

Treatment heterogeneity in comparative effectiveness of teriparatide vs bisphosphonates: multi-database cohort study

Alexandros Rekkas¹, Annika M. Jödicke², David van Klaveren³, Daniel Prieto-Alhambra^{1,2}, Peter R. Rijnbeek¹

¹ Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, Netherlands

² Pharmaco- and Device Epidemiology, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

³ Department of Public Health, Erasmus University Medical Center, Rotterdam, Netherlands

1 Introduction

2 Osteoporosis is a common condition characterised by decreased bone density and associated with increased risk for
3 fragility fractures, which affects almost 30% of women aged ≥ 50 years [1]. It was estimated that in 2005, over 2
4 million incident fractures occurred in the United States, with hip fractures accounting for more than 70% of the
5 costs. By 2025, annual fractures and costs could rise by almost 50% [2].

6 Fracture prevention is thus the key focus of anti-osteoporotic therapy, reducing the disease burden both for the
7 affected patient but also on the population level. Several pharmacological agents are available for primary and
8 secondary prevention. The choice of anti-osteoporotic agent largely depends on history of fragility fracture and
9 anticipated fracture risk. Oral bisphosphonates (BP) are first line treatments for postmenopausal patients with
10 increased fracture risk in clinical guidelines (#AACE/ACE 2020) as for their favourable cost-effectiveness and
11 safety profile. Teriparatide, a parathyroid hormone analogue administered as a daily injection, was approved by the
12 FDA in 2002 as the first anabolic agent for treatment of severe postmenopausal osteoporosis. While clinical trials
13 among patients with previous vertebral fractures showed a substantial risk reduction for new vertebral fracture
14 comparing teriparatide to placebo [3] and risedronate [4], its efficacy on low-incident major osteoporotic fractures,
15 especially hip fractures, is less well established: Previous randomised trials assessing hip fracture comprised only
16 few events, thus not providing sufficient power to show differences between treatment groups when studying
17 comparative effectiveness. A recent meta-analysis by Diez-Perez and colleagues assessed the effect of teriparatide
18 on hip fracture, indicating a significant 80% risk reduction compared to placebo and a non-significant 46% risk
19 reduction when compared to active controls [5].

20 With teriparatide-containing biosimilars being launched in recent years, treatment costs dropped significantly.
21 Subsequently, the discussion of cost-effectiveness in patients with less severe forms of osteoporosis is restarting.
22 Therefore, evaluation of the effectiveness in hip fracture prevention is required. In particular, observational studies
23 assessing comparative effectiveness in a real-world setting are required, as effectiveness may differ from clinical trial
24 as for differences in patient's persistence and adherence to their anti-osteoporotic treatment. While hip fracture is a
25 comparatively rare outcome, it is unambiguously defined and reliably recorded in routinely collected data, allowing
26 for a direct comparison to results from clinical trial meta-analyses. While results from non-controlled, observational
27 studies suggest that teriparatide may reduce the risk of hip fractures [6,7], this is the first observational study
28 assessing teriparatide in hip fracture risk reduction compared to oral bisphosphonate users in postmenopausal
29 women using multiple large real-world databases.

30 **2 Methods**

31 **2.1 Study design**

32 We performed a retrospective new user cohort study to estimate the effectiveness of teriparatide compared to
33 oral bisphosphonates in patients with osteoporosis[8]. We included female participants above the age of 50 with
34 established osteoporosis (any condition occurrence of hip, wrist, spine or shoulder/humerus fracture in their history,
35 prior to treatment initiation) and at least 365 days of continuous observation period before the index event. Patients
36 were considered to be new-users if they received no anti-osteoporosis drugs (raloxifene, bazedoxifene, denosumab,
37 abaloparatide, romosozumab) 365 days prior to treatment initiation with teriparatide or an oral bisphosphonate.

38 Our primary efficacy outcome was hip fracture. Vertebral fracture and a composite major osteoporotic fracture,
39 defined as hip, vertebral or wrist/forearm/proximal humerus fracture, were our secondary efficacy outcomes. We
40 excluded patients that had experienced the outcome under study any time prior to treatment initiation. Patient
41 time-at-risk started 1 day after treatment initiation and finished 730 days after treatment initiation.

