Clinical Epidemiology

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Real-World Comparability of Antiresorptive Osteoporosis Treatment Groups Among Treatment-Naïve and Treatment-Experienced Women Ages 55 and Older in the United States

Kathleen E Hurwitz[1], Lydia Feinstein[1], David A Pritchard[1], Peter Samai[1], Ying Yu[1], Min Kim[2], Akhila Balasubramanian[2], Tzu-Chieh Lin[2], Brian D Bradbury[2], M Alan Brookhart[3]

[1]Epidemiology, Target RWE, Durham, NC, USA; [2]Center for Observational Research, Amgen Inc., Thousand Oaks, CA, USA; [3]Department of Population Health Sciences, Duke University, Durham, NC, USA

Correspondence: Kathleen E Hurwitz, Epidemiology, Target RWE, Road Ste 120, Box 377, Durham, NC, 27713, USA, Email KHurwitz@targetrwe.com

Purpose: Selection of osteoporosis (OP) treatment is affected by patients' disease severity and fracture risk, potentially confounding real-world comparative effectiveness and safety studies of antiresorptive medications. To inform the choice of valid treatment contrasts for subsequent real-world comparative studies, we assessed comparability of antiresorptive OP treatment groups using negative control outcomes (NCOs).

Patients and Methods: Women aged ≥55 years in Optum© Clinformatics® Data Mart from October 2010 through June 2019 who received denosumab, zoledronic acid (ZA), or oral bisphosphonates (BPs) were included. We estimated the 1-year cumulative risks for 12 NCOs by treatment group among treatment-naïve and treatment-experienced women using augmented inverse-probability of treatment and censoring weighted (AIPW) estimation. A Bayesian sensitivity analysis was conducted to aggregate estimates and associated variances into a form characterized by magnitude and probability.

Results: Women in both treatment-naïve (n = 199,335) and treatment-experienced (n = 33,296) cohorts initiated treatment at a mean age of 71.8 years. Treatment-naïve women initiating denosumab had similar 1-year risks of most NCOs compared with initiators of ZA (maximum observed RD = 2.47% for colon cancer screening). However, significant risk differences were observed for seven NCOs when comparing ZA or denosumab with oral BPs. Among treatment-experienced women, all NCOs indicated similar risks when comparing denosumab to alendronate alone. Only one NCO (dementia: RD = 0.42%) was associated with treatment when comparing denosumab to oral BPs, and one (influenza vaccine: RD = 3.57%) was associated with treatment when comparing ZA to oral BPs. Results of the Bayesian analysis aligned with our qualitative interpretations.

Conclusion: Comparative studies including denosumab or ZA versus oral BPs among treatment-experienced, commercially insured women aged ≥55 years in the United States are likely valid with respect to comparability of treatment groups. Our results do not support conducting observational studies examining these treatment contrasts in the overall treatment-naïve population. However, comparison of ZA and oral BPs among treatment-naïve women with a prior fracture may be undertaken with minimal expected residual bias.

Plain Language Summary: Osteoporosis is a chronic condition that makes bones weaker and more likely to break, especially in older women. There are several medications available to prevent fractures. These include tablets called oral bisphosphonates and injections like zoledronic acid and denosumab. Doctors choose which treatment to prescribe based on many factors such as the person's health and fracture risk. However, these choices can make it difficult for researchers to fairly compare how well treatments work in real-world settings.

To better understand which treatments can be reliably compared, we looked at insurance claims data from more than 230,000 US women aged 55 years and older who started or switched osteoporosis medications. We compared groups of women taking different treatments by measuring their risk of having unrelated health events—such as getting a flu shot or a check-up—that should not be

https://doi.org/10.2147/CLEP.S562445
Received: 22 August 2025
Accepted: 20 January 2026
Published: 3 February 2026

affected by any osteoporosis drug. If one group experienced more of these events compared to another, it might mean the two groups are different in ways that could bias future research.

We found that women who switched from one osteoporosis medication to another were similar enough to allow for fair comparisons across treatments. However, among women starting treatment for the first time, only comparisons between denosumab and zoledronic acid appeared valid. Comparisons involving oral bisphosphonates showed signs of bias, unless we restricted analyses to women who had recently broken a bone.

These findings will help guide future studies that aim to accurately evaluate the real-world effectiveness and safety of osteoporosis medications.

Keywords: osteoporosis treatment, comparative effectiveness, negative control outcomes, real-world evidence, Bayesian framework

Introduction

Osteoporosis (OP) is a systemic skeletal disorder characterized by low bone mass and compromised bone strength, predisposing individuals, particularly older women, to an increased risk of fracture.[1] Management of OP includes reducing modifiable risk factors through dietary and lifestyles changes (eg, quitting smoking, reducing alcohol consumption, increasing calcium and vitamin D intake, exercise) and treatment with pharmacologic therapy.[2] Pharmacologic treatments available for OP include antiresorptives such as oral bisphosphonates (BPs) (alendronate, risedronate, and ibandronate), intravenous BPs such as zoledronic acid (ZA), and RANKL inhibitors such as denosumab.

Real-world comparative effectiveness studies of antiresorptive medications may be confounded by disease severity, as selection of OP treatment is likely affected by measured and unmeasured factors related to patients' fracture risk. Previous research suggests that treatment-naïve initiators of oral BPs are comparable to each other (eg, alendronate versus risedronate).[3] This would be expected because oral BPs are well established in clinical practice, have the same labeling/indications and route of administration, and are commonly used as first line therapies. However, treatment-naïve initiators of denosumab were not comparable to treatment-naïve initiators of oral BPs,[3] likely due to increased relative ease of use of denosumab compared to oral BPs (ie, single administration every six months vs daily or weekly administration, respectively) and reimbursement constraints that restrict access to denosumab to patients at higher risk of fracture or among patients that have failed on various classes of therapy.

It is commonly accepted that all studies using real-world data have some amount of residual confounding. Prior to conducting a comparative effectiveness study between antiresorptives, evidence of residual confounding using negative control outcomes (NCOs) needs to be evaluated.[4,5] For an NCO to be valid it must: 1) be causally unrelated to the treatments under study; and 2) share common bias structures with the treatments and outcomes of interest.[4,6] A lack of an association between the treatment and NCO provides evidence of comparability of treatment groups, supporting further comparative analyses using relevant disease endpoints and mitigating concerns about residual confounding. NCOs have been used to assess potential bias in several recent studies, including assessments of methylphenidate treatment and seizures,[7] antidiabetics and COVID-19 diagnosis or death,[8] and the effectiveness of mammography screening.[9] The use of NCOs is also being adopted by the FDA for real-world evidence as part of the Prescription Drug User Fee Act VII Commitment.[10]

Using 12 NCOs, we sought to assess the comparability of antiresorptive OP treatment groups among both treatment-naïve and treatment-experienced women in a large, national commercial insurance claims database. Collectively, the findings from this study inform the choice of valid treatment contrasts for subsequent comparative effectiveness and safety studies of OP medications and fracture outcomes.

