

Title

A longitudinal analysis of the current status of published clinical practice guidelines and their effect on disease burden

Supplementary Methods. Search strategy for CPGs and studies that evaluate the quality of CPGs using the Appraisal of Guidelines for Research & Evaluation II (AGREE II) instrument.

Search strategy for CPGs

PubMed

#1 "Practice Guideline"[Publication Type] NOT "Randomized Controlled Trial"[Publication Type]) NOT "Review"[Publication Type]) NOT "Systematic Review"[Publication Type]) NOT "Meta-Analysis"[Publication Type]

#2 guideline*[Title] NOT consensus [Title] NOT review*[Title]

#3 #1 AND #2

Embase

#1 'practice guideline'/exp

#2 guideline*:ti NOT consensus:ti NOT review:ti

#3 #1 AND #2

#4 #3 AND 'article'/it

#5 #4 AND 'practice guideline'/de

Search strategy for articles that appraise CPGs with AGREE II instrument

PubMed

#1 Guidelines as Topic (MeSH)

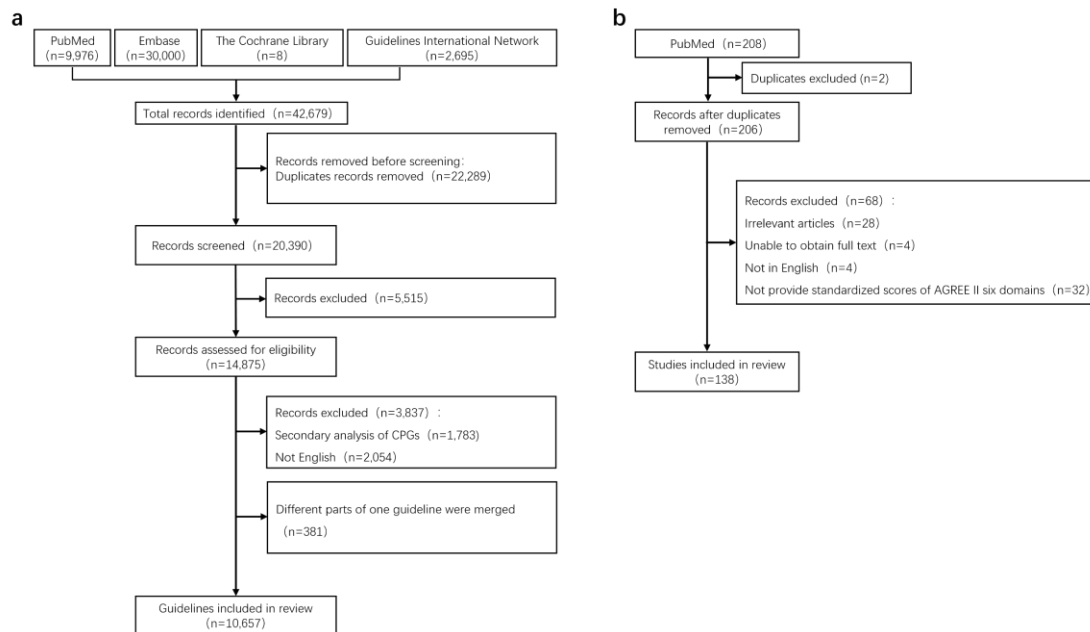
#2 (Clinical [Title] OR clinic [Title]) AND (guideline*[Title] OR guidelines [Title] OR recommendation [Title] OR recommendations [Title])

#3 AGREE II [Title] OR the Appraisal of Guidelines for Research and Evaluation [Title]

#4 #1 OR #2

#5 #3 AND #4

Supplementary Figures



a. Flow chart of CPGs selection.

b. Flow chart of the selection for articles that evaluated CPGs using AGREE II.

Supplementary Figure 1. Flow chart of study inclusion.

Supplementary Tables

Supplementary Table 1. Trends for the number of different categories of clinical practice guidelines.

Category	Number of CPGs			<i>t</i>	<i>df</i>	EAPC (%; 95%CI)	<i>p</i> -value
	Total	1995	2023				
Europe	4285	25	221	11.156	27	7.36 (6.03 to 8.71)	1.29×10^{-11}
America	3445	61	139	6.752	27	3.15 (2.23 to 4.09)	3.00×10^{-7}
Asia	1590	9	169	17.387	27	16.54 (14.54 to 18.56)	3.43×10^{-16}
Cross-region	558	9	39	7.894	27	11.40 (8.46 to 14.43)	1.74×10^{-8}
Neoplasms	2147	6	136	10.796	27	9.31 (7.56 to 11.09)	2.68×10^{-11}
Cardiovascular diseases	1493	15	69	7.410	27	5.30 (3.87 to 6.75)	5.70×10^{-8}
Digestive diseases	730	4	47	8.332	27	7.79 (5.90 to 9.70)	6.09×10^{-9}
Musculoskeletal disorders	438	2	30	8.626	27	8.28 (6.34 to 10.25)	3.06×10^{-9}
Neurological disorders	401	4	26	5.991	27	5.96 (3.97 to 7.99)	2.17×10^{-6}
Chronic respiratory disease	319	8	17	5.683	27	5.66 (3.68 to 7.69)	4.90×10^{-6}
Treatment	4793	48	258	12.429	27	5.92 (4.96 to 6.88)	1.11×10^{-12}
Diagnosis	2669	13	159	11.247	27	7.35 (6.03 to 8.68)	1.08×10^{-11}
Prevention	895	25	39	5.155	27	3.86 (2.38 to 5.37)	2.01×10^{-5}
Nursing	568	24	25	3.972	27	4.03 (2.02 to 6.07)	4.77×10^{-4}

EAPC, estimated annual percentage change; CI, confidence interval. CPGs that contain zero annual publication counts could not have their EAPC calculated. Statistical significance of trends was assessed using linear regression of the log-transformed annual rates. The test statistic (*t*), degrees of freedom (*df*), and exact *p*-values are provided for each trend. The EAPC and its 95% CI serve as the effect size. All tests were two-sided. No adjustments for multiple comparisons were made.

Supplementary Table 2. The update of different categories of clinical practice guidelines.

Category	Updated CPGs	Total CPGs	Ratio
America	608	3445	17.65%
Europe	565	4285	13.19%
Asia	260	1590	16.35%
Cross-region	87	558	15.59%
Oceania	12	250	4.80%
Africa	12	84	14.29%
Neoplasms	527	2147	24.55%
Cardiovascular diseases	248	1493	16.61%
Digestive diseases	97	730	13.29%
Chronic respiratory diseases	64	319	20.06%
Diabetes and kidney diseases	62	408	15.20%
Musculoskeletal disorders	57	438	13.01%
Neurological disorders	54	401	13.47%
Skin and subcutaneous diseases	38	291	13.06%
Mental disorders	23	106	21.70%
Sense organ diseases	17	79	21.52%

Supplementary Table 3. Comparison of AGREE II scores for CPGs across time periods. Values are median (interquartile range)

Domain	Before 2001 (n=14)	2001-2005 (n=52)	2006-2010 (n=242)	2011-2015 (n=643)	2016-2020 (n=624)	After 2020 (n=58)	H	df	ε^2 (95%CI)	p-value
Scope and purpose	63.91 (57.26, 78.25)	66.94 (52.78, 79.99)	69.00 (46.17, 85.75)	75.00 (58.00, 88.00)	77.90 (62.50, 89.00)	77.09 (61.11, 89.96)	31.72	5	0.016 (0.003-0.028)	6.76×10^{-6}
Stakeholder involvement	42.34 (25.66, 50.00)	43.66 (30.54, 56.91)	40.30 (20.83, 58.83)	50.00 (31.50, 67.75)	51.90 (37.00, 72.00)	52.39 (30.67, 67.71)	41.84	5	0.023 (0.008-0.037)	6.35×10^{-8}
Rigor of development	30.47 (19.55, 52.50)	28.05 (18.07, 51.95)	37.50 (12.25, 64.44)	51.00 (27.09, 72.91)	56.48 (34.38, 75.13)	57.10 (31.12, 71.11)	58.69	5	0.033 (0.015-0.051)	2.27×10^{-11}
Clarity of presentation	51.39 (44.40, 70.83)	65.97 (49.48, 75.35)	70.90 (47.05, 83.33)	76.00 (58.33, 89.00)	83.00 (69.00, 91.44)	86.00 (71.15, 91.67)	89.57	5	0.052 (0.031-0.073)	8.28×10^{-18}
Applicability	24.50 (11.46, 46.83)	30.38 (18.80, 42.93)	29.09 (15.15, 47.92)	35.40 (18.90, 57.00)	42.00 (20.83, 61.10)	32.80 (18.81, 50.51)	30.30	5	0.016 (0.004-0.027)	1.29×10^{-5}
Editorial independence	6.25 (4.16, 37.88)	20.41 (0.00, 46.30)	31.46 (0.00, 64.35)	53.00 (20.81, 79.00)	70.80 (42.00, 88.00)	71.61 (43.48, 91.92)	171.96	5	0.103 (0.075-0.133)	2.78×10^{-35}

Statistical significance for each domain was determined using a two-sided Kruskal-Wallis test, followed by Dunn's post-hoc test with Bonferroni adjustment for multiple comparisons.

Supplementary Table 4. Comparison of AGREE II scores for CPGs (Pre- vs Post-2011). Values are median (interquartile range)

Domain	Pre-2011 (n=308)	Post-2011 (n=1325)	W	δ (95% CI)	p-value
Scope and purpose	68.1 (50.0, 83.3)	76.7 (60.0, 88.9)	167797	-0.178 (-0.249 to -0.104)	1.15×10^{-6}
Stakeholder involvement	42.5 (22.0, 58.0)	51.0 (33.3, 69.4)	158853	-0.222 (-0.289 to -0.151)	1.33×10^{-9}
Rigor of development	34.7 (13.5, 61.2)	54.7 (30.2, 74.0)	150094	-0.264 (-0.335 to -0.191)	4.55×10^{-13}
Clarity of presentation	69.0 (47.2, 83.0)	80.6 (63.0, 90.3)	147869	-0.275 (-0.343 to -0.205)	4.78×10^{-14}
Applicability	29.1 (15.6, 47.9)	38.0 (19.0, 59.6)	168901.5	-0.172 (-0.239 to -0.104)	2.41×10^{-6}
Editorial independence	29.1 (0.0, 60.6)	60.4 (29.2, 83.3)	120831	-0.408 (-0.468 to -0.344)	5.31×10^{-29}

Statistical significance for each domain was determined using a two-sided Wilcoxon rank-sum test. No adjustments for multiple comparisons were made.

Supplementary Table 5. Comparison of AGREE II scores for CPGs across different regions. Values are median (interquartile range)

Domain	Africa (n=37)	America (n=563)	Asia (n=326)	Cross-region (n=163)	Europe (n=496)	Oceania (n=48)	H	df	ε^2 (95%CI)	p-value
Scope and purpose	65 (54-78)	80.6 (65.3-91.7)	65 (37-79)	74.1 (64.3-84.7)	79 (58-90)	75.2 (62.5-94.1)	121.7603	5	0.072 (0.051-0.100)	1.33×10^{-24}
Stakeholder involvement	46 (36-54)	52.8 (38-71)	33 (15-54.1)	51.4 (40.9-64.7)	53.5 (34.7-74)	49.2 (31-68.2)	126.5649	5	0.075 (0.051-0.103)	1.27×10^{-25}
Rigor of development	4 (1-26)	59 (35-78)	26.8 (5-56.3)	60 (42.1-66.8)	54.8 (32.3-75)	40.9 (25.4-65.2)	219.7498	5	0.132 (0.101-0.166)	1.68×10^{-45}
Clarity of presentation	80.6 (74-89)	81 (67-92)	62.1 (42.2-81)	83 (70.8-87)	79.1 (63-90.3)	76.3 (57.7-89.3)	120.3858	5	0.071 (0.048-0.102)	2.60×10^{-24}
Applicability	19 (11-42)	38.9 (21-57.5)	25 (12.5-41.9)	50 (31.3-62.5)	38 (19.8-62.5)	35.2 (18-63.5)	95.58378	5	0.056 (0.040-0.081)	4.50×10^{-19}
Editorial independence	8 (0-28)	63 (33-83)	17 (0-56.2)	81 (51-94.4)	58.3 (33.3-79.2)	48.6 (24.2-71.5)	223.8273	5	0.134 (0.105-0.169)	2.25×10^{-46}

Statistical significance for each domain was determined using a two-sided Kruskal-Wallis test, followed by Dunn's post-hoc test with Bonferroni adjustment for multiple comparisons.

Supplementary Table 6. Comparison of AGREE II scores for CPGs across different diseases. Values are median (interquartile range)

Domain		Cardiovascular diseases	Chronic respiratory diseases	Diabetes and kidney diseases	Digestive diseases	Mental disorders	Musculoskeletal disorders	Neoplasms	Neurological disorders	Sense organ diseases	Skin and subcutaneous diseases	H	df	ε^2 (95%CI)	p- value
Scope and purpose		69.4 (58.3–86.1)	76 (68.5 – 85.5)	77.5 (47.2 – 94)	83.2 (53.5– 94.4)	70 (50 – 85.5)	80.6 (60–92.5)	73.6 (59.7– 83)	75.4 (61–85.8)	81 (53.5 – 86.6)	83 (68–89)	20.30936	9	0.012 (0.004-0.049)	1.61×10 ⁻²
Stakeholder involvement		48.8 (37.1–58.3)	52 (37–67)	32.8 (16.8 – 72.5)	34.7 (19.2– 67.4)	48.3 (29.5– 67)	50 (34.7–67)	50 (38.9 – 59.8)	47.1 (29–66.3)	51 (30.8 – 68.5)	44.4 (31.9 – 64.8)	21.67396	9	0.014 (0.003-0.049)	9.97×10 ⁻³
Rigor of development		32.2 (20.8–58.3)	61 (44–75)	39 (6.2 – 63.2)	52 (25.8 – 71.9)	49 (18.5 – 69)	60 (36–79)	54.1 (35.3– 67.2)	43.4 (24–70.7)	25 (18–53)	65.6 (39–73)	47.62955	9	0.042 (0.027-0.082)	3.00×10 ⁻⁷
Clarity of presentation		66.7 (50–85.7)	75 (58.5 – 85)	83 (57.5 – 100)	70.7 (51.5– 91.7)	73.1 (62.5– 83)	79.8 (61–91.8)	83.3 (67.8– 87)	77.9 (61.1–88)	68.1 (54.7– 90.3)	83 (68.1–89)	21.10365	9	0.013 (0.005-0.049)	1.22×10 ⁻²
Applicability		41.7 (28.9–54.1)	31 (16.5 – 50)	24 (9.8 – 52.3)	22.9 (6.2 – 43.8)	32.2 (16.5– 57)	34 (17–56.5)	40.3 (19.6– 59.5)	32.1 (20.2 – 62.5)	31 (21.4 – 37.8)	47 (33–60)	40.78899	9	0.035 (0.020-0.074)	5.46×10 ⁻⁶
Editorial independence		50.6 (23.3–78.8)	48 (31–75)	43 (17–83.9)	53.5 (15.6– 83.3)	58 (33 – 72.1)	66.7 (35.5–83.3)	70.1 (47.9– 94.4)	47.5 (19.5 – 76.6)	50 (12.5 – 88.2)	44 (25–65)	47.43507	9	0.042 (0.027-0.082)	3.26×10 ⁻⁷

Statistical significance for each domain was determined using a two-sided Kruskal-Wallis test, followed by Dunn’s post-hoc test with Bonferroni adjustment for multiple comparisons.

Supplementary Table 7. Sensitivity analysis of national-level associations between the annual number of CPGs and disease burden metrics across one- to five-year lags.

Lag (years)	RR (DALYs rate)	95%CI	<i>p</i> -value	RR (Deaths rate)	95%CI	<i>p</i> -value
1	0.9997	0.9965-1.0028	0.836	0.9980	0.9939-1.0021	0.330
2	0.9997	0.9965-1.0029	0.858	0.9980	0.9937-1.0022	0.343
3	0.9998	0.9966-1.0030	0.887	0.9980	0.9936-1.0023	0.358
4	0.9998	0.9966-1.0031	0.920	0.9980	0.9935-1.0025	0.376
5	0.9999	0.9966-1.0032	0.937	0.9979	0.9933-1.0026	0.378

Generalized Estimating Equations (GEE) are used for data analysis. All *p*-values are two-sided. No adjustments for multiple comparisons were made.

Supplementary Table 8. National-level associations between the annual proportion of CPGs scoring >70% in each AGREE II domain and disease burden with a two-year lag.

Domain	RR (DALYs rate)	95%CI	<i>p</i> -value	RR (Deaths rate)	95%CI	<i>p</i> -value
Scope and purpose	1.0089	0.9746-1.0443	0.616	1.0006	0.9493-1.0547	0.982
Stakeholder involvement	1.0159	0.9710-1.0628	0.495	1.0284	0.9588-1.1031	0.433
Rigor of development	1.0142	0.9713-1.0589	0.523	0.9918	0.9262-1.0620	0.814
Clarity of presentation	1.0232	0.9872-1.0604	0.209	1.0067	0.9531-1.0633	0.811
Applicability	1.0457	0.9955-1.0985	0.075	1.0613	0.9735-1.1570	0.177
Editorial independence	1.0098	0.9701-1.0512	0.632	0.9955	0.9329-1.0623	0.891

Generalized Estimating Equations (GEE) are used for data analysis. All *p*-values are two-sided. No adjustments for multiple comparisons were made.

Supplementary Table 9. National-level associations between the annual proportion of CPGs scoring >70% in each AGREE II domain and disease burden with a three-year lag.

Domain	RR (DALYs rate)	95%CI	<i>p</i> -value	RR (Deaths rate)	95%CI	<i>p</i> -value
Scope and purpose	1.0057	0.9673-1.0455	0.776	0.9891	0.9297-1.0521	0.727
Stakeholder involvement	1.0167	0.9690-1.0668	0.499	1.0338	0.9569-1.1169	0.399
Rigor of development	1.0167	0.9713-1.0642	0.478	0.9968	0.9255-1.0736	0.933
Clarity of presentation	1.0193	0.9807-1.0593	0.331	0.9981	0.9390-1.0610	0.952
Applicability	1.0519	0.9968-1.1101	0.065	1.0784	0.9792-1.1876	0.125
Editorial independence	1.0074	0.9650-1.0516	0.737	0.9981	0.9302-1.0709	0.957

Generalized Estimating Equations (GEE) are used for data analysis. All *p*-values are two-sided. No adjustments for multiple comparisons were made.

Supplementary Table 10. National-level associations between the annual proportion of CPGs scoring >70% in each AGREE II domain and disease burden with a four-year lag.

Domain	RR (DALYs rate)	95%CI	p-value	RR (Deaths rate)	95%CI	p-value
Scope and purpose	1.0076	0.9673-1.0496	0.715	0.9866	0.9267-1.0504	0.673
Stakeholder involvement	1.0022	0.9588-1.0476	0.923	1.0035	0.9400-1.0714	0.916
Rigor of development	0.9974	0.9535-1.0433	0.909	0.9622	0.8996-1.0291	0.262
Clarity of presentation	1.0143	0.9750-1.0552	0.48	0.9987	0.9400-1.0610	0.966
Applicability	1.0308	0.9812-1.0830	0.228	1.0285	0.9466-1.1174	0.507
Editorial independence	1.0072	0.9621-1.0543	0.76	0.9913	0.9246-1.0629	0.807

Generalized Estimating Equations (GEE) are used for data analysis. All *p*-values are two-sided. No adjustments for multiple comparisons were made.

Supplementary Table 11. National-level associations between the annual proportion of CPGs scoring >70% in each AGREE II domain and disease burden with a five-year lag.

Domain	RR (DALYs rate)	95%CI	p-value	RR (Deaths rate)	95%CI	p-value
Scope and purpose	1.0142	0.9715-1.0588	0.521	0.9948	0.9307-1.0633	0.879
Stakeholder involvement	0.9923	0.9475-1.0392	0.743	0.9842	0.9215-1.0511	0.634
Rigor of development	0.9947	0.9472-1.0447	0.833	0.953	0.8868-1.0242	0.19
Clarity of presentation	1.0107	0.9697-1.0535	0.614	0.9964	0.9347-1.0622	0.912
Applicability	1.0085	0.9594-1.0601	0.741	0.9897	0.9132-1.0726	0.8
Editorial independence	1.0088	0.9595-1.0606	0.731	0.9877	0.9161-1.0650	0.748

Generalized Estimating Equations (GEE) are used for data analysis. All *p*-values are two-sided. No adjustments for multiple comparisons were made.

Supplementary Table 12. Sensitivity analysis of national-level associations between the annual proportion of CPGs scoring >70% in each AGREE II domain and disease burden with a one-year lag.

Domain	RR (DALYs rate)	95%CI	p-value	RR (Deaths rate)	95%CI	p-value
Scope and purpose	0.9983	0.9848-1.0120	0.806	0.9968	0.9716-1.0227	0.806
Stakeholder involvement	1.0011	0.9931-1.0091	0.791	1.0025	0.9852-1.0200	0.782
Rigor of development	0.9999	0.9937-1.0061	0.973	0.9999	0.9861-1.0140	0.992
Clarity of presentation	0.9995	0.9895-1.0095	0.919	0.9988	0.9761-1.0221	0.921
Applicability	1.0009	0.9945-1.0073	0.792	1.002	0.9877-1.0166	0.782
Editorial independence	1.0001	0.9900-1.0103	0.979	1.0023	0.9736-1.0318	0.877

Generalized Estimating Equations (GEE) are used for data analysis. All *p*-values are two-sided. No adjustments for multiple comparisons were made.

Supplementary Table 13. Sensitivity analysis of national-level associations between the annual proportion of CPGs scoring >70% in each AGREE II domain and disease burden with a two-year lag.

Domain	RR (DALYs rate)	95%CI	p-value	RR (Deaths rate)	95%CI	p-value
Scope and purpose	0.9983	0.9851-1.0117	0.805	0.9967	0.9709-1.0232	0.806
Stakeholder involvement	1.0011	0.9934-1.0088	0.786	1.0024	0.9853-1.0199	0.782
Rigor of development	0.9999	0.9934-1.0065	0.981	0.9999	0.9860-1.0141	0.992
Clarity of presentation	0.9995	0.9896-1.0094	0.914	0.9988	0.9762-1.0220	0.921
Applicability	1.0008	0.9946-1.0070	0.803	1.0020	0.9877-1.0165	0.785
Editorial independence	1.0000	0.9902-1.0099	0.998	1.0024	0.9729-1.0328	0.877

Generalized Estimating Equations (GEE) are used for data analysis. All *p*-values are two-sided. No adjustments for multiple comparisons were made.

Supplementary Table 14. Sensitivity analysis of national-level associations between the annual proportion of CPGs scoring >70% in each AGREE II domain and disease burden with a three-year lag.

Domain	RR (DALYs rate)	95%CI	p-value	RR (Deaths rate)	95%CI	p-value
Scope and purpose	0.9985	0.9867-1.0105	0.809	0.9971	0.9744-1.0203	0.805
Stakeholder involvement	1.0009	0.9942-1.0076	0.795	1.0023	0.9864-1.0184	0.782
Rigor of development	0.9999	0.9945-1.0053	0.960	0.9999	0.9874-1.0125	0.986
Clarity of presentation	0.9995	0.9907-1.0083	0.903	0.9989	0.9786-1.0198	0.92
Applicability	1.0006	0.9954-1.0059	0.812	1.0018	0.9888-1.0150	0.784
Editorial independence	0.9997	0.9916-1.0079	0.944	1.0019	0.9763-1.0282	0.887

Generalized Estimating Equations (GEE) are used for data analysis. All *p*-values are two-sided. No adjustments for multiple comparisons were made.

Supplementary Table 15. Sensitivity analysis of national-level associations between the annual proportion of CPGs scoring >70% in each AGREE II domain and disease burden with a four-year lag.

Domain	RR (DALYs rate)	95%CI	p-value	RR (Deaths rate)	95%CI	p-value
Scope and purpose	0.9986	0.9869-1.0103	0.809	0.997	0.9736-1.0210	0.807
Stakeholder involvement	1.0007	0.9952-1.0063	0.794	1.0021	0.9872-1.0172	0.782
Rigor of development	0.9999	0.9952-1.0045	0.950	0.9998	0.9879-1.0119	0.980
Clarity of presentation	0.9994	0.9912-1.0077	0.895	0.9989	0.9789-1.0192	0.912
Applicability	1.0005	0.9960-1.0050	0.819	1.0017	0.9894-1.0142	0.786
Editorial independence	0.9995	0.9923-1.0068	0.900	1.0018	0.9771-1.0271	0.888

Generalized Estimating Equations (GEE) are used for data analysis. All *p*-values are two-sided. No adjustments for multiple comparisons were made.

Supplementary Table 16. Sensitivity analysis of national-level associations between the annual proportion of CPGs scoring >70% in each AGREE II domain and disease burden with a five-year lag.

Domain	RR (DALYs rate)	95%CI	<i>p</i> -value	RR (Deaths rate)	95%CI	<i>p</i> -value
Scope and purpose	0.9989	0.9897-1.0082	0.813	0.9975	0.9778-1.0177	0.810
Stakeholder involvement	1.0005	0.9964-1.0045	0.818	1.0017	0.9895-1.0140	0.784
Rigor of development	0.9998	0.9965-1.0032	0.929	0.9998	0.9900-1.0097	0.975
Clarity of presentation	0.9996	0.9938-1.0054	0.891	0.9990	0.9821-1.0162	0.909
Applicability	1.0003	0.9970-1.0036	0.868	1.0014	0.9912-1.0116	0.793
Editorial independence	0.9991	0.9922-1.0060	0.799	1.0009	0.9826-1.0197	0.920

Generalized Estimating Equations (GEE) are used for data analysis. All *p*-values are two-sided. No adjustments for multiple comparisons were made.