TFOS Lifestyle - Evidence quality report: Advancing the evaluation and synthesis of research evidence

Laura E. Downie [a,*], Alexis Ceecee Britten-Jones [a], Ruth E. Hogg [b], Isabelle Jalbert [c], Tianjing Li [d], Gareth Lingham [e], Su-Hsun Liu [d], Riaz Qureshi [d], Ian J. Saldanha [f], Sumeer Singh [a], Jennifer P. Craig [g]

[a] Department of Optometry and Vision Sciences, The University of Melbourne, Parkville, Victoria, Australia
[b] Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Belfast, United Kingdom
[c] School of Optometry and Vision Sciences, UNSW Sydney, Australia
[d] Department of Ophthalmology and Epidemiology, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[e] Centre for Eye Research Ireland, Technological University Dublin, Dublin, Ireland
[f] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
[g] Department of Ophthalmology, New Zealand National Eye Centre, The University of Auckland, Auckland, New Zealand

ARTICLE INFO

ABSTRACT

Evidence-based practice is a dominant paradigm in healthcare that emphasizes the importance of ensuring the translation of the best available, relevant research evidence into practice. An Evidence Quality Subcommittee was established to provide specialized methodological support and expertise to promote rigorous and evidence-based approaches for the Tear Film and Ocular Surface Society (TFOS) Lifestyle Epidemic reports. The present report describes the purpose, scope, and activity of the Evidence Quality Subcommittee in the undertaking of high-quality narrative-style literature reviews, and leading prospectively registered, reliable systematic reviews of high priority research questions, using standardized methods for each topic area report. Identification of predominantly low or very low certainty evidence across the eight systematic reviews highlights a need for further research to define the efficacy and/or safety of specific lifestyle interventions on the ocular surface, and to clarify relationships between certain lifestyle factors and ocular surface disease. To support the citation of reliable systematic review evidence in the narrative review sections of each report, the Evidence Quality Subcommittee curated topic-specific systematic review databases and relevant systematic reviews underwent standardized reliability assessment. Inconsistent methodological rigor was noted in the published systematic review literature, emphasizing the importance of internal validity assessment. Based on the experience of implementing the Evidence Quality Subcommittee, this report makes suggestions for incorporation of such initiatives in future international taskforces and working groups. Content areas broadly relevant to the activity of the Evidence Quality Subcommittee, including the critical appraisal of research, clinical evidence hierarchies (levels of evidence), and risk of bias assessment, are also outlined.

1. Introduction

This report is part of the Tear Film & Ocular Surface Society (TFOS; www.tearfilm.org) Workshop, entitled 'A Lifestyle Epidemic: Ocular Surface Disease,' which was undertaken to establish the direct and indirect impacts that everyday lifestyle choices and challenges have on ocular surface health. Across eight subcommittees, the Workshop considered how the ocular surface is affected by the digital environment, cosmetics, nutrition, elective medications and procedures, environmental conditions, lifestyle challenges, contact lens wear, and societal challenges. The main outputs of the Workshop are a set of published reports. Intended to provide an evidence-based evaluation of the available research, the reports identified and summarized the evidence relevant to each topic area.

The previous TFOS Workshops [1–4] have been influential in informing ocular surface research and practice, globally. With the intent

---

* Corresponding author. Department of Optometry and Vision Sciences, The University of Melbourne, Australia.
E-mail address: ldownie@unimelb.edu.au (L.E. Downie).

of providing specialized methodological support and expertise in evidence appraisal and synthesis for the TFOS Lifestyle Epidemic Workshop, the Evidence Quality Subcommittee was convened as a new initiative for the current Workshop. The present report describes the purpose, scope, and activity of the Evidence Quality Subcommittee, which contributed to two main aspects of each topic area report: (i) narrative review: defining best practices for conducting and reporting narrative-style literature reviews, including the citation and appropriate description of relevant and reliable systematic review evidence; and (ii) systematic review: leading the undertaking of a prospectively registered, reliable systematic review for a rigorous evaluation of a high priority research question.

Herein, the Evidence Quality Subcommittee discusses considerations associated with implementing these initiatives as part of the TFOS Lifestyle Epidemic Workshop. From lessons learned, suggestions are made that may guide the incorporation of an Evidence Quality Subcommittee to facilitate an evidence-based approach to future international taskforces and working groups. The present report also discusses content areas broadly relevant to the remit of the Evidence Quality Subcommittee, including evidence-based practice, critical appraisal of research, clinical evidence hierarchies (levels of evidence), and the assessment of risk of bias.

2. Evidence-based practice and critical appraisal of research

2.1. Definitions

Evidence-based practice has been defined as "the conscientious, explicit and judicious use of current best (research) evidence in making decisions about the care of individual patients" [5]. Evidence-based practice is now the dominant paradigm in healthcare that emphasizes the importance of translating relevant, robust research evidence into practice as a foundation for providing high-quality care to patients. More broadly, 'evidence-based' is used to describe approaches founded on the highest quality scientific evidence. Inherent to these definitions is the concept of identifying the "best" evidence, which requires an evaluation of whether the research methodology is sound, to minimize potential biases and errors (i.e., an assessment of the internal validity of the research). Although the accepted benchmark for publishing research involves independent critique of a paper via peer review, this process does not guarantee scientific rigor. Published studies can vary substantially in their quality. In 2015, a high-profile paper published in *PLoS Medicine* contentiously claimed that up to half of research findings in peer-reviewed papers could be false [6]. Key factors proposed to undermine the scientific validity of studies were risks of bias, inadequate sample sizes, small study effects, and the use of inappropriate statistical analyses. It is, thus, not sufficient to rely on publication status alone as a measure of research quality.

Table 1
National Health and Medical Research Council Evidence Hierarchy: designations of 'level of evidence' according to the type of research question [9].

Level of evidence	Research question type				
	Intervention	Diagnostic accuracy	Prognosis	Etiology	Screening intervention
I	Systematic review of Level II studies	Systematic review of Level II studies	Systematic review of Level II studies	Systematic review of Level II studies	Systematic review of Level II studies
II	Randomized controlled trial	Study of test accuracy with an independent, blinded comparison with a valid reference standard among consecutive persons with a defined clinical presentation	Prospective cohort study[d]	Prospective cohort study	Randomized controlled trial
III-1	Pseudo-randomized controlled trial[a]	Study of test accuracy with an independent, blinded comparison with a valid reference standard, among non-consecutive persons with a defined clinical presentation	All or none[e]	All or none[e]	Pseudo-randomized controlled trial[a]
III-2	Comparative study with concurrent controls: ■ Non-randomized, experimental trial[b] ■ Cohort study ■ Case-control study ■ Interrupted time series with a control group	Comparison with reference standard that does not meet the criteria required for Level II and III-1 evidence	Analysis of prognostic factors amongst persons in a single arm of a randomized controlled trial	Retrospective cohort study	Comparative study with concurrent controls: ■ Non-randomized, experimental trial[b] ■ Cohort study ■ Case-control study
III-3	Comparative study without concurrent controls: ■ Historical control study ■ Two or more single arm study[c] ■ Interrupted time series without a parallel control group	Diagnostic case-control study	Retrospective case study	Case-control study	Comparative study without concurrent controls: ■ Historical control study ■ Two or more single arm study
IV	Case series with either post-test or pre-test/post-test outcomes	Study of diagnostic yield (no reference standard)	Case series, or cohort study of persons at different stages of disease	Cross-sectional study or case series	Case series

[a] Pseudo-randomized controlled trials assign participants to the intervention(s) using alternate allocation or some other non-randomized method.

