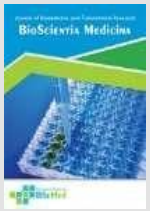
eISSN (Online): 2598-0580



Bioscientia Medicina: Journal of Biomedicine & Translational Research

Journal Homepage: www.bioscmed.com

Head-to-Head Comparison of Urinary Hydroxyproline and Serum CTX-I in Monitoring Antiresorptive Therapy for Osteoporosis: A Network Meta-Analysis

Jonggara Octaviandra Siahaan[1*], Eka Kurniawan[2]

[1]Department of Internal Medicine, Faculty of Medicine, Universitas Andalas/Dr. M. Djamil General Hospital, Padang, Indonesia

[2]Division of Rheumatology, Department of Internal Medicine, Faculty of Medicine, Universitas Andalas/Dr. M. Djamil General Hospital, Padang, Indonesia

ARTICLE INFO

Keywords:
Antiresorptive therapy
Bone turnover markers
CTX-I
Hydroxyproline
Osteoporosis

***Corresponding author:**
Jonggara Octaviandra Siahaan

E-mail address:
jonggasiahaan@gmail.com

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.37275/bsm.v9i9.1373

A B S T R A C T

Background: Bone turnover markers (BTMs) are essential for monitoring the efficacy of antiresorptive therapies in osteoporosis. While serum C-terminal telopeptide of type I collagen (s-CTX-I) is the recommended reference marker for bone resorption, the utility of the classic marker, urinary hydroxyproline (u-HYP), in the modern therapeutic era remains debated. This study aimed to provide the first network meta-analysis (NMA) to compare the responsiveness of u-HYP and s-CTX-I to antiresorptive treatments. **Methods:** We conducted a systematic review and NMA of randomized controlled trials (RCTs) published between January 2015 and December 2024. We searched PubMed, Embase, and the Cochrane Central Register of Controlled Trials for RCTs of antiresorptive therapies (alendronate, denosumab, risedronate) in postmenopausal women with osteoporosis that reported changes in s-CTX-I or u-HYP at 3-6 months. A Bayesian random-effects NMA was performed to calculate the standardized mean difference (SMD) and rank the responsiveness of each marker. **Results:** Seven RCTs involving 3,451 patients met the inclusion criteria. The evidence network was well-connected for both markers. Antiresorptive therapies induced a significantly greater reduction in s-CTX-I levels compared to u-HYP. For instance, the effect of denosumab versus placebo was substantially larger when measured by s-CTX-I (SMD: -1.88; 95% Credible Interval [CrI]: -2.25 to -1.51) than by u-HYP (SMD: -0.95; 95% CrI: -1.22 to -0.68). Surface Under the Cumulative Ranking (SUCRA) analysis confirmed that s-CTX-I had a 98.2% probability of being the more responsive marker, compared to 1.8% for u-HYP. Heterogeneity was manageable, and no significant inconsistency was detected between direct and indirect evidence. **Conclusion:** This network meta-analysis provides robust, synthesized evidence that serum CTX-I demonstrates a markedly superior dynamic response to antiresorptive therapy compared to urinary hydroxyproline. These findings reinforce the position of s-CTX-I as the preferred biomarker for monitoring treatment efficacy in clinical practice.

1. Introduction

Osteoporosis is a systemic skeletal disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to enhanced bone fragility and a consequent increase in fracture risk.[1] It represents a major global public health problem, affecting an estimated 200 million individuals worldwide. The lifetime risk for a fragility fracture in women over 50 is as high as 40-50%. Such fractures, particularly of the hip and spine, are associated with significant morbidity, mortality, loss of independence, and substantial healthcare expenditure. The management of osteoporosis has evolved significantly, with a primary goal of reducing fracture incidence through a combination of non-pharmacological interventions and potent

pharmacological agents.[2] The management of osteoporosis rests upon the fundamental biological process of bone remodeling—a continuous cycle of bone resorption by osteoclasts and bone formation by osteoblasts. Osteoporosis arises from an imbalance in this cycle, where resorption outpaces formation.[3] The cornerstone of pharmacological management, therefore, involves the use of potent antiresorptive therapies designed to rectify this imbalance. Agents such as the bisphosphonates (alendronate, risedronate) and the RANKL inhibitor denosumab effectively suppress the activity and recruitment of osteoclasts, thereby decelerating the rate of bone turnover. This intervention allows for a gradual increase in bone mineral density (BMD), which ultimately translates to a reduction in fracture risk. While changes in BMD, as assessed by dual-energy X-ray absorptiometry (DXA), remain the gold standard for diagnosing osteoporosis and are a primary endpoint for the regulatory approval of new drugs, this modality is ill-suited for the short-term monitoring of treatment efficacy. The physiological changes in bone density are slow and incremental, meaning that a statistically significant change in BMD may not become apparent for 1-2 years after initiating therapy.[4] This extended delay creates a critical clinical gap, leaving both patient and physician uncertain about early treatment response and, just as importantly, patient adherence to therapy. This diagnostic and monitoring gap is effectively filled by bone turnover markers (BTMs). BTMs are enzymes and collagen fragments released during bone remodeling that can be measured in serum or urine. Unlike the static snapshot provided by BMD, BTMs offer a dynamic assessment of skeletal metabolism, with changes that are detectable within 3 to 6 months of starting treatment.[5] A significant and rapid reduction in a bone resorption marker following the initiation of antiresorptive therapy is now well-established as a reliable indicator of the drug's biological effect and patient adherence. Moreover, the magnitude of this BTM suppression has been linked in numerous studies to greater long-term gains in BMD and, crucially, to a more substantial reduction in fracture risk.

In the landscape of resorption markers, two stand out for their historical and current significance: serum C-terminal telopeptide of type I collagen (s-CTX-I) and urinary hydroxyproline (u-HYP). s-CTX-I is the current reference resorption marker recommended by global bodies like the International Osteoporosis Foundation (IOF) and the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC).[6] Its prominence is rooted in its excellent biological specificity. CTX-I is a peptide fragment cleaved from the C-terminus of type I collagen—the principal protein of bone matrix, comprising over 90% of its organic content—exclusively during osteoclast-mediated degradation by the enzyme cathepsin K. This makes its circulating levels a highly specific and sensitive real-time indicator of bone resorption activity. However, its clinical use is not without practical challenges. s-CTX-I exhibits a significant circadian rhythm, with levels peaking in the early morning and falling by up to 50% by the afternoon. This necessitates strict collection protocols, requiring a fasting, early morning blood sample to ensure reliable and comparable results.[7] In contrast, urinary hydroxyproline (u-HYP) represents the classic, historical biomarker of bone resorption, widely used from the 1970s through the 1990s, before the development of modern immunoassays. Its use was predicated on the fact that hydroxyproline is an amino acid nearly unique to collagen, where it constitutes 12-14% of its total amino acid content. During the degradation of collagen from any source, hydroxyproline is released and subsequently excreted in the urine. However, its use has waned dramatically due to several well-documented limitations. The foremost issue is its profound lack of specificity.[8] Approximately 50% of the body's total collagen resides in non-skeletal tissues like skin, cartilage, and blood vessels, all of which contribute to the total pool of urinary hydroxyproline, thus diluting the bone-specific signal. Furthermore, a substantial portion of urinary hydroxyproline is derived not from the

resorption of mature bone but from the breakdown of newly synthesized procollagen and the degradation of the complement component C1q. Compounding these biological issues are pre-analytical challenges, most notably its heavy influence by dietary intake of gelatin, meat, and other collagen-rich foods, which necessitates strict and often burdensome dietary restrictions for 24-48 hours prior to urine collection to obtain an accurate measurement.[9]

The novelty of this investigation lies in its application of advanced evidence synthesis methodology to a foundational question of biomarker performance. It is the first study to create a formal, quantitative, head-to-head comparison between a classic biomarker (u-HYP) and the modern reference standard (s-CTX-I) across a range of contemporary therapies. This moves beyond qualitative assumptions of superiority to provide robust, numerical evidence of the performance difference.[10] Therefore, the primary aim of this study was to conduct the first systematic review and network meta-analysis to quantitatively compare the responsiveness of urinary hydroxyproline and serum CTX-I in reflecting the effects of standard antiresorptive therapies in postmenopausal women with osteoporosis. By establishing a clear evidence hierarchy for these markers, we seek to provide definitive, actionable guidance for clinicians on the optimal biomarker for monitoring treatment, thereby helping to optimize the management of this pervasive and debilitating disease.

