Supplementary Table S3

Consolidated response-category distributions, code 9 handling, and model–reference disagreement patterns

Supplementary Table S3. Response-category distributions and model–expert reference agreement patterns, including the handling of responses coded 9 ("Unable to assess").

Panel	Item	AI session / group	Model or data source	n	Summary
Code 9 responses	All applicable items	—	Overall frequency	7 entries	All code 9 entries occurred in P2a, P2b, or P2c. No code 9 response occurred for S1, S2, S4, O2, P1, S3, or O1a.
Code 9 responses	P2a	—	Periodontist 1	Case 2	Retained as a separate nominal category in α. Three different categories → no majority reference; case excluded from AI–reference P2a comparison. P2b/P2c for this rater: not applicable.
Code 9 responses	P2a	—	Periodontist 1	Case 18	Same as above; case excluded from AI–reference P2a comparison (n = 46).
Code 9 responses	P2a	—	ChatGPT-5.6	Case 48	Retained as a category; counted as disagreement with the reference (0). P2b/P2c for ChatGPT in this case: not applicable.
Code 9 responses	P2b	—	Periodontist 1	Case 2	Rater's P2a = 9 → not applicable; not included in P2b denominator.
Code 9 responses	P2b	—	Periodontist 1	Case 9	Rater's P2a = 1 → retained in P2b denominator; not counted as selection of any option.
Code 9 responses	P2b	—	ChatGPT-5.6	Case 48	Model's P2a = 9 → not applicable.
Code 9 responses	P2c	—	ChatGPT-5.6	Case 48	Model's P2a = 9 → not applicable; not included in P2c analysis.
Category distributions	S1	Orthodontist	Three orthodontists (individual responses)	144	0=34; 1=48; 2=54; 3=8
Category distributions	S1	Orthodontist	Expert-based reference (≥4/6)	34	0=10; 1=7; 2=15; 3=2
Category distributions	S1	Orthodontist	ChatGPT-5.6	48	0=9; 1=17; 2=22
Category distributions	S1	Orthodontist	Gemini Pro	48	0=12; 1=17; 2=19
Category distributions	S1	Orthodontist	Claude Opus 4.8 High	48	0=11; 1=11; 2=26
Category distributions	S2	Orthodontist	Three orthodontists (individual responses)	144	0=21; 1=44; 2=60; 3=19
Category distributions	S2	Orthodontist	Expert-based reference (≥4/6)	39	0=8; 1=8; 2=17; 3=6
Category distributions	S2	Orthodontist	ChatGPT-5.6	48	0=2; 1=12; 2=27; 3=7
Category distributions	S2	Orthodontist	Gemini Pro	48	0=8; 1=13; 2=18; 3=9
Category distributions	S2	Orthodontist	Claude Opus 4.8 High	48	1=17; 2=19; 3=12
Category distributions	S3	Orthodontist	Three orthodontists (individual responses)	144	0=46; 1=22; 2=12; 3=64
Category distributions	S3	Orthodontist	Expert-based reference (≥4/6)	37	0=15; 1=2; 2=2; 3=18
Category distributions	S3	Orthodontist	ChatGPT-5.6	48	0=2; 1=10; 3=36
Category distributions	S3	Orthodontist	Gemini Pro	48	0=10; 1=8; 2=2; 3=28
Category distributions	S3	Orthodontist	Claude Opus 4.8 High	48	0=10; 1=11; 3=27