[b] This also includes controlled before-and-after (pre-test/post-test) studies, as well as adjusted indirect comparisons (i.e., utilize A vs B and B vs C, to determine A vs C with statistical adjustment for B).

[c] Comparing single arm studies (i.e., case series from two studies). This would also include unadjusted indirect comparisons (i.e., utilize A vs B and B vs C, to determine A vs C but where there is no statistical adjustment for B).

[d] At study inception, the cohort is either non-diseased or all at the same stage of the disease. A randomized controlled trial with persons either non-diseased or at the same stage of the disease in both arms of the trial would also meet the criterion for this level of evidence.

[e] All or none of the people with the risk factor(s) experience the outcome; and the data arises from an unselected or representative case series which provides an unbiased representation of the prognostic effect.

Critical appraisal of research involves the careful and comprehensive evaluation of the scientific rigor of a study to determine the level to which its results can be trusted, and the extent to which its findings should influence practice [7]. Critical appraisal requires an understanding of study design and an appreciation for how bias might affect the internal validity of different research methods [8]. For clinical research, which is the focus of this report, the first critical appraisal step typically involves identifying the adopted design (e.g., systematic review, randomized controlled trial, cohort study) and determining its appropriateness to answer the research question (see Table 1) [9]. These principles are often visually represented in 'evidence hierarchies' (see Section 2.2), which rank study designs based on the maximal possible rigor of the method to answer a certain type of research question; this ranking is often described as the 'level of evidence'. However, relying only on the evidence level to judge the rigor of a given study can be problematic, as other factors beyond its design are important (e.g., risk of bias, precision) [10]. Beyond considering the design of a study, critical appraisal of research involves evaluating how well a particular study has been conducted to minimize various types of biases. This should involve a structured approach, whereby 'risk of bias' is assessed using a validated tool that is appropriate for the study design (see Table 2). It is thus possible for studies to be 'ranked' at the same evidence hierarchy level, but to have different methodological rigor. These concepts are elaborated upon in the following sub-sections.

2.2. Clinical evidence hierarchies

A 'hierarchy of evidence' seeks to rank study designs from best (least potential for bias) to worst (highest potential for bias) in terms of their appropriateness to answer a specific type of research question. Evidence hierarchies are frequently presented as pyramids, where each study design type corresponds to a 'level'. A representative evidence hierarchy for intervention-type research questions is shown in Fig. 1; it should be noted that the study designs constituting Level II-IV evidence will differ for other types of research questions. Whilst there remains some debate about the optimal configuration of evidence hierarchies [10], the approach taken in the TFOS Lifestyle Epidemic Workshop was to adopt a common framework, defined by the National Health and Medical Research Council of Australia [9]. The National Health and Medical Research Council evidence hierarchy (Table 1) defines the study designs best suited to answer each of the five key types of clinical research questions (i.e., intervention, etiology, diagnostic accuracy, prognosis and screening intervention) [9] based on the likelihood that the study design has minimized the impact of bias. The higher the 'level' of a study design, the more likely it is to yield valid findings. A separate hierarchy of evidence has also been proposed for the assessment of qualitative research studies, but its use was not required for the literature evaluated for the TFOS Lifestyle reports [18].

As summarized in Table 1, Level II evidence is defined by the use of the most appropriate (best) design for an *individual* primary research study to answer a research question; the specific methodology depends
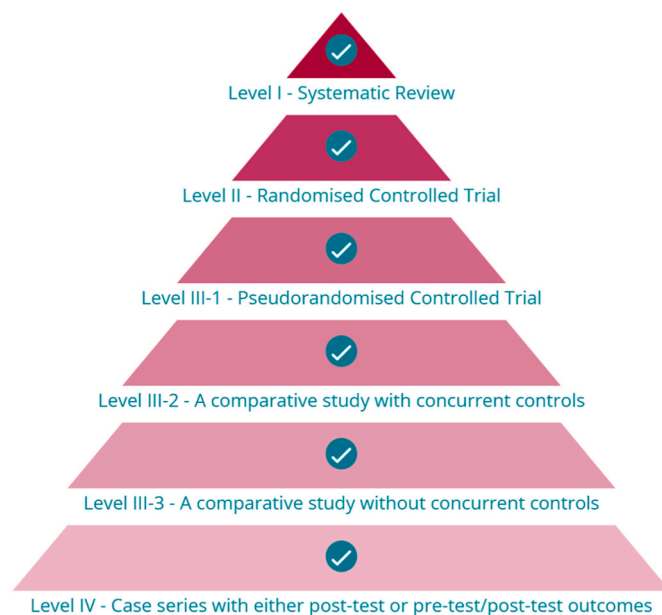
Table 2
Examples of tools used to evaluate risk of bias in different study design types.

Study type	Risk of bias assessment tool(s)
Systematic review	AMSTAR-2 [11]
	ROBIS [12]
Randomized controlled trial	Cochrane RoB 2 [13]
Non-randomized intervention study	ROBINS-I [14]
Diagnostic-test accuracy	QUADAS-2 [15]
Case control and cohort studies	Newcastle-Ottawa scale [16]
Pre-post intervention studies with no control group	NIH assessment tool [17]

Abbreviations: AMSTAR, A MeaSurement Tool to Assess systematic Review; NIH, National Institute of Health; ROBIS, Risk of Bias in Systematic Reviews; RoB, Risk of Bias; ROBINS-I, Risk of Bias in Non-Randomized Studies of Interventions; QUADAS, Quality Assessment of Diagnostic Accuracy Studies.

Fig. 1. Representative evidence hierarchy for Intervention research questions, based on the National Health and Medical Research Council of Australia schema [9]. Image taken from the CrowdCARE database [7] tutorial.

on the question to be addressed. For intervention-type questions, Level II evidence is a randomized controlled trial, whereas for other research questions, such as for studies that examine the etiology or prognosis of a condition, Level II evidence is a prospective cohort study. Study designs that are less robust constitute progressively lower levels of evidence (i. e., Level III and Level IV), and thus their findings should be viewed less definitively.