42 **2.2 Data sources**

43 We ran our analyses on three US observational databases mapped to OMOP-CDM version. More specifically:

- 44 ▪ **IMB MarketScan Commercial Database (CCAE)** includes health insurance claims across the continuum
45 of care (e.g. inpatient, outpatient, outpatient pharmacy, carve-out behavioral healthcare) as well as enrollment
46 data from large employers and health plans across the United States who provide private healthcare coverage
47 for more than 155 million employees, their spouses, and dependents. This administrative claims database
48 includes a variety of fee- for-service, preferred provider organizations, and capitated health plans.
- 49 ▪ **Optum De-Identified Clininformatics® Data Mart Database – Date of Death (Optum-DOD)**: is derived
50 from a database of administrative health claims for members of large commercial and Medicare Advantage
51 health plans. The database includes approximately 17-19 million annual covered lives, for a total of over
52 65 million unique lives over a 12 year period (1/2007 through 12/2019). The population is geographically
53 diverse, spanning all 50 states.
- 54 ▪ **Optum de-identified Electronic Health Record Dataset (Optum-EHR)** is derived from dozens of
55 healthcare provider organizations in the United States, that include more than 700 Hospitals and 7000
56 Clinics; treating more than 102 million patients receiving care in the United States.

57 2.3 Statistical analyses

58 We carried out two sets of analyses. First, we derived overall treatment effect estimates of teriparatide compared to
59 oral bisphosphonates regarding the three outcomes of interest. To account for potential measured confounding we
60 developed separate large-scale propensity score models within each database based on LASSO logistic regression
61 using the same predefined set of measured covariates [8]. To assume that treatment and comparator cohorts
62 stand in equipoise we required that the majority of the patients in both have preference scores between 0.3 and
63 0.7. We assumed that covariate balance had been reached, if all after-adjustment baseline characteristics had
64 standardized absolute mean differences of less than 0.1. We estimated treatment effects using Cox proportional
65 hazards models with treatment as the sole covariate. To account for observed confounding we used variable ratio
66 matching (1:10) on the propensity score using a caliper of 0.2 on the logit scale. Hazard ratios (HR) derived in
67 different databases were summarized using random effects meta-analysis.

68 For the second set of analyses, we used the standardized framework for risk-based assessment of treatment effect
69 heterogeneity, that follows the directives of the PATH statement [9,10]. Following data extraction, the framework
70 was applied in two steps. First, we derived individualized risk predictions for the three efficacy outcomes. We built
71 the prediction models using LASSO logistic regression on the propensity score matched (1:10) subpopulation of
72 the pooled treatment arms, aiming to remove any effect of measured confounding that may cause differential
73 fit of our models across treatment arms. We considered the same large set of candidate covariates as for the
74 development of the propensity score models.

75 For each outcome we used the derived prediction models to divide the population in 4 equally-sized subgroups,
76 based on the quarters of the predicted risk distribution. Within each of these risk-based subgroups we developed a
77 new propensity score model. Our analyses were performed on the propensity score matched (1:4) subset of the risk
78 subgroup subset. We derived relative effect estimates using Cox proportional hazards models only with treatment
79 as a predictor. Absolute effect estimates were calculated based on the difference of the Kaplan-Meier estimates,
80 on day 365 after treatment initiation.

81 For each outcome we used the derived prediction models to stratify the study population using two separate
82 stratification schemes. In the first approach, we divide the population into the lowest 75% and the top 25% hip
83 fracture risk subgroups. In the second approach, the population is divided using existing guidelines on teriparatide
84 treatment based on major fracture risk (REF!!). Within each of these risk-based subgroups we developed a new
85 propensity score model. Our analyses were performed on the propensity score matched (1:10) subset of the risk
86 subgroup subset. We derived relative effect estimates using Cox proportional hazards models only with treatment
87 as a predictor. Absolute effect estimates were calculated based on the difference of the Kaplan-Meier estimates,

⁸⁸ on day 365 after treatment initiation.