Materials and Methods

Study Population

The study population included women aged ≥55 years in the Optum© Clinformatics® Data Mart database, a de-identified adjudicated administrative health claims database that includes 15–18 million individuals covered annually by commercial insurance in all 50 US states, including privately insured individuals (<65 years) and those with employer-sponsored

Medicare Advantage plans (≥65 years). The data capture reimbursement records for inpatient and outpatient diagnoses, procedures, and medications.

We used data from October 2010 through June 2019, with patients entering the cohort from February 2012 through December 2018. Women were included if they received denosumab, zoledronic acid, risedronate, alendronate, or oral ibandronate, and had at least 15-months of continuous enrollment preceding treatment initiation (index date). We excluded women with a history of Paget's disease, cancer (excluding non-melanoma skin cancer), cancer treatment (chemotherapy, hormonal treatment, radiation therapy), or dementia, as well as those prescribed more than one study treatment at index. For the analysis of early hip fracture, we additionally excluded women who had a documented hip fracture within 90 days prior to index.

Treatments, Outcomes, and Follow-Up

Exposures of interest included new use of the following four OP treatments: alendronate alone, oral BPs (alendronate, risedronate, and ibandronate), ZA, and denosumab. Alendronate was an individual exposure of interest because it is the most commonly used first line OP treatment among oral BPs. Medication-specific dosing intervals were as follows: oral BPs, 30 days; ZA, 365 days; denosumab, 182 days. Women who received a different study medication within 455 days prior to index were considered treatment-experienced and those with no prior OP study medication use (based on all available patient history) were categorized as treatment-naïve. Teriparatide and abaloparatide were not included because they are used infrequently and romosozumab was not included because of its recent approval. Within the treatment-experienced group, oral BP users were required to switch from an alternate oral BP, while denosumab or ZA initiators could have switched from any other study drug during the 455 days prior to index. Denosumab and ZA were identified using Healthcare Common Procedure Coding System (HCPCS) and National Drug Code (NDC) codes. The HCPCS codes for these drugs were not specific to the indication for postmenopausal osteoporosis; however, we excluded patients with any evidence of cancer (ie, diagnoses, procedures, treatments) to remove other possible indications (eg, bone metastases). Oral BPs were identified using NDCs. The exposure contrasts of interest were denosumab versus ZA, denosumab versus oral BPs, denosumab versus alendronate, and ZA versus oral BPs.

The 12 NCOs were selected a priori by an independent panel of subject matter experts and identified in the data using diagnosis, procedure, and drug codes. The selected conditions had to be identifiable in the data, expected to have no causal association with the OP medications of interest, and were thought to have some association with a domain of confounding, such disease severity, frailty, access to care, and health-seeking behavior.[11] The pre-specified outcomes included: decubitus ulcer, dementia, transfusions, accidents, annual wellness visit, pelvic exams for cancer screening, influenza vaccine use, Mohs surgery, herpes zoster vaccine, colon cancer screening, visual field tests, and early hip fracture. Supplementary Table S1 summarizes the domain(s) of confounding addressed by each of the selected NCOs.

Women were followed from treatment initiation until experiencing the first of the following: NCO of interest, death, treatment discontinuation (defined as ≥30-day gap), disenrollment, diagnosis or treatment related to cancer or Paget's disease, end of risk period (90 days for early hip fracture; 365 days for all other outcomes), or end of study data on December 31, 2019.

Covariates

A total of 53 covariates that may influence the provision of OP treatment or which are known or hypothesized risk factors for the outcome were included to control for confounding. Covariates were measured during the 15-month lookback period (inclusive of the index date) and included: demographics (age, geographic region, calendar year of index), healthcare utilization (number of physician office and emergency room visits and number of hospital visits), and the Charlson comorbidity index score, as well as the individual chronic diseases and medications specified in Supplementary Table S2. Each covariate was identified using a case ascertainment algorithm, provided in Supplementary Table S3, that utilized individual or combinations of inpatient and outpatient ICD-9-CM or ICD-10-CM diagnosis or procedure codes, CPT-4 or HCPCS procedure codes, and/or NDCs.

Statistical Analysis

We estimated the 1-year cumulative risk of each NCO (apart from early hip fracture, which was evaluated at 90 days) using an augmented inverse-probability of treatment and censoring weighted (AIPW) estimation function, with death treated as a competing risk.[12–14] This approach, as described by Ozenne et al (2020), uses models for both treatment/censoring to address confounding in initial treatment assignment and potentially informative censoring by loss to follow-up, augmented with an assumed model for the outcome. The AIPW estimator is doubly robust: if either the treatment/censoring model or the outcome model is properly specified, then the estimator will be consistent. We employed the AIPW estimator to mitigate bias due to model misspecification given that NCO analyses may indicate lack of comparability due to residual confounding or model misspecification. The propensity score for each exposure contrast was estimated using a multinomial logistic regression model that included all baseline covariates and history of the NCOs, except binary covariates with a prevalence of less than 0.1% within any index drug group. Covariate balance for each treatment comparison was assessed in the corresponding inverse-probability of treatment weighted population using the standardized mean difference (SMD). An SMD >0.1 suggested residual imbalance for a given covariate. For treatment contrasts with residual imbalances, we collapsed categories where appropriate. Inverse probability of censoring weights was estimated by modeling the composite risk of censoring (cancer or Paget's disease, treatment discontinuation, disenrollment) for each outcome using a Cox proportional hazards model with age as a predictor. The outcome model was also fit using Cox proportional hazards regression and contained the same covariates as the propensity score model. We note that both Cox regression models evaluated time in intervals according to the observed occurrence of events in the data (as opposed to a prespecified coarsening of time, eg, 30-day intervals). For the subgroup analyses, we used a minimal set of covariates for the outcome model to avoid issues with model convergence because of the smaller sample size. All model specifications are available in Supplementary Table S2.

We conducted subgroup analyses among: 1) patients older than age 75 years; 2) patients with a recent history of fracture (within 12 months prior to index); and 3) patient results stratified by calendar time (before and after switching from ICD-9 to ICD-10 coding system on October 1, 2015). We also conducted a sensitivity analysis that emulated an ITT design, where we did not censor for treatment discontinuation, treatment switches, or for patients starting another treatment while still on their index treatment. We did not adjust for multiple comparisons when reviewing outcome estimates, because the purpose of our study was to identify possible bias rather than detect true associations.[15]

We additionally implemented a Bayesian framework that aggregates multiple NCO estimates and their associated variances into a new form characterized by two summary parameters: magnitude and probability of residual bias. As conclusions from an NCO study are strengthened by evaluating multiple negative controls, the results are typically a collection of estimates and their corresponding variance estimates. However, when these effect measures range in both size and precision, interpreting the totality of results is not straightforward.