Level I evidence is provided by systematic reviews of appropriate Level II evidence. Systematic reviews aim to identify, appraise, collate, and synthesize findings from studies that are relevant to a specific research question. In general, key steps involved in conducting a systematic review include: (1) Developing a research question, often using the 'PICO' (Participant, Intervention Comparator, Outcomes) framework (or similar) to define the scope of the question. Outcome measures, appropriate to answer the research question, should also be prospectively defined; (2) Defining the study eligibility (inclusion/exclusion) criteria; (3) Developing comprehensive literature search strategies, typically in multiple electronic databases, and with at least one supplementary search method (e.g., searching the reference lists of included studies to identify other potentially relevant research), to minimize the risk of incomplete evidence capture; (4) Executing the searches and developing a citation database for subsequent citation screening; (5) Selecting relevant studies for inclusion in the review based on the predefined eligibility criteria; this is a two-stage process, involving initial title/abstract screening of all non-duplicate citations, followed by full-text evaluations of potentially relevant studies; (6) Extracting data from the included studies using standardized templates or software; (7) Evaluating the risk of bias in included studies using established tools that are appropriate to the study design (see Section 2.3 and Table 2); (8) Summarizing and synthesizing findings; this may include a meta-analysis, involving the application of suitable statistical approaches to combine the results of two more studies that address the same research question, to derive an overall effect estimate [19]. Meta-analyses will generally have greater precision than the individual studies within them; (9) Presenting the outcomes of the systematic review; this often includes an assessment of the certainty of the body of the evidence, using an approach such as Grading of Recommendations, Assessment, Development and Evaluations (GRADE) [20]. This system provides a framework for developing and presenting evidence summaries, and an

approach for making clinical practice recommendations based on the available evidence. The specifics of each of these steps should be defined *a priori,* in a systematic review protocol and/or in a publicly-accessible systematic review registry to maximize the transparency and replicability of the methodology. Steps 5 to 7 should be independently performed by at least two of the systematic review authors, with a consensus process instituted for any disagreements; this process minimizes the potential for error(s) and bias.

Benefits of robust systematic review methods include comprehensively collating all relevant evidence on a research topic, enabling research gaps to be identified and prioritized, and potentially identifying publication bias. It should be noted that systematic reviews that incorporate primary research studies of lower level evidence are more likely to be influenced by bias, and thus it has been recommended that a systematic review should only be assigned a level of evidence as high as the studies it includes, except where those studies are all Level II evidence (in which case the systematic review is assigned as Level I evidence [9]). Outcomes of systematic reviews that are based on primary studies of lower levels of evidence will also typically have lower scoring on the Grading of Recommendations, Assessment, Development and Evaluations certainty assessments.

The level of evidence available to answer a research question will guide what is considered the 'best' evidence in a specific context [21, 22]. An evidence-based approach might be limited to lower level evidence (such as case series) if this is all that is available in the literature. For example, many of the associations between cosmetic use and ocular surface adverse events derive from case series (Level IV etiology studies) [23]. In the absence of higher level evidence, these findings may constitute the 'best' current evidence, but should be viewed with less certainty than if they were reported from well-conducted prospective cohort studies (as the optimal primary study design to answer etiological clinical questions). In contrast, there may be clinical questions where an abundance of higher-level evidence exists. This can pose its own challenges when attempting to interpret apparently divergent outcomes across studies. In these cases, systematic reviews (Level I studies) can be useful for synthesizing and summarizing the effects of multiple studies, and potentially reconciling different study findings by exploring reasons for heterogeneity. For example, multiple randomized controlled trials (Level II intervention studies) have been conducted to evaluate the efficacy and safety of oral polyunsaturated fatty acid supplements for treating dry eye disease [24]. In this scenario, the 'best' evidence would derive from well-conducted systematic reviews, rather than relying on the results of a single randomized controlled trial [25]. Fundamental to adopting an evidence-based approach is not disregarding robust high-level evidence in favor of lower level evidence that might align with a particular view, belief, or anecdotal perception.

2.3. Risk of bias

The 'level' of evidence considers the overall design features of a study, but is not sufficient to determine whether the findings from a particular study are robust. The internal validity of an individual study is reflected by the extent to which bias has been minimized or, ideally, eliminated. Recognizing that distinct types of bias may be most relevant to certain study designs, different 'risk of bias' assessment tools exist for specific study designs [26]; some common tools, used in the systematic reviews for the TFOS Lifestyle Workshop, are summarized in Table 2. Tools to assess the risk of bias in systematic reviews evaluate aspects of the methods, such as the completeness of the literature search strategies, the extent to which the study eligibility criteria were defined, and whether the methodological rigor of included studies was assessed. In contrast, tools used to evaluate the risk of bias in randomized controlled trials focus on factors relevant to this primary research study design, such as how the participant randomization sequence was generated, whether allocation to the intervention was concealed, and whether participants, study personnel, and/or outcome assessors were masked

(blinded) to the intervention(s). All these factors, and others, influence the internal validity of a study and, thus, the level of confidence that should be placed in its findings. While studies with commercial involvement may be seen to be at higher risk of bias, evaluations of bias in these instances should be nuanced and specific to individual studies. Certain aspects that may lead to bias can potentially be addressed by investigators; for example, by having the researchers undertaking the study to control its design, execution, data analysis and reporting, as is common for investigator-initiated research. Another measure that can assist with reducing bias is prospective registration or publication of the study protocol, including secondary analyses.

The importance of study-level risk of bias assessment is demonstrated by the vast heterogeneity in internal validity that can exist for studies at an equivalent level on the evidence hierarchy (Fig. 1). For example, positioned at the peak of the evidence pyramid and often described as 'gold-standard' evidence [27], systematic reviews are frequently used to inform medical and public health decision making [28]. It is therefore vital that systematic reviews are conducted using approaches that minimize the potential for biased or inaccurate conclusions, and are reported completely and transparently. With the intent of improving and standardizing the reporting of systematic reviews, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [29] was introduced in 2009, and updated in 2020 [30]. This globally accepted guideline provides a framework for transparent and complete reporting of systematic reviews. In addition, akin to the prospective registration of clinical trials, in 2011 an international register for systematic reviews in health and social care, PROSPERO, was established with the goal of reducing research waste associated with poor quality and/or unnecessarily duplicated systematic review efforts [31].

Despite these initiatives, published systematic reviews are not consistently of high quality. A well-conducted systematic review should be: (i) systematic (in its approach, including the identification of relevant literature); (ii) explicit (in its objectives and methods), and (iii) reproducible (with respect to its methodology and findings). Systematic reviews have been shown to vary in their methodological quality across multiple medical areas, including pediatric surgery [32], emergency medicine [33], radiology [34], gynaecology [35] and pulmonary medicine [36]. In eye care, investigations into the reliability of systematic reviews have highlighted a spectrum of quality across several subtopic areas, including age-related macular degeneration [37,38], retinal and vitreous conditions [39], glaucoma [40], cataract [41], corneal and conjunctival disease [42], and refractive error management [43]. Whilst the number of published systematic reviews is increasing over time, with a 20-fold increase in indexed reports over the past two decades [44], there is some evidence that there has not been a concomitant improvement in their reliability [38]. In general terms, the methodological limitations are varied, but include not adhering to an *a priori* protocol, inadequate literature searches, insufficient appraisal of the included studies, and lack of reporting of review author conflicts of interest. Although journal impact factor and article citation rate have shown to correlate with systematic review methodological rigor, these features should not be viewed as a surrogate measure of study quality [38]. Across several disciplines, systematic reviews published in the *Cochrane Database of Systematic Reviews* have been assessed to be of generally higher reliability [32,36,38,45–47] than reviews in other journals; these findings align with the generally accepted position that *Cochrane* has led the international standard for undertaking and reporting systematic reviews. Together, these findings highlight the importance of study-level risk of bias assessments to determine whether a piece of Level I evidence is reliable before considering the application of its findings into practice.