⁸⁹ Residual study bias from unmeasured confounding can still be present in observational studies, which often may
⁹⁰ not be visible when evaluating propensity score adjustment. To account for that, we considered a set of 126
⁹¹ negative control outcome experiments, where a null effect was assumed to be true. We used the estimated relative
⁹² effects for these outcomes to derive an empirical approximation to the true null distribution, which was then used
⁹³ to calibrate the hazard ratios and their 95% confidence intervals for the 3 outcomes of interest [11,12].

94 **3 Results**

95 **3.1 Treatment effect heterogeneity**

96 **[Introductory stuff for Table 1 !!]**

97 In all databases adequate equipoise of the preference score distributions was achieved (Figure 1). After matching
98 all baseline covariates (>35,000 in each database) were well balanced (Figure 2). This indicates that we were able
99 account for the majority of the observed confounding present in all databases. However, negative control analyses
100 time falsely favored oral bisphosphonates over teriparatide, which suggests the presence of residual unobserved
101 confounding (Figure 3).

102 In terms of hip fracture risk we estimated calibrated hazard ratios of 1.1 (0.73 to 1.7; 95% CI), 0.85 (0.65 to 1.1;
103 95% CI) and 0.93 (0.64 to 1.3; 95% CI), in CCAE, Optum-DOD and Optum-EHR respectively. The calibrated
104 meta-analytic hazard ratio of the overall effect of teriparatide compared to oral bisphosphonates was 0.92 (0.76 to
105 1.1; 95% CI). Results for major osteoporotic and vertebral fractures are provided in the supplement.

106 When using the internally developed hip fracture risk prediction models to divide the population into lower 75%
107 and top 25% risk subgroups, we observed. . . .

108 **4 Discussion**

109 **5 References**

- 110 1 Wright N, Saag K, Dawson-Hughes B *et al.* The impact of the new national bone health alliance (nbha)
111 diagnostic criteria on the prevalence of osteoporosis in the usa. *Osteoporosis International* 2017;**28**:1225–32.
- 112 2 Cole ZA, Dennison EM, Cooper C. Osteoporosis epidemiology update. *Current Rheumatology Reports* 2008;**10**:92–
113 6. doi:10.1007/s11926-008-0017-6
- 114 3 Neer RM, Arnaud CD, Zanchetta JR *et al.* Effect of parathyroid hormone (1-34) on fractures and bone mineral
115 density in postmenopausal women with osteoporosis. *New England Journal of Medicine* 2001;**344**:1434–41.
116 doi:10.1056/nejm200105103441904
- 117 4 Kendler DL, Marin F, Zerbini CAF *et al.* Effects of teriparatide and risedronate on new fractures in post-
118 menopausal women with severe osteoporosis (VERO): A multicentre, double-blind, double-dummy, randomised
119 controlled trial. *The Lancet* 2018;**391**:230–40. doi:10.1016/s0140-6736(17)32137-2
- 120 5 Díez-Pérez A, Marin F, Eriksen EF *et al.* Effects of teriparatide on hip and upper limb fractures in patients with
121 osteoporosis: A systematic review and meta-analysis. *Bone* 2019;**120**:1–8. doi:10.1016/j.bone.2018.09.020
- 122 6 Burge RT, Disch DP, Gelwicks S *et al.* Hip and other fragility fracture incidence in real-world teriparatide-treated
123 patients in the united states. *Osteoporosis International* 2016;**28**:799–809. doi:10.1007/s00198-016-3888-9
- 124 7 Silverman S, Langdahl BL, Fujiwara S *et al.* Reduction of hip and other fractures in patients receiving teriparatide
125 in real-world clinical practice: Integrated analysis of four prospective observational studies. *Calcified Tissue
126 International* 2018;**104**:193–200. doi:10.1007/s00223-018-0485-2
- 127 8 Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and
128 synthetic data experiments. *International Journal of Epidemiology* 2018;**47**:2005–14. doi:10.1093/ije/dyy120
- 129 9 Kent DM, Paulus JK, Klaveren D van *et al.* The predictive approaches to treatment effect heterogeneity (PATH)
130 statement. *Annals of Internal Medicine* 2019;**172**:35. doi:10.7326/m18-3667
- 131 10 Kent DM, Klaveren D van, Paulus JK *et al.* The predictive approaches to treatment effect heterogeneity (PATH)
132 statement: Explanation and elaboration. *Annals of Internal Medicine* 2019;**172**:W1. doi:10.7326/m18-3668
- 133 11 Schuemie MJ, Ryan PB, DuMouchel W *et al.* Interpreting observational studies: Why empirical calibration is
134 needed to correct p -values. *Statistics in Medicine* 2013;**33**:209–18. doi:10.1002/sim.5925
- 135 12 Schuemie MJ, Hripcak G, Ryan PB *et al.* Robust empirical calibration of p -values using observational data.
136 *Statistics in Medicine* 2016;**35**:3883–8. doi:10.1002/sim.6977

