Assessment of quality of systematic reviews in dermatology using AMSTAR 2 Part 2 of 2. Validity and reliability testing of AMSTAR 2

Rowena Natividad S.Flores-Genuino,[1] Maria Christina Filomena R. Batac,[2]
Anne Julienne M. Genuino,[3] Ian Theodore G. Cabaluna[4]

ABSTRACT

BACKGROUND

AMSTAR 2 enables a more detailed assessment of systematic reviews and includes non-randomised studies of healthcare interventions, compared to its earlier version, AMSTAR. We validated AMSTAR 2 in a group of systematic reviews in dermatology in the Philippines.

METHODS

We used a cohort of systematic reviews (SRs) in dermatology from the Philippine that were previously described in Part 1 of this 2-part series. The SRs included clinical trials on any intervention for the treatment or prevention of a dermatologic disease or for maintenance of healthy skin, hair or nails. Two reviewers independently extracted data and used AMSTAR 2 to appraise the methodological quality of each included SR. We determined construct validity by comparing the number of critical flaws between a set of non-Cochrane and matched Cochrane reviews, using Wilcoxon rank sum test. We tested for interrater reliability of the AMSTAR 2 tool using Gwet's AC1 statistic.

RESULTS:

We included 20 non-Cochrane systematic reviews in dermatology by Philippine-based authors, and a set of 20 reviews from the Cochrane skin group, matched by year and randomly chosen. Construct validity testing showed a significantly greater number of AMSTAR 2 critical flaws (median 4.5 vs 0.0; z=3.64; P=0.000) and non-critical weaknesses (5 vs 2.0; z-score=3.10; P-value=0.001) by non-Cochrane reviews compared to a matched set of Cochrane skin group reviews. There was good interrater reliability (average Gwet's AC1 statistic = 0.87) with the lowest agreement (0.62) for discussion of heterogeneity (item 14), and the highest agreement (0.97) for study selection criteria (item 3).

CONCLUSION

The AMSTAR 2 was a valid and reliable tool for assessing systematic reviews using a cohort of

[1] Department of Anatomy, College of Medicine, University of the Philippines, Manila, Philippines
[1] Department of Dermatology, Makati Medical Center, Makati, Philippines
[1] Section of Dermatology, Department of Medicine, Manila Doctors Hospital
[2] Department of Dermatology, University of the Philippines Manila-Philippine General Hospital
[3] Health Technology Assessment Unit, Department of Health, Manila, Philippines
[4] Department of Clinical Epidemiology, College of Medicine, University of the Philippines-Manila; Asia Pacific Center for Evidence Based Healthcare

UPM-RGAO Registration No. 2019-0681

Source of funding: Philippine Dermatological Society

Conflict of interest: none

reviews by dermatology reviews, both non-Cochrane and Cochrane. Further validation of the AMSTAR 2 is needed to determine if it can be applied to a wide variety of systematic reviews.

Key words: AMSTAR, AMSTAR 2, dermatology, validity, reliability, systematic reviews, meta-analysis

INTRODUCTION

The AMSTAR (A Measurement Tool to Assess Systematic Reviews), initially developed in 2007 to evaluate the methodological quality of systematic reviews of randomised trials, was recently updated to AMSTAR 2 in 2017. In addition to a more detailed assessment of systematic reviews, it included non-randomised studies of healthcare interventions.[1] The 16-item tool includes assessing the research question and inclusion criteria, protocol, study design selection, search strategy, study selection and data extraction process, statistical analysis, risk of bias analysis, source of funding and conflict of interest disclosure. An overall confidence rating in the results of a systematic review (high, moderate, low or critically low) can be determined based on seven identified critical domains. The AMSTAR 2 has been initially validated by the developers and showed that most items had moderate to substantial level of agreement. Further validation is being encouraged by the developers to improve the usability of AMSTAR 2.[1] In a multi-center study (N=30 RCTs), interrater reliability varied by center, but across all centers was substantial (AC1 0.61 to 0.80) to almost perfect (AC1 0.81 to 0.99) for 8/11 (73%) AMSTAR, and 8/16 (50%) AMSTAR 2. Inter-center reliability was substantial to almost perfect for 6/11 (55%) AMSTAR, and 12/16 (75%) AMSTAR 2 items. Agreement on confidence in the results of the review (AMSTAR 2) ranged from slight (AC1 0.05, 95% confidence interval (CI) −0.17 to 0.27) to perfect (1.00) between review authors and moderate (AC1 0.58, 95% CI 0.30 to 0.85) to substantial (AC1 0.74, 95% CI 0.30 to 0.85) across centers.[2] Another study showed moderate agreement for AMSTAR 2 (median kappa, 0.51), and a substantial agreement for AMSTAR (median kappa, 0.62) for two groups of four raters each assigned one of two samples of systematic reviews (N=30).[3] A recently completed validation study that was presented at the 2019 Cochrane Colloqium showed substantial to almost perfect agreement in 12/16 (75%) items; with poor agreement in four items (#6, 12, 13 and 14) on duplicate data extraction, impact of risk of bias and accounting for it in interpretation of results, and satisfactory explanation of heterogeneity.[2,4] The authors recommended the need for improved clarity and guidance, transparent reporting of decision rules by authors of overviews of reviews,[5] and caution when using the AMSTAR 2 quality of risk of bias ratings as inclusion criteria for systematic reviews.

There is a need to validate the AMSTAR 2 to enable a more reliable assessment of systematic reviews and a more robust evidence base for researchers and clinicians, which will eventually lead to rational clinical practice.

OBJECTIVES

1. To determine the construct validity of AMSTAR 2 - Construct validity refers to whether or not a proposition assumed to exist is confirmed with use of the tool.[6]

2. To determine the inter-rater reliability of AMSTAR 2 – Inter-rater reliability is the extent to which two or more raters agree, addressing the issue of consistency in implementing the tool.[7]

METHODS

A registered protocol for this validation study is available upon request from the author.

Details of inclusion and exclusion criteria for study eligibility, list of databases and secondary sources searched, and the screening and data extraction process were described in Part 1 of this paper series.[8] Since there was only one Cochrane review in the included studies, we searched for a matched cohort of reviews (matched by year of publication) published by the Cochrane Skin Group, as comparator group for the non-Cochrane reviews to test construct validity.

A. AMSTAR 2 tool

There are 16 items in the tool that are assessed as follows (Table 1):

Table 1. 16 items of AMSTAR 2 tool

No.	Item	Responses
1	Did the research questions and inclusion criteria for the review include the components of PICO?	'Yes' or 'No'
2	**Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?**	**'Yes', 'Partial Yes', or 'No'**
3	Did the review authors explain their selection of the study designs for inclusion in the review?	'Yes' or 'No'
4	**Did the review authors use a comprehensive literature search strategy?**	**'Yes', 'Partial Yes', or 'No'**
5	Did the review authors perform study selection in duplicate?	'Yes' or 'No'
6	Did the review authors perform data extraction in duplicate?	'Yes' or 'No'
7	**Did the review authors provide a list of excluded studies and justify the exclusions?**	**'Yes', 'Partial Yes', or 'No'**
8	Did the review authors describe the included studies in adequate detail?	'Yes', 'Partial Yes', or 'No'
9	**Did the review authors use a satisfactory technique for assessing the risk of bias in individual studies that were included in the review?**	**'Yes', 'Partial Yes', or 'No'**
10	Did the review authors report on the sources of funding for the studies included in the review?	'Yes' or 'No'
11	**If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?**	**'Yes', 'Partial Yes', or 'No meta-analysis conducted'**
12	If meta-analysis was performed, did the review authors assess the potential impact of risk of bias in individual studies on the results of the meta-analysis or other evidence synthesis?	**'No meta-analysis conducted'**
13	**Did the review authors account for risk of bias in individual studies when interpreting/ discussing the results of the review?**	'Yes' or 'No'
14	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	'Yes' or 'No'
15	**If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?**	**'No meta-analysis conducted'**
16	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	'Yes' or 'No'

Note: Critical items are in bold font

There are 7 critical items (2, 4, 7, 9, 11, 13, and 15) while the remaining 9 are non-critical items. A response of 'no' to critical items constitute a critical flaw; while for non-critical items, a non-critical weakness. The number of these critical flaws and non-critical weaknesses are used to rate the overall confidence in the results of the review (Table 2)

Table 2. Overall confidence rating

Rating	Description
High	No or one non-critical weakness: the systematic review provides an accurate and comprehensive summary of the results of the available studies that address the question of interest
Moderate	More than one non-critical weakness*: the systematic review has more than one weakness but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review
Low	One critical flaw with or without non-critical weaknesses: the review has a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest
Critically low	More than one critical flaw with or without non-critical weaknesses: the review has more than one critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies

B. Data collection

The method of data collection is the same as in Part 1 of this paper series: independent assessments using the AMSTAR-2 tool by two reviewers, with resolution of disagreements by consensus or a third reviewer, using a pre-tested data collection form. We compared two sets of systematic reviews: Non-Cochrane (n=20) and Cochrane (n=20). Data for the Non-Cochrane group was collected from 20 of 21 dermatology reviews by Philippine-based authors in Part 1 of this paper series. The remaining review was a Cochrane review, and was added to 19 Cochrane skin group reviews that were matched as to year. We hypothesized that the Cochrane reviews have higher compliance with AMSTAR 2 items, less critical flaws and non-critical weaknesses, and higher overall confidence rating than non-Cochrane reviews, as a test for construct validity.

C. Outcomes

a. Construct validity – whether the cohort of non-Cochrane systematic reviews had higher median number of critical flaws than a matched cohort of Cochrane systematic reviews

b. Interrater reliability – whether the pair of reviewers agreed on responses to each AMSTAR 2 item using Gwet's AC1 statistic (Table 3).

Table 3. Gwet's AC1 statistic interpretation

Gwet's AC1	Agreement
<0.00	Poor
0.00 to 0.20	Slight
0.21 to 0.40	Fair
0.41 to 0.60	Moderate
0.61 to 0.80	Substantial
0.81 to 0.99	Almost perfect
1.00	Perfect

D. Data analysis

Descriptive analysis such as means and SD for continuous data, and frequency and percentage distribution for categorical data was done using Microsoft Excel. We used Wilcoxon rank sum test to compare median number of critical flaws and non-critical weaknesses between non-Cochrane and Cochrane reviews. We used Z-test to compare proportions of studies that reported each AMSTAR 2 item, as well as the proportion of studies for each overall confidence rating. We computed for Gwet's AC1 statistic to determine interrater reliability. Compared to Cohen's Kappa, Gwet's AC1 was shown to provide a more stable inter-rater reliability coefficient and is less affected by prevalence and marginal probability.[9]

RESULTS:

Pretesting

We pretested the AMSTAR 2 tool among three assessors using three systematic reviews (two non-Cochrane,[10,11] and one Cochrane[12]) and discussed items wherein there was disagreement in responses until we came to a consensus on how to harmonize our assessments.

Search results

We included a total of 40 SRs (20 non-Cochrane SRs; and 20 Cochrane SRs) in this validation study (Flow diagram for non-Cochrane reviews, Fig 1; Flow diagram for Cochrane reviews, Fig. 2).

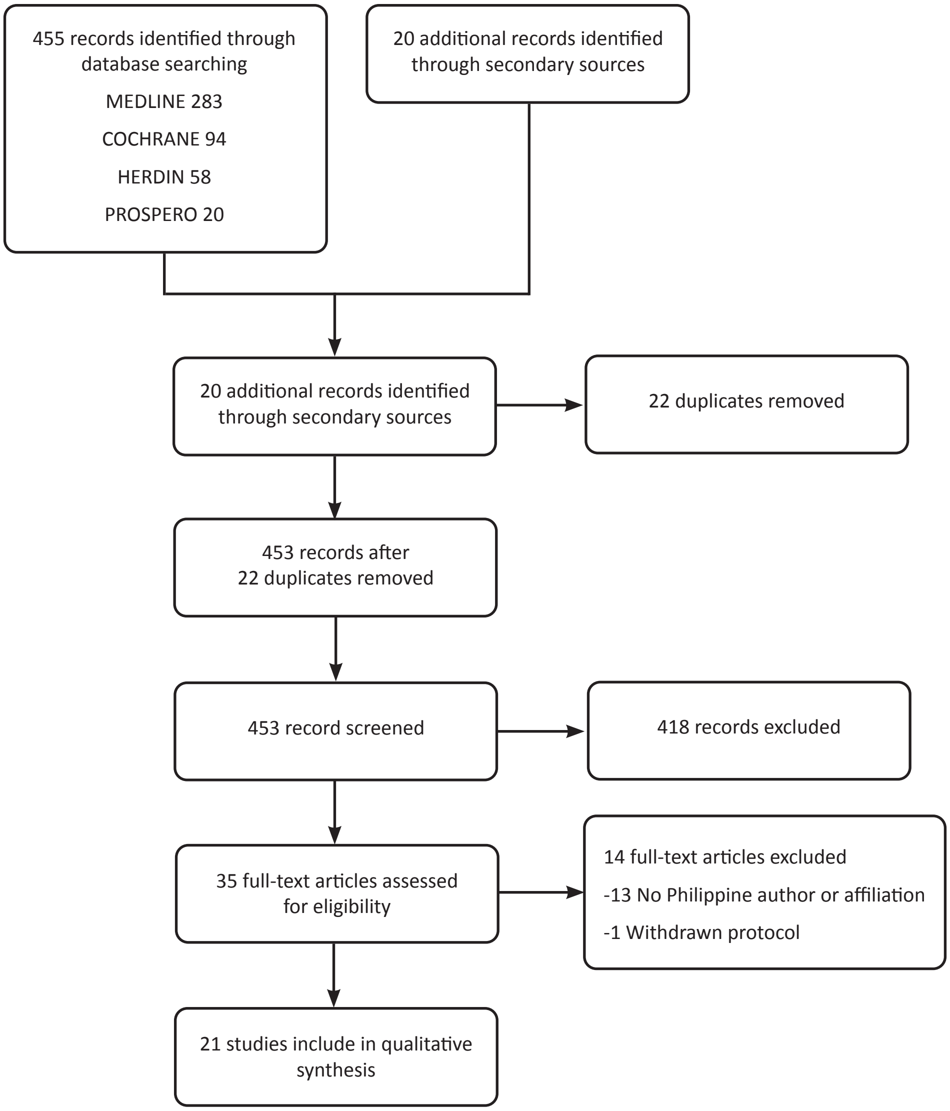
455 records identified through database searching
MEDLINE 283
COCHRANE 94
HERDIN 58
PROSPERO 20
20 additional records identified through secondary sources
20 additional records identified through secondary sources
22 duplicates removed
453 records after 22 duplicates removed
453 record screened
418 records excluded
35 full-text articles assessed for eligibility
14 full-text articles excluded
-13 No Philippine author or affiliation
-1 Withdrawn protocol
21 studies include in qualitative synthesis

Figure 1. Study flow diagram (20 non-Cochrane reviews)

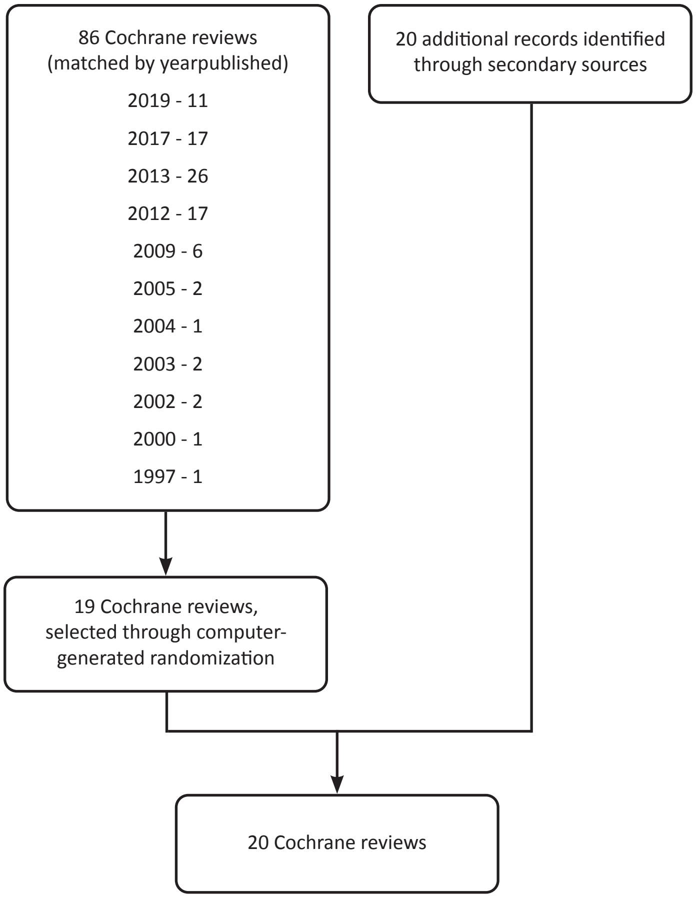
86 Cochrane reviews
(matched by yearpublished)
2019 - 11
2017 - 17
2013 - 26
2012 - 17
2009 - 6
2005 - 2
2004 - 1
2003 - 2
2002 - 2
2000 - 1
1997 - 1
20 additional records identified through secondary sources
19 Cochrane reviews, selected through computer-generated randomization
20 Cochrane reviews

Figure 2. Study flow diagram (20 Cochrane reviews)

Characteristics of Included Studies

The characteristics of the 40 included SRs (20 non-Cochrane and 20 Cochrane)are shown in Table 3. The non-Cochrane and Cochrane groups were similar in the decade of publication (60 vs 65% published in the 2010s), number of authors (median, 3 vs 4), percentage of university-based authors (85 vs 100%), disease category (60 vs 50% were infections and infestations, and eczemas), route of administration (50–55% were topical; 45% were oral), number of included studies (median, 4 vs 6), and number of participants (424 vs 328). However, they differed in the specialty of authors since in the non-Cochrane group, majority (85%) were in dermatology while the Cochrane group had more varied specialties (dermatology, 35%; researchers/statisticians 30%; public health/primary care/general practice, 20%). The Cochrane group had 100% of the studies being published in an indexed journal (i.e., Cochrane library) versus only 60% of the non-Cochrane group. However, the non-Cochrane reviews had a greater percentage of studies that mentioned PRISMA in the report (4/20 or 20%) despite the fact that only 8/11 (73%) of the journals in which they were published instructed authors to use PRISMA reporting checklist.

Table 3. Characteristics of Included Studies

Study characteristic	Non-Cochrane SRs (n=20) No. (%)	Cochrane SRs (n=20) No. (%)
Year of publication		
1990s	2 (10)	0
2000s	6 (30	7 (35)
2010s	12 (60)	13 (65)
No. of authors, median (range)	3 (1, 13)	4 (2, 9)
Specialty of authors*		
Dermatology	17 (85)	7 (35)
Community/Family Medicine	2 (10)	1 (5)
Researchers/Statisticians	1 (5)	6 (30)
Cardiology	1 (5)	0
Health sciences/Medicine	1 (5)	3 (15)
Public health/Primary care/GP	0	4 (20)
Pediatrics	0	2 (10)
Infectious disease/Tropical medicine	0	3 (15)
Clinical social medicine	0	2 (10)
Immuno-allergy	0	2 (10)
Preventive medicine	0	2 (10)
Type of institutional affiliation		
University-based	17 (85)	20 (100)
Not university-based	2 (10)	0
No information	1 (5)	0
Disease category		
Infections/Infestations	8 (40)	5 (25)
Eczemas	4 (20)	5 (25)
Papulosquamous disorders	2 (10)	3 (15)
Diseases of hair follicle	2 (10)	0
Wounds/burns	0	3 (15)
Others	4 (20)	4 (20)
Route of administration*		
Topical	10 (50)	11 (55)
Oral	9 (45)	9 (45)
Phototherapy/Hyperbaric oxygen	1 (5)	1 (5)
Intralesional/Intravenous	2 (10)	1 (5)
Surgical	0	2 (10)
No. of included studies, Median (range)	4 (2, 94)	6 (0, 89)
No. of participants, Median (range)	425 (44, 42588)	329 (0, 10583)
No. of studies that mentioned PRISMA	4 (20)	1 (5)
Indexed journals	12(60)	20 (100)
Journal instruction to use PRISMA †	8/11 (73)	15/15 (100)

**Total exceeds n=20 since some studies had more than 1 route of administration; † published after 2009*

Compliance with AMSTAR 2 Items

The Cochrane reviews had a higher reporting rate (70 to 100%) for 15/16 items than the non-Cochrane reviews (0 to 85%) Figure 2). The largest differences in reporting rate were for two items: item #2 (having a pre-registered review protocol; 95% vs 5%) and item #15 (publication bias presence and impact; 100% vs 60%). however, both differences were not statistically significant (Chi-squared value 3.28, P=0.07; Chi-squared value, 2.12 and P=0.15, respectively). The smallest differences were for item #8 (list of excluded studies; 95% vs 85%) and item #1 (PICO; 95% vs 75%).

Figure 3. Percentage distribution of studies that reported each AMSTAR 2 item

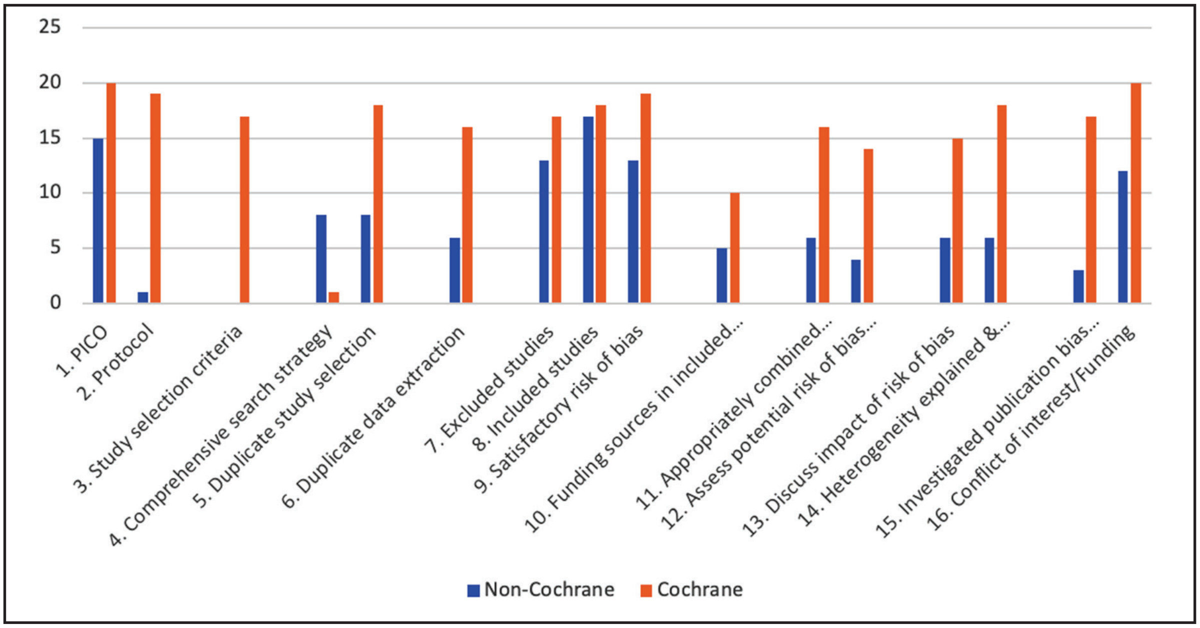

Table 4. Proportions of studies that reported each AMSTAR 2 item

No.	AMSTAR 2 Item	Non-Cochrane (n=20)		Cochrane (n=20)		Z-score	P value
		No.	%	No.	%		
1	PICO	15	75	19	95	-1.7712	0.07672
2	Protocol	1	5	18	90	0.53826	<0.00001*
3	Study selection criteria	0	0	0	0	NA	NA
4	Comprehensive search strategy	8	40	18	90	-3.315	0.00094*
5	Duplicate study selection	8	40	16	80	-2.582	0.00988*
6	Duplicate data extraction	6	30	20	100	-4.641	<0.0001
7	Excluded studies	13	65	18	90	-1.8932	0.05876
8	Included studies	17	85	19	95	-1.0541	0.23972
9	Satisfactory risk of bias	13	65	19	95	-2.3717	0.1778
10	Funding sources in included studies	5	25	10	50	-1.633	1.031
11	Appropriately combined studies in meta-analysis	6	30	16	80	-3.1782	0.00148*
12	Assess potential risk of bias impact on meta-analysis	4	20	14	70	-3.1782	0.00148*
13	Discuss impact of risk of bias	6	30	15	75	-3.1782	0.00438*
13	Heterogeneity explained & discussed	6	30	18	90	-3.873	0.0001*
15	Investigated publication bias presence and impact	3	15	17	85	-4.4272	0.00001*
16	Conflict of interest/Funding	12	60	20	100	-3.1623	0.00158*

PICO, Population/Intervention/Comparison/Outcome

Construct validity testing

Construct validity testing showed a significantly greater number of AMSTAR 2 critical flaws (median 4.5 vs 0.0; z=3.64; P=0.000) and non-critical weaknesses (5 vs 2.0; z-score=3.10; P-value=0.001) by non-Cochrane reviews compared to a matched set of Cochrane skin group reviews (Table 5).

Table 5. Comparison of median number of critical flaws and non-critical weaknesses between Non-Cochrane and Cochrane reviews

	Non-Cochrane reviews (n=20)	Cochrane reviews (n=20)	Z-score	P-value
No. of critical flaws	4.5	0	3.64	0.000
No. of non-critical weaknesses	5	2.0	3.10	0.001

The confidence rating was significantly higher in Cochrane reviews with 40% having high rating (P=0.00158) and 20% having moderate rating (P=0.03846) (vs none in the non-Cochrane reviews). On the other hand, there was a significantly higher proportion of non-Cochrane reviews with a critically low rating (95%) vs 30% of Cochrane reviews; P=0.00001) (Table 6).

Table 6. Frequency distribution of studies based on overall rating

Overall rating	Non-Cochrane reviews (n=20) No. (%)	Cochrane reviews (n=20) No. (%)	Z-score	P-value
High	0	8 (40)	3.1623	0.00158*
Moderate	0	4 (20)	2.1082	0.03846*
Low	1 (5)	2 (10)	0.6003	0.5485
Critically low	19 (95)	6 (30)	4.2458	0.00001*

Inter-reliability testing

Between two independent reviewers, there was good interrater reliability (average Gwet's AC1 statistic = 0.87; range 0.60 to 1.00). There was almost perfect to perfect agreement for 12/16 (75%) AMSTAR 2 items; almost perfect in 8 items (#1, 6 to 10, 12, and 16), perfect agreement in 4 items (#2, 3, 5, and 15), and substantial agreement in 3 items (#4, 11, and 14). Only one item had fair agreement (#14 on satisfactory explanation of heterogeneity) (Table 7).

Table 7. Interrater reliability of AMSTAR 2 between two raters (using Gwet's AC1 statistic)

Item No.	Description	Gwet's AC1 statistic (n=20)	Reliability
1	PICO	0.84	Almost perfect
2	Protocol	1.00	Perfect
3	Study selection criteria	1.00	Perfect
4	Comprehensive search strategy	0.77	Substantial
5	Duplicate study selection	1.00	Perfect
6	Duplicate data extraction	0.89	Almost perfect
7	Excluded studies	0.82	Almost perfect
8	Included studies	0.95	Almost perfect
9	Satisfactory risk of bias	0.95	Almost perfect
10	Funding sources in included studies	0.84	Almost perfect
11	Appropriately combined studies in meta-analysis	0.78	Substantial
12	Assess potential risk of bias impact on meta-analysis	0.84	Almost perfect
13	Discuss impact of risk of bias	0.78	Substantial
14	Heterogeneity explained & discussed	0.60	Fair
15	Investigated publication bias presence and impact	1.00	Perfect
16	Conflict of interest/Funding	0.89	Almost perfect
	Average	**0.87**	Almost perfect

DISCUSSION

The AMSTAR 2 tool showed acceptable construct validity and interrater reliability in a set of dermatology reviews. The non-Cochrane and Cochrane reviews were comparable in number of authors, type of institutional affiliation, disease category, route of administration of intervention, number of included studies and participants. However, they differed in the methods, since the Cochrane reviews strictly followed the Cochrane collaboration methods. In addition, the Cochrane reviews were all published in the Cochrane library, and had a more diverse background of its authors, including patient advocates and non-medical personnel.

Using the AMSTAR 2 tool to assess the methodological quality of the reviews, non-Cochrane reviews had a significantly greater number of critical flaws, which explains its lower overall rating in confidence level in the results. The greatest advantage of Cochrane reviews are the required publication of a study protocol prior to the review, and the intensive expert guidance and review by the Cochrane editorial team. However, it is notable that for item # 3 ("Did the review authors explain their selection of study designs for inclusion in the review?"), both Cochrane and non-Cochrane reviews had 0% compliance. This may be due to the fact that, in general, Cochrane intervention reviews, restrict their study design to RCTs as a practical way to deal with the fact that non-randomized studies are harder to track and identify and will delay their regular updates. Thus, Cochrane authors may no longer feel the need to explain their choice of study design as it is inherent in the Cochrane methods. On the other hand, non-Cochrane reviews are not necessarily limited by this study design restriction but still do not fulfill this AMSTAR 2 item adequately. Preregistration of a review protocol (item #2), reporting funding sources of included studies (item #10), and assessment of publication bias (item #15) were poorly reported in non-Cochrane reviews in our study. In a previous methodological review comparing 30 primary non-Cochrane reviews and their updated reviews, the authors also noted one item that was not reported by any of the reviews: item #3, on explanation of study design selection. There were also five poorly reported items (only in 33% or less of studies) — #2, review methods/protocol prior to the review; #7, list of excluded studies; #12, potential impact of risk of bias on meta-analysis; #15 adequate investigation of publication bias and its impact; and #16 potential sources of conflicts of interest (Gao 2019). A preregistered review protocol is important to ensure transparency and reduce risk of bias (Stewart 2012). The impact of the risk of bias must be considered in the overall quality of evidence as it may affect our confidence in the results (Guyatt 2011). Publication bias may lead to an overestimated treatment effect or suggest non-existing effects. Industry-sponsored trials may lead to more favorable efficacy results and conclusions than sponsorship by other sources (Lundh 2017).[13]

Interrater reliability was almost perfect to perfect in 75% of the items in our validation study, with greater agreement than a recent study presented at the Cochrane Colloqium 2019 that sampled 30 reviews and reported substantial to almost perfect agreement in only 56% (Gates 2019). In addition, unlike the previous validation study that had poor agreement for four items, our study only had fair agreement as the lowest score and only for item #14 (satisfactory explanation of heterogeneity). A probable reason why we also disagreed on whether heterogeneity was satisfactorily explained in the discussion is due to the various types of heterogeneity. Statistical, methodological and clinical heterogeneity may have been interpreted differently by the reviewers in our study.[14]

With the AMSTAR 2 being shown as valid and reliable tool to assess the methodologic quality of the set of dermatology systematic reviews in this study, it may then be used by authors, peer reviewers, and journal editors in determining the merit of systematic review reports submitted for publication. For dermatology clinicians who practice evidence-based medicine, knowing the methodologic quality of a systematic review report may help in deciding whether to trust the evidence.

CONCLUSION

The AMSTAR 2 tool showed acceptable construct validity when we compared a group of non-Cochrane dermatology systematic reviews from the Philippines, and a matched group of Cochrane systematic reviews. Non-Cochrane reviews had more critical flaws and non-critical weaknesses and had lower overall confidence rating than Cochrane reviews. There was good interrater reliability between two independent reviewers. However, items that are more subjective and need expertise, such as discussing impact of risk of bias in results of meta-analyses and assessing heterogeneity, need more clarity and guidance from the developer. The AMSTAR 2 tool should be further validated in other specialties and settings to determine its generalizability.

RECOMMENDATIONS

Improved clarity and guidance should be provided to increase level of agreement between raters using AMSTAR 2 tool. Raters should be explicit and transparent by reporting basis for judgments for each AMSTAR 2 item.

Quality or risk of bias ratings using AMSTAR 2 tool should be used cautiously since there are still items that are highly subjective and expertise-dependent.

REFERENCES

1. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *Br Med* J. 2017;358:1-9. doi:10.1136/bmj.j4008

2. Gates M, Gates A, Duarte G, et al. The reliability, usability, and applicability of tools to appraise quality and risk of bias in systematic reviews: A prospective evaluation of AMSTAR, AMSTAR 2 and ROBIS. In: *26th Cochrane Colloqium. 23 October 2019*.

3. Lorenz RC, Matthias K, Pieper D, et al. A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool. *J Clin Epidemiol*. 2019;114:133-140. doi:10.1016/j.jclinepi.2019.05.028

4. Gates A, Gates M, Duarte G, et al. Evaluation of the reliability, usability, and applicability of AMSTAR, AMSTAR 2, and ROBIS: Protocol for a descriptive analytic study. 2018:1-7.

5. Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: Planning a relevant research question and objective for an overview. *Syst Rev*. 2018;7(1):1-9. doi:10.1186/s13643-018-0695-8

6. MW C. Quantitative Survey Methods in Health Research. In: Saks M, Allsop J, eds. *Researching Health: Qualitative, Quantitative and Mixed Methods. Third Edition*. 3rd ed. London: SAGE Publications Ltd,; 2019:253.

7. Lange R. Inter-rater Reliability. In: Kreutzer J, DeLuca J, Caplan B, eds. *Encyclopedia of Clinical Neuropsychology*. New York: Springer; 2011:98. doi:10.1007/978-0-387-79948-3_1203

8. Flores-Genuino R, Batac M, Genuino A, Cabaluna I. Assessing quality of systematic reviews in dermatology from the Philippines using AMSTAR 2 Part 1: Methodologic quality of dermatological systematic reviews from the Philippines. *J Phil Dermatol Soc*. 2020.

9. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. *BMC Med Res Methodol.* 2013;13(1):1-7. doi:10.1186/1471-2288-13-61

10. Ganeshkumar P, Murhekar M V, Poornima V, et al. Dengue infection in India: A systematic review and meta-analysis. *PLoS Negl Trop Dis*. 2018:2-3. doi:10.1371/journal.pntd.0006618

11. Canlas KM, Macalintal-Canlas RA, Sakai F. Efficacy of calcitonin gene-related peptide antagonists in the treatment of acute migraine: A systematic review and meta-analysis. *Acta Med Philipp*. 2019;53(1):44-51.

12. Lazzerini M, Wanzira H. Oral zinc for treating diarrhoea in children (Review). *Cochrane Database Syst Rev.* 2016;(12). doi:10.1002/14651858.CD005436.pub5

13. Lundh A, Lexchin J, Mintzes B, Jb S, Bero L. Industry sponsorship and research outcome (Review). *Cochrane Database Syst Rev*. 2017;Issue 2. doi:10.1002/14651858.MR000033.pub3.www.cochranelibrary.com

14. Gagnier JJ, Moher D, Boon H, Beyene J, Bombardier C. Investigating clinical heterogeneity in systematic reviews: A methodologic review of guidance in the literature. *BMC Med Res Methodol 2012*,. 2012;12:111. practice in Nigeria. *Int J Dermatol* 2012;51(9):1086-9.