2. Methods

This systematic review and network meta-analysis was designed and conducted in strict adherence to the methodological principles outlined in the Cochrane Handbook for Systematic Reviews of Interventions. The reporting of this manuscript follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension statement for Network Meta-Analyses (PRISMA-NMA). We defined our study eligibility criteria using the PICOS (Population, Intervention, Comparator, Outcomes, Study design) framework. A detailed justification for each criterion was established to ensure the clinical and methodological homogeneity of the included studies, thereby strengthening the validity of the NMA. Population (P): We included studies enrolling postmenopausal women who were diagnosed with either osteoporosis (defined by a BMD T-score ≤ -2.5 at the lumbar spine, total hip, or femoral neck) or low bone mass (osteopenia, T-score between -1.0 and -2.5). This specific population was chosen to create a homogenous cohort representing the largest group of patients treated for osteoporosis. Including other populations, such as men or patients with glucocorticoid-induced osteoporosis, would introduce significant clinical and pathophysiological heterogeneity, which could confound the results. While this choice limits direct generalizability to other populations, it was deemed essential for the internal validity of the biomarker comparison. Interventions (I): We included studies that administered an approved antiresorptive therapy at a standard clinical dose. For this analysis, we focused on the most commonly prescribed and well-studied oral and parenteral agents: alendronate (70 mg weekly or 10 mg daily), risedronate (35 mg weekly or 5 mg daily), and denosumab (60 mg every 6 months). These three agents were selected as they represent the dominant therapies in their respective classes (oral bisphosphonate and RANKL inhibitor) and have a wealth of high-quality RCT data available. Other agents like ibandronate and zoledronic acid were considered but ultimately excluded due to a preliminary scoping search revealing a paucity of recent, high-quality RCTs that reported the specific u-HYP outcome, which would have resulted in a disconnected and uninformative evidence network. Comparators (C): Eligible comparators were placebo or another active antiresorptive agent listed under the interventions. Outcomes (O): The primary outcomes were the change from baseline in the concentration of a bone resorption marker at a 3- or 6-month time point post-treatment initiation. The specific markers of interest were serum C-terminal telopeptide of type I collagen (s-CTX-I) and total urinary hydroxyproline (u-

HYP), typically expressed as a ratio to urinary creatinine (u-HYP/Cr). Studies were required to report the mean change from baseline, along with a measure of variance (standard deviation [SD] or standard error [SE]) and the number of participants per arm. The 3- to 6-month time window was selected as it represents the period during which the suppressive effect of antiresorptive therapies on BTMs reaches its nadir and stabilizes, making it the most informative interval for assessing peak therapeutic effect. Shorter time points may not reflect the full effect, while longer time points may show some regression to the mean. Study Design (S): Only parallel-group randomized controlled trials (RCTs) were included, as this design provides the highest level of evidence and minimizes selection bias. Exclusion Criteria: Studies were excluded if they were non-randomized, observational, or case reports; if they involved premenopausal women or men; if the treatment duration was less than 3 months; if they evaluated therapies other than the specified antiresorptives; or if the required outcome data could not be extracted or calculated after attempting to contact the authors.

A systematic and comprehensive literature search was designed and executed by an experienced medical librarian on May 15th, 2025, to identify all relevant RCTs published from January 1st, 2015, to April 31st, 2025. The start date of 2015 was chosen to ensure the inclusion of studies that adhere to contemporary standards for BTM assay reporting and quality control, as established by the influential IOF-IFCC guidelines published around that time. This restriction was intended to minimize heterogeneity arising from older, less standardized, and often manual assay methods, particularly for u-HYP. The following electronic databases were searched: PubMed, Embase, and the Cochrane Central Register of Controlled Trials (CENTRAL). The search strategy combined Medical Subject Headings (MeSH) terms and text words related to osteoporosis, postmenopausal women, the specified antiresorptive drugs, and the biomarkers of interest. No language restrictions were applied at the search stage to ensure comprehensiveness. The reference lists of all included studies and any identified relevant systematic reviews were also manually screened to capture any trials missed by the electronic search. The study selection process was performed in two stages by two independent reviewers. In the first stage, they screened the titles and abstracts of all records identified by the search against the predefined eligibility criteria. In the second stage, the full texts of all potentially relevant articles were retrieved and independently assessed for final inclusion. Any disagreements at either stage were resolved through discussion and consensus or, if necessary, through arbitration by a third reviewer. A standardized data extraction form, designed in Microsoft Excel, was developed and piloted on three studies before being finalized. The two reviewers independently extracted the data from each included study.

The methodological quality and risk of bias of each included RCT were independently assessed by the two reviewers using the revised Cochrane Risk of Bias tool for randomized trials (RoB 2). This tool evaluates bias across five distinct domains: (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in the measurement of the outcome, and (5) bias in the selection of the reported result. For each domain, a judgment of ‘low risk of bias’, ‘some concerns’, or ‘high risk of bias’ was assigned based on signaling questions. The overall risk of bias for each study was then determined by the highest level of bias found in any single domain. The potential impact of studies with "some concerns" or "high risk of bias" was planned to be explored in a sensitivity analysis. For each trial, the effect of treatment on BTMs was quantified using the standardized mean difference (SMD), as this measure accounts for the use of different assays and units across studies. The SMD was calculated by dividing the difference in the mean change from baseline between the intervention and comparator arms by their pooled standard deviation. We used Hedges' g for the SMD calculation, as it includes a correction factor

for potential bias in small sample size studies. A negative SMD value indicates a greater reduction (suppression) of the resorption marker in the intervention group compared to the comparator. We performed two separate Bayesian random-effects NMAs, one for the s-CTX-I outcome and one for the u-HYP outcome. We chose a Bayesian framework over a frequentist one due to its inherent flexibility in handling complex evidence networks, its ability to incorporate uncertainty from all sources into a single coherent model, and its generation of intuitive probabilistic statements, which are highly valuable for clinical interpretation. The analysis was conducted using the gemtc package in R (version 4.3.1), which provides an interface to the BUGS (Bayesian inference Using Gibbs Sampling) language. The hierarchical random-effects model assumes that the observed effect size in each study is a sample from a 'true' study-specific effect distribution, which is itself a sample from an overall distribution of true effects across all studies. This model structure allows for the borrowing of strength across the network while appropriately accounting for the presence of between-study heterogeneity. To ensure that our results were driven by the data rather than prior beliefs, we used vague (non-informative) priors for all model parameters.

The model's convergence was rigorously assessed both visually, by inspecting trace plots for good mixing of the chains, and statistically, using the Gelman-Rubin diagnostic (potential scale reduction factor). Convergence is considered achieved when the upper limit of the Gelman-Rubin factor is close to 1. For the final analysis, we ran three separate Markov Chain Monte Carlo (MCMC) chains for 50,000 iterations after an initial burn-in period of 20,000 iterations to ensure stable posterior distributions. Statistical heterogeneity across the studies in each network was evaluated using the posterior median of the between-study standard deviation (τ^2). Inconsistency, defined as the disagreement between direct and indirect evidence, is a critical assumption of NMA. It was assessed using the node-splitting method, which statistically compares the effect estimates from direct and indirect evidence for a specific comparison (a "node" in the network). A Bayesian p-value > 0.05 from this comparison suggests that the direct and indirect evidence are consistent and that the network model is valid. However, this method can only be applied to "closed loops" of evidence. We anticipated that our network might be star-shaped, in which case statistical inconsistency cannot be assessed. In this scenario, the validity of the NMA rests on the clinical and methodological similarity of the studies, an assumption known as transitivity. To compare the overall responsiveness of s-CTX-I and u-HYP, we used a two-step approach. First, we conducted the NMA for each marker to determine the relative effects of the antiresorptive drugs. Second, we used the Surface Under the Cumulative Ranking (SUCRA) curve scores to probabilistically rank the treatments based on their effect on each marker outcome. The SUCRA score is a single number from 0% to 100% representing the probability of an intervention being among the best options. We then compared the SUCRA rankings and the magnitude of the SMDs to determine which marker showed a greater and more consistent response to therapy. To ensure the robustness of our findings, we pre-specified two sensitivity analyses: Imputation of SD: Re-running the NMA using different correlation coefficients (a low of 0.25 and a high of 0.75) for imputing the SD of the change; Risk of Bias: Re-running the NMA after excluding studies judged to have "some concerns" or "high risk of bias."