¹³⁷ **6 Tables and figures**

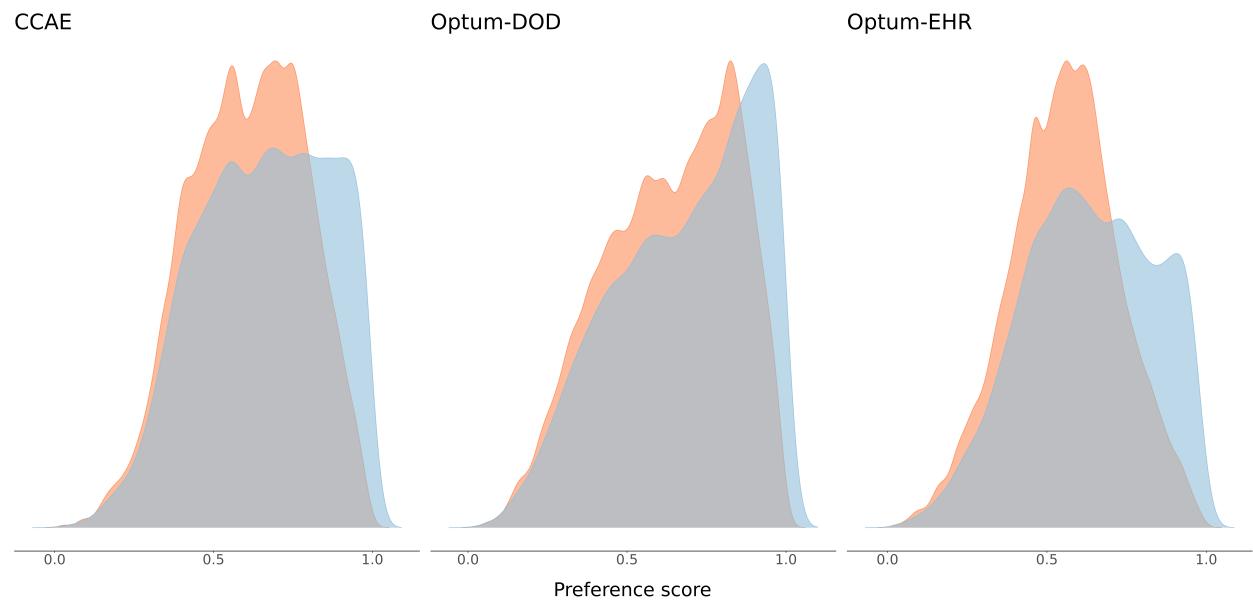


Figure 1: Calibrated overall results

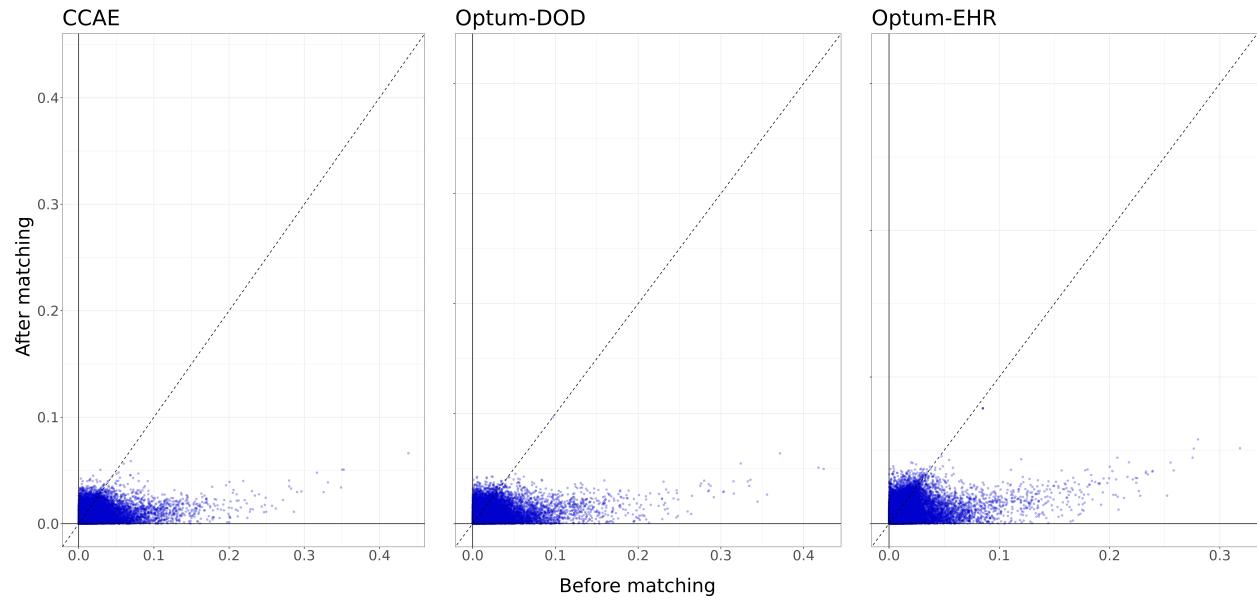