A final consideration is that the terms 'risk of bias' and 'quality assessment' are sometimes used interchangeably in the epidemiological literature [48]. *Cochrane* has suggested that 'quality assessment' refers to the extent to which a study was designed, conducted, analyzed,

interpreted and reported to avoid systematic errors, while 'risk of bias' assessment focuses specifically on what flaws in the design, conduct and analysis affect the study results [49]. As such, risk of bias can be viewed as related to, but distinct from, study quality. Even a well designed and implemented study can be at risk of bias [48]. As an example, participants are highly unlikely to be successfully masked (blinded) in a randomized controlled trial that compares a surgical intervention with a pharmacological intervention. The researchers may have done the best they can (i.e., performed a 'high quality' study), but this does not guarantee that the results are free from bias (i.e., due to lack of participant blinding).

3. Reviewing the literature: systematic versus narrative reviews

3.1. Definitions

In general terms, literature reviews seek to identify and summarize potentially large volumes of evidence into an accessible and useable format for end-users (e.g., scientists, clinicians, policy-makers). They "create a firm foundation for advancing knowledge. A successful literature review facilitates theory development, closes areas where a plethora of research exists, and uncovers areas where research is needed" [50]. The current report has thus far considered systematic reviews (as defined in Section 2.2), and the expectation that they follow established methodological approaches to minimize bias. Many other types of literature reviews, such as scoping reviews, rapid reviews and overviews of reviews, also exist; these are beyond the scope of the present report and differ in their structure, methodological approach, and reporting requirements [51]. Another main type of literature review, of relevance to the present report, is a 'narrative review'. A comparison of systematic and narrative reviews is provided in Table 3, highlighting their relative strengths and weaknesses, and thus how they can provide complementary information within each TFOS topic area report.

Narrative reviews are the most common form of literature review in the medical literature [52]. This type of review typically seeks to identify and summarize literature on a topic, but often does not have a pre-defined methodology. In contrast to systematic reviews, which are generally limited to a focused (narrow) research question, narrative reviews provide considerable latitude for the author(s) to explore the breadth of a topic, often integrating findings from a range of different study designs (and thus levels of evidence), including preclinical and clinical studies. This approach has the advantage of permitting the exploration and development of potentially complex, narrative arguments to cover topics where a systematic review may be inappropriate. Narrative reviews can be highly influential [50]; although, they traditionally do not have pre-defined methods or include systematic internal validity assessments, and may selectively include studies of which the review authors are aware. These factors, which relate to reliability, should be considered when interpreting findings from narrative reviews [53]. These considerations are also reflected by the absence of narrative reviews as a 'level' of evidence on traditional clinical evidence hierarchies (Table 1).

3.2. Scale for the quality assessment of narrative reviews

With the intent of improving the conduct and reporting of narrative reviews, and to address a need for an instrument to evaluate the quality of these types of reviews, the Scale for the Quality Assessment of Narrative Review Articles (SANRA) [52] was developed in 2019. It comprises six domains, scored from 0 (low standard – not at all) to 2 (high standard – thoroughly); a score of 1 (partially) indicates an intermediary level of performance. These six domains consider the extent to which: (1) the article's importance is justified to the readership; (2) concrete aims and/or questions are formulated; (3) the literature searches are described; (4) key statements are supported by references; (5) there is sound scientific reasoning, with the use of

Table 3
Comparison of the typical characteristics of narrative and systematic reviews.

	NARRATIVE REVIEW	SYSTEMATIC REVIEW
FEATURES		
Scope of review question	Broad and overarching	Narrow and specific
Review protocol	Generally not developed	Should be established *a priori*
Literature sources and search strategies	Unlikely to be comprehensive, and may not be explicitly reported	Aims to be comprehensive, and should involve multiple databases, with explicitly defined and reproducible search strategies (including search dates)
Study selection process	Often not specified	Should be specifically detailed; best practice involves two independent review author assessments
Study selection criteria	Often not specified	Explicitly defined *a priori*
Risk of bias assessment of included studies	Generally not performed	Risk of bias assessment using established tools
Data extraction/ summary process	Generally not defined	Required to be systematic and pre-specified
Evidence synthesis	Qualitative	Qualitative ± quantitative (meta-analyses)
CONSIDERATIONS		
Strengths	● Breadth of consideration of the subject matter ● Scope to integrate preclinical and clinical findings ● Development of narrative arguments	● Comprehensive synthesis of all evidence relevant to a specific question ● Structured approaches and reporting guidelines aim to minimize bias ● Certainty of the body of evidence can be determined using established approaches ● Allows for assessment of publication bias
Weaknesses	● Typically, lack of pre-defined methods and lack of reproducibility increase risks of bias	● Restricted in scope (answers a focused question) ● Resource intensive ● Reporting biases may be amplified

appropriate evidence; and (6) relevant outcome data are presented appropriately. The sum score for an article (out of 12) is then proposed to provide a measure of the "construct quality of a narrative review article" [52].

4. Contributions of the Evidence Quality Subcommittee to the TFOS Lifestyle Epidemic Workshop

In prior TFOS Workshops, assessments of evidence were generally performed using a 'level of evidence' schema based on that defined in the American Academy of Ophthalmology Preferred Practice Pattern guidelines [54] (Table S1). This grading classification defines three levels of evidence 'quality' for both clinical and basic science studies, with categorizations primarily in the context of intervention-type research questions (i.e., where randomized controlled trials are at the peak of the hierarchy). In addition, wherever possible, peer-reviewed publications, rather than conference abstracts, were cited.

For the TFOS Lifestyle Epidemic Workshop, the Evidence Quality Subcommittee was a new initiative that sought to provide specific expertise and support in research evidence appraisal and synthesis to each of the topic area subcommittees. Evidence Quality Subcommittee members were invited to contribute based on their recognized expertise in evidence-based practice; many hold leadership positions in evidence synthesis organizations, such as *Cochrane* and the *Agency for Healthcare*

Research and Quality Evidence-based Practice Center Program. Evidence Quality Subcommittee members contributed to the two sub-components of each topic area report, which consisted of a broad narrative review and a systematic review on a focused clinical question. The following subsections of this report discuss the specific roles and functions of the Evidence Quality Subcommittee and the practical implementation of this support.