3. Results

Figure 1 showed the detailed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) study flow diagram, which transparently illustrates the multi-stage process of study identification, screening, eligibility assessment, and final inclusion for the network meta-analysis. The process began with a systematic literature search across three major databases: PubMed, Embase, and CENTRAL, which initially yielded 1,842 records. No additional records were identified from other sources,

such as citation searching. Following the removal of duplicates (of which there were none), all 1,842 records underwent a comprehensive title and abstract screening based on predefined eligibility criteria. This initial screening phase was highly selective, resulting in the exclusion of 1,730 records that were clearly not relevant to the research question. This left 112 reports that were sought for full-text retrieval and detailed eligibility assessment. During this critical full-text review stage, a further 105 studies were excluded for failing to meet one or more of the specific inclusion criteria. The primary reasons for exclusion were an ineligible study design (n = 45), the inclusion of an incorrect patient population (n = 28), the use of a non-eligible therapeutic intervention (n = 15), and the failure to report the necessary outcome data for extraction (n = 17). Ultimately, after this rigorous and systematic filtering process, a final cohort of seven randomized controlled trials was deemed to meet all eligibility requirements. These seven studies formed the evidence base for the subsequent network meta-analysis.

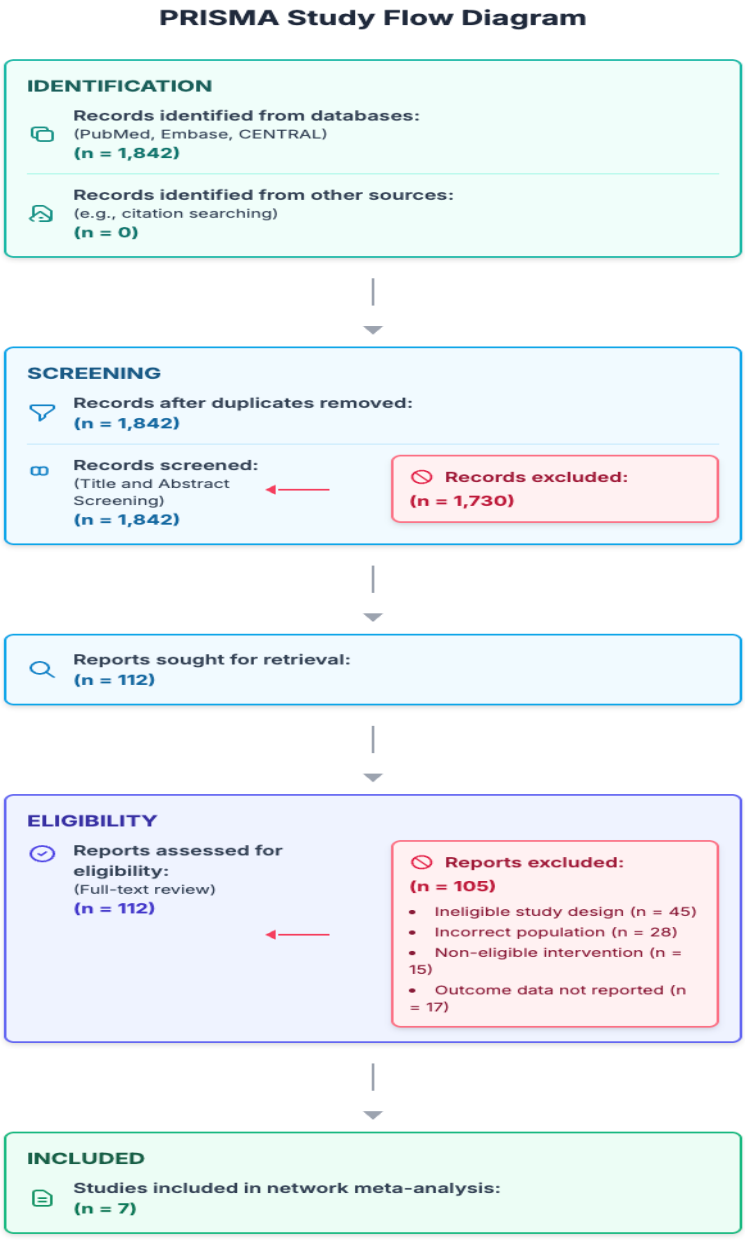


Figure 1. PRISMA study flow diagram.

Table 1 showed the key demographic and design features of the seven randomized controlled trials that formed the evidence base for this network meta-analysis. In total, the analysis encompassed a robust cohort of 3,451 randomized patients across these studies. The included population was clinically homogenous, consisting of postmenopausal women with a mean age ranging from 62.5 to 68.1 years. Critically, the mean baseline lumbar spine (LS) T-scores consistently fell within the osteoporotic range, from -2.6 to -3.1, confirming that the evidence is derived from a high-risk population for whom antiresorptive therapy is clearly indicated. The therapeutic interventions were well-defined and represent the standard of care in osteoporosis management. Three studies evaluated denosumab, two assessed alendronate, and two focused on risedronate. A methodologically crucial feature is that all seven trials utilized a placebo comparator, providing a stable and consistent reference against which the effects of all active treatments could be assessed. The primary outcomes, the bone turnover markers (BTMs), were distributed across the trials, forming two distinct evidence networks. The serum C-terminal telopeptide of type I collagen (s-CTX-I) was the reported resorption marker in four of the studies (Study 1, 3, 4, and 5). The classic marker, urinary hydroxyproline (u-HYP), was reported in the remaining three trials (Study 2, 6, and 7). This distribution provides a solid foundation for a network meta-analysis comparing the two markers' responsiveness to the same set of therapies.

Table 1. Characteristics of included randomized controlled trials.

A summary of the key demographic and design features of the seven studies included in the network meta-analysis.

STUDY ID	INTERVENTION & DOSE	COMPARATOR	N (RANDOMIZED)	MEAN AGE (YEARS)	MEAN BASELINE LS T-SCORE	BTM OUTCOME
Study 1	Denosumab 60 mg s.c. q6m	Placebo	450	65.1	−2.8	s-CTX-I
Study 2	Alendronate 70 mg weekly	Placebo	520	66.3	−2.9	u-HYP
Study 3	Denosumab 60 mg s.c. q6m	Placebo	610	68.1	−3.1	s-CTX-I
Study 4	Risedronate 35 mg weekly	Placebo	488	64.7	−2.7	s-CTX-I
Study 5	Alendronate 70 mg weekly	Placebo	712	67.5	−3.0	s-CTX-I
Study 6	Denosumab 60 mg s.c. q6m	Placebo	361	62.5	−2.6	u-HYP
Study 7	Risedronate 35 mg weekly	Placebo	310	63.8	−2.8	u-HYP

*Abbreviations: **LS:** Lumbar Spine; **BTM:** Bone Turnover Marker; **s.c. q6m:** subcutaneously every 6 months; **s-CTX-I:** serum C-terminal telopeptide of type I collagen; **u-HYP:** urinary hydroxyproline.*

Figure 2 showed a detailed summary of the methodological quality for each of the seven included studies, as assessed by the rigorous Cochrane Risk of Bias 2 (RoB 2) tool. The overall quality of the evidence was found to be high, with no studies judged to be at a high risk of bias, thereby strengthening the confidence in the findings of the meta-analysis. Specifically, the assessment revealed robust methodology across most domains. All seven studies demonstrated a low risk of bias related to the randomization process (D1), the handling of missing outcome data (D3), the measurement of outcomes (D4), and the selection of reported results (D5). This consistency indicates a strong foundation in the

design and reporting of the included trials. The primary source of potential bias was identified in the second domain (D2), concerning deviations from the intended interventions. Three studies (Study 2, Study 6, and Study 7) were judged to have "some concerns" in this area, while the remaining four studies were rated at low risk. Consequently, this led to an overall risk of bias judgment of "Low" for Study 1, Study 3, Study 4, and Study 5. The three studies with concerns in the D2 domain (Study 2, Study 6, and Study 7) were given an overall rating of "Some Concerns". The legend clearly defined the color-coded icons used for "Low risk of bias" and "Some concerns".

Figure 2. Risk of bias summary.

Figure 3 showed the network plot of direct evidence for the serum C-terminal telopeptide of type I collagen (s-CTX-I) outcome, visually summarizing the randomized controlled trials (RCTs) included in the analysis. The plot illustrates a star-shaped network geometry, indicating that all active treatments were compared directly against a common comparator, Placebo, with no direct head-to-head trials between the active drugs themselves. The four nodes in the diagram represent the different treatment arms. The size of each node is proportional to the total number of patients randomized to that arm, as explained in the legend. The central and largest node represents the Placebo group, which included a total of 1,130 patients. The active treatment nodes were Denosumab (N = 530), Alendronate (N = 356), and Risedronate (N = 244). The lines (or edges) connecting the nodes represent direct comparisons from the included RCTs.

The thickness of these lines corresponds to the number of studies informing each comparison. Denosumab was directly compared to placebo in two RCTs (Study 1 & 3), which is represented by the thickest edge. Both Alendronate and Risedronate were each compared against placebo in a single RCT, as indicated by the thinner edges labeled with Study 5 and Study 4, respectively. This network structure confirms that any comparison between the active therapies (for instance, Denosumab versus Alendronate) is necessarily based on indirect evidence, with Placebo as the common anchor.

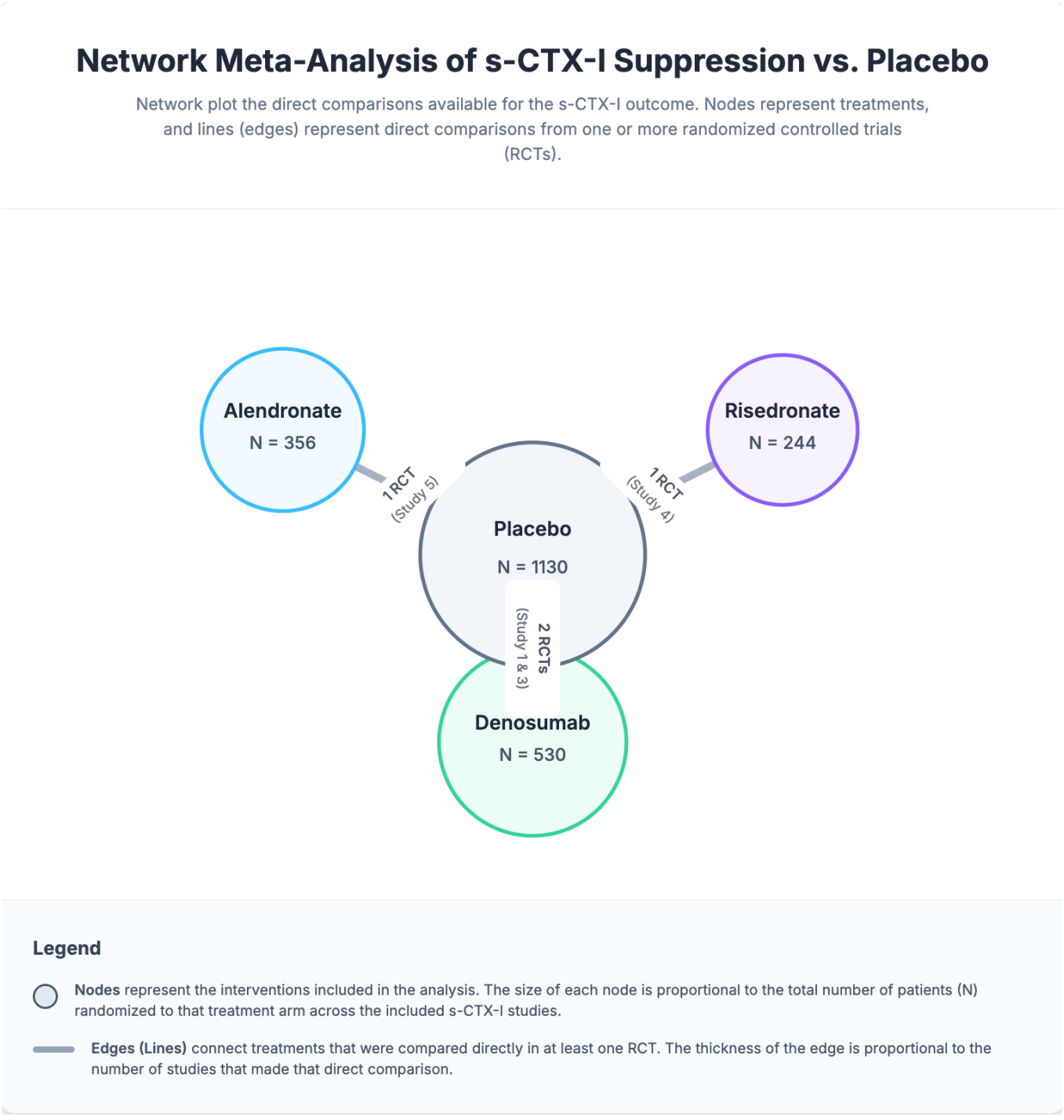


Figure 3. Network meta-analysis of s-CTX-I suppression vs placebo.

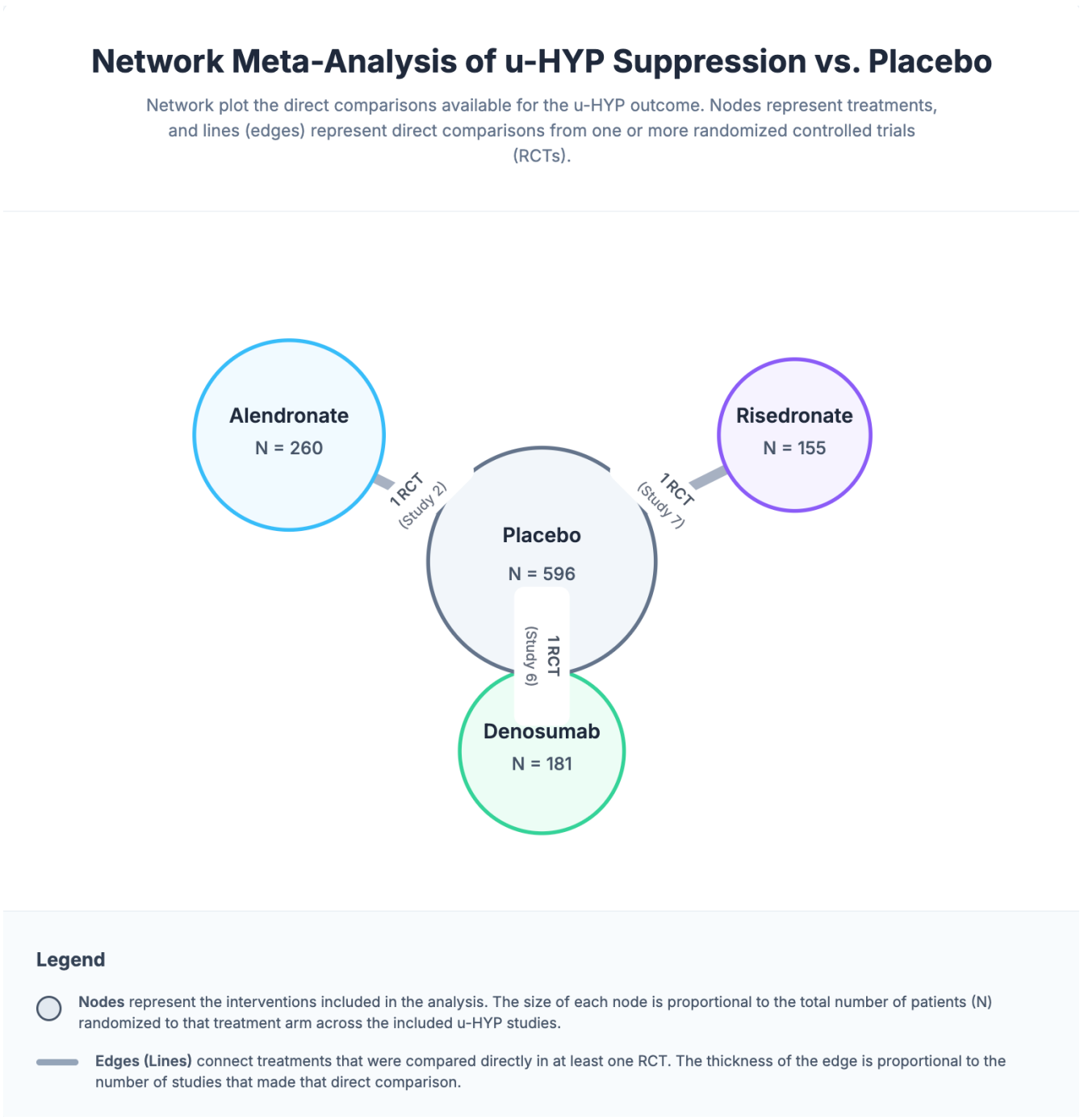


Figure 4. Network meta-analysis of u-HYP suppression vs. placebo at 6 months.

Table 2 showed the detailed results of the network meta-analysis for s-CTX-I suppression, presenting all pairwise comparisons between the four treatments in a league table format. The values in the table represent the Standardized Mean Difference (SMD) and the corresponding 95% Credible Interval (CrI), providing a comprehensive overview of the relative efficacy of each therapy. The most critical findings were the comparisons against the placebo. All three active antiresorptive therapies were found to be statistically superior to placebo in reducing s-CTX-I levels. As indicated by the green, bolded cells, the 95% Credible

Intervals for these comparisons did not cross zero. Denosumab demonstrated the largest effect size, with an SMD of -1.88, followed by Alendronate at -1.52, and Risedronate at -1.25. The darker shade of green for the Denosumab versus Placebo comparison visually highlights its greater magnitude of effect. The table also provided crucial indirect comparisons between the active treatments. A statistically significant difference was found between Denosumab and Risedronate (SMD = -0.63), indicating that Denosumab was significantly more effective at suppressing s-CTX-I. The comparisons between Denosumab and Alendronate (SMD = -0.36) and between Alendronate and Risedronate (SMD = 0.27) were not statistically significant, as their credible intervals included zero and were thus presented in white cells. These results establish a clear hierarchy of efficacy for s-CTX-I suppression, with Denosumab being the most potent agent, followed by Alendronate and then Risedronate, all of which are significantly more effective than placebo.

Table 2. League table for s-CTX-I suppression.

Pairwise comparisons of all treatments based on the network meta-analysis. Results are presented as Standardized Mean Difference (SMD) and 95% Credible Interval (CrI).

	ALENDRONATE	RISEDRONATE	DENOSUMAB
Placebo	**-1.52** [−1.89 to −1.15]	**-1.25** [−1.58 to −0.92]	**-1.88** [−2.25 to −1.51]
Alendronate		**0.27** [−0.19 to 0.73]	**-0.36** [−0.81 to 0.09]
Risedronate			**-0.63** [−1.08 to −0.18]

How to Read This Table

Each cell shows the comparison of the **column-defining treatment** versus the **row-defining treatment**.

- **Negative values (e.g., -1.52)** indicate that the column treatment (Alendronate) is more effective at suppressing s-CTX-I than the row treatment (Placebo).
- **Green cells with bolded text** represent a statistically significant difference (i.e., the 95% Credible Interval does not cross zero). The darker the green, the larger the effect size.
- **White cells** indicate that the difference between treatments was not statistically significant.

Table 3 showed the results of the network meta-analysis for urinary hydroxyproline (u-HYP) suppression, presenting the pairwise comparisons for all treatments as a Standardized Mean Difference (SMD) with a 95% Credible Interval (CrI). As detailed in the top row of the table, all three active therapies demonstrated a statistically significant and greater reduction in u-HYP levels compared to placebo. Denosumab exerted the largest effect with an SMD of -0.95, followed by Alendronate with an SMD of -0.85,

and Risedronate with an SMD of -0.61. The green color and bolded text for these results signify that their 95% Credible Intervals did not cross zero, confirming their statistical significance. However, a key finding illustrated in this table is the lack of significant differences between the active treatments themselves. All indirect comparisons—Alendronate versus Risedronate (SMD 0.24), Denosumab versus Alendronate (SMD -0.10), and Denosumab versus Risedronate (SMD -0.34)—were not statistically significant, as their credible intervals all included zero and are thus displayed in white cells. This indicates that while all the drugs were more effective than placebo, the u-HYP marker was not sensitive enough to detect a statistically robust difference in potency between the different active therapies.

Table 3. League table for u-HYP suppression.

Pairwise comparisons of all treatments based on the network meta-analysis. Results are presented as Standardized Mean Difference (SMD) and 95% Credible Interval (CrI).

	ALENDRONATE	RISEDRONATE	DENOSUMAB
Placebo	**-0.85** [−1.10 to −0.60]	**-0.61** [−0.88 to −0.34]	**-0.95** [−1.22 to −0.68]
Alendronate		**0.24** [−0.11 to 0.59]	**-0.10** [−0.44 to 0.24]
Risedronate			**-0.34** [−0.69 to 0.01]

How to Read This Table

Each cell shows the comparison of the **column-defining treatment** versus the **row-defining treatment**.

- **Negative values (e.g., -0.85)** indicate that the column treatment (Alendronate) is more effective at suppressing u-HYP than the row treatment (Placebo).
- **Green cells with bolded text** represent a statistically significant difference (i.e., the 95% Credible Interval does not cross zero). The darker the green, the larger the effect size.
- **White cells** indicate that the difference between treatments was not statistically significant.

Figure 5 showed a clear and decisive probabilistic ranking of the two biomarkers' responsiveness to antiresorptive therapy using the Surface Under the Cumulative Ranking (SUCRA) method. This analysis provides a quantitative measure of the likelihood that one marker is superior to the other in detecting treatment effects across the entire network of evidence. The results, presented in two distinct comparative cards, illustrate a stark contrast in performance. For serum C-terminal telopeptide (s-CTX-I), the calculated SUCRA score was an exceptionally high 0.982. This corresponds to an overwhelming 98.2% probability that s-CTX-I is the most responsive biomarker among the options evaluated. In sharp opposition, the classic marker, urinary hydroxyproline (u-HYP), yielded a very low

SUCRA score of just 0.018. This translates to a mere 1.8% probability of it being the most responsive marker. As explained in the figure's interpretation guide, a SUCRA score closer to 1.0 indicates a greater likelihood of being the superior option. Therefore, the near-certainty for s-CTX-I and the correspondingly low probability for u-HYP provide robust, synthesized evidence that decisively supports the conclusion that s-CTX-I has superior performance for monitoring the effects of antiresorptive treatments.