Figure 2: Calibrated overall results

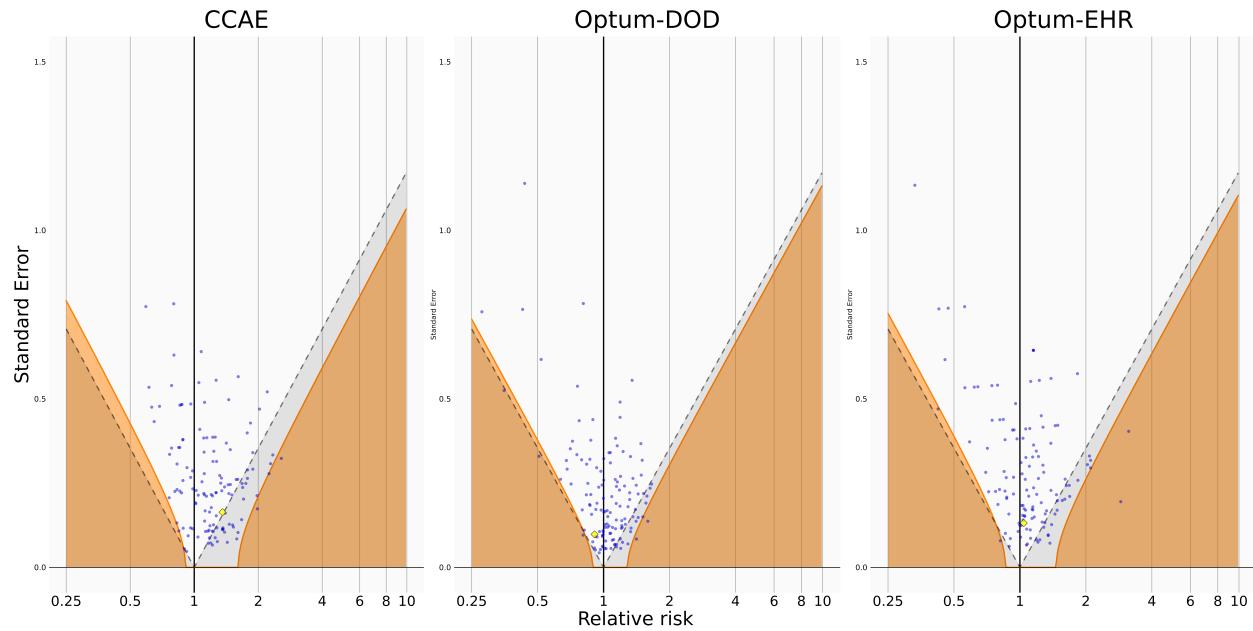


Figure 3: Overall negative control plot.

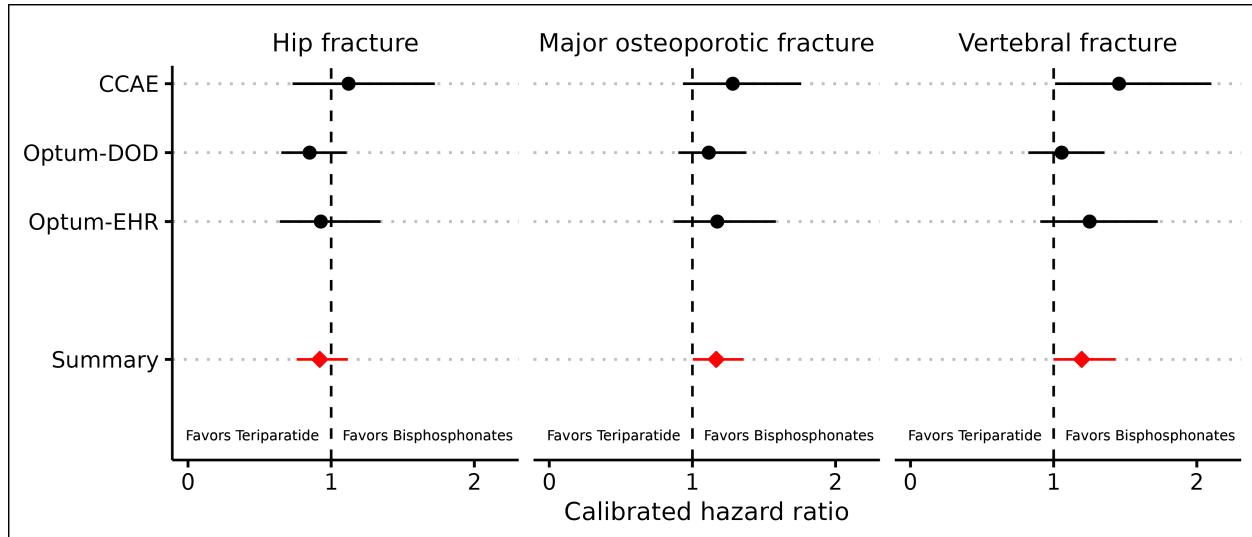


Figure 4: Calibrated hazard ratios for the 3 outcomes of interest across the 3 considered databases using a set of 126 negative controls. Values below 1 favour teriparatide, while values above 1 favour bisphosphonates.