4.1. Narrative reviews

As for previous TFOS Workshop initiatives [1–4], the central focus of each topic area report was an extensive narrative literature review. Led by the Chair of each subcommittee, and in consultation with their members, detailed narrative review outlines were developed *a priori* for each topic area. These outlines were reviewed for coherence, completeness, and scope (including potential overlap across reports) by the TFOS Lifestyle Epidemic Steering Committee and full Workshop membership, including Evidence Quality Subcommittee members. After this feedback, the outlines were refined, reviewed by the TFOS Executive Committee, and returned to the topic area Chairs for amendment. These revised versions were re-circulated to the Steering Committee for final approval. This multi-step, internal peer review activity is a key feature of the TFOS Workshop quality control process, which seeks to minimize potential biases, including reducing the risk of possible content omissions or undue overlap across the reports, through extensive internal, constructive review. The Evidence Quality Subcommittee made several additional distinct contributions to the narrative review process for the current Workshop, as follows:

(i) Video presentation to guide best practice for narrative reviews

To promote optimal and consistent approaches by Workshop members writing the narrative reviews, the Evidence Quality Subcommittee Chair recorded a presentation that was shared with the TFOS Workshop membership before the commencement of writing of the narrative reviews. The 25-minute video recording provided a summary of the following content, considered in the present report: (a) Narrative versus systematic literature reviews (Section 3.1); (b) Features of high-quality narrative reviews and how these relate to the Scale for the Quality Assessment of Narrative Review Articles tool domains (Section 3.2); (c) Principles of comprehensive literature searching in electronic databases (Section 4.1(iii)); (d) The National Health and Medical Research Council evidence hierarchy (Section 2.2); (e) Research critical appraisal (Sections 2.3 and 4.1(ii)); and (f) How to incorporate reliable and relevant systematic review evidence into a narrative review (Section 4.2(iv)).

(ii) CrowdCARE: research critical appraisal training

A component of the best-practice narrative review video presentation included instructions on the use of CrowdCARE [7] (Crowdsourcing Critical Appraisal of Research Evidence, crowdcare.unimelb.edu.au), a free, novel digital platform that uses crowdsourcing to teach the critical appraisal skills underpinning evidence-based practice. CrowdCARE provides access to interactive tutorials that provide detailed information relating to the National Health and Medical Research Council levels of evidence [9], and the clinical study designs that relate to each evidence level. The platform also includes worked examples on how to critically appraise research studies, including systematic reviews and randomized controlled trials, using validated risk of bias tools. After completing the online tutorials, CrowdCARE users contribute by appraising studies from the PubMed database and can then compare their appraisals to those submitted by others to obtain immediate feedback on their contributions. TFOS member engagement with the CrowdCARE platform was optional.

(iii) Standardized electronic database search strategies

To support Workshop members in performing comprehensive literature searches for the narrative reviews, the Evidence Quality Subcommittee developed standardized electronic searches for each of PubMed, Medline, and Embase (Table S2). Using a combination of keywords and controlled vocabulary terms, these standardized searches aimed to identify research articles broadly related to ocular surface structures, and conditions that may affect these tissues. For the Workshop, the 'ocular surface' was defined as the cornea, limbus, conjunctiva, eyelids and eyelashes, lacrimal apparatus, and tear film, along with their associated glands and muscular, vascular, lymphatic, and neural support. 'Ocular surface disease' included established diseases affecting any of the listed structures, as well as etiologically related perturbations and responses associated with these diseases. Disease was considered from an etiological perspective, to include infection, inflammation, allergy, trauma, neoplasia, dysfunction, degeneration, and inherited conditions. The searches were provided for optional use by contributors to the narrative reviews, with a recommendation to link (and limit) these foundational searches (using relevant Boolean operators, such as "AND") with other keywords relevant to the subject matter.

(iv) Curation of reliable and relevant systematic review databases for each topic area report

To ensure reliable and relevant systemic review evidence was appropriately cited in each topic area report, the Evidence Quality Subcommittee curated topic-focused databases of systematic review evidence that were distributed, by the relevant Chair, to the members of each subcommittee. The curation of each database involved several steps, which are summarized in Fig. 2.

To curate each database, the Evidence Quality Subcommittee Chair first accessed the full Cochrane Eyes and Vision United States (CEV@US) Database of Systematic Reviews in Eyes and Vision [55] (updated August 2021). This database provides a repository of the citation details of all systematic reviews and meta-analyses published in Medline and Embase for all eye and vision conditions meeting the following criteria: (i) the publication reports on at least one eye and vision condition; and (ii) the publication describes itself as one or more of a systematic review and/or meta-analysis. To meet the latter criterion, included reports were those that described themselves as a systematic review and/or meta-analysis in their title, abstract, or full-text report, or that met the Institute of Medicine's definition of a systematic review [56], when these terms were not used. To create the database, pairs of researchers from the CEV@US worked independently to screen titles and abstracts from a search that had combined 'eyes' and 'vision' keywords and controlled vocabulary terms with a validated search filter [57]. Any disagreements in classification were resolved by consensus, and then full-text reports were evaluated using a similar process, to confirm eligibility.

In August 2021, in alignment with the Workshop timeframe, the CEV@US Database of Systematic Reviews in Eyes and Vision [55] comprised 5871 unique records. Using an Excel spreadsheet format of the database, the Evidence Quality Subcommittee Chair screened all titles and abstracts to identify citations broadly relevant to the topic of the 'ocular surface'. Following this initial screening, the list of articles initially marked as irrelevant were randomized, and re-screened by the same author, to minimize the likelihood that any potentially relevant records were inadvertently excluded. Following this two-stage screening process, a total of 1113 systematic review records were deemed 'broadly relevant' to the Workshop.

To identify citations potentially relevant to each of the eight TFOS Workshop Ocular Surface Lifestyle Epidemic topic areas, the Evidence Quality Subcommittee member assigned to each topic area subcommittee screened the 'ocular surface' list (n = 1113 citations) at title/abstract level for relevance to their specific topic focus based on the finalized narrative report outlines. This process excluded an additional 169 articles, leaving a total of 944 citations deemed 'broadly relevant' to
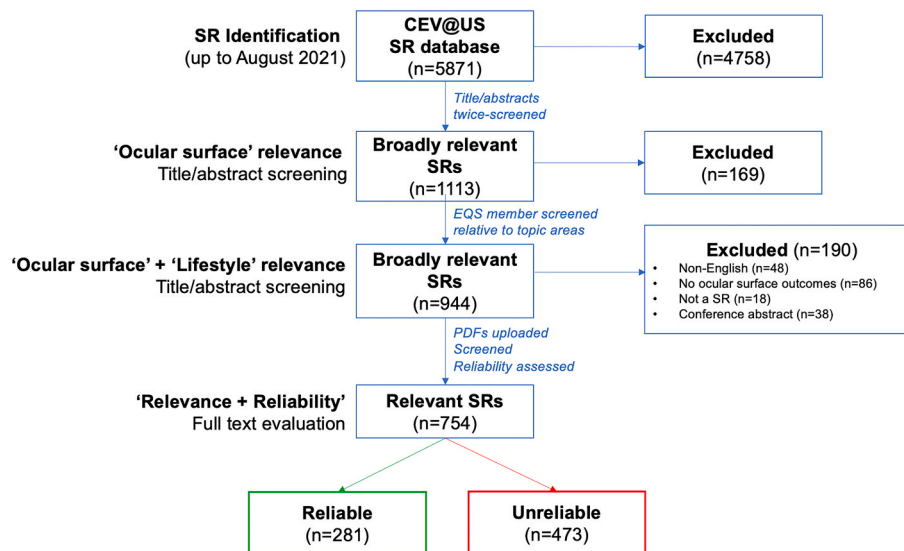
Fig. 2. **PRISMA flow diagram** summarizing the identification and reliability assessment of systematic review citations derived from the Cochrane Eyes and Vision United States Project (CEV@US) Database of Systematic Reviews in Eyes and Vision to populate systematic review databases for the TFOS Lifestyle Epidemic Workshop topic area narrative reviews.

both the 'ocular surface' and 'lifestyle factors'. Full texts were manually retrieved for the 944 articles, which were re-evaluated by each Evidence Quality Subcommittee topic area subcommittee member, for relevance to their topic area report, and categorized as either 'highly relevant' (i. e., the systematic review topic is clearly within the scope of the report) or 'possibly relevant' (i.e., the systematic review topic may be relevant to the scope of the report, depending on the content covered in the narrative review); 190 additional articles were deemed irrelevant at this stage, with reasons for exclusion provided in Fig. 2.

The remaining 754 'possibly' or 'highly' relevant systematic reviews were published between 1995 and 2021. Using their full text articles, these reviews underwent a 'rapid reliability assessment' using a five-item tool (Table 4) [42] that has been used to appraise systematic reviews to inform the development of clinical guidelines, including the American Academy of Ophthalmology Preferred Practice Patterns and European Glaucoma Guidelines [41]. These criteria are considered a minimum set of methodological requirements for reliability. Each article was appraised by one Evidence Quality Subcommittee member; any uncertainties in classification were discussed with the Evidence Quality Subcommittee Chair. For a systematic review to be judged as 'reliable' it needed to satisfy *all* five criteria. Following these assessments, a total of 281 broadly relevant systematic reviews were categorized as reliable (37%). The remaining 473 articles (63%) were deemed unreliable; of these, 138 systematic reviews failed one reliability criterion, 135 failed two criteria, 138 failed three criteria, 54 failed four criteria, and 12 failed all five criteria. As shown in Fig. 3, over the past decade there has been a steady increase in the annual number of systematic reviews published on topics broadly relevant to the ocular surface and lifestyle factors. However, many of these systematic reviews were deemed unreliable, and their overall methodological rigor, considered as the proportion of reliable reviews published in a given year, has remained similar over time (Fig. 3).

The output of this process was a suite of systematic review databases, individually curated by the Evidence Quality Subcommittee Chair, for the eight topic area reports. These databases were shared with members of each topic area subcommittee, who were advised to use the database as a resource to support the consistent use and interpretation of systematic review evidence in each narrative review. For each article considered 'highly' or 'possibly relevant' to each topic area, the database included the full citation details, article abstract, and reliability assessment details (including judgments for each of the five reliability

Table 4
Criteria for rapid assessment of the reliability of systematic reviews.

Criterion	Definition applied to systematic review reports
1. Defined eligibility criteria [41]	Described inclusion and/or exclusion criteria for eligible studies.
2. Conducted comprehensive literature search [58]	(1) Described an electronic search of two or more bibliographic databases; AND (2) Used a search strategy comprising a mixture of controlled vocabulary and keywords; AND (3) Reported using at least one other method of searching, such as searching of conference abstracts; identified ongoing trials; complemented electronic searching by hand search methods (e.g., checking reference lists); and contacted included study authors or experts.
3. Assessed risk of bias of included studies [58]	Used any method (e.g., scales, checklists, or domain-based evaluation) designed to assess the methodologic rigor of included studies.
4. Used appropriate methods for meta-analysis[a] (if performed) [41, 58,59]	Used quantitative methods that: (1) Were appropriate for the study design analyzed (e.g., maintained the randomized nature of trials; used adjusted estimates from observational studies); (2) Correctly computed the weight of studies in meta-analyses.
5. Observed concordance between review findings and conclusions [41,60]	Authors' reported conclusions were consistent with findings, provided a balanced consideration of benefits and harms, and did not favor a specific intervention if there was lack of evidence.

[a] If no meta-analysis was performed, the review was considered to automatically meet this criterion.

criteria). To facilitate potential incorporation into the narrative review, each systematic review was also classified into one of four categories (Table 5) to indicate its reliability and relevance to the topic area. The total number of systematic review articles in each topic area database ranged from 106 (Cosmetics [23]) to 434 (Elective Medications and Procedures [61]).

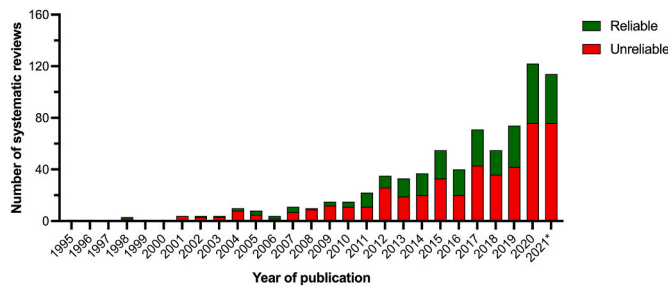After initial drafting of the narrative reviews, the Evidence Quality

Fig. 3. Number of systematic reviews published on topics broadly relevant to the ocular surface and lifestyle factors, stratified by whether they were assessed as reliable or unreliable using a 'rapid reliability assessment tool' [42]. *Note: 2021 is an incomplete year (up to August 2021, in alignment with the Workshop time frame).

Subcommittee member contributing to each topic area reviewed their relevant report to evaluate the citation of systematic review evidence. This evaluation focused on ensuring that systematic reviews that were judged to be 'reliable and highly relevant' were appropriately cited in the narrative review, and when other categories of reviews were cited (see Table 5), their findings were suitably qualified in light of any limitations of the review. Evidence Quality Subcommittee members also identified any additional systematic reviews that had been cited in the narrative reviews but were not in the systematic review database generated for the TFOS Lifestyle Workshop (for example for the Environmental Conditions Report [62]), a systematic review relating to aerosol-generating procedures [63] was appropriately cited given its broad relevance to the topic area). For these additional articles, Evidence Quality Subcommittee members performed the five-item reliability assessment (Table 4) and provided the topic area Chair with these assessments, along with any suggestions for revising the description of this evidence in the narrative review, as appropriate.

4.2. Systematic reviews

A novel feature of the current TFOS Lifestyle Epidemic Workshop was the addition of a systematic review (with or without a meta-analysis, depending on the research question, data availability, and study heterogeneity) as part of each topic area report. The purpose of this initiative was to provide a comprehensive synthesis of research evidence relevant to a specific, high-priority question using structured approaches and reporting frameworks.

Eight systematic reviews were generated across the Workshop, in parallel with the narrative review writing. Each systematic review focused on a single research question, prioritized by each topic area subcommittee (Table 6). To ensure a focused and feasible question [64], subcommittees drafted questions that were then shared with key members of the TFOS Executive, including the Evidence Quality Subcommittee Chair, for feedback, primarily with respect to scope. Once the questions were finalized, each systematic review was coordinated by the relevant topic area Evidence Quality Subcommittee member (Table 6), in consultation with the Evidence Quality Subcommittee Chair. In addition, at least two additional members from each topic area subcommittee participated in the undertaking of each systematic review.

The Evidence Quality Subcommittee sought to apply consistent, best practice methods across all of the systematic reviews [58]. Table 7 summarizes the key methodological features implemented across the reports, which meet all of the reliability criteria defined in Table 4. Five of the eight systematic reviews (Cosmetics [23], Digital environment [66], Lifestyle challenges [67], Contact Lenses [65] and Nutrition [24]) incorporated meta-analyses. For the other topic areas there was insufficient data and/or the data were deemed too heterogeneous for meta-analyses to be meaningful; in these cases, tabulated and narrative summaries of the evidence were provided. The number of unique records double-screened at the title/abstract screening stage in the individual systematic reviews ranged from 1417 (Lifestyle challenges [67]) to 9338 (Societal challenges [68]). The final number of unique studies

Table 5
Overall systematic review classifications for narrative topic area reviews.

Reliability assessment	Relevance to topic area subcommittee	Interpretation and use in narrative review
Reliable	High	**Reliable, High Relevance** systematic reviews have been: i) assessed as reliable using the five criteria and ii) deemed <u>likely</u> to be highly relevant to the topic area. These articles should be cited in the narrative review (unless deemed irrelevant, which should be discussed with the Evidence Quality Subcommittee member).
Reliable	Possibly	**Reliable, Possibly Relevant** systematic reviews have been: i) assessed as reliable using the five criteria (Table 4) and ii) deemed <u>possibly</u> relevant to the topic area. Whether these articles are cited in the narrative review will depend on the specific content covered.[a]
Unreliable	High	**Unreliable, High Relevance** systematic reviews have been: i) assessed as <u>not</u> reliable using the five criteria but ii) deemed <u>likely</u> to be highly relevant to the topic area. Findings from these articles need to be interpreted with caution as methodological concerns affect their validity.[b] It is generally recommended that these reviews are <u>not</u> cited in the narrative review; if they are, they need to be discussed in the context of their limitations and the implications of the methodological weakness(es) on their interpretation. If a Reliable, High Relevance review on the same content is available, it should be preferentially cited.
Unreliable	Possibly	**Unreliable, Possibly Relevant** systematic reviews have been: i) assessed as <u>not</u> reliable using the five criteria but ii) deemed <u>potentially</u> relevant to the topic area. Findings from these articles need to be interpreted with caution as methodological concerns affect their validity.[b] It is strongly recommended that these articles are <u>not</u> cited in the narrative review.[a]

[a]Many of the reviews in this 'possibly' category had general relevance to the management of ocular surface disease. [b]Areas of methodological concern could be identified by considering which one or more of the five criteria were not satisfied. An article meeting fewer criteria generally had greater methodological concerns, but each criterion is not necessarily equally weighted.

Table 6

Summary of prioritized clinical questions evaluated using systematic review methods, and the Evidence Quality Subcommittee member(s) who led each review.

Subcommittee	Clinical question	PROSPERO registration number	Evidence Quality Subcommittee member(s)
Contact lenses [65]	What lifestyle factors are associated with people dropping out of contact lens wear?	CRD42022297616	Jalbert, I
Cosmetics [23]	Is the use of eyelash growth serums associated with symptoms and/or signs of ocular surface disease?	CRD42022296378	Liu, S-H
Digital environment [66]	Which ocular surface disease management approaches reduce symptoms associated with digital device use?	CRD42022296735	Lingham, G
Elective medications and procedures [61]	What is the impact of refractive surgery on quality of life?*	CRD42022301818	Hogg, R; Qureshi, R
Environmental conditions [62]	What is the association between outdoor environment pollution and dry eye disease symptoms and/or signs in humans?	CRD42021297238	Saldanha, IJ
Lifestyle challenges [67]	Are chronic primary pain disorders associated with dry eye disease?	CRD42021296994	Britten-Jones, AC
Nutrition [24]	What are the effect(s) of different forms of intentional food restriction on ocular surface health?	CRD42022297045	Downie, LE; Singh, S
Societal challenges [68]	Has the COVID-19 pandemic changed the severity or outcome of ocular surface disease?	CRD42022299681	Li, T; Qureshi, R

Abbreviation: COVID-19, Coronavirus Disease 2019. *limited to small incision lenticule extraction (SMILE) in the report.

included in the individual systematic reviews ranged from 14 (Cosmetics [23]) to 40 (Societal challenges [68]). As informed by the research question and availability of evidence, a range of study designs were included across the reviews, including randomized controlled trials, non-randomized intervention trials, cohort studies, case-control studies, cross-sectional studies, and pre-post intervention studies (with no control group).

5. Reflections on the contribution and implementation of the Evidence Quality Subcommittee to the TFOS Lifestyle Epidemic Workshop

The Evidence Quality Subcommittee was formed with the goal of promoting the adoption of consistent and advanced literature review and synthesis methods across the TFOS Lifestyle Epidemic Workshop

Table 7

Methodological features of the topic area systematic reviews undertaken as part of the TFOS Lifestyle Workshop.

Methodological domain	Description	Purpose
Internal peer review process	Draft systematic review protocols, and review outputs underwent internal peer review by the Evidence Quality Subcommittee Chair ± a second Evidence Quality Subcommittee member, with revisions made further to this feedback, before broader circulation of the systematic review report to the full TFOS Membership alongside the topic area narrative review.	Internal quality control measure to enhance alignment and consistency across the reports.
A priori protocol, including defining study eligibility criteria	Systematic review protocols were prospectively registered on PROSPERO (https://www.crd.york.ac.uk/prospero/); registration details are provided in Table 6.	Reduces bias in the conduct and reporting of the review, and increases transparency.
Systematic literature searches	Systematic literature searches were crafted in consultation with the Evidence Quality Subcommittee Chair and/or information specialist, based on the standardized search strategies for the narrative review (Table S2). At least two different electronic databases were searched.	Maximizes the likelihood of complete literature capture.
Study flow	A PRISMA flow diagram [29] was included to depict the flow of citations through the different phases of the systematic review.	Provides a visual summary to transparently report the study selection process.
Risk of bias assessments	Included studies were assessed independently for risk of bias by at least two contributors, using validated tools appropriate to the study design (see Table 2). Any disagreements in assessment were resolved by consensus. The presentation of risk of bias assessment was customized for each report, using software and figure formats deemed appropriate to best represent the information.	Establishes internal validity of the evidence included in the review.
Data extraction	Data were extracted from eligible studies by either two contributors independently, or by a single contributor and verified by a second contributor [69]. Any disagreements were resolved by consensus.	Seeks to ensure the consistency and accuracy of data derived from individual studies.
Evidence syntheses	Planned methods to synthesize eligible studies were prospectively defined, including the criteria for undertaking meta-analyses, and outcome measures.	Aims to reduce bias in the conduct and reporting of the review, and to increase transparency.

(continued on next page)

Table 7 (*continued*)

Methodological domain	Description	Purpose
Certainty of the body of the evidence	When pre-specified in the review protocol, the GRADE approach [20] was used to assess the certainty of the body of the evidence for individual outcomes.	Application of GRADE seeks to guide the rating of the quality of the overall evidence base, and the strength of this evidence.

Abbreviations: TFOS, Tear Film and Ocular Surface Society; GRADE, Grading of Recommendations, Assessment, Development and Evaluations; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

subcommittee topic area reports. The focus was on ensuring the appropriate evaluation and presentation of clinical evidence. From the outset, a key consideration in defining the scope and activity of the Evidence Quality Subcommittee was how to best achieve a desired balance of rigor and feasibility. All contributors to the TFOS Workshop, including the Evidence Quality Subcommittee members, were volunteers and committed to tight timelines to complete each stage of the Workshop process. Additional factors that were pertinent to the current Workshop were the need for fully remote engagement across multiple countries (due to international travel restrictions related to the COVID-19 pandemic), the diversity of expert contributors and breadth of knowledge in evidence synthesis methods among the Workshop membership (n = 153 members), and the time-intensive nature of performing systematic reviews. It has been estimated that conducting one systematic review requires, on average, 30 person-weeks of full-time work [70].

For the narrative review component of each topic area report, the principal gains from the involvement of the Evidence Quality Subcommittee related to: (i) improving general awareness about different types of literature reviews (i.e., narrative vs systematic reviews), clinical research questions, and study designs and rigor; (ii) defining the factors that influence the quality of narrative reviews, including the value of using a quality assessment tool (the Scale for the Quality Assessment of Narrative Review Articles [52]); and (iii) facilitating incorporation of high-quality, relevant systematic review evidence into the narrative reviews by providing curated citation databases to each topic area subcommittee.

Incorporating a systematic review within each topic area report achieved a systematic and rigorous evaluation of the body of evidence for a specific research/clinical question that was judged to be of importance by global experts comprising each subcommittee. This contribution adds a new dimension to the consideration of evidence across the breadth of the Workshop. It was a priority for the eight systematic reviews, performed in parallel, with oversight from the Evidence Quality Subcommittee Chair, to adopt consistent methodological approaches. For most outcomes evaluated across the eight systematic reviews, only low or very low certainty evidence (using the Grading of Recommendations, Assessment, Development and Evaluations approach [20]) was identified. These findings highlight an ongoing need for high-quality research to examine relationships between specific lifestyle factors and ocular surface disease (e.g., the association between outdoor environmental pollution and dry eye disease) and define the efficacy and/or safety of specific interventions on ocular surface parameters (e.g., eye lash growth serums, interventions targeted to treat digital eye strain symptoms, intentional dietary restriction).

Some limitations were recognized to the approaches adopted to identify, appraise, and summarize the available evidence for both the narrative and systematic review subcomponents. Despite implementing the described processes to support the appropriate description and citation of high-quality, relevant evidence in the narrative reviews, it is acknowledged that these remain non-systematic literature reviews.

Standardized search strategies were provided to all Workshop members and, although their use was encouraged, it was not mandatory. Individual Workshop members independently undertook electronic literature searches, on different days, and in distinct databases. Given the breadth of each topic area report, ensuring the incorporation of all relevant evidence was simply not possible. Nonetheless, the approach that was used to identify and ensure the citation of reliable and relevant systematic reviews to inform the evidence summaries is a well-accepted method that has been previously adopted for clinical practice guideline development [43]. Furthermore, the robust internal peer review process by all members of the full set of reports that occurs within the TFOS Workshop minimizes the potential for important research findings to be omitted or less critical research to be over-emphasized.

In general, the methods adopted for the systematic review subcomponents of each report were robust and standardized (see Table 7), although several potential areas for improvement are acknowledged. Whilst there was general enthusiasm to contribute to the systematic reviews, considerable training was often needed for non-Evidence Quality Subcommittee contributors to perform standardized systematic review tasks, and this had to occur over a relatively short time frame. A key outcome of the engagement of the Evidence Quality Subcommittee was to upskill a larger number of contributors in these areas as a foundation for the future. It is also noted that the literature searches in the eight systematic reviews were typically restricted to English-language articles obtained from two electronic databases, which may have led to some non-English literature not being captured.

6. Key learnings and future considerations

In this section, learnings from implementing the Evidence Quality Subcommittee in the current Workshop are summarized (as described in Section 5), and suggestions that may guide evidence evaluation in future similarly large, international research synthesis initiatives are proposed (Table 8). Whether certain features are relevant to a particular future initiative will likely depend on the expertise and make-up of the contributors, breadth of topics to be reviewed, timelines for the activity, and extent to which the reports are intended to inform clinical practice and/or future research in the field. The suggestions are based on first-hand experience that may assist international bodies when pursuing similar initiatives.

7. Conclusions

In conclusion, evidence-based approaches consider the best-available, current evidence and should be foundational to healthcare and policy decision-making. Essential to this principle are that conclusions should be based on a comprehensive consideration of the evidence (i.e., that the representation of studies is as complete and unbiased as possible) and that greater emphasis is placed on research that has been appraised and shown to have used robust methods that are appropriate to the scientific question being addressed.

The Evidence Quality Subcommittee was established to support evidence-based methodological approaches for the TFOS Lifestyle Epidemic international Workshop. This report has outlined the purpose, scope, activity and contributions of the Evidence Quality Subcommittee to both the narrative and systematic review components of each topic area report. In addition, the present report describes the key methodological considerations associated with aiming to deliver high-quality literature reviews in each format, and reflects on the advantages and potential challenges associated with implementing standardized evidence evaluation methods for large, international evidence synthesis endeavours. These descriptions and suggestions may be useful for other global bodies contemplating similar undertakings.

Table 8

Suggestions for incorporating an Evidence Quality Subcommittee in future international evidence synthesis initiatives.

Feature	Process(es) implemented	Future suggestion(s) and the rationale
1. Formation of the Evidence Quality Subcommittee	• An Evidence Quality Subcommittee was an entirely new initiative and (to our knowledge) a novel undertaking for a global evidence synthesis endeavour. • Supported by the overall Workshop Chair, a written proposal for an Evidence Quality Subcommittee was submitted and approved by the Workshop Leadership Group. This occurred after the topic area subcommittees were established. • Due to timeline considerations, Evidence Quality Subcommittee members with established track records in evidence appraisal and/or synthesis in the eye care field were invited by the Evidence Quality Subcommittee Chair following consultation with the Workshop Leadership Group. To promote diversity and inclusion, representation was sought from a range of contributors in terms of geographic location, topic expertise, career stage and demographic factors.	• To optimize planning and coordination, establish the Evidence Quality Subcommittee concurrently with other subcommittees. • For consistency, run an open application call for the Evidence Quality Subcommittee, whereby individuals apply for membership, with competitive selection as for the other subcommittees. • Consider resource needs in the context of the planned timelines. • The expertise of Evidence Quality Subcommittee members should include individuals with expertise in systematic review and meta-analysis methods (e.g., developing and executing systematic search strategies, risk of bias assessment).
2. Scope of the sub-committee	• The evidence appraisal and synthesis approaches to be adopted by the Evidence Quality Subcommittee were developed in parallel with the early-stage planning of the scope of each topic area report, in consultation with the Workshop Steering Committee. This was an iterative process, involving several collaborative online meetings to reach consensus on how the Evidence Quality Subcommittee might be best engaged to contribute to the literature evaluations and syntheses.	• Establish the role, scope, and methodological approaches for the Evidence Quality Subcommