conduct this study or prepare this manuscript. Janssen Research and Development is the sponsor of Open Access.

**Conflicts of interest** Jesse A. Berlin, Daniel Fife, Amy Matcho, Kayur Patel, Martijn Schuemie, Joel Swerdel, Patrick Ryan, and Rachel Weinstein were full-time employees of Johnson & Johnson, or a subsidiary, at the time the study was conducted. Johnson & Johnson manufactures several medications that contain paracetamol or ibuprofen. Jesse A. Berlin, Daniel Fife, Amy Matcho, Kayur Patel, Martijn Schuemie, Joel Swerdel, Patrick Ryan, and Rachel Weinstein own stock, stock options and pension rights from the company.

**Ethical approval** This study is based on data from the CPRD obtained under license from the UK Medicines and Healthcare products Regulatory Agency. The protocol for this study is available at ClinicalTrials.gov (NCT identifier: NCT02830178) and was approved by the data provider's ISAC (protocol 15\_173R).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Kurth T, Glynn RJ, Walker AM, et al. Analgesic use and change in kidney function in apparently healthy men. *Am J Kidney Dis.* 2003;42:234–44.
2. Curhan GC, Willett WC, Rosner B, et al. Frequency of analgesic use and risk of hypertension in younger women. *Arch Intern Med.* 2002;162:2204–8.
3. Dedier J, Stampfer M, Hankinson S, et al. Nonnarcotic analgesic use and the risk of hypertension in US women. *Hypertension.* 2002;40:604–8.
4. Chan AT, Manson JE, Albert CM, et al. Nonsteroidal anti-inflammatory drugs, paracetamol, and the risk of cardiovascular events. *Circulation.* 2006;113:1578–87.
5. Evans M, Fore D, Belloc R, et al. Acetaminophen, aspirin and progression of advanced chronic kidney disease. *Nephrol Dial Transplant.* 2009;24:1908–18.
6. Lipworth L, Friis S, Mellekjaer L, et al. A population-based cohort study of mortality among adults prescribed paracetamol in Denmark. *J Clin Epidemiol.* 2003;56:796–801.
7. de Vries F, Setakis E, van Staa TP, et al. Concomitant use of ibuprofen and paracetamol and the risk of major clinical safety outcomes. *Br J Clin Pharmacol.* 2010;70:429–38.
8. Roberts E, Delgado Nunes V, Buckner S, Latchem S, Constanti M, Miller P, Doherty M, Zhang W, Birrell F, Porcheret M, Dziedzic K, Bernstein I, Wise E, Conaghan PG. Paracetamol: not as safe as we thought? A systematic literature review of observational studies. *Ann Rheum Dis.* 2016;75(3):552–9.
9. Herret E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol.* 2015;44(3):827–36.
10. Matcho A, Ryan P, Fife D, Reich C. Fidelity assessment of a Clinical Practice Research Datalink conversion to the OMOP common data model. *Drug Saf.* 2014;37:945–59.

11. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc B*. 1996;58(1):267–88.
12. Walker AM, Patrick AR, Lauer MS, Hornbrook MC, Marin MG, Platt R, Roger VL, Stang P, Schneeweiss S. A tool for assessing the feasibility of comparative effectiveness research. *Comp Eff Res*. 2013;2013(3):11–20.
13. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. *J Clin Epidemiol*. 2015;68:627–36.
14. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. *Drug Saf*. 2013;36(Suppl 1):S33–47. doi:[10.1007/s40264-013-0097-8](https://doi.org/10.1007/s40264-013-0097-8).
15. Tisdale JE, Miller DA. Drug-induced diseases: prevention, detection, and management. Bethesda: American Society of Health-System Pharmacists; 2010.
16. Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med*. 2014;33(2):209–18.