
Analysis of spin in the reporting of studies of topical treatments of photoaged skin



Catherine C. Motosko, BS,^{a,b} Anna K. Ault, BA,^a Laura L. Kimberly, MSW, MBE,^a George A. Zakhem, BS,^{a,b} M. David Gothard, MS,^c Roger S. Ho, MD, MS, MPH,^b and Alexes Hazen, MD^a
New York, New York, and East Canton, Ohio

Background: Spin—reporting that distorts the interpretation of results—is not unusual within scientific literature.

Objective: To appraise strategies of spin among placebo-controlled double-blind clinical trials of topical treatments for photoaged skin.

Methods: A systematic review of the literature was performed to identify placebo-controlled double-blind clinical trials of topical treatments for photoaged skin. A survey of spin strategies was developed and applied to the cohort of identified studies.

Results: The systematic review led to the identification of 20 studies in which various types of spin strategies, broadly classified as either inappropriate statistical analyses or inappropriate interpretation of results, were used. The most commonly used strategies included use of multiple primary outcomes (95%), inappropriate extrapolation of results from specific outcomes to global improvements (95%), focus on within-group comparisons (75%), and focus on interim analyses to give more weight to nonsignificant findings (65%).

Limitations: Classification of spin strategies was subjective and might not encompass all the methods used in the published literature.

Conclusion: Findings in this study inform efforts to reduce spin in the dermatologic literature. (J Am Acad Dermatol 2019;80:516-22.)

Key words: cosmetics; ethics; general dermatology.

Chronologic aging results in several skin changes, which can be a significant aesthetic concern for many patients. The benefits of therapies, such as retinoids and sunscreen, are well established, but the evidence behind newer compounds is more ambiguous. The financial incentives in this \$130 billion industry might cause sponsors to put pressure on authors to present their data in the most positive light and spin results.¹ Spin is defined as reporting that distorts the interpretation of results,

thereby suggesting a more favorable outcome and misleading readers.²

The widespread use of spin within the medical literature has led to the development of validated instruments to assess distorted interpretations of results within randomized controlled trials,³ non-randomized intervention studies,⁴ diagnostic accuracy studies,⁵ and systematic reviews.⁶ Modifying these instruments to accommodate the study design has led to evaluations of spin in biomedical literature

From the Hansjörg Wyss Department of Plastic Surgery^a and Ronald O. Perleman Department of Dermatology,^b New York University Langone Health; and Biostats Inc, East Canton.^c

Funding sources: None.

Conflicts of interest: None disclosed.

Accepted for publication April 17, 2018.

Reprints not available from the authors.

Correspondence to: Alexes Hazen, MD, Hansjörg Wyss Department of Plastic Surgery, NYU Langone Health, 550 1st Ave, New York, NY 10016. E-mail: alexes.hazen@nyumc.org.

Published online April 22, 2018.

0190-9622/\$36.00

© 2018 by the American Academy of Dermatology, Inc.

<https://doi.org/10.1016/j.jaad.2018.04.034>

across multiple disciplines of medicine.⁷⁻¹⁰ Variation in the definition of outcomes for treatments of photoaged skin further opens the rhetoric of articles to spin, as even histologic evidence might not translate into clinical results sought by patients opting for elective cosmetic treatments.

The goal of this study is to appraise strategies of spin used in the most rigorously performed studies on topical treatments for photoaged skin—placebo-controlled, double-blind clinical trials with both clinical measurements and objective instrumental measurements of results. By highlighting the strategies of spin used within this cohort, we hope to improve the scientific rigor of studies within the field of dermatology.

METHODS

Selection of studies

The MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials databases were systematically reviewed for clinical trials examining treatments of photoaged skin published during January 1, 2007–October 17, 2017. The following search terms were used: skin* AND (photoag* OR photodamag* OR wrinkl* OR hyperpigment* OR sundamag*). Searches were filtered to include randomized clinical studies performed on humans and were limited to English language articles. Inclusion and exclusion criteria are summarized in Table I.

After duplicates were removed, titles and abstracts of the remaining articles were reviewed by 2 authors (Ms Motosko and Ault) to exclude trials that did not meet eligibility criteria. The full text of the remaining articles was reviewed, and inclusion criteria was applied.

Evaluation of spin

A literature review of studies evaluating spin was performed to develop a preliminary classification of spin strategies on the basis of previously published instruments.^{3-6,11} All identified strategies and available instruments were applied to the included studies by 2 reviewers, and revisions were made by all authors, drawing from extensive experience within plastic surgery, dermatology, statistics, and ethics. The final instrument was reapplied to the study set by both reviewers, with any discrepancies being resolved by arbitration by the senior author (Dr Hazen).

Evaluation of the spin and spin strategies used in the results sections of articles were prioritized. Studies that were determined to be without any significant between-group differences in their results sections at the final time point were evaluated for the extent of spin in the entirety of the full text, using classifications proposed by Boutron and colleagues (Table II).³

CAPSULE SUMMARY

- Reporting and interpretation of findings in the medical literature might be inconsistent with study methods and results.
- Even high-quality study designs within the dermatologic literature offers opportunities to spin results.
- Physicians, scientists, editors, and peer-reviewers should be mindful of spin techniques to improve the quality of research and care of our patients.

RESULTS

All 20 studies¹²⁻³¹ meeting the inclusion criteria (Fig 1) were reviewed for each of 10 spin strategies, which were subdivided into either incorrect use of statistical analysis or inappropriate interpretation of results (Table III). Studies and the strategies identified are summarized in Supplementary Table I (available at <http://www.jaad.org>).

The results sections of 15 (75%) studies focused on within-group comparisons. In 11 (55%) studies, there was a failure to report findings that were claimed to be measured in either the methods or results sections. In 7 (35%) studies, improvements were reported as a percent change from baseline. In 6 (30%) studies, there was a reliance on photographs or data from single subjects to substantiate results that were otherwise unsupported by statistical analysis of the sample's data. In 19 (95%) studies, multiple primary outcomes were used. In most studies, a median of 3 (range 1-17) clinically assessed measurements and a median of 5 (range 5-17) instrumentally assessed measurements were reported, and the number of measurements could not be determined in 3 studies. In 2 (10%) studies, multiple statistical analyses on the same data were performed.

The discussion of the results was focused on significant findings of interim analyses in 13 (65%) studies when findings at the final time point were either nonsignificant or not reported. Most interim analyses were performed at each of 3 (median; range 1-6) different time points. In 10 (50%) studies, the interventions were concluded to cause positive effects despite statistically nonsignificant results. Similar claims were made on the basis of findings unsupported by statistical analysis in 12 (60%) studies. In 19 (95%) studies, the results of specific outcomes were extrapolated to overall global improvement. No significant primary outcomes

Table I. Inclusion and exclusion criteria of the primary literature review

Inclusion criteria	Exclusion criteria
<ul style="list-style-type: none"> • Double-blind, placebo-controlled, randomized clinical trial • Topical intervention compared with either vehicle or placebo • Evaluation of photoaging (ie, wrinkling, roughness, texture, tone, and hydration) • Use of both in-person blinded clinician evaluation and instrumental assessment (not including histology) 	<ul style="list-style-type: none"> • Treatments indicated for precancerous lesions (actinic keratosis), scarring (ie, acne scars), or pigmentation alone (ie, melasma) • Control group was given no treatment • Clinical evaluation relied on photographs • Topical treatment was combined with nontopical intervention (ie, laser, microneedling, injections)

Table II. Levels of spin adapted from Boutron et al³

Spin level	Explanation
None	No spin observed
Low	Spin with acknowledgment of the statistically nonsignificant results for the primary outcome or spin in the conclusions section with no acknowledgment of the statistically nonsignificant results for the primary outcome but reported with uncertainty and recommendations for further trials
Moderate	Spin in the conclusions section with no acknowledgment of the statistically nonsignificant results for the primary outcome but reported with uncertainty or recommendations for further trials
High	Spin in the conclusions section without reporting uncertainty, no recommendations for further trials, and no acknowledgment of the statistically nonsignificant results for the primary outcome or spin with recommendation to use the treatment in clinical practice

were reported in 5 studies,^{14,18,23,26,30} all of which had high levels of spin as defined by Table II.

A secondary analysis was performed to examine potential motives of spin. Conflicts of interest were evaluated by assessing sources of funding, disclosures, or authors' affiliations with manufacturers of the tested product. Disclosures were reported in 14 studies,^{12-15,18-22,24-28} of which 7 studies^{12,13,15,18,26-28} reported authors' affiliations with the same companies that manufactured the tested products. Of the remaining studies that did not report any disclosures, bylines in 2 studies^{16,17} listed authors' affiliations with manufacturers of the tested product. Washout periods, in which subjects could not use specific

products, such as retinoids, for a specified period of time, were used in 14 studies,^{12-19,22,25-29} and the washout period ranged from 1 week to 2 years. In 1 study,²⁶ there was a failure to report both the types of restricted treatments and length of the washout period. Of the 20 included studies, 6^{14-16,18,26,27} appeared in the *Journal of Drugs in Dermatology* and 4^{12,13,25,28} appeared in the *International Journal of Cosmetic Science*. The remainder of the studies all appeared in separate peer-reviewed journals, which published 1 study each.

DISCUSSION

Spin is common and has a profound impact on both the research community and patients.³ Previous studies have described strategies, such as the use of inappropriate statistical tests to establish significance, manipulation of nonsignificant results to make them seem significant, and suggesting a beneficial effect despite nonsignificance.^{3,32} Similar strategies were identified in the articles of this review, all of which concluded that the topical application of the studied compounds showed promise in treating photoaged skin.

Increasing commercialization of research and normalization of academic-industry partnerships are creating an environment in which the incentive to translate scientific results into commercially viable products exerts influence on the research agenda. Within the dermatologic literature, studies with reported conflicts of interest have been shown to be more likely to report positive results and, in turn, are more likely to be accepted for publication.³³

Statistical manipulation of the data might be used to achieve positive results. Analyzing the active and placebo groups separately neglects the role of the placebo as a benchmark against which to compare the study results and as a control for the placebo effect.³⁴ Further, analyzing and reporting data as percent change from baseline fails to account for baseline variation, lowers the statistical power, and

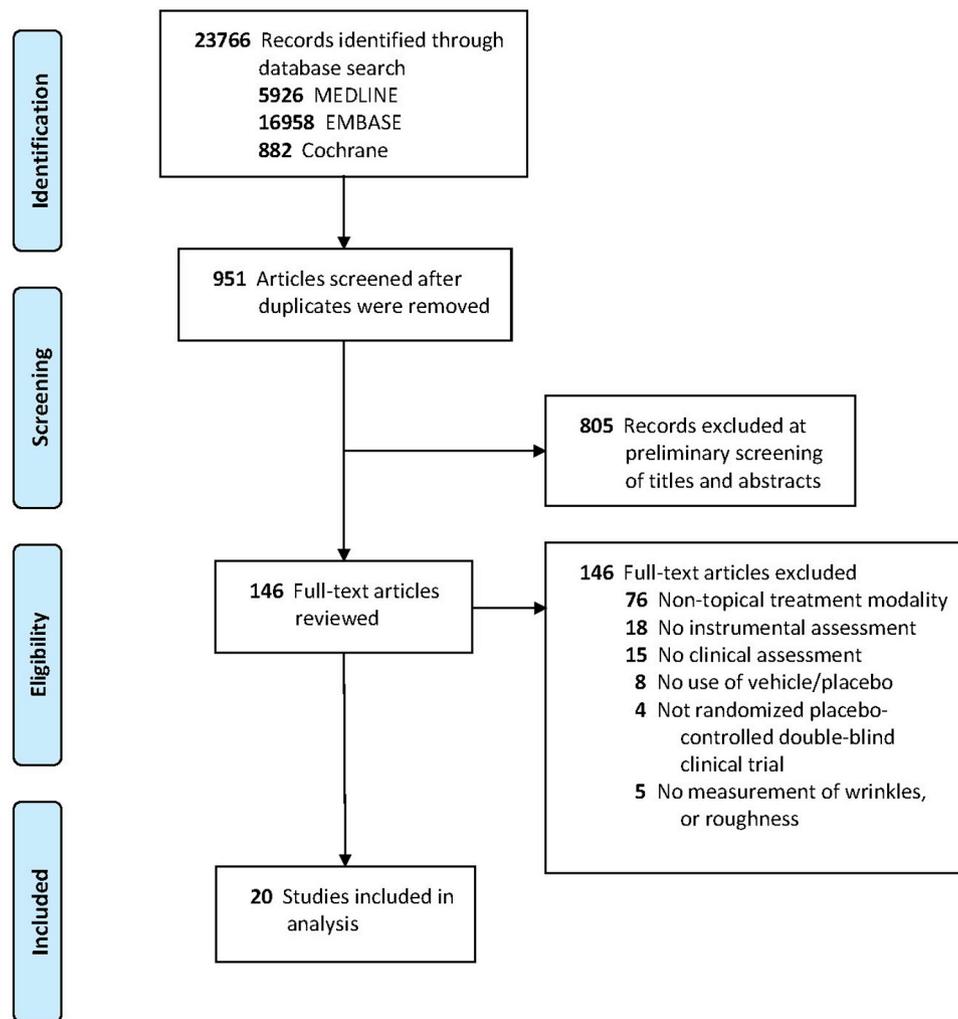


Fig 1. Flow chart depicting study selection for systematic review.

leaves fractions highly sensitive to changes in variance.³⁵ To more accurately report percent change, significance established using another method, such as analysis of covariance, should first be reported; then, baseline and posttreatment scores should be used to convert results into percent change.³⁵

Using multiple measurements, applying multiple statistical tests to a data set, and performing interim analyses also yields misleading and erroneous results.³² Two studies^{19,20} reported 36 *P* values from 6 outcomes, compared at 3 time points, using 2 statistical models; this analysis leaves an 84% chance of finding a false-positive result by random chance due to multiple tests inflating the rate of type I errors. Interim data, if properly handled, may be incorporated into statistical models,³⁶ but interpretation should not rely on interim results alone.¹¹ Use of multiple measurements might reflect the challenge in quantifying cosmetic improvement, with 1 instrument (Skin-Visiometer, Courage + Khazaka

electronic GmbH, Köln, Germany) reporting at least 5 parameters per measured line.³⁷ All study results should be reported, yet reporting all measured parameters increases the rate of type I errors and challenges interpretation of the results.

Success contrived by manipulated data puts patients at risk, imperils the integrity of the research enterprise, and leads to the propagation of future studies that waste time, money, and resources.^{38,39} The research community has an obligation to address the root causes of spin and ensure that research is conducted ethically, with sound and reliable reporting of results. Pressures from external funding sources and potential for promotion or recognition might influence the ways data are collected, interpreted, and portrayed.^{2,32} Further, the competitive nature of research often prioritizes quantity of publications over quality.³⁸ Researchers must be supported in adhering to widely accepted standards for appropriate application of statistical

Table III. Survey used to evaluate strategies of spin with examples extracted from sample

Strategy	Studies, N (%)	Example
Inappropriate statistical analysis		
Focus on within-group comparisons	15 (75)	"The global score improved progressively from 28.1 ± 3.2 at baseline to 18.1 ± 3.0 at 6 months of treatment with the tested cream." ¹⁷
Selective reporting of results	11 (55)	"A replica from the right and left periorbital areas was taken at weeks 0 (baseline), 4, 8 ([excluded study]), and 12 ([included study])..." Only results of excluded study (non-placebo-controlled) were described.
Reporting of percent change from baseline	7 (35)	"Nasolabial folds showed statistically greater improvement versus vehicle using the WSRS beginning at week 12 (10.3% improvement vs 0.9%)..." ¹⁶
Relying on single subject (n = 1) to claim results of entire sample	6 (30)	"Furthermore, the average depth of wrinkles, deepest wrinkle, total wrinkle volume, total wrinkle area, total form factor wrinkles, and total length of wrinkle are all decreased after the treatment in two subjects." ³⁰
Use of multiple primary outcomes	19 (95)	
Use of multiple statistical analyses on same data	2 (10)	
Inappropriate interpretation of results		
Focus on interim analyses to give more weight to nonsignificant findings	13 (65)	"The analysis of silicone replicas by cast shadows method revealed that the mean wrinkle length was significantly reduced compared to placebo from 56 days of use ($P = 0.012$, Fig. 6)" ²⁸ Results at final time point were nonsignificant.
Claim an effect for statistically nonsignificant results	10 (50)	"Major-line roughness for the active group showed a strong trend ($P < .1$) favoring active group over vehicle group." ²⁶
Claim an effect for nonstatistically analyzed results	12 (60)	"Clinical photography showed improvement in the visual signs of aging (Figures 3a and 4). Silicone replicas of the crow's feet area demonstrated improvement in skin texture and reduction in fine lines and wrinkles that corresponded to the visual changes seen in clinical photographs (Figure 3b)." ¹⁵
Inappropriate extrapolation of results from specific outcome to global improvement	19 (95)	"This study demonstrates that addition of growth factors and cytokines to a basic skin care regimen reduces the signs of photoaging." ²⁶ Extrapolation is based largely on statistically nonsignificant trends.

analytic methods and data reporting. Moreover, every effort should be made to improve the publishing and peer-review process to realign incentives and promote both critical appraisal of study methodology and publication of negative or nonsignificant findings.

This study focused only on several of the most common and apparent strategies of spin, which might not reflect all spin strategies due to the subjective nature of identifying spin. The intention was to assess the study methods used in data reporting and the ramifications for the conclusions presented in the studies concerning the efficacy of topical treatments in photoaged skin rather than to question the motives or character of the studies' authors.

Conclusion

Authors might use these identified spin strategies to report a positive result, even without evidence to support the claim. By highlighting strategies used to spin results, we hope physicians, scientists, editors, and peer-reviewers might be mindful of these techniques before recommending treatments to patients, pursuing further research, publishing studies, or being swayed by publication bias or career pressures that could lead them to distort results and mislead readers.

REFERENCES

1. Euromonitor International. The future of skin care: game-changing trends and influencers. Available at: <http://www.euromonitor.com/the-future-of-skin-care-game-changing-trends-and-influencers/report>; 2015.

2. Chiu K, Grundy Q, Bero L. 'Spin' in published biomedical literature: a methodological systematic review. *PLoS Biol.* 2017; 15(9):e2002173.
3. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA.* 2010; 303(20):2058-2064.
4. Lazarus C, Haneef R, Ravaud P, Boutron I. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. *BMC Med Res Methodol.* 2015;15: 85.
5. Ochoado EA, de Haan MC, Reitsma JB, Hooft L, Bossuyt PM, Leeftang MM. Overinterpretation and misreporting of diagnostic accuracy studies: evidence of "spin". *Radiology.* 2013; 267(2):581-588.
6. Yavchitz A, Ravaud P, Altman DG, et al. A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity. *J Clin Epidemiol.* 2016;75: 56-65.
7. Arunachalam L, Hunter IA, Killeen S. Reporting of randomized controlled trials with statistically nonsignificant primary outcomes published in high-impact surgical journals. *Ann Surg.* 2017;265(6):1141-1145.
8. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. *J Clin Oncol.* 2014;32(36): 4120-4126.
9. Cordoba G, Schwartz L, Woloshin S, Bae H, Gotzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ.* 2010;341: c3920.
10. Hernandez AV, Pasupuleti V, Deshpande A, Thota P, Collins JA, Vidal JE. Deficient reporting and interpretation of non-inferiority randomized clinical trials in HIV patients: a systematic review. *PLoS One.* 2013;8(5):e63272.
11. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c869.
12. Berardesca E, Ardigo M, Cameli N, Mariano M, Agozzino M, Matts PJ. Randomized, double-blinded, vehicle-controlled, split-face study to evaluate the effects of topical application of a gold silk sericin/niacinamide/Signaline complex on biophysical parameters related to skin ageing. *Int J Cosmet Sci.* 2015;37(6):606-612.
13. Bertin C, Zunino H, Lanctin M, et al. Combined retinol-lactose-glycolic acid effects on photoaged skin: a double-blind placebo-controlled study. *Int J Cosmet Sci.* 2008;30(3): 175-182.
14. Fabi SG, Cohen JL, Peterson JD, Kiripolsky MG, Goldman MP. The effects of filtrate of the secretion of the *Cryptomphalus aspersa* on photoaged skin. *J Drugs Dermatol.* 2013;12(4): 453-457.
15. Farris PK, Edison BL, Brouda I, Weinkauff RL, Green BA. A high-potency, multimechanism skin care regimen provides significant antiaging effects: results from a double-blind, vehicle-controlled clinical trial. *J Drugs Dermatol.* 2012;11(12): 1447-1454.
16. Farris PK, Edison BL, Weinkauff RL, Green BA. A novel, volumizing cosmetic formulation significantly improves the appearance of target Glabellar lines, nasolabial folds, and crow's feet in a double-blind, vehicle-controlled clinical trial. *J Drugs Dermatol.* 2014;13(1):41-46.
17. Haftek M, Mac-Mary S, Le Bitoux MA, et al. Clinical, biometric and structural evaluation of the long-term effects of a topical treatment with ascorbic acid and madecassoside in photoaged human skin. *Exp Dermatol.* 2008;17(11):946-952.
18. Hsu J, Skover G, Goldman MP. Evaluating the efficacy in improving facial photodamage with a mixture of topical antioxidants. *J Drugs Dermatol.* 2007;6(11):1141-1148.
19. Hwang E, Park SY, Jo H, et al. Efficacy and safety of enzyme-modified *Panax* ginseng for anti-wrinkle therapy in healthy skin: a single-center, randomized, double-blind, placebo-controlled study. *Rejuvenation Res.* 2015;18(5): 449-457.
20. Hwang E, Park SY, Yin CS, Kim HT, Kim YM, Yi TH. Antiaging effects of the mixture of *Panax* ginseng and *Crataegus pinnatifida* in human dermal fibroblasts and healthy human skin. *J Ginseng Res.* 2017;41(1):69-77.
21. Kikuchi K, Suetake T, Kumasaka N, Tagami H. Improvement of photoaged facial skin in middle-aged Japanese females by topical retinol (vitamin A alcohol): a vehicle-controlled, double-blind study. *J Dermatolog Treat.* 2009;20(5):276-281.
22. Kim H, Kim N, Jung S, et al. Improvement in skin wrinkles from the use of photostable retinyl retinoate: a randomized controlled trial. *Br J Dermatol.* 2010;162(3):497-502.
23. Kim HM, An HS, Bae JS, et al. Effects of palmitoyl-KVK-L-ascorbic acid on skin wrinkles and pigmentation. *Arch Dermatol Res.* 2017;309(5):397-402.
24. Kim YH, Chung CB, Kim JG, et al. Anti-wrinkle activity of ziyuglycoside I isolated from a *Sanguisorba officinalis* root extract and its application as a cosmeceutical ingredient. *Biosci Biotechnol Biochem.* 2008;72(2):303-311.
25. Lee J, Jung E, Lee H, Seo Y, Koh J, Park D. Evaluation of the effects of a preparation containing asiaticoside on periorcular wrinkles of human volunteers. *Int J Cosmet Sci.* 2008;30(3): 167-173.
26. Mehta RC, Smith SR, Grove GL, et al. Reduction in facial photodamage by a topical growth factor product. *J Drugs Dermatol.* 2008;7(9):864-871.
27. Roure R, Nollent V, Dayan L, Camel E, Bertin C. A Double-blind, 12-week study to evaluate the antiaging efficacy of a cream containing the NF- κ B inhibitor 4-hexyl-1, 3-phenylenediol and ascorbic acid-2 glucoside in adult females. *J Drugs Dermatol.* 2016;15(6):750-758.
28. Sohm B, Cenizo V, Andre V, Zahouani H, Pailleur-Mattei C, Vogelgesang B. Evaluation of the efficacy of a dill extract in vitro and in vivo. *Int J Cosmet Sci.* 2011; 33(2):157-163.
29. Sommerfeld B. Randomised, placebo-controlled, double-blind, split-face study on the clinical efficacy of Tricutan on skin firmness. *Phytomedicine.* 2007;14(11):711-715.
30. Wang Y, Wang M, Xiao XS, Pan P, Li P, Huo J. The anti-wrinkle efficacy of argireline, a synthetic hexapeptide, in Chinese subjects: a randomized, placebo-controlled study. *Am J Clin Dermatol.* 2013;14(2):147-153.
31. Watanabe F, Hashizume E, Chan GP, Kamimura A. Skin-whitening and skin-condition-improving effects of topical oxidized glutathione: a double-blind and placebo-controlled clinical trial in healthy women. *Clin Cosmet Investig Dermatol.* 2014;7: 267-274.
32. Marco CA, Larkin GL. Research ethics: ethical issues of data reporting and the quest for authenticity. *Acad Emerg Med.* 2000;7(6):691-694.
33. Perlis CS, Harwood M, Perlis RH. Extent and impact of industry sponsorship conflicts of interest in dermatology research. *J Am Acad Dermatol.* 2005;52(6):967-971.
34. Bland JM, Altman DG. Comparisons against baseline within randomised groups are often used and can be highly misleading. *Trials.* 2011;12:264.

35. Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Med Res Methodol.* 2001;1:6.
36. Berry DA. Interim analyses in clinical trials: classical vs. Bayesian approaches. *Stat Med.* 1985;4(4):521-526.
37. Courage + Khazaka electronic GmbH. Skin-Visiometer® SV 700 USB- Assessing the Skin Microrelief. Köln, Germany 2007.
38. Biros MH, Fish SS, Taggart P. Research fundamentals VI: misconduct in biomedical research. *Acad Emerg Med.* 1999; 6(8):840-848.
39. Stern AM, Casadevall A, Steen RG, Fang FC. Financial costs and personal consequences of research misconduct resulting in retracted publications. *eLife.* 2014;3: e02956.

Supplementary Table I. Summary of studies' results of clinical and objective measures of improvement

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Berardesca et al, 2015 ¹²	<i>International Journal of Cosmetic Science</i> (1.581)	Gold silk sericin, extracted from irodori silk cocoons was combined at 2% (w/v) with 5% niacinamide (w/v) and 0.1% a blend of 1,2-diacylglycerol from olive oil and fatty alcohols from jojoba wax applied twice daily for 8 weeks	Wrinkle, texture, tone	Stratum corneum hydration, transepidermal water loss, gross elasticity, net elasticity, wrinkling volume, smoothness, uniformity	NA	Use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; inappropriate extrapolation of results from specific outcome to global improvement
Bertin et al, 2008 ¹³	<i>International Journal of Cosmetic Science</i> (1.581)	0.1% retinol, 5% lactose, 4% glycolic acid applied twice daily for 12 weeks	Total number of wrinkles, total length of surface with wrinkles	Total length of surface with wrinkles, total surface with wrinkles, roughness, mean amplitude	Total number of wrinkles, skin microrelief, viscoelastic properties, skin moisturization, skin radiance, skin firmness, presence of wrinkles	Focus on within-group comparison; selective reporting of results; use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; claim an effect for statistically nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Fabi et al, 2013 ¹⁴	<i>Journal of Drugs in Dermatology</i> (1.708)	Secretions from the snail <i>Cryptomphalus aspersa</i> in a 8% emulsion (Tensage Contour Cream; Biopelle, Inc, Ferndale, MI) applied every morning; 40% liquid serum (Tensage Intensive Ampoules; Biopelle, Inc, Ferndale, MI) applied every evening for 12 weeks	NA		Photodamage, mean periocular and perioral rhytid severity scores, number of coarse lines, average maximum difference in luminance value, length of coarse lines, improvement in fine lines	Focus on within-group comparison; selective reporting of results; relying on single subject (n = 1) to claim results of entire sample; use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; claim an effect for statistically nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Farris et al, 2012 ¹⁵	<i>Journal of Drugs in Dermatology</i> (1.708)	Facial cleanser (NeoStrata Skin Activating Wash), day cream (NeoStrata Skin Active Matrix Support SPF 20), night cream (NeoStrata Skin Active Cellular Restoration), and eye cream (NeoStrata Skin Active Intensive Eye Therapy) (all from NeoStrata Company Inc, Princeton, NJ) with numerous active ingredients including alpha-hydroxy and polyhydroxy acids, maltobionic acid, N-acetyl-glucosamine, retinol, peptides, Swiss apple stem cell extract, broad-spectrum sunscreens, and antioxidants such as Chardonnay grape seed extract applied twice daily for 16 weeks	Fine lines, wrinkles, clarity, laxity, visual roughness, tactile roughness, dyschromia, pore size, sallowness, eyelid crepiness, lifted appearance of eyelid, overall appearance	Skin thickness	Dermal density, skin texture, fine lines, wrinkles	Focus on within-group comparison; selective reporting of results; reporting of percent change from baseline; relying on single subject (n = 1) to claim results of entire sample; use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; claim an effect for statistically nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Farris et al, 2014 ¹⁶	<i>Journal of Drugs in Dermatology</i> (1.708)	A serum of cosmetic benefit ingredients, N-acetyl-tyrosinamide, N-acetyl hydroxyproline, and glycolic acid and a cream of N-acetyl glucosamine, triethyl citrate, palmitoyl oligo- and tetrapeptides (Matrixy; Sederma, France) applied twice daily for 16 weeks	Glabellar lines, nasolabial folds, under eye lines and wrinkles, crow's feet, nasolabial folds	Pinch recoil time	Dermal density	Focus on within-group comparison; selective reporting of results; reporting of percent change from baseline; relying on single subject (n = 1) to claim results of entire sample; use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; claim an effect for statistically nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Haftek et al, 2008 ¹⁷	<i>Experimental Dermatology</i> (2.532)	5% stabilized vitamin C with 0.1% madecassoside (Redermic; La Roche-Posay Laboratoire Pharmaceutique, France) applied twice daily for 6 months	Hydration, roughness, laxity, suppleness, wrinkles, radiance, pigmented spots	NA	Hydration, roughness/laxity, suppleness, fine/course wrinkles, radiance, brown spots, skin homogeneity, skin hydration, crows' feet volume, cutaneous elasticity	Focus on within-group comparison; selective reporting of results; reporting of percent change from baseline; use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; claim an effect for statistically nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Hsu et al, 2007 ¹⁸	<i>Journal of Drugs in Dermatology</i> (1.708)	Green and white teas, mangosteen, and pomegranate extract applied twice daily for 60 days	NA	NA	Skin smoothness, wrinkle scale	Focus on within-group comparison; selective reporting of results; reporting of percent change from baseline; relying on single subject (n = 1) to claim results of entire sample; use of multiple primary outcomes; claim an effect for statistically nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement
Hwang et al, 2015 ¹⁹	<i>Rejuvenation Research</i> (2.827)	0.50 (unreported units) of enzyme modified <i>Panax ginseng</i> applied twice daily for 12 weeks	Global photodamage scores	Skin roughness, smoothness depth, arithmetic average roughness	Maximum roughness, average roughness	Focus on within-group comparison; use of multiple primary outcomes; use of multiple statistical analyses on same data; focus on interim analyses to give more weight to nonsignificant findings; inappropriate extrapolation of results from specific outcome to global improvement

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Hwang et al, 2017 ²⁰	<i>Journal of Ginseng Research</i> (4.082)	0.50 (unreported units) of 1:1 mixture (by weight) of <i>Panax ginseng</i> and <i>Cratagis pinnatifida</i> applied twice daily for 12 weeks	Global photodamage scores	Skin roughness, maximum roughness, average roughness, smoothness depth, arithmetic average roughness	NA	Focus on within-group comparison; use of multiple primary outcomes; use of multiple statistical analyses on same data; focus on interim analyses to give more weight to nonsignificant findings; inappropriate extrapolation of results from specific outcome to global improvement
Kikuchi et al, 2017 ²¹	<i>Journal of Dermatological Treatment</i> (1.890)	0.075% retinol cream applied once nightly for 26 weeks; 50% of participants applied cream once every other day for the first 2 weeks	Fine and deep wrinkles	Skin conductance	Pigmentation, rosy erythema	Focus on within-group comparison; selective reporting of results; use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement
		0.04% retinol cream once nightly for 13 weeks; 50% of participants applied cream once every other day for the first 2 weeks	NA	NA	Pigmentation, fine wrinkles, deep wrinkles, rosy erythema, skin conductance	

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Kim et al, 2008 ²⁴	<i>Bioscience, Biotechnology, and Biochemistry</i> (1.295)	0.03% ziyuglycoside I cream applied for 12 weeks	Photodamage	Average roughness	Depth of roughness, maximum roughness, smoothness depth, arithmetic average roughness	Selective reporting of results; use of multiple primary outcomes; inappropriate extrapolation of results from specific outcome to global improvement
Kim et al, 2010 ²²	<i>The British Journal of Dermatology</i> (4.706)	0.06% retinyl retinoate applied twice daily for 12 weeks	Photodamage scores	Maximum roughness, average roughness	Skin roughness, smoothness depth, arithmetic average roughness	Focus on within-group comparison; use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; inappropriate extrapolation of results from specific outcome to global improvement
Kim et al, 2017 ²³	<i>Archives of Dermatological Research</i> (2.327)	0.075% palmatoylKVK amino-propyl ascorbyl phosphate applied twice daily for 12 weeks	NA	NA	Photodamage score, skin roughness, maximum roughness, average roughness, dermal density	Use of multiple primary outcomes; claim an effect for statistically nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Lee et al, 2008 ²⁵	<i>International Journal of Cosmetic Science</i> (1.581)	0.1% asiaticoside applied for 12 weeks	NA	Density of skin microrelief, depths of deep furrows, skin roughness, average roughness, arithmetic average roughness	Hydration, small wrinkles, wrinkles, glare, brown spots, roughness, suppleness, small wrinkles, maximum roughness, smooth roughness	Focus on within-group comparison; selective reporting of results; use of multiple primary outcomes; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement
Mehta et al, 2008 ²⁶	<i>Journal of Drugs in Dermatology</i> (1.708)	Gel containing unreported ingredients (NouriCel-MD; SkinMedica, Allergan, Madison, NJ) applied twice daily for 6 months	NA	NA	Major line average roughness; fine lines and textures average roughness; major line shadow; fine lines and textures shadow; fine wrinkling; mottled pigmentation; tactile roughness; sallowness; telangiectasia	Focus on within-group comparison; use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; claim an effect for statistically nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Roure et al, 2016 ²⁷	<i>Journal of Drugs in Dermatology</i> (1.708)	4-Hexyl-1,3-phenylenediol and L-ascorbic acid 2-glucoside in a cream applied twice daily for 12 weeks	Wrinkle parameters	Spot coloration	Light variable and individual typological angle; skin elasticity	Reporting of percent change from baseline; relying on single subject (n = 1) to claim results of entire sample; use of multiple primary outcomes; inappropriate extrapolation of results from specific outcome to global improvement
Sohm et al, 2011 ²⁸	<i>International Journal of Cosmetic Science</i> (1.581)	1% dill seed extract to whole face twice daily for 84 days (3 months)	Skin elasticity	Mean wrinkle length, mean wrinkle area, skin elasticity	Slackness of jaw line, wrinkle number, total wrinkle area, total wrinkle length	Focus on within-group comparison; reporting of percent change from baseline; use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; claim an effect for statistically nonsignificant findings; inappropriate extrapolation of results from specific outcome to global improvement

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Sommerfeld, 2007 ²⁹	<i>Phytotherapy Research</i> (3.526)	Extract of <i>Rosmarinus officinalis</i> with ursolic acid, extract of <i>Centella asiatica</i> , tetrahydrocurcumin, dimethylaminoethanol applied twice daily for 28 days	Effects of treatment	Skin firmness	NA	Selective reporting of results
Wang et al, 2013 ³⁰	<i>Journal of Cosmetic and Laser Therapy</i> (1.113)	Emulsion containing 10% acetyl hexapeptide-3 argireline (McEit [Tianjin] International Trade Co Ltd, China), formulated to mimic botulinum neurotoxin type A, applied twice daily for 4 weeks	NA	NA	Average wrinkle height, difference from peak to lowest point of all wrinkles in region, average wrinkle height, overall wrinkles in region	Focus on within-group comparison; selective reporting of results; relying on single subject (n = 1) to claim results of entire sample; use of multiple primary outcomes; focus on interim analyses to give more weight to nonsignificant findings; claim an effect for statistically nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement

Continued

Supplementary Table I. Cont'd

Study	Journal (impact factor in 2016)	Intervention	Clinically significant improvements in treatment vs placebo group*	Instrumentally significant improvements in treatment vs placebo group*	Nonsignificant findings or findings of unreported significance*	Spin used
Watanabe et al, 2014 ³¹	<i>Clinical, Cosmetic and Investigational Dermatology</i> (NA)	Lotion containing oxidized glutathione applied twice daily for 10 weeks	NA	Melanin index, curvature index value, keratin index value	Skin elasticity, skin whitening, wrinkle reduction, skin smoothing efficacy	Focus on within-group comparison; reporting of percent change from baseline; use of multiple primary outcomes; claim an effect for statistically nonsignificant findings; claim an effect for nonstatistically analyzed results; inappropriate extrapolation of results from specific outcome to global improvement

NA, Not available; *SPF*, sun protection factor; *w/v*, weight per volume.

*Ambiguity of language in articles preclude definitive reporting of results and might be incomplete.