Panel	Item	AI session / group	Model or data source	n	Summary
Category distributions	S4	Orthodontist	Three orthodontists (individual responses)	144	2=69; 3=75
Category distributions	S4	Orthodontist	Expert-based reference (≥4/6)	32	2=6; 3=26
Category distributions	S4	Orthodontist	ChatGPT-5.6	48	2=48
Category distributions	S4	Orthodontist	Gemini Pro	48	2=21; 3=27
Category distributions	S4	Orthodontist	Claude Opus 4.8 High	48	1=14; 2=33; 3=1
Category distributions	O1a	Orthodontist	Three orthodontists (individual responses)	144	0=71; 1=73
Category distributions	O1a	Orthodontist	Expert-based reference	48	0=23; 1=25
Category distributions	O1a	Orthodontist	ChatGPT-5.6	48	0=24; 1=24
Category distributions	O1a	Orthodontist	Gemini Pro	48	0=23; 1=25
Category distributions	O1a	Orthodontist	Claude Opus 4.8 High	48	0=27; 1=21
Category distributions	O2	Orthodontist	Three orthodontists (individual responses)	144	0=34; 1=49; 3=61
Category distributions	O2	Orthodontist	Expert-based reference	48	0=12; 1=16; 3=20
Category distributions	O2	Orthodontist	ChatGPT-5.6	48	0=11; 1=16; 3=21
Category distributions	O2	Orthodontist	Gemini Pro	47	0=9; 1=19; 2=1; 3=18
Category distributions	O2	Orthodontist	Claude Opus 4.8 High	48	0=9; 1=13; 3=26
Category distributions	S1	Periodontist	Three periodontists (individual responses)	144	0=31; 1=24; 2=74; 3=15
Category distributions	S1	Periodontist	Expert-based reference (≥4/6)	34	0=10; 1=7; 2=15; 3=2
Category distributions	S1	Periodontist	ChatGPT-5.6	48	0=7; 1=18; 2=23
Category distributions	S1	Periodontist	Gemini Pro	48	0=12; 1=11; 2=25
Category distributions	S1	Periodontist	Claude Opus 4.8 High	48	0=11; 1=11; 2=26
Category distributions	S2	Periodontist	Three periodontists (individual responses)	144	0=30; 1=27; 2=56; 3=31
Category distributions	S2	Periodontist	Expert-based reference (≥4/6)	39	0=8; 1=8; 2=17; 3=6
Category distributions	S2	Periodontist	ChatGPT-5.6	48	1=15; 2=25; 3=8
Category distributions	S2	Periodontist	Gemini Pro	48	0=8; 1=14; 2=21; 3=5
Category distributions	S2	Periodontist	Claude Opus 4.8 High	48	1=17; 2=19; 3=12
Category distributions	S3	Periodontist	Three periodontists (individual responses)	144	0=50; 1=17; 2=34; 3=43
Category distributions	S3	Periodontist	Expert-based reference (≥4/6)	37	0=15; 1=2; 2=2; 3=18
Category distributions	S3	Periodontist	ChatGPT-5.6	48	0=1; 1=11; 2=1; 3=35
Category distributions	S3	Periodontist	Gemini Pro	48	0=5; 1=14; 2=6; 3=23

Panel	Item	AI session / group	Model or data source	n	Summary
Category distributions	S3	Periodontist	Claude Opus 4.8 High	48	0=10; 1=11; 3=27
Category distributions	S4	Periodontist	Three periodontists (individual responses)	144	1=3; 2=29; 3=112
Category distributions	S4	Periodontist	Expert-based reference (≥4/6)	32	2=6; 3=26
Category distributions	S4	Periodontist	ChatGPT-5.6	48	2=48
Category distributions	S4	Periodontist	Gemini Pro	48	2=19; 3=29
Category distributions	S4	Periodontist	Claude Opus 4.8 High	48	1=14; 2=33; 3=1
Category distributions	P1	Periodontist	Three periodontists (individual responses)	144	0=55; 1=29; 2=17; 3=43
Category distributions	P1	Periodontist	Expert-based reference	48	0=23; 1=8; 2=6; 3=11
Category distributions	P1	Periodontist	ChatGPT-5.6	48	0=29; 1=9; 2=10
Category distributions	P1	Periodontist	Gemini Pro	48	0=20; 1=18; 2=8; 3=2
Category distributions	P1	Periodontist	Claude Opus 4.8 High	48	0=27; 1=10; 2=6; 3=5
Category distributions	P2a	Periodontist	Three periodontists (individual responses)	144	0=85; 1=57; 9=2
Category distributions	P2a	Periodontist	Expert-based reference	46	0=28; 1=18
Category distributions	P2a	Periodontist	ChatGPT-5.6	48	0=42; 1=5; 9=1
Category distributions	P2a	Periodontist	Gemini Pro	48	0=35; 1=13
Category distributions	P2a	Periodontist	Claude Opus 4.8 High	48	0=37; 1=11
Category distributions	P2c	Periodontist	Three periodontists (responses with P2a = 1)	57	1=39; 2=1; 3=10; 4=7
Category distributions	P2c	Periodontist	Expert-based reference	18	1=14; 3=3; 4=1
Category distributions	P2c	Periodontist	ChatGPT-5.6	4	1=4
Category distributions	P2c	Periodontist	Gemini Pro	13	1=8; 2=3; 4=2
Category distributions	P2c	Periodontist	Claude Opus 4.8 High	11	1=5; 4=6
Ordinal deviations (AI − reference)	S1	Orthodontist	ChatGPT-5.6	34	≤−2: 0; −1: 2; 0: 30; +1: 2; ≥+2: 0; mean deviation: ±0.00
Ordinal deviations (AI − reference)	S1	Orthodontist	Gemini Pro	34	≤−2: 0; −1: 7; 0: 25; +1: 2; ≥+2: 0; mean deviation: −0.15
Ordinal deviations (AI − reference)	S1	Orthodontist	Claude Opus 4.8 High	34	≤−2: 0; −1: 2; 0: 31; +1: 1; ≥+2: 0; mean deviation: −0.03
Ordinal deviations (AI − reference)	S2	Orthodontist	ChatGPT-5.6	39	≤−2: 0; −1: 1; 0: 27; +1: 11; ≥+2: 0; mean deviation: +0.26
Ordinal deviations (AI − reference)	S2	Orthodontist	Gemini Pro	39	≤−2: 0; −1: 2; 0: 34; +1: 3; ≥+2: 0; mean deviation: +0.03

Panel	Item	AI session / group	Model or data source	n	Summary
Ordinal deviations (AI − reference)	S2	Orthodontist	Claude Opus 4.8 High	39	≤−2: 0; −1: 0; 0: 26; +1: 13; ≥+2: 0; mean deviation: +0.33
Ordinal deviations (AI − reference)	S4	Orthodontist	ChatGPT-5.6	32	≤−2: 0; −1: 26; 0: 6; +1: 0; ≥+2: 0; mean deviation: −0.81
Ordinal deviations (AI − reference)	S4	Orthodontist	Gemini Pro	32	≤−2: 0; −1: 8; 0: 21; +1: 3; ≥+2: 0; mean deviation: −0.16
Ordinal deviations (AI − reference)	S4	Orthodontist	Claude Opus 4.8 High	32	≤−2: 7; −1: 19; 0: 6; +1: 0; ≥+2: 0; mean deviation: −1.03
Ordinal deviations (AI − reference)	O2	Orthodontist	ChatGPT-5.6	48	≤−2: 1; −1: 2; 0: 40; +1: 3; ≥+2: 2; mean deviation: +0.06
Ordinal deviations (AI − reference)	O2	Orthodontist	Gemini Pro	47	≤−2: 3; −1: 2; 0: 37; +1: 4; ≥+2: 1; mean deviation: −0.02
Ordinal deviations (AI − reference)	O2	Orthodontist	Claude Opus 4.8 High	48	≤−2: 0; −1: 1; 0: 38; +1: 3; ≥+2: 6; mean deviation: +0.31
Ordinal deviations (AI − reference)	S1	Periodontist	ChatGPT-5.6	34	≤−2: 0; −1: 2; 0: 29; +1: 3; ≥+2: 0; mean deviation: +0.03
Ordinal deviations (AI − reference)	S1	Periodontist	Gemini Pro	34	≤−2: 0; −1: 3; 0: 30; +1: 1; ≥+2: 0; mean deviation: −0.06
Ordinal deviations (AI − reference)	S1	Periodontist	Claude Opus 4.8 High	34	≤−2: 0; −1: 2; 0: 31; +1: 1; ≥+2: 0; mean deviation: −0.03
Ordinal deviations (AI − reference)	S2	Periodontist	ChatGPT-5.6	39	≤−2: 0; −1: 0; 0: 28; +1: 10; ≥+2: 1; mean deviation: +0.31
Ordinal deviations (AI − reference)	S2	Periodontist	Gemini Pro	39	≤−2: 0; −1: 5; 0: 31; +1: 3; ≥+2: 0; mean deviation: −0.05
Ordinal deviations (AI − reference)	S2	Periodontist	Claude Opus 4.8 High	39	≤−2: 0; −1: 0; 0: 26; +1: 13; ≥+2: 0; mean deviation: +0.33
Ordinal deviations (AI − reference)	S4	Periodontist	ChatGPT-5.6	32	≤−2: 0; −1: 26; 0: 6; +1: 0; ≥+2: 0; mean deviation: −0.81
Ordinal deviations (AI − reference)	S4	Periodontist	Gemini Pro	32	≤−2: 0; −1: 7; 0: 22; +1: 3; ≥+2: 0; mean deviation: −0.12
Ordinal deviations (AI − reference)	S4	Periodontist	Claude Opus 4.8 High	32	≤−2: 7; −1: 19; 0: 6; +1: 0; ≥+2: 0; mean deviation: −1.03
Ordinal deviations (AI − reference)	P1	Periodontist	ChatGPT-5.6	48	≤−2: 7; −1: 5; 0: 35; +1: 1; ≥+2: 0; mean deviation: −0.50
Ordinal deviations (AI − reference)	P1	Periodontist	Gemini Pro	48	≤−2: 6; −1: 4; 0: 31; +1: 7; ≥+2: 0; mean deviation: −0.27
Ordinal deviations (AI − reference)	P1	Periodontist	Claude Opus 4.8 High	48	≤−2: 7; −1: 1; 0: 36; +1: 3; ≥+2: 1; mean deviation: −0.33
Nominal discordances	S3	Orthodontist	ChatGPT-5.6	37	Agreement: 20; disagreement: 17; reference→model pairs (cases): 0→1: 7; 0→3: 6; 1→3: 2; 2→3: 2

Panel	Item	AI session / group	Model or data source	n	Summary
Nominal discordances	S3	Orthodontist	Gemini Pro	37	Agreement: 28; disagreement: 9; reference→model pairs (cases): 0→1: 3; 0→3: 2; 0→2: 1; 2→3: 1; 1→3: 1; 3→1: 1
Nominal discordances	S3	Orthodontist	Claude Opus 4.8 High	37	Agreement: 27; disagreement: 10; reference→model pairs (cases): 0→1: 6; 0→3: 2; 2→1: 1; 2→0: 1
Nominal discordances	O1a	Orthodontist	ChatGPT-5.6	48	Agreement: 43; disagreement: 5; reference→model pairs (cases): 1→0: 3; 0→1: 2
Nominal discordances	O1a	Orthodontist	Gemini Pro	48	Agreement: 42; disagreement: 6; reference→model pairs (cases): 0→1: 3; 1→0: 3
Nominal discordances	O1a	Orthodontist	Claude Opus 4.8 High	48	Agreement: 42; disagreement: 6; reference→model pairs (cases): 1→0: 5; 0→1: 1
Nominal discordances	S3	Periodontist	ChatGPT-5.6	37	Agreement: 19; disagreement: 18; reference→model pairs (cases): 0→1: 9; 0→3: 4; 1→3: 2; 2→3: 2; 0→2: 1
Nominal discordances	S3	Periodontist	Gemini Pro	37	Agreement: 22; disagreement: 15; reference→model pairs (cases): 0→1: 7; 0→3: 2; 0→2: 2; 3→1: 2; 2→1: 1; 1→2: 1
Nominal discordances	S3	Periodontist	Claude Opus 4.8 High	37	Agreement: 27; disagreement: 10; reference→model pairs (cases): 0→1: 6; 0→3: 2; 2→1: 1; 2→0: 1
Nominal discordances	P2a	Periodontist	ChatGPT-5.6	46	Agreement: 32; disagreement: 14; reference→model pairs (cases): 1→0: 13; 0→9: 1
Nominal discordances	P2a	Periodontist	Gemini Pro	46	Agreement: 39; disagreement: 7; reference→model pairs (cases): 1→0: 6; 0→1: 1
Nominal discordances	P2a	Periodontist	Claude Opus 4.8 High	46	Agreement: 37; disagreement: 9; reference→model pairs (cases): 1→0: 8; 0→1: 1
Nominal discordances	P2c	Periodontist	ChatGPT-5.6	4	Agreement: 4; disagreement: 0; reference→model pairs (cases): none
Nominal discordances	P2c	Periodontist	Gemini Pro	12	Agreement: 7; disagreement: 5; reference→model pairs (cases): 1→4: 2; 1→2: 2; 3→2: 1
Nominal discordances	P2c	Periodontist	Claude Opus 4.8 High	10	Agreement: 5; disagreement: 5; reference→model pairs (cases): 1→4: 4; 3→4: 1

Notes. For S1–S4, the expert-based reference was the multidisciplinary category selected by at least four of the six experts. The reference distributions and model–reference comparisons were based on 34, 39, 37, and 32 cases for S1, S2, S3, and S4, respectively. Model category distributions included all 48 cases, except for Gemini Pro on O2, for which no response was provided for case 33 ($n = 47$). In rows reporting pooled individual expert responses, n denotes ratings rather than unique cases; for non-conditional items, $n = 144$ (three experts × 48 cases). For other items, the specialty-specific reference was derived from the three experts in the corresponding specialty. Ordinal deviations were calculated as the model code minus the reference code; positive values indicate a higher category on the item's ordered scale. P2b was treated as a multiple-response item, and code 9 was not counted as selection of any response option.

P2c was considered applicable only when the same rater or model had indicated in P2a that surgical intervention was necessary. Responses that did not meet this condition were treated as structurally missing. Five expert P2c entries recorded despite P2a ≠ 1 were excluded, leaving 57 applicable expert ratings. For Gemini Pro and Claude Opus 4.8 High, model category distributions included a response indicating that surgery was necessary for case 48, although none of the periodontists indicated that surgery was necessary; consequently, no expert-based P2c reference was available for that case. The corresponding model–reference comparison denominators were therefore 12 and 10, compared with model category-distribution denominators of 13 and 11. ChatGPT-5.6 indicated that surgery was necessary in five cases but did not provide a P2c response for case 29; this response was treated as missing, resulting in $n = 4$ for its model–reference comparison. For P2a, code 9 was retained as a separate nominal category. Cases 2 and 18, for which no majority reference could be established, were excluded from the AI–reference comparisons ($n = 46$). Categories follow the study codebook.