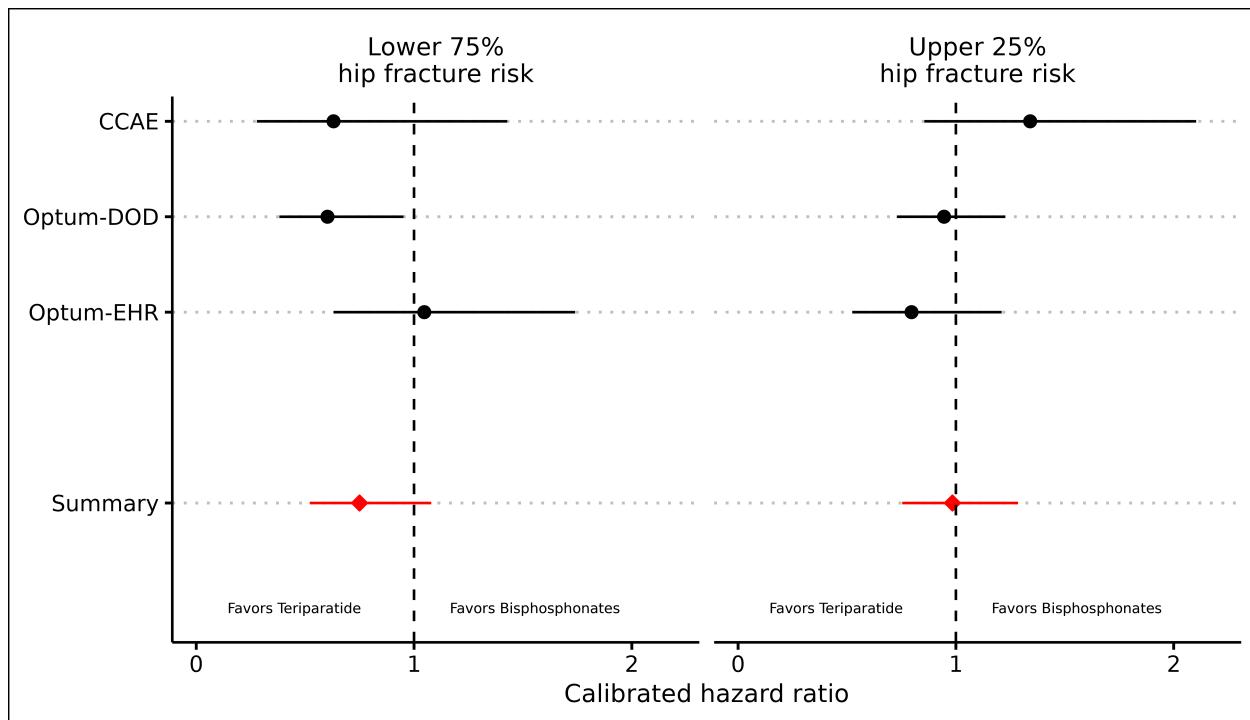


Figure 5: Calibrated hazard ratios for the 3 outcomes of interest across the 3 considered databases using a set of 126 negative controls. Values below 1 favour teriparatide, while values above 1 favour bisphosphonates.

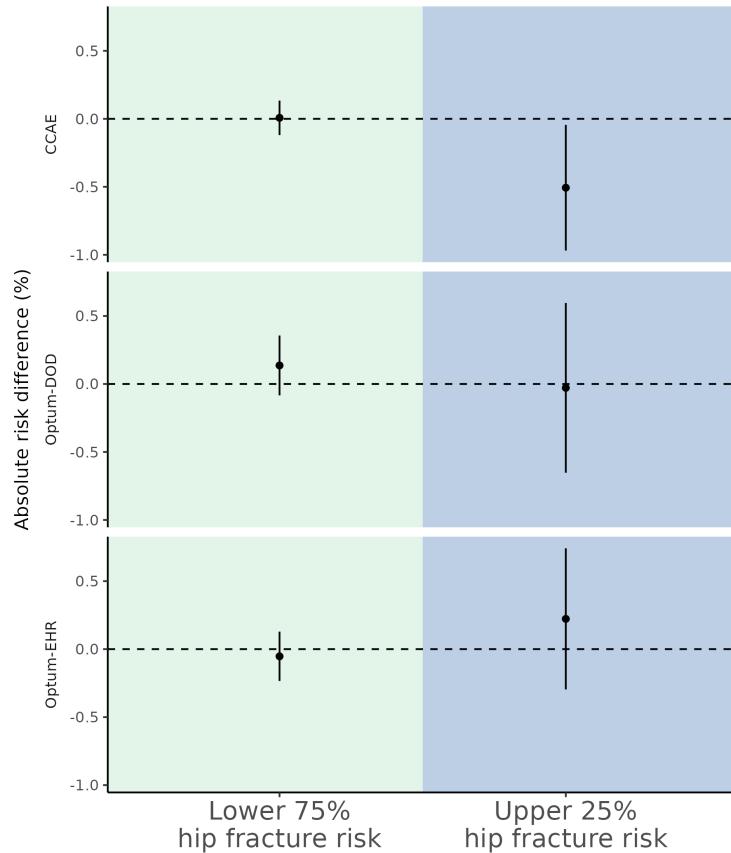


Figure 6: Calibrated hazard ratios for the 3 outcomes of interest across the 3 considered databases using a set of 126 negative controls. Values below 1 favour teriparatide, while values above 1 favour bisphosphonates.