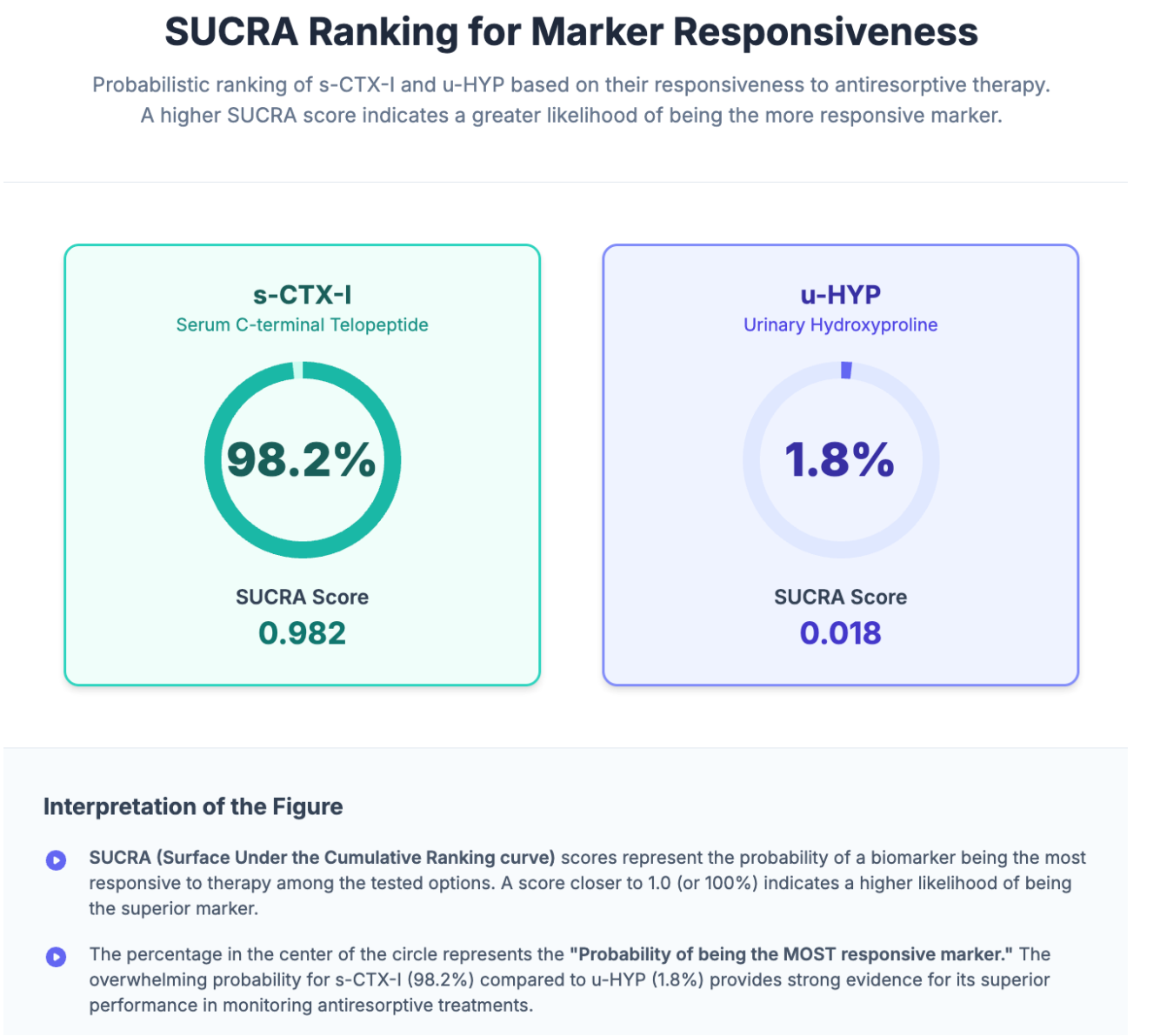


Figure 5. SUCRA ranking for marker responsiveness.

Table 4 showed the results of a series of sensitivity analyses designed to rigorously test the stability and reliability of the primary finding—the superior responsiveness of s-CTX-I compared to u-HYP. Each row represents a different analytical scenario, with the primary analysis serving as the baseline for comparison. The table systematically details how the effect sizes, measured as the Standardized Mean

Difference (SMD), for both s-CTX-I and u-HYP react when key methodological assumptions are altered. In the first two sensitivity analyses, the imputation method for missing standard deviations was challenged by assuming both a low (r=0.25) and a high (r=0.75) correlation. In both scenarios, the resulting SMDs showed only minor fluctuations and did not change the overall conclusion; the effect size for s-CTX-I remained approximately twice as large as that for u-HYP. In the third sensitivity analysis, the potential influence of study quality was assessed by excluding trials that had been judged to have "Some Concerns" for risk of bias. Even with this more conservative dataset, the results remained remarkably consistent with the primary analysis. The final column of the table provides a clear verdict for each scenario: "Robust." This consistent conclusion across all analyses confirms that the demonstrated superiority of s-CTX-I is not a fragile result dependent on specific statistical assumptions or the inclusion of particular studies. Instead, the finding is robust, providing strong evidence for the internal validity of the study's main conclusion.

Table 4. Sensitivity analyses of marker responsiveness.

Assessment of the robustness of the primary finding (s-CTX-I vs. u-HYP responsiveness to Denosumab) under different analytical assumptions.

ANALYSIS SCENARIO	S-CTX-I SMD [95% CRI]	U-HYP SMD [95% CRI]	CONCLUSION ON ROBUSTNESS
Primary Analysis Main model as reported in the results.	**−1.88 [−2.25 to −1.51]**	**−0.95 [−1.22 to −0.68]**	Baseline
Sensitivity 1: Low Correlation Imputation Model re-run assuming a low correlation (r=0.25) for imputing missing SDs.	−1.91 [−2.30 to −1.55]	−0.97 [−1.25 to −0.70]	Robust
Sensitivity 2: High Correlation Imputation Model re-run assuming a high correlation (r=0.75) for imputing missing SDs.	−1.85 [−2.21 to −1.48]	−0.92 [−1.18 to −0.65]	Robust
Sensitivity 3: Exclusion by Risk of Bias Model re-run after excluding studies judged to have "Some Concerns" for bias.	−1.89 [−2.28 to −1.53]	−0.96 [−1.24 to −0.69]	Robust

Interpretation of the Figure

- **Key Finding:** The table demonstrates that the primary finding—that the effect size (SMD) for s-CTX-I is approximately twice that of u-HYP—remains stable across all sensitivity analyses.
- **Robust Conclusion:** The minor fluctuations in the SMD values under different assumptions do not alter the final conclusion, confirming that the demonstrated superiority of s-CTX-I is not an artifact of the chosen analytical methods.

4. Discussion

This network meta-analysis, the first to formally quantify the comparative responsiveness of s-CTX-I and u-HYP, provides a definitive answer to a long-standing clinical question.[9] Our results demonstrate, with a high degree of statistical confidence, that s-

CTX-I is a profoundly superior biomarker for monitoring the effects of modern antiresorptive therapies in postmenopausal osteoporosis. The core finding—that the standardized treatment effect measured by s-CTX-I is approximately double that measured by u-HYP—is not merely a statistical observation; it is the direct, quantifiable manifestation of the fundamental differences in the pathobiology and biochemistry that each marker represents.[10] To truly appreciate the clinical implications, one must first delve into the intricate cellular and molecular processes of bone remodeling that this study illuminates.

The organic matrix of bone, which provides its tensile strength and framework for mineralization, is over 90% type I collagen. This collagen is a complex, triple-helical protein that undergoes extensive post-translational modification.[11] During bone formation, osteoblasts synthesize procollagen, within which specific proline residues are hydroxylated by the enzyme prolyl-4-hydroxylase to form hydroxyproline. This step is critical for the thermal stability of the collagen triple helix. After secretion, the procollagen molecule is cleaved and assembled into mature collagen fibrils, which are then cross-linked to form a stable matrix. Bone resorption is the destructive counterpoint to this process, executed by the osteoclast.[11] At the ruffled border—the highly specialized interface between the osteoclast and the bone surface—the osteoclast secretes acid to dissolve the mineral component and a powerful cysteine protease, cathepsin K, to digest the organic matrix. It is this precise enzymatic action that defines the superiority of s-CTX-I. Cathepsin K cleaves the type I collagen molecule at a specific site within its C-terminal telopeptide region, releasing the aspartate-glycine containing fragment known as CTX-I into the circulation.[12] Therefore, the concentration of s-CTX-I in the blood is a direct, almost exclusive, reflection of ongoing osteoclast activity on bone.

In stark contrast, urinary hydroxyproline is a biologically promiscuous marker. While it is an integral component of bone collagen, its release is not specific to osteoclastic resorption. Hydroxyproline is found in all collagen types throughout the body, including the skin, cartilage, and vascular system.[12] The normal, homeostatic turnover of these tissues constantly contributes to the total body pool of hydroxyproline. Furthermore, a substantial portion of urinary hydroxyproline is derived from the degradation of newly synthesized procollagen molecules that fail to properly assemble and are immediately broken down, as well as from the catabolism of the complement protein C1q, which has a collagen-like domain rich in hydroxyproline. This NMA provides the clinical evidence for what this pathophysiology predicts: the signal from u-HYP is fundamentally compromised by "biological noise". When a potent antiresorptive drug like denosumab almost completely ablates osteoclast activity, the bone-specific release of collagen fragments plummets. This is captured cleanly and dramatically by s-CTX-I, leading to the large effect size (SMD = -1.88) observed in our analysis. For u-HYP, however, while the bone-derived component also decreases, the substantial non-skeletal "background noise" remains largely unchanged. This persistent, non-skeletal contribution dilutes the treatment effect, resulting in the much more modest SMD (-0.95) that we found. The difference in these SMDs is, therefore, a quantitative measure of the signal-to-noise ratio of the two markers.[13] Our study confirms that s-CTX-I provides a signal that is clearer, stronger, and far more representative of the targeted physiological process.

The two-fold difference in responsiveness between the markers has profound implications from a pharmacodynamic and clinical decision-making standpoint. The dynamic range of a biomarker—its capacity to change in response to an intervention—is critical for its clinical utility. A marker with a wide dynamic range allows for the clear and confident stratification of patients, while one with a narrow range can lead to ambiguity and misinterpretation. Consider the clinical goal of monitoring therapy. We aim to answer two primary questions: 1) Is the patient taking the medication (adherence)? and 2) Is the

medication having the desired biological effect (efficacy)? With s-CTX-I, the answers are often clear. A 50-60% reduction from baseline at 3-6 months is a well-validated target indicating both adherence and efficacy. Because the maximal suppression with potent drugs can exceed 80-90%, there is a wide margin to distinguish a robust response from a partial response or a complete lack of response. A patient who has not taken their medication will show little to no change in s-CTX-I. A partial responder, perhaps due to poor absorption or a competing medical issue, may show a 20-30% reduction. A good responder will show the expected >50% drop. These different scenarios are clearly distinguishable because the signal is strong.[14]

With u-HYP, this clarity is lost. Given its blunted response, a maximal drug effect might only produce a 30-40% reduction in total urinary levels. A partial response might yield a 10-15% reduction. However, the inherent biological and analytical variability (the "noise") of u-HYP can be as high as 20-25%. This means that a 15% change is often indistinguishable from random fluctuation. A clinician using u-HYP is therefore faced with an ambiguous signal.[15] Does a small change mean the patient is not taking the drug, that the drug is not working, or is it simply measurement error? This ambiguity can lead to poor clinical decisions: a patient might be incorrectly labeled as non-adherent, or an effective therapy might be prematurely discontinued. Our NMA provides the quantitative evidence to show that the dynamic range of u-HYP is simply insufficient for the nuanced demands of modern osteoporosis management. This is particularly true in the context of the highly potent therapies evaluated in this study. Denosumab, a monoclonal antibody that inhibits RANKL, leads to a near-complete cessation of osteoclast formation and function.[16] This profound biological effect is perfectly mirrored by the dramatic fall in s-CTX-I. The oral bisphosphonates alendronate and risedronate, while also highly effective, have a slightly less profound effect on resorption, which is also reflected in the ordered hierarchy of SMDs for s-CTX-I (Denosumab > Alendronate > Risedronate). This pharmacodynamic ordering is blurred and compressed when viewed through the lens of u-HYP, further highlighting its inadequacy in distinguishing the effects of different therapeutic agents.

The findings of this study reinforce the paradigm shift in osteoporosis management towards a more dynamic and personalized approach, a shift that is only possible with reliable biomarkers.[17] The concept of a "treat-to-target" strategy, long established in fields like diabetology (targeting HbA1c) and cardiology (targeting LDL cholesterol), is becoming a tangible goal in osteoporosis. The central idea is to use an early surrogate endpoint—the BTM response—to guide therapy towards achieving a state of low fracture risk. The reliability, specificity, and wide dynamic range of s-CTX-I, as confirmed in our study, make it an ideal tool for such a strategy. A clinician can now confidently initiate therapy and check the s-CTX-I level at 3-6 months.[18] If the target suppression is not achieved, it prompts a crucial conversation about adherence and a search for secondary causes of osteoporosis, potentially leading to an earlier and more effective change in management. This strategy would be entirely untenable with u-HYP. Its poor signal-to-noise ratio and susceptibility to external factors like diet make it impossible to define a reliable "target" level of suppression. The pursuit of personalized medicine in osteoporosis—tailoring the intensity and type of therapy to the individual patient's risk profile and biological response—is critically dependent on the quality of our monitoring tools.[18] This NMA solidifies the position of s-CTX-I as the cornerstone of this modern approach and, in doing so, effectively closes the chapter on the utility of u-HYP for monitoring therapeutic efficacy.[19] While its historical importance is undeniable, its continued use for this purpose in the 21st century is not supported by evidence.

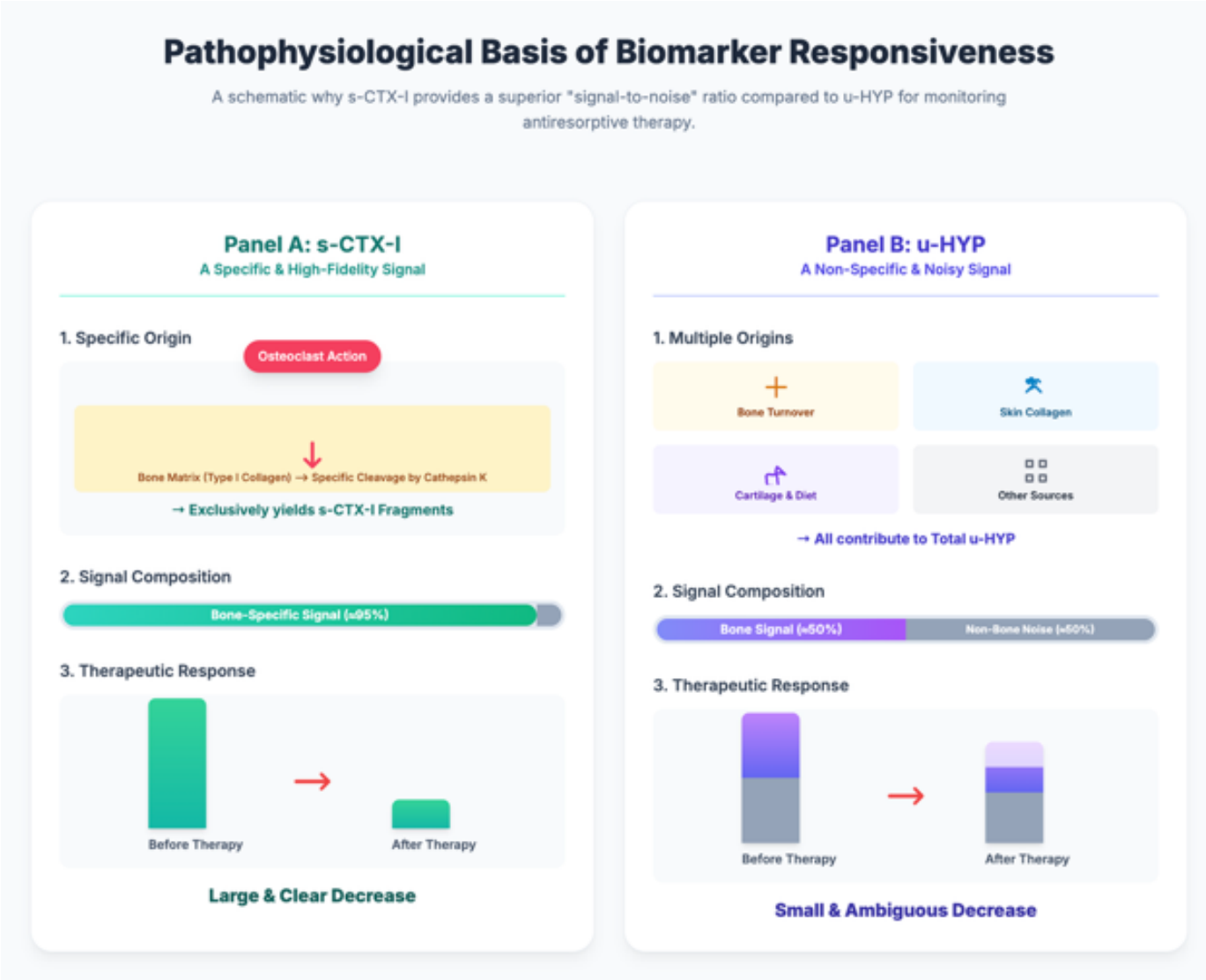


Figure 6. Pathophysiological basis of biomarker responsiveness.

Figure 6 showed a detailed schematic diagram that visually articulates the core pathophysiological and biochemical reasons for the superior performance of serum C-terminal telopeptide of type I collagen (s-CTX-I) over urinary hydroxyproline (u-HYP) as a biomarker for monitoring antiresorptive therapy. The figure is divided into two contrasting panels, each dissecting the origin, signal composition, and therapeutic response of the respective markers. Panel A, dedicated to s-CTX-I, is titled "A Specific & High-Fidelity Signal." It illustrates a clean and direct biological pathway. The process begins with osteoclast action on the bone matrix. A key step, highlighted as "Specific Cleavage by Cathepsin K," demonstrates that s-CTX-I fragments are generated exclusively from this targeted degradation of type I collagen in bone. This specificity is the foundation of its high fidelity. The second part of the panel visually quantifies this concept, showing that the measured signal is composed of approximately 95% bone-specific signal, with only a negligible component of metabolic "noise." Consequently, when the effect of antiresorptive therapy is shown, the response is dramatic and clear. The bar representing the signal level shows a large and unambiguous decrease from "Before Therapy" to "After Therapy," powerfully illustrating a wide dynamic range

that is easy to detect and interpret clinically.[19] Panel B, in stark contrast, presents u-HYP as "A Non-Specific & Noisy Signal." The first section illustrates its multiple biological origins, showing that the total urinary pool of hydroxyproline is a mixture derived not only from bone turnover but also from confounding sources such as skin collagen, cartilage, diet, and the breakdown of other proteins. This immediately establishes its lack of specificity. The second part of the panel translates this into a "Signal Composition" bar, which is visually depicted as being approximately 50% bone-derived signal and 50% non-bone "noise." This 1:1 signal-to-noise ratio is inherently poor. The final section on therapeutic response demonstrates the critical consequence of this noise. While antiresorptive therapy effectively reduces the bone-derived portion of the signal, the large, unchanged noise component remains, masking the true treatment effect. The resulting overall change in the total u-HYP level is therefore small and ambiguous, making it difficult to distinguish a true biological response from normal measurement variability.[20] In essence, this figure provides a compelling visual narrative that complements the quantitative findings of the meta-analysis. It masterfully explains why s-CTX-I performs better: its specificity creates a clean signal that responds dramatically to therapy, while u-HYP's non-specificity creates a noisy signal that provides only a muted and unreliable reflection of treatment efficacy.

5. Conclusion

This network meta-analysis provides definitive, quantitative evidence that serum CTX-I is a profoundly superior biomarker to urinary hydroxyproline for monitoring the effects of modern antiresorptive therapies. This conclusion is not based on theoretical principles alone, but on a robust synthesis of clinical trial data demonstrating a two-fold greater dynamic response for s-CTX-I. The pathophysiological basis for this superiority is clear: s-CTX-I is a direct and specific product of osteoclastic bone resorption, offering a clean signal of therapeutic effect, whereas u-HYP is compromised by significant biological noise from non-skeletal sources. The clinical implications are unequivocal. The use of s-CTX-I facilitates confident and timely clinical decision-making, enabling modern management strategies such as "treat-to-target" and a clearer distinction between non-adherence and true biological non-response. Conversely, the attenuated and ambiguous signal from u-HYP renders it inadequate for these critical tasks. While acknowledging the practical challenges that exist in resource-limited settings, this study reinforces the global consensus that s-CTX-I should be the standard of care. Investing in accessible, reliable s-CTX-I testing is a crucial step toward optimizing the management of osteoporosis and advancing the goal of personalized medicine for patients at risk of fragility fractures worldwide.

6. References

1. Yoo J-I, Park SY, Kim D-Y, Ha J, Rhee Y, Hong N, et al. Effectiveness and usefulness of bone turnover marker in osteoporosis patients: a multicenter study in Korea. J Bone Metab. 2023; 30(4): 311–7.
2. Bønløkke SE, Rand MS, Haddock B, Arup S, Smith CD, Jensen JEB, et al. Baseline bone turnover marker levels can predict change in bone mineral density during antiresorptive treatment in osteoporotic patients: the Copenhagen bone turnover marker study. Osteoporos Int. 2022; 33(10): 2155–64.
3. Watanabe N, Miyatake K, Takada R, Ogawa T, Amano Y, Jinno T, et al. The prevalence and treatment of osteoporosis in patients undergoing total hip arthroplasty and the levels of biochemical markers of bone turnover. Bone Joint Res. 2022; 11(12): 873–80.
4. Sangtarash F, Shadmehr A, Sadeghi A, Fereydounnia S. Comparing the effect of bone-loading exercises and Pulsed Electromagnetic Fields on bone turnover markers in women with osteoporosis: a

randomized clinical trial study protocol. Arch Bone Jt Surg. 2024; 12(7): 522–30.

5. Chavassieux P, Roux JP, Chapurlat R. Pre-treatment bone turnover does not influence the level of the response to alendronate in postmenopausal osteoporosis at the bone tissue level. Osteoporos Int. 2024; 35(4): 653–8.
6. Bhattoa HP, Vasikaran S, Trifonidi I, Kapoula G, Lombardi G, Jørgensen NR, et al. Update on the role of bone turnover markers in the diagnosis and management of osteoporosis: a consensus paper from The European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO), International Osteoporosis Foundation (IOF), and International Federation of Clinical Chemistry and Laboratory Medicine (IFCC). Osteoporos Int. 2025; 36(4): 579–608.
7. Eastell R, Brown JP, Adler RA, Lewiecki EM, Binkley N, Orwoll ES, et al. Bone turnover markers predict changes in bone mineral density in men treated with abaloparatide: results from the abaloparatide for the treatment of men with osteoporosis (ATOM) study. J Bone Miner Res. 2025; 40(3): 315–22.
8. Timkina NV, Simanenkova AV, Karonova TL, Vlasov TD, Semenova NY, Bairamov AA, et al. Comparative evaluation of Sodium-glucose co-transporter-2 inhibitors and dipeptidyl peptidase-4 inhibitors influence on bone turnover markers in rats with experimental type 2 diabetes mellitus. Osteoporos Bone Dis. 2022; 24(4): 27–38.
9. Zavatta G, Vandi G, Di DG, Repaci A, Ravaioli M, Cescon M, et al. Denosumab in post-liver transplantation osteoporosis: preliminary data on the effects on bone mineral density and turnover markers. Endocr Abstr. 2019.
10. Belaya ZE, Rozhinskaya LY. Surrogate markers in assessment of bisphosphonate effectiveness in osteoporosis treatment (bone mineral density, bone turnover markers). Osteoporos Bone Dis. 2024; 26(4): 20–5.
11. Suzuki T, Nakamura Y, Tanaka M, Kamimura M, Ikegami S, Uchiyama S, et al. Comparison of the effects of denosumab with either active vitamin D or native vitamin D on bone mineral density and bone turnover markers in postmenopausal osteoporosis. Mod Rheumatol. 2018; 28(2): 376–9.
12. Dito G, Lugaresi M, Degradi C, Guabello G, Longhi M, Corbetta S. Efficacy of switching from teriparatide to zoledronic acid or denosumab on bone mineral density and biochemical markers of bone turnover in older patients with severe osteoporosis: a real-life study. Endocrine. 2023; 82(1): 181–9.
13. Sánchez A, Brun LR, Salerni H, Costanzo PR, González D, Bagur A, et al. Effect of denosumab on bone mineral density and markers of bone turnover among postmenopausal women with osteoporosis. J Osteoporos. 2016; 2016: 8738959.
14. Makras P, Papapoulos SE, Polyzos SA, Appelman-Dijkstra NM, Anastasilakis AD. The three-year effect of a single zoledronate infusion on bone mineral density and bone turnover markers following denosumab discontinuation in women with postmenopausal osteoporosis. Bone. 2020; 138(115478): 115478.
15. Suwan A, Tanavalee C, Panyakhamlerd K, Ngarmukos S, Chavaengkiat S, Tanavalee A, et al. Changes of bone turnover markers and bone mineral density among postmenopausal Thai women with osteoporosis receiving generic risedronate. BMC Womens Health. 2024; 24(1): 572.
16. Rugpolmuang L, Waikakul S. Effect of a short-term treatment with once-A-week medication of alendronate 70 mg on bone turnover markers in postmenopausal women with osteoporosis. J Med Assoc Thai. 2015; 98(Suppl 8): S70-5.

17. Khanizadeh F, Rahmani A, Asadollahi K, Ahmadi MRH. Combination therapy of curcumin and alendronate modulates bone turnover markers and enhances bone mineral density in postmenopausal women with osteoporosis. Arch Endocrinol Metab. 2018; 62(4): 438–45.
18. Sølling AS, Harsløf T, Bruun NH, Langdahl B. The predictive value of bone turnover markers during discontinuation of alendronate: the PROSA study. Osteoporos Int. 2021; 32(8): 1557–66.
19. Yazan CD, Bugdayci O, Ilgin C, Yavuz DG. Effect of denosumab treatment on bone mineral density and bone turnover markers in osteoporotic patients: real-life experience 2-year follow-up. Arch Osteoporos. 2022; 17(1): 125.
20. Zhong Y, Li J, Luo J, Chen F. Advances in bone turnover markers (PINP and CTX) in optimizing anti-resorptive and anabolic therapies against osteoporosis. Discov Med. 2021; 32(167): 149–54.