

**Web only file 2:** Table of pooled results of adherence per individual STARD item comparing studies published pre- and post-STARD.

STARD item	Pre-STARD			Post-STARD		Overall
	Studies n	Evaluated articles n	Item reported n (%)	Evaluated articles n	Item reported n (%)	Effect estimate OR (95%CI)
1. Study identified as test accuracy study	4	396	108 (27.3%)	220	64 (29.1%)	1.80 [0.63, 5.08]
2. Research questions/aims	4	396	339 (85.6%)	220	204 (92.7%)	1.52 [0.70, 3.31]
3. Study population	5	516	290 (56.2%)	340	147 (43.2%)	0.94 [0.63, 1.40]
4. Participant recruitment	5	516	395 (76.6%)	340	316 (92.9%)	2.89 [1.41, 5.92]
5. Participant sampling	5	516	234 (45.3%)	340	223 (65.6%)	1.89 [1.33, 2.69]
6. Data collection	5	516	359 (69.6%)	340	266 (78.2%)	1.44 [1.02, 2.03]
7. Reference standard and rationale	5	516	325 (63.0%)	340	315 (92.6%)	1.72 [0.79, 3.74]
<i>8. Technique of:</i>						
8a. Index test	5	516	355 (68.8%)	340	286 (84.1%)	1.32 [0.64, 2.72]
8b. Reference standard	5	546	194 (35.5%)	340	190 (55.9%)	1.28 [0.65, 2.53]
<i>9. Units/cut-offs/categories for:</i>						
9a. Index test	5	516	437 (84.7%)	340	289 (85.0%)	1.37 [0.43, 4.34]
9b. Reference standard	4	428	276 (64.5%)	324	218 (67.3%)	1.37 [0.99, 1.89]
<i>10. Persons executing:</i>						
10a. Index test	5	516	113 (21.9%)	340	120 (35.3%)	1.21 [0.86, 1.71]
10b. Reference standard	5	516	57 (11.0%)	340	72 (21.2%)	1.22 [0.82, 1.83]
<i>11. Blinding of results of:</i>						
11a. Index test	5	516	307 (59.5%)	340	171 (50.3%)	0.93 [0.58, 1.47]
11b. Reference test	5	516	112 (21.7%)	340	110 (32.4%)	1.36 [0.87, 2.13]
12. Methods for statistics used	4	396	123 (31.1%)	220	58 (26.4%)	1.49 [0.54, 4.09]
<i>13. Methods for test reproducibility for:</i>						
13a. Index test	3	308	56 (18.2%)	203	64 (31.5%)	1.01 [0.28, 3.62]
13b. Reference standard	3	308	11 (3.6%)	203	10 (4.9%)	0.55 [0.07, 4.65]
14. Dates of study	5	516	332 (64.3%)	340	241 (70.9%)	1.69 [1.23, 2.33]
15. Characteristics of study population	5	516	293 (56.8%)	340	250 (73.5%)	1.85 [1.00, 3.43]
16. Eligible patients not undergoing either test	5	516	296 (57.4%)	340	198 (58.2%)	1.02 [0.76, 1.38]
<i>16a. Flow diagram</i>	4	492	15 (3.0%)	313	45 (14.4%)	4.79 [1.39, 16.50]
17. Time interval between tests	4	396	127 (32.1%)	220	75 (34.1%)	1.38 [0.69, 2.76]
18. Distribution of severity of disease	5	516	315 (61.0%)	340	215 (63.2%)	2.18 [0.92, 5.15]
19. Cross tabulation of results	4	396	235 (59.3%)	220	159 (72.3%)	0.90 [0.53, 1.53]
20. Adverse events	3	236	35 (14.8%)	185	20 (10.8%)	0.74 [0.40, 1.37]
21. Estimates of diagnostic accuracy	4	396	177 (44.7%)	220	96 (43.6%)	1.46 [0.67, 3.18]
22. How uninterpretable results were handled	4	396	194 (49.0%)	220	113 (51.4%)	1.10 [0.75, 1.61]
23. Estimates of variability of accuracy	4	396	133 (33.6%)	220	119 (54.1%)	2.62 [1.57, 4.37]
<i>24. Estimates of test reproducibility, for:</i>						
24a. Index test	3	308	71 (23.1%)	203	70 (34.5%)	0.84 [0.31, 2.26]
24b. Reference standard	2	148	8 (5.4%)	168	8 (4.8%)	0.87 [0.32, 2.40]
25. Clinical applicability of findings	4	396	384 (97.0%)	220	215 (97.7%)	1.31 [0.16, 10.57]