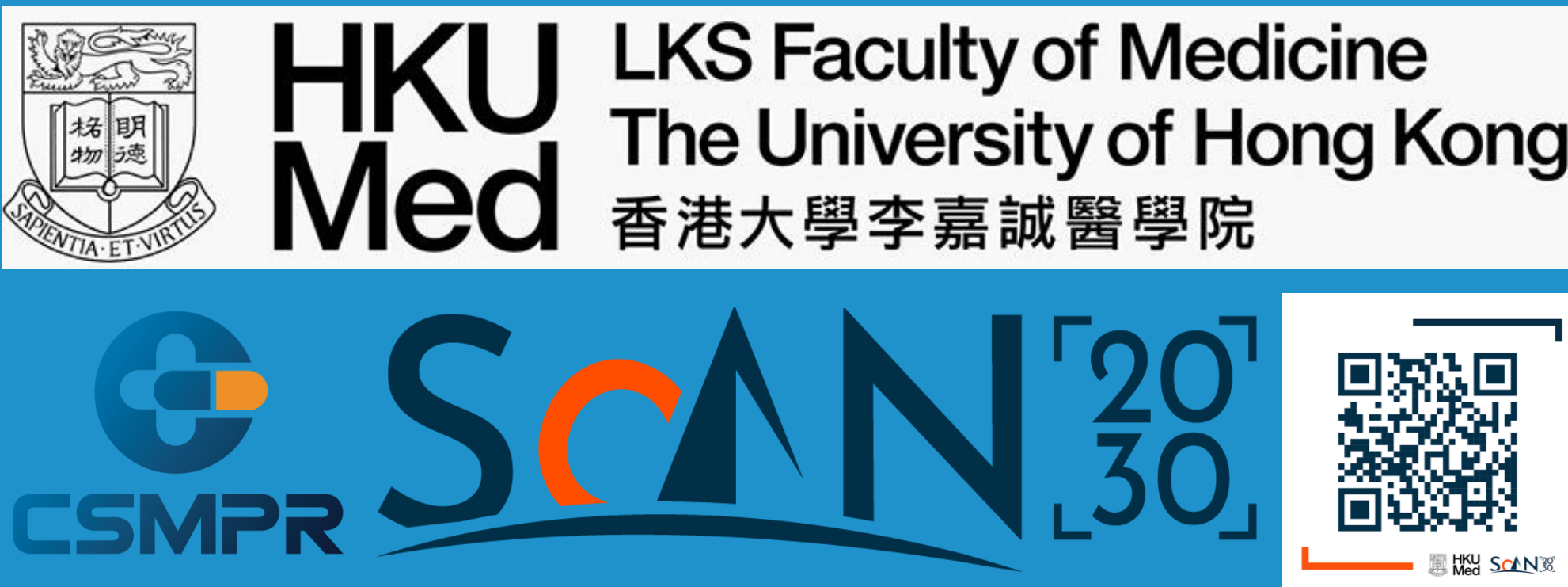# Artificial Intelligence in Evidence Synthesis: A Systematic Review and Meta-Analysis of Emerging Biologics for Improving Skeletal Health in Osteogenesis Imperfecta

Chengfei Li, MPH[1]*, **Zonglin Dai, PhD[2]***, **Wing Chung Tang, PhD[3]**, **Zesen Gao, MSC[4]**, **Vivien K Y Chan, PhD[1]**, **Mariana Ramirez-Posada, PhD[5]**, **Jiyeong Kim, PhD[5]**, **CL CHEUNG, PhD[1]**, **Ian Wong, PhD[1]**, **Dong Dong, PhD[6]**, **Michael To, PhD[7]**, **Dawn Craig, PhD[8]**, **Xue Li, PhD[2]#**

[1]Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, Hong Kong
[2]Department of Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, Hong Kong
[3]Yu Chun Keung Medical Library, The University of Hong Kong, Hong Kong SAR, Hong Kong
[4]The University of Hong Kong Libraries, The University of Hong Kong, Hong Kong SAR, Hong Kong
[5]Stanford Center for Digital Health, Department of Medicine, Stanford University, Stanford, CA, USA
[6]The Jockey Club School of Public Health and Primary Care (JCSPHPC), The Chinese University of Hong Kong, Hong Kong SAR, Hong Kong
[7]Department of Orthopaedics and Traumatology, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, Hong Kong
[8]Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

* Co-first authors
# Corresponding author: sxueli@hku.hk

EPH19

INTRODUCTION
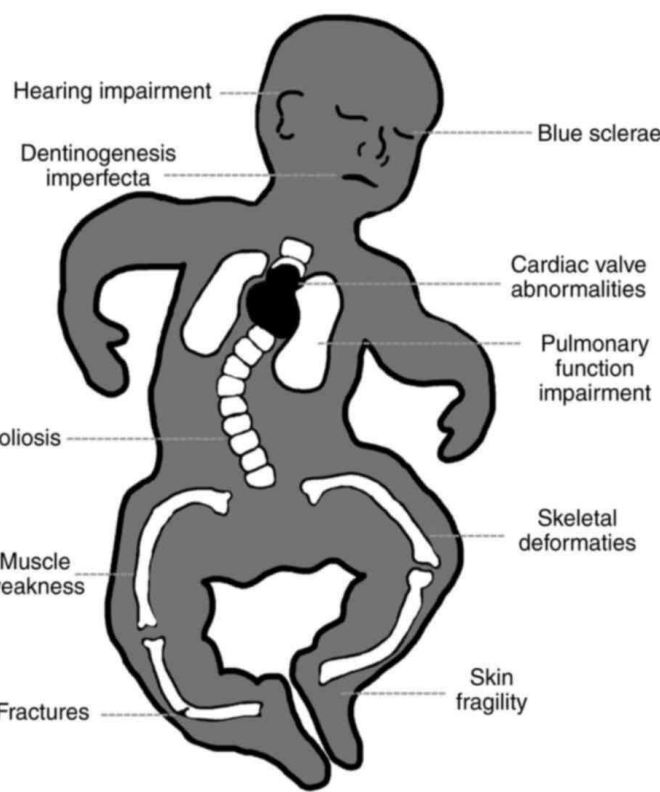
Osteogenesis imperfecta (OI)
- Rare genetic disorder
- Standard of care: bisphosphonates
- Emerging therapies: biologics (mechanism-targeted; evidence fragmented)
- Objective: To systematically evaluate the effectiveness and safety of biologics in OI using artificial intelligence (AI)-assisted evidence synthesis.
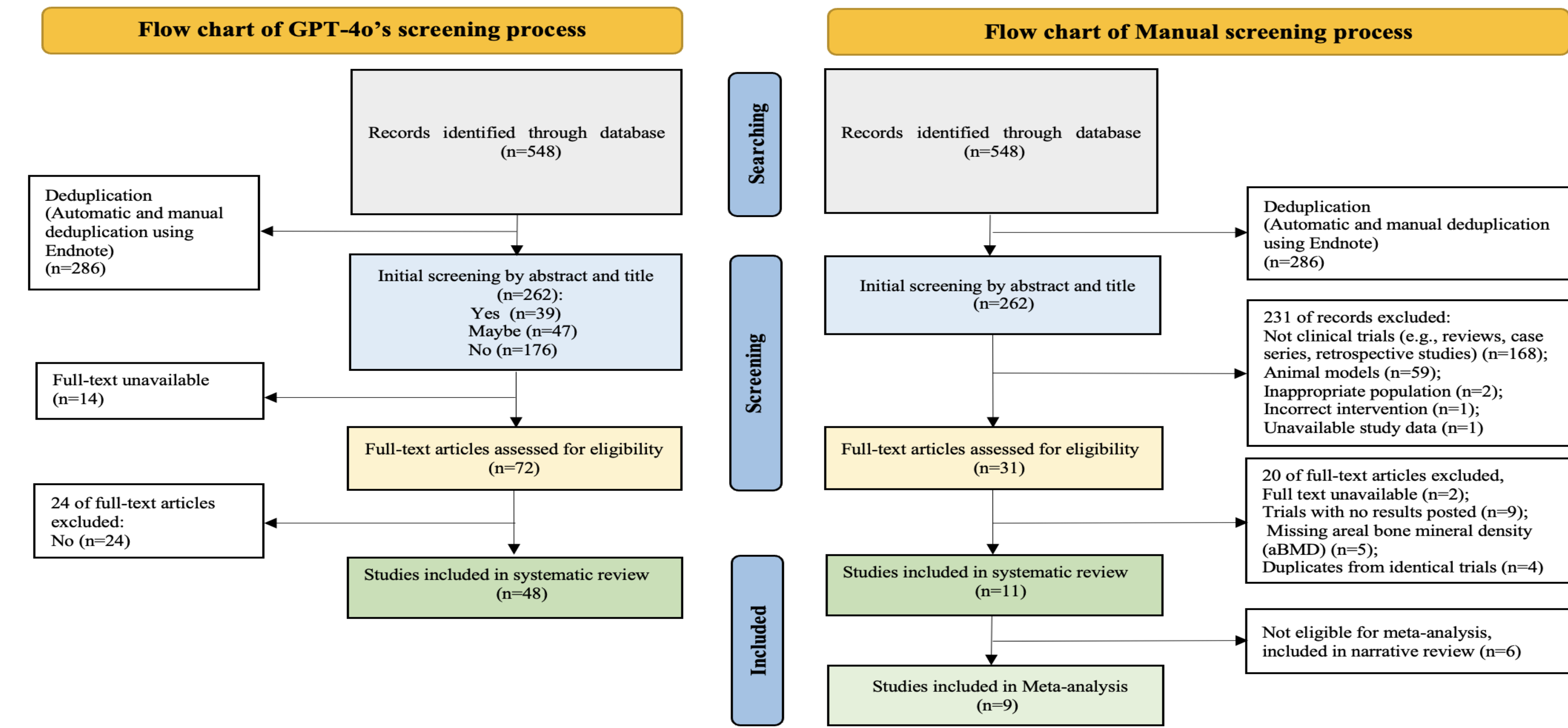
METHOD

- Systematic review and random-effects meta-analysis of trials on denosumab, setrusumab, teriparatide, and fresolimumab in OI.
- **Primary outcomes**: change (%) in areal bone mineral density (aBMD); fracture incidence; safety profiles.
- **AI-assisted workflow**: GPT-4o (2024-08-06 API) assisted title/abstract screening and risk-of-bias appraisal using rule-based prompts.
- **Performance validation**: Compared AI to human using sensitivity, specificity, and weighted Cohen's kappa.

RESULTS

- AI showed high sensitivity in abstract (96.8%) and full-text screening (90.9%), cut screening time by over 95%, and was ~100 times faster per article than humans.
- Agreement with humans in quality assessment was substantial (kappa = 0.806).

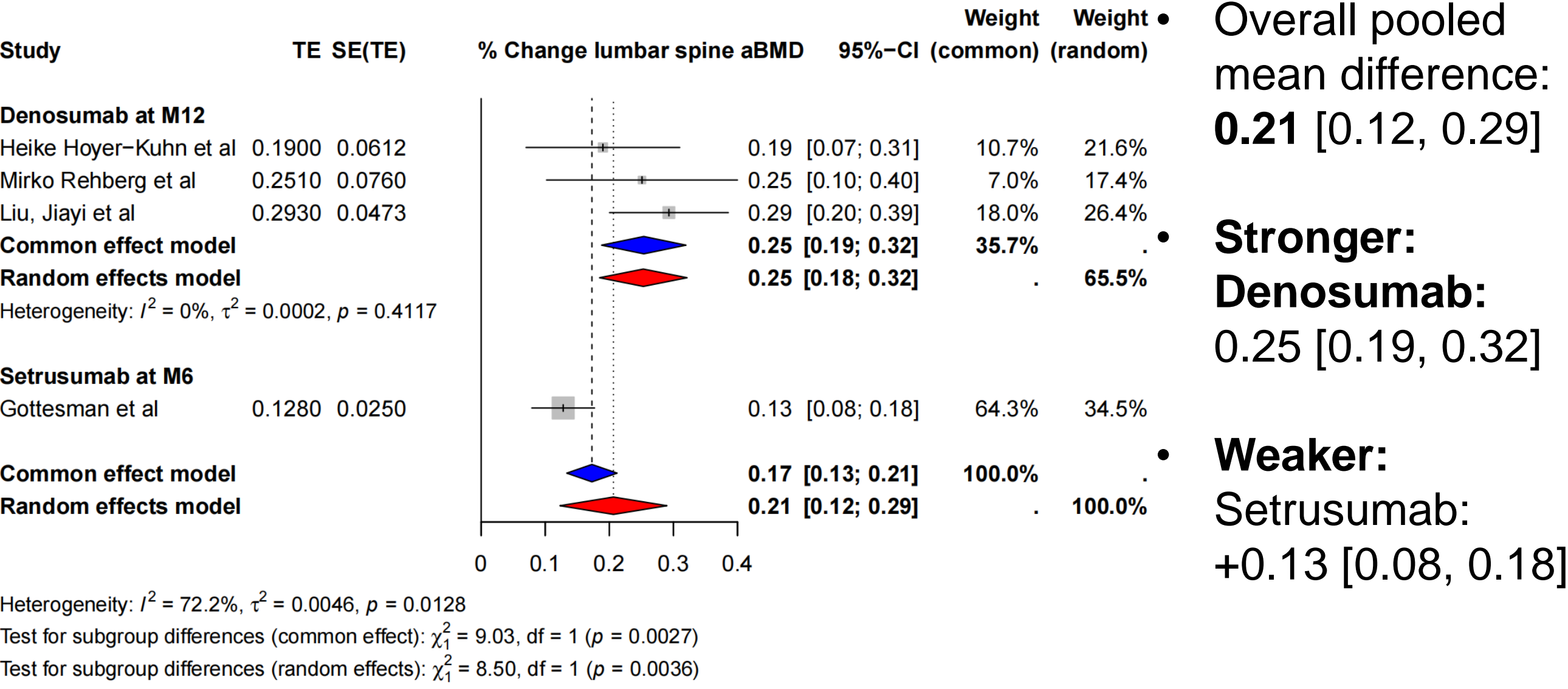### Comparative PRISMA flow diagram of AI-assisted and Manual Screening

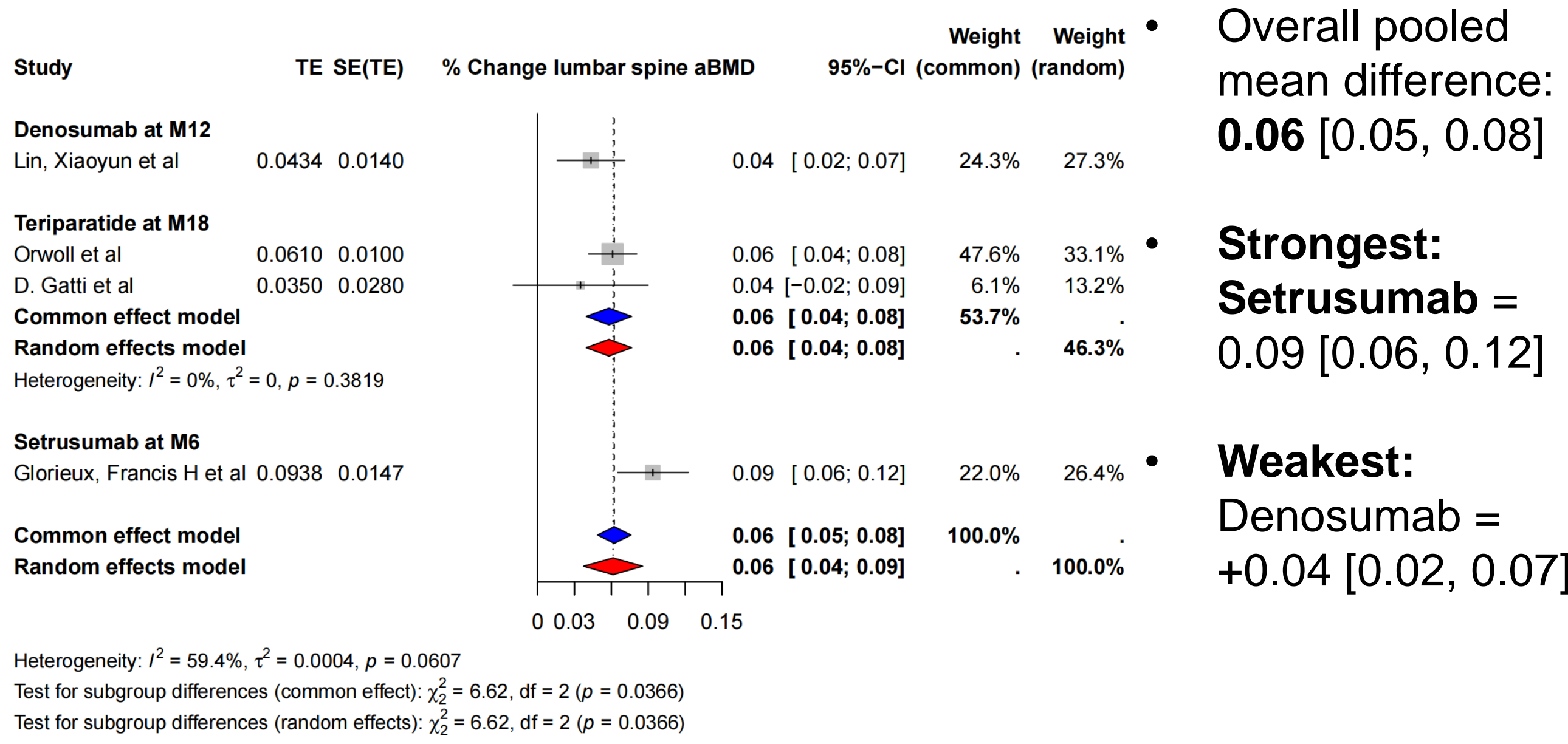### Quality assessment results of human and GPT-based tools

Note*: Assignment: effect of assignment to intervention; Missing data: missing outcome data; Measurement: measurement of the outcome; Selection: selection of the reported result.

Biologics on lumbar aBMD change (%) in OI children

- Overall pooled mean difference: **0.21 [0.12, 0.29]**
- **Stronger: Denosumab:** 0.25 [0.19, 0.32]
- **Weaker: Setrusumab:** +0.13 [0.08, 0.18]

Biologics on lumbar aBMD change (%) in OI adults

- Overall pooled mean difference: **0.06 [0.05, 0.08]**
- **Strongest: Setrusumab =** 0.09 [0.06, 0.12]
- **Weakest: Denosumab =** +0.04 [0.02, 0.07]

AI performance in literature Screening

Document type			Human			Sensitivity (TP/[TP+FN])	Specificity (TN/[TN+FP])	Screening time (seconds/paper, mean ± SD)	Estimated time for manual screening (minutes/paper)
			Positive	Negative	Total				
Title & Abstract	GPT-4o	Positive	30	56	86	96.8%	75.8%	3.50±0.14	~5
		Negative	1	175	176				
		Total	31	231	262				
Full-text		Positive	10	38	48	90.9%	84.9%	9.67±1.85	~15
		Negative	1	213	214				
		Total	11	251	262				

Abbreviations: TP, true positive; FN, false negative; TN, false negative; FP, false positive

CONCLUSIONS

- GPT-4o enhances evidence synthesis by increasing screening efficiency and improving quality assessment, providing a scalable way to reduce manual workload; human oversight remains crucial for tasks that require contextual understanding and clinical reasoning.
- Denosumab and setrusumab effectively improve lumbar spine aBMD in OI, but current evidence does not confirm a reduction in fracture risk with biologics.

CONTACT INFORMATION

- Contact person: Xue LI (sxueli@hku.hk)
- LinkedIn: Shirley, Xue LI