A complete description of the Bayesian approach is provided in the Supplemental Materials (Supplementary Methods 1). Briefly, we constructed the posterior predictive distribution of the expected RD for the collection of 12 NCOs (as identified a priori by subject matter experts) and reported the 80%, 90%, and 95% quantiles of the absolute value of the distribution. The modeling approach specified an underlying distribution with modes on either side of the null value of 0, reflecting the possibility that residual bias may manifest itself positively or negatively away from the null. The resulting posterior distributions may then be used to consider whether the magnitude and probability of bias is acceptably low enough to perform a comparative safety or effectiveness study under the same conditions. For example, a Bayesian summary RD close to the null value of 0 (ie, the interval between the 2.5 and 97.5% quantiles of the posterior predictive distribution contains the null) may indicate adequate comparability between treatment groups. It can also be used to contextualize existing comparative results. Posterior predictive distributions were compiled for each combination of exposure contrast of interest, patient treatment history type (ie, naïve vs experienced), and population sub-group and/or sensitivity analysis described above for a total of 66 scenarios.

This study was a retrospective analysis of deidentified, secondary data and determined to be exempt from oversight by an Institutional Review Board (IRB) under 45 CFR 46.101(b)(4) of the US Department of Health and Human Services regulations (Advarra IRB, Columbia, Maryland, waived ethical approval for this work). All statistical analyses were performed using R software, version 3.5.2.

Results

Baseline Characteristics

There were 199,335 eligible treatment-naïve women (67% alendronate alone, 81% oral BPs, 6% ZA, 13% denosumab) and 33,296 treatment-experienced women (26% alendronate alone, 51% oral BPs, 12% ZA, 37% denosumab) (Figure 1). In both cohorts, women initiated treatment at a mean age of 71.8 years (SD = 8.2). The denosumab and ZA groups were generally similar with regard to baseline characteristics but were older and had a higher prevalence of comorbidities compared to oral BP initiators (both naïve and experienced patients) (Table 1). Before weighting, denosumab and ZA initiators had a higher prevalence of prior preventive measures (eg, influenza vaccine, colon screening) than oral BP initiators in both cohorts (Table 2). In addition, accidents were more common in ZA and denosumab treated subjects compared to oral BP treated subjects. After weighting, the measured covariates were balanced for all treatment comparisons (Supplementary Table S4).

Treatment-Naïve Comparisons

Women initiating denosumab had similar 1-year risks of most NCOs compared with initiators of ZA (Figure 2). However, when comparing ZA or denosumab with oral BPs, we observed significant differences on the risk difference scale for 7 out of 12 NCOs, suggesting that residual bias exists between ZA or denosumab and oral BPs. Therefore, in a treatment-naïve population, comparative studies between ZA or denosumab and oral BPs are not justified. For example,

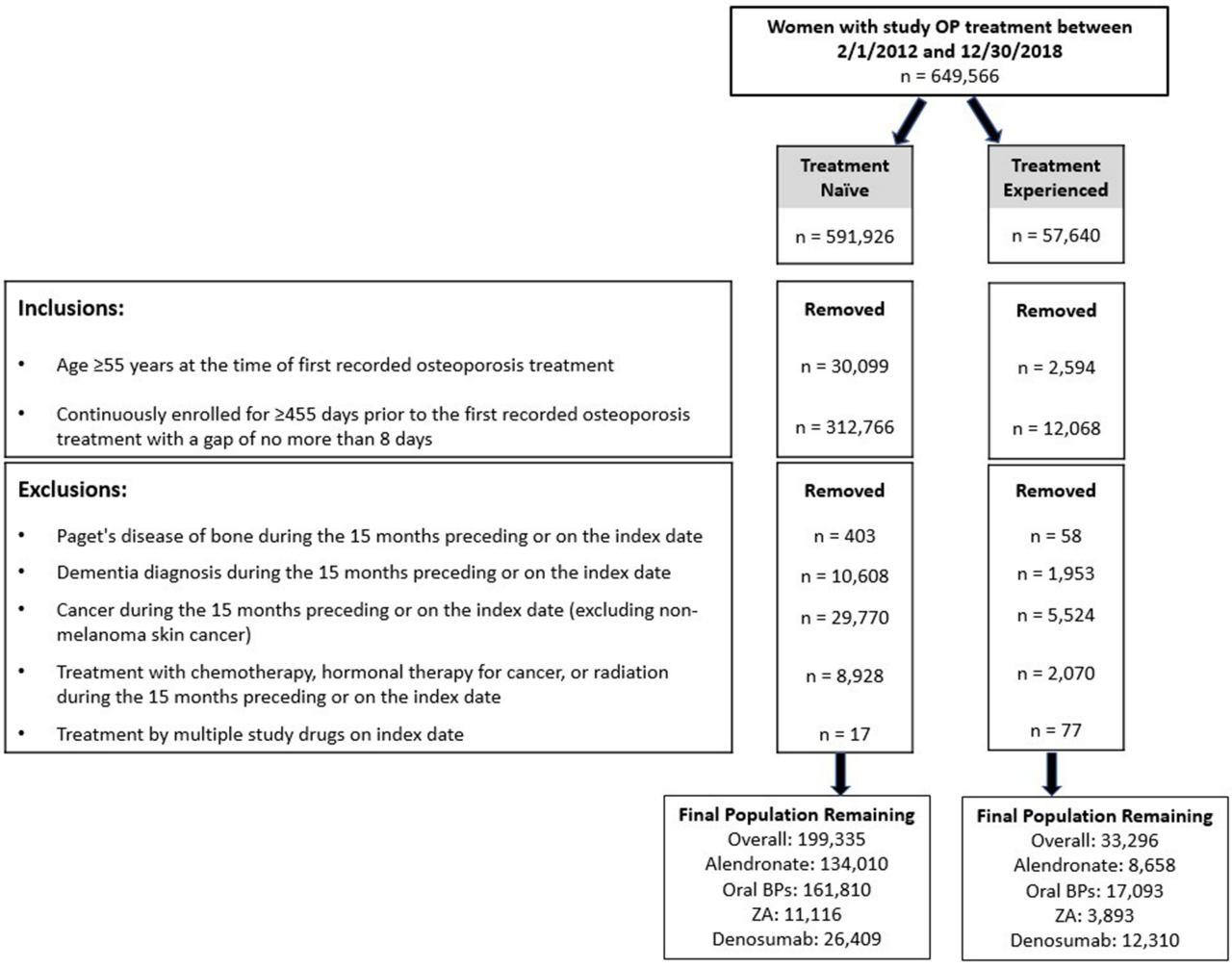

Figure 1 Study attrition diagram.

https://doi.org/10.2147/CLEP.S562445

Table 1 Baseline Characteristics (%) Assessed During the 15 Months Prior to Treatment Initiation Among Treatment-Naïve and Treatment-Experienced Women

Variable	Treatment-Naive				Treatment-Experienced			
	Overall	Denosumab	ZA	Oral BPs	Overall	Denosumab	ZA	Oral BPs
Total, N	199,335	26,409	11,116	161,810	33,296	12,310	3,893	17,093
Demographics/Healthcare Utilization								
Age, years								
55-64	19.2	14.7	15.7	20.2	20.5	15.9	16.7	24.7
65-74	43.9	39.5	43.5	44.7	40.4	40.5	42.5	39.8
75+	36.9	45.9	40.8	35.1	39.1	43.6	40.7	35.5
ER visits								
0	69.1	65.1	64.6	70.1	70.8	67.8	64.8	74.2
1+	30.9	34.9	35.4	29.9	29.2	32.2	35.2	25.8
Hospital days								
0	85.6	82.2	81.8	86.4	86.7	85.0	82.9	88.8
1-7	8.0	9.4	9.9	7.6	7.5	8.3	9.2	6.5
8+	6.5	8.5	8.3	6.0	5.8	6.7	7.9	4.7
Comorbidities								
History of fracture	15.0	21.0	18.7	13.8	14.5	18.5	18.6	10.7
Osteoarthritis	21.4	28.2	28.4	19.7	24.0	27.2	29.9	20.3
Diabetes	22.1	20.4	18.5	22.6	20.2	19.8	16.9	21.3
Hyperlipidemia	49.9	52.6	49.6	49.5	52.6	53.6	51.5	52.2
COPD	14.9	16.6	17.9	14.4	14.7	16.1	17.1	13.1
Asthma	5.7	6.8	7.9	5.4	5.8	6.3	6.3	5.3
Allergies	19.6	22.0	22.2	19.0	21.9	22.5	23.7	21.1
Vitamin D deficiency	27.2	36.0	34.9	25.2	30.0	34.6	33.9	25.7
Colorectal polyps	7.8	8.5	9.4	7.5	8.5	8.7	10.5	7.9
Charlson score								
0	25.7	22.7	22.9	26.4	24.9	23.2	24.9	26.1
1	8.7	8.3	7.8	8.8	8.8	8.5	9.2	9.0
2	13.1	13.5	14.8	12.9	13.0	13.1	13.6	12.8
≥3	52.6	55.5	54.6	51.9	53.3	55.3	52.3	52.1
Medications								
Corticosteroids	39.5	45.0	48.6	38.0	42.5	45.6	50.6	38.4
Antidepressants	29.3	28.3	31.1	29.4	31.0	32.1	36.6	29.0
Antidiabetics	12.6	9.1	8.5	13.4	11.0	10.3	9.3	11.9
Proton pump inhibitors	22.8	24.7	28.3	22.1	27.6	28.1	32.1	26.3
Systemic antibiotics	54.6	52.0	52.2	55.2	60.0	61.1	61.5	58.9

Abbreviations: BP, oral bisphosphonate; COPD, chronic obstructive pulmonary disease; ZA, zoledronic acid.

compared with oral BPs, patients on denosumab had a higher risk of having an annual wellness exam (RD = 4.7%, 95% CI = 3.5, 5.8%). Overall, patterns were similar on the risk ratio scale (Supplementary Figure S1).

Treatment-Experienced Comparisons

Among treatment-experienced women, all NCOs when comparing denosumab to alendronate alone indicated similar risks (Figure 3). Significant differences in risk were observed for only 2 of 12 NCOs (dementia and wellness) when comparing denosumab to oral BPs, and only 1 NCO (influenza vaccine) was associated with treatment when comparing ZA to oral BPs. Estimates on the risk ratio scale were similar and are available in Supplementary Figure S2.

Table 2 Baseline History of Negative Control Outcomes (%) Assessed During the 15 Months Prior to Treatment Initiation Among Treatment-Naïve and Treatment-Experienced Women

Variable	Treatment-Naive				Treatment-Experienced			
	Overall	Denosumab	ZA	Oral BPs	Overall	Denosumab	ZA	Oral BPs
Total, N	199,335	26,409	11,116	161,810	33,296	12,310	3,893	17,093
Decubitus ulcer	2.3	3.4	3.1	2.1	2.3	2.7	2.6	1.9
Transfusion	2.1	3.1	3.0	1.9	1.9	2.5	2.8	1.3
Accident	14.3	18.7	18.1	13.3	12.9	16.1	16.5	9.8
Wellness visit	56.1	55.8	55.3	56.3	55.5	57.1	58.3	53.6
Pelvic screening	7.9	9.8	10.1	7.5	9.0	9.6	8.8	8.7
Colon screening	41.1	45.2	46.4	40.1	43.3	44.7	47.1	41.3
Influenza vaccine	46.3	50.5	53.1	45.2	51.6	53.4	56.5	49.3
Herpes zoster vaccine	5.5	5.3	6.0	5.5	6.8	7.2	7.4	6.3
Mohs surgery	1.3	1.9	1.8	1.2	1.4	1.8	1.4	1.1
Visual test	1.0	1.2	1.3	0.9	1.2	1.3	1.5	1.1
Hip fracture	2.2	3.0	2.3	2.1	1.7	2.2	2.1	1.3

Notes: History of dementia is not shown, as we excluded patients with a prior history of dementia.
Abbreviations: BP, oral bisphosphonate; ZA, zoledronic acid.

Subgroup Analyses

Figures 4 and 5 present RDs for the 12 NCOs for subgroups within treatment-naïve and treatment-experienced women, respectively. Across all subgroup analyses in treatment-naïve women, comparisons between denosumab and ZA were more consistent than comparisons between denosumab and oral BPs. Like the main analysis, there was more residual confounding in the comparisons against oral BP. Treatment-naïve women aged >75 years experienced similar patterns of risk differences for all NCOs as the main cohort. However, among treatment-naïve patients with a recent fracture, the RD between treatment and NCOs was attenuated, suggesting decreased residual bias in this subgroup compared to the overall cohort. For example, among treatment-naïve patients with a recent fracture, the RD for wellness exams when comparing denosumab or ZA with oral BPs moved closer to the null. Similarly, among treatment-experienced patients initiating denosomab (as compared to oral BPs) RDs for the wellness exam and dementia NCOs were also attenuated. When stratifying by calendar period, we observed stronger associations before October 2015 than after among treatment-experienced women (eg, RD for influenza vaccine for denosumab versus oral BPs, before 2015: 3.9% [95% CI: 1.7–6.2%] versus after 2015: 2.6% [95% CI: −0.1–5.4%]). However, risks were similar in both time periods for the

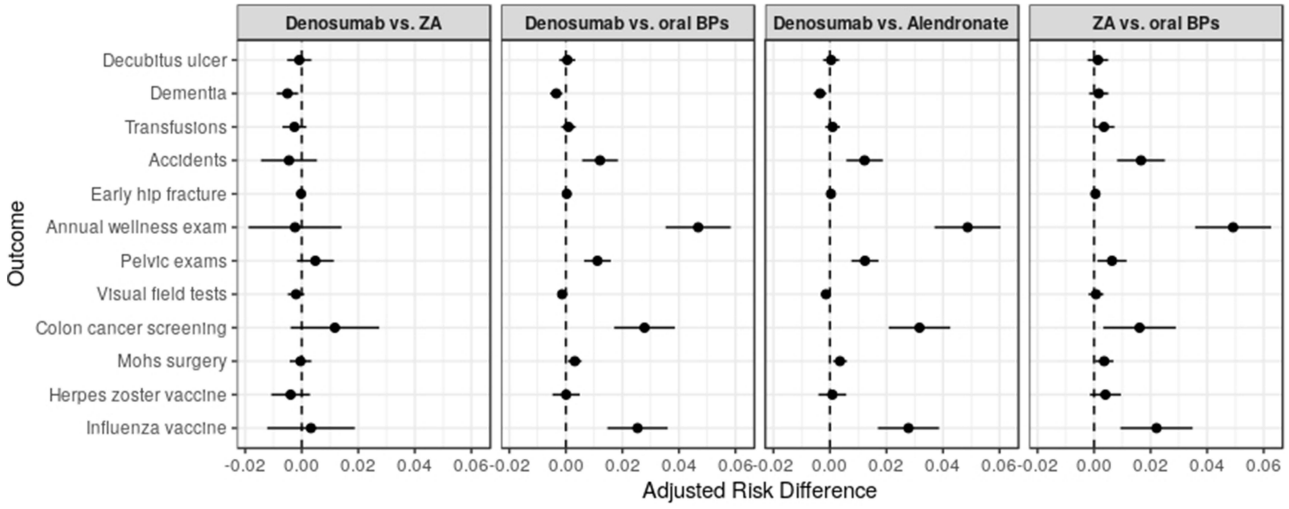

Figure 2 1-year cumulative risk differences of negative control outcomes for treatment-naïve women initiating an OP treatment.

https://doi.org/10.2147/CLEP.S562445

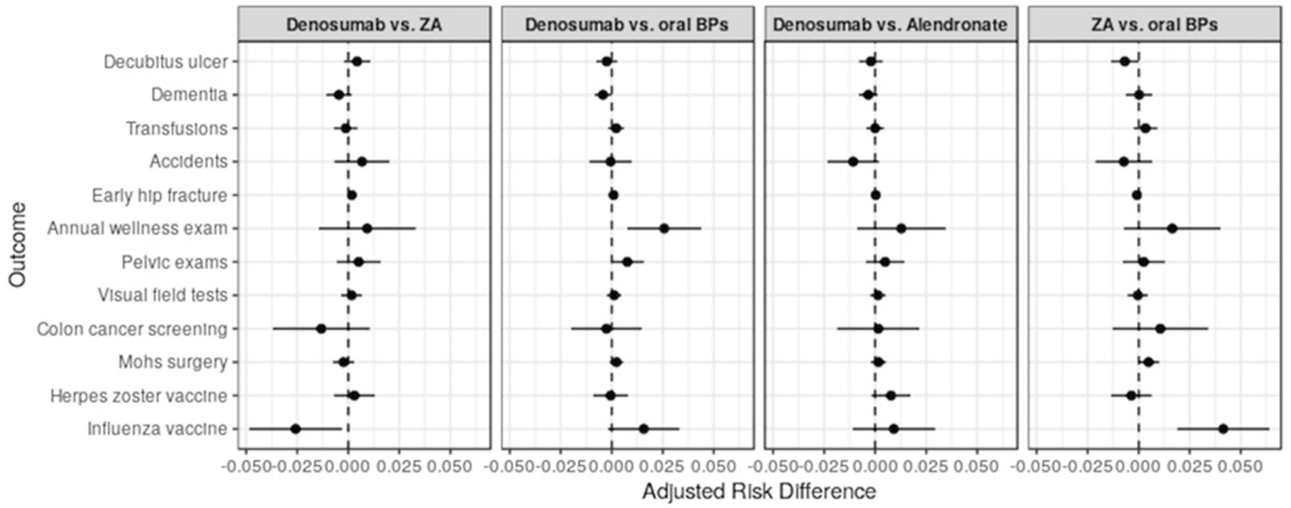

Figure 3 1-year cumulative risk differences of negative control outcomes for treatment-experienced women switching to a new OP treatment.

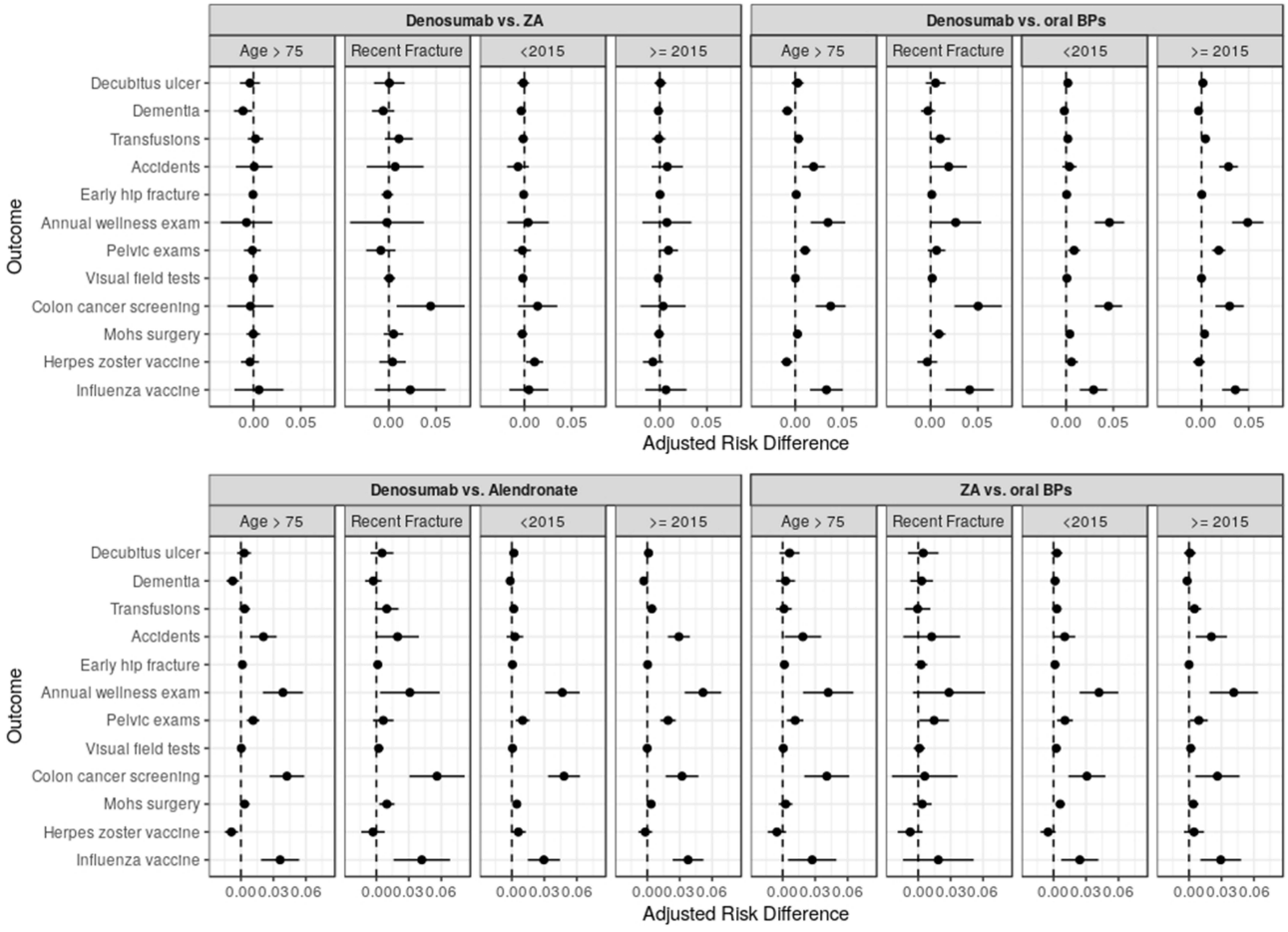

Figure 4 1-year cumulative risk differences of negative control outcomes for treatment-naïve women initiating an OP treatment, by subgroup.

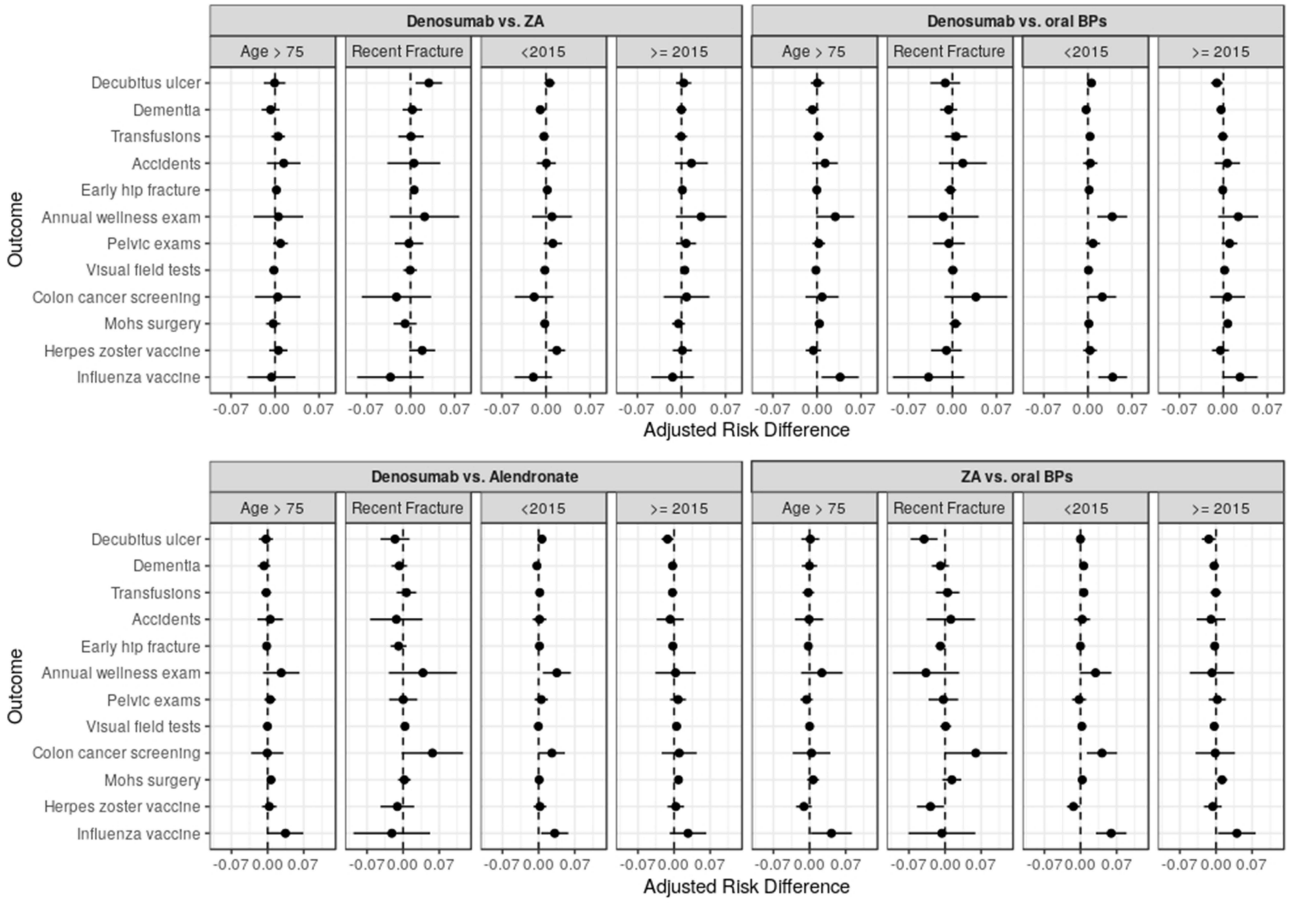

Figure 5 1-year cumulative risk differences of negative control outcomes for treatment-experienced women switching to a new OP treatment, by subgroup.

treatment-naïve cohort. RDs for the 12 NCOs in the ITT sensitivity analysis are provided in Supplementary Figure S3 (treatment-naïve) and Supplementary Figure S4 (treatment-experienced). Overall, associations with individual NCOs in the sensitivity analysis were consistent with the results in the overall cohort.

Bayesian NCO Bias Characterization

The 80%, 90%, and 95% quantiles of the absolute value of the posterior distributions estimated for all treatment contrasts on the risk difference scale are provided in Supplementary Table S5. Overall, the results of the Bayesian analysis aligned with our qualitative interpretation. For example, for the denosumab vs ZA contrasts in the treatment-naïve cohort, only one RD (dementia) had a 95% CI that excluded the null. Findings from the Bayesian analysis were consistent with the low levels of residual bias indicated by the individual NCO results; the upper bound of the 95% quantile-based interval of the absolute RD was 5 events in 1000. In scenarios with higher residual bias, the results from the Bayesian analysis were also consistent. When considering the denosumab vs alendronate contrasts in the treatment-naïve cohort, more than one-half of the 95% CIs excluded the null, suggestive of greater residual bias. The Bayesian analysis produced a 95% quantile-based interval of the absolute RD with an upper bound of 37 events in 1000.

Discussion

Among a population of 199,335 treatment-naïve and 33,296 treatment-experienced women, we estimated residual bias of different OP treatments using 12 NCOs to assess the comparability of the treatment groups for comparative effectiveness and safety studies. After adjusting for over 50 possible measured confounders using a doubly robust estimator, we found that denosumab and ZA treatment-naïve users had similar risks of NCOs, suggesting that these two treatments could be

https://doi.org/10.2147/CLEP.S562445

reasonably compared with respect to safety and effectiveness outcomes of interest. Based on the NCO analyses, no other treatment groups were considered comparable among treatment-naïve women overall. After aggregating the results across all NCOs, the absolute magnitude of residual bias for comparisons of treatment-naïve users of ZA and oral BP was 2.9% (as indicated by the 95% credible interval), an amount that was considered to be a high level of bias compared to any hypothesized treatment effect for a potential comparative effectiveness follow-up study. However, restriction to patients with severe disease, indicated by a recent fracture, improved comparability for treatment-naïve women initiating ZA as compared to oral BPs. Among treatment-experienced women, it appeared that all women initiating denosumab, ZA, and oral BPs (including alendronate alone), were comparable (95% credible intervals of the absolute risk differences having between as many as 1.1% and 2.5%, depending on the contrast). Restriction by age showed no improvement in residual confounding among treatment-naïve or treatment-experienced women. However, restriction by calendar time suggested better comparability in the later period among treatment-experienced women; however, there was little impact in the treatment-naïve population.

Our results in the treatment-naïve population were similar to results previously reported among women enrolled in MarketScan commercial or Medicare Advantage plans.[3] Women initiating denosumab were comparable to those initiating ZA, but residual bias was likely present in the comparison with oral BPs. While we found residual confounding to be associated with women taking denosumab vs oral BPs, the estimates were generally in the same direction. For example, the RD estimate for wellness exam in our study was 4.7% (95% CI: 3.5–5.8%) and was 1.15% (95% CI: 0.41–1.89%) in the MarketScan population.[3] It is encouraging that we found similar results in two different data sources, as our study population was generally older (mean age = 71 years vs 67 years in MarketScan) and had more comorbidities (15% vs 10% in MarketScan had a history of fracture).[3]

It is likely that women initiating an injectable medication as their first observed treatment for OP are meaningfully different from women initiating an oral BP. Indeed, we observed that, compared to women initiating an oral BP, women starting denosumab or ZA had increased healthcare utilization (eg, outpatient visits and hospitalizations) and appeared to be in worse health (eg, higher prevalence of comorbidities). Elevated differences in the risk of NCOs indicated that unmeasured confounding remained after adjustment. This is further evidence that using claims data alone to estimate comparative effectiveness or safety of first line injectable medications as compared to an oral BP is likely to result in a biased comparison among treatment-naïve, commercially insured women in the United States.

Among women starting second line treatment, the estimates were closer to the null, suggesting less evidence of residual confounding. Even when comparing women switching to denosumab with women switching to an oral BP, the risk of most NCOs was similar. Patients can switch medications for a variety of reasons, including more advanced disease (such as fractures while on treatment), the emergence of side effects, and physician or patient preference for specific dosing schedules.[16] This could implicitly make the treatment groups more comparable even in the absence of clinical measurements, such as low bone mineral density, which is an important predictor of fracture.[17]

Restricting the cohort to the recent fracture subgroup appeared to make the treatment groups more comparable. For example, among treatment-naïve women, there were marginal differences in the risk for wellness visits when comparing denosumab to oral BP users in the full cohort (4.7% RD) which were further reduced in the recent fracture subgroup (2.6% RD). There were fewer NCOs that were associated in the high-risk subgroup for both treatment-naïve and treatment-experienced patients. It has been shown that restriction can often reduce unmeasured confounding where analytic methods for adjustment fall short.[18,19] Restricting to women aged >75 years did not appear to reduce residual confounding, as estimates were similar between the subgroup and the full cohort, likely attributable to the median age at treatment initiation being 71 years. This could indicate that recent fracture is a better indicator of severe/advanced disease in this population, which is supported by research showing that prior fractures are predictive of future fractures.[20]

We observed a difference in the amount of possible residual confounding among the treatment-experienced group stratified by calendar time—there were more NCOs that were associated with treatment (for all treatment groups) before October 2015 than after October 2015. This could indicate channeling bias that may have been more apparent in the years soon after denosumab was approved for post-menopausal osteoporosis in 2010.[21,22] Physicians may preferentially administer denosumab to sicker patients with more advanced OP, resulting in unmeasured confounding. However, we did not see similar trends suggestive of channeling bias among the treatment-naïve cohort. Similarly, oral

bisphosphonates represent a heterogeneous group of medications with numerous generic formulations whereas no generic versions of denosumab were available during the time period of the current study (the first interchangeable biosimilar to denosumab was approved by the FDA in March 2024).[23] Although this heterogeneity may limit comparability when grouping oral bisphosphonates together, our findings did not change substantially when we restricted comparisons to alendronate alone.

Our study had some limitations. The 12 NCOs were identified by clinical subject matter experts with specific consideration of various hypothetical confounding mechanisms, increasing our confidence that the observed null associations indicate the relevant treatment groups are comparable. However, as in all real-world studies, residual confounding may remain due to undefined confounding domains or unmeasured variables. For example, the data did not contain clinical measures of disease severity, such as bone mineral density or body mass index. Bone mineral density and fracture history while on therapy are important determinants of whether clinicians initiate or escalate to denosumab or zoledronate, particularly among patients with more severe osteoporosis or fractures occurring despite oral bisphosphonate treatment. Prescribers may also avoid oral bisphosphonates in patients using proton pump inhibitors (PPIs) due to concerns about gastrointestinal absorption. Additionally, PPIs themselves have been associated with increased fracture risk—which may result in channeling higher-risk patients toward IV or injectable agents.[24,25] Additionally, the confounders we included were measured at baseline and thus our assessment did not capture information on potential time-varying confounding mechanisms. While we required patients in our study sample to be continuously enrolled for at least 15 months prior to index and used all available look-back to distinguish treatment naïve patients from experienced users, some patients may have been misclassified if their prior OP medication use was not captured in the available data. Prior research suggests that BP users may have long gaps in treatment due to "treatment holidays",[26,27] and that denosumab and ZA users may have extended intervals between sequential treatments. Results from the Bayesian analysis could benefit from a formal "gating" approach where an a priori rule determining whether a particular contrast is comparable is defined based on some maximum allowed observed absolute risk difference relative to a given level of probability. The level of the allowed absolute risk difference could be determined relative to a hypothesized risk difference for a comparative effectiveness study or a required level of confidence for a safety study. Lastly, our cohort was comprised of women with commercial or Medicare Advantage insurance; therefore, our findings may not generalize to other populations, such as the general Medicare population or those residing outside the United States.

Conclusion

Overall, our findings support the conclusion that comparative effectiveness and safety studies of fracture outcomes and denosumab or ZA versus oral BPs among treatment-experienced, commercially insured women in the United States are likely valid with respect to the comparability of treatment groups, particularly for high-risk subgroups and in later calendar periods. However, our results do not support conducting observational studies examining these treatment contrasts among treatment-naïve patients in this commercially insured study population.

Data Sharing Statement

The data for this study were licensed by Amgen Inc. from Optum Health, a commercial data provider in the US. The data use agreement does not permit the authors to share the raw data externally.

Funding

This study was funded by Amgen Inc.

Disclosure

DP, YY, and KH are employees of and own equity in Target RWE. At the time, this study was conducted, LF and PS were employees and owned equity in Target RWE. MAB serves as a consulting scientist and owns equity in Target RWE. He has served as a scientific advisor to AbbVie, Amgen, Atara Biosciences, Brigham and Women's Hospital, Gilead/Kite, Merck, and Vertex. MK, AB, JL, LS, RS, and BB are employees of and stockholders in Amgen Inc. The authors report no other conflicts of interest in this work.

References

1. NIH Consensus Development Panel on Osteoporosis Prevention D, Therapy. Osteoporosis prevention, diagnosis, and therapy. *JAMA*. 2001;285(6):785–795. doi:10.1001/jama.285.6.785.
2. Blake J, Cosman FA, Lewiecki EM, et al. Management of osteoporosis in postmenopausal women: the 2021 position statement of The North American Menopause Society. *Menopause*. 2021;28(9):973–997. doi:10.1097/GME.0000000000001831
3. McGrath LJ, Spangler L, Curtis JR, et al. Using negative control outcomes to assess the comparability of treatment groups among women with osteoporosis in the United States. *Pharmacoepidemiol Drug Saf*. 2020;29(8):854–863. doi:10.1002/pds.5037
4. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383–388. doi:10.1097/EDE.0b013e3181d61eeb
5. Dusetzina SB, Brookhart MA, Maciejewski ML. Control outcomes and exposures for improving internal validity of nonrandomized studies. *Health Serv Res*. 2015;50(5):1432–1451. doi:10.1111/1475-6773.12279
6. Shi X, Miao W, Tchetgen ET. A selective review of negative control methods in epidemiology. *Curr Epidemiol Rep*. 2020;7(4):190–202. doi:10.1007/s40471-020-00243-4
7. Man KKC, Lau WCY, Coghill D, et al. Association between methylphenidate treatment and risk of seizure: a population-based, self-controlled case-series study. *Lancet Child Adolesc Health*. 2020;4(6):435–443. doi:10.1016/S2352-4642(20)30100-0
8. Sainsbury C, Wang J, Gokhale K, et al. Sodium-glucose co-transporter-2 inhibitors and susceptibility to COVID-19: a population-based retrospective cohort study. *Diabetes Obes Metab*. 2021;23(1):263–269. doi:10.1111/dom.14203
9. Lousdal ML, Lash TL, Flanders WD, et al. Negative controls to detect uncontrolled confounding in observational studies of mammographic screening comparing participants and non-participants. *Int J Epidemiol*. 2020;49(3):1032–1042. doi:10.1093/ije/dyaa029
10. Understanding the use of negative controls to assess the validity of non-interventional studies of treatment using real-world evidence. Duke Margolis Center for Health Policy. Available from: https://healthpolicy.duke.edu/events/understanding-use-negative-controls-assess-validity-non-interventional-studies-treatment. Accessed January 22, 2025.
11. Brookhart MA, Sturmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care*. 2010;48(6 Suppl):S114–20. doi:10.1097/MLR.0b013e3181dbebe3
12. Hubbard AE, Laan MJ, Robins JM. Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies. In: *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer; 2000:135–177.
13. Ozenne BMH, Scheike TH, Staerk L, Gerds TA. On the estimation of average treatment effects with right-censored time to event outcome and competing risks. *Biom J*. 2020;62(3):751–763. doi:10.1002/bimj.201800298
14. Breskin A, Butler AM, Sahrmann J, Brookhart MA. Estimating vaccine effects: a comparison of causal methods [abstract]. *Pharmacoepidemiol Drug Saf*. 2020;29(Suppl 3):S361–S362.
15. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology*. 1990;1(1):43–46.
16. Goldshtein I, Rouach V, Shamir-Stein N, Yu J, Chodick G. Role of side effects, physician involvement, and patient perception in non-adherence with oral bisphosphonates. *Adv Ther*. 2016;33(8):1374–1384. doi:10.1007/s12325-016-0360-3
17. Johnell O, Kanis JA, Oden A, et al. Predictive value of BMD for Hip and other fractures. *J Bone Miner Res*. 2005;20(7):1185–1194. doi:10.1359/jbmr.050304
18. Zhang HT, McGrath LJ, Ellis AR, Wyss R, Lund JL, Stürmer T. Restriction of pharmacoepidemiologic cohorts to initiators of medications in unrelated preventive drug classes to reduce confounding by frailty in older adults. *Am J Epidemiol*. 2019;188(7):1371–1382. doi:10.1093/aje/kwz083
19. McGrath LJ, Ellis AR, Brookhart MA. Controlling time-dependent confounding by health status and frailty: restriction versus statistical adjustment. *Am J Epidemiol*. 2015;182(1):17–25. doi:10.1093/AJE/KWU485
20. Kanis JA. Diagnosis of osteoporosis and assessment of fracture risk. *Lancet*. 2002;359(9321):1929–1936. doi:10.1016/S0140-6736(02)08761-5
21. Lobo FS, Wagner S, Gross CR, Schommer JC. Addressing the issue of channeling bias in observational studies with propensity scores analysis. *Res Soc Adm Pharm*. 2006;2(1):143–151. doi:10.1016/j.sapharm.2005.12.001
22. Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of β 2-agonists. *Am J Epidemiol*. 1996;144(12):1161–1169. doi:10.1093/oxfordjournals.aje.a008895
23. U.S. Food and Drug Administration. FDA approves first interchangeable biosimilars to Prolia and Xgeva to treat certain types of osteoporosis and prevent bone events in cancer [news release]. U.S. Food and Drug Administration; 2024. Available from: https://www.fda.gov/drugs/news-events-human-drugs/fda-approves-first-interchangeable-biosimilars-prolia-and-xgeva-treat-certain-types-osteoporosis-and. Accessed December 9, 2025.
24. Heilmeier U, Siegenthaler A, Meyer U, et al. Regular proton-pump inhibitor intake is associated with deterioration of peripheral bone mineral density, microarchitecture, and strength in older patients as assessed by high-resolution peripheral quantitative computed tomography (HR-pQCT). *Calcif Tissue Int*. 2025;16(1):131. doi:10.1007/s00223-025-01420-7
25. Ngamruengphong S, Leontiadis GI, Radhi S, Dentino A, Nugent K. Proton pump inhibitors and risk of fracture: a systematic review and meta-analysis of observational studies. *Am J Gastroenterol*. 2011;106(7):1209–1218. doi:10.1038/ajg.2011.113
26. Hayes KN, Winter EM, Cadarette SM, Burden AM. Duration of bisphosphonate drug holidays in osteoporosis patients: a narrative review of the evidence and considerations for decision-making. *J Clin Med*. 2021;10(5):1140. doi:10.3390/JCM10051140
27. Brookhart MA, Avorn J, Katz JN, et al. Gaps in treatment among users of osteoporosis medications: the dynamics of noncompliance. *Am J Med*. 2007;120(3):251–256. doi:10.1016/J.AMJMED.2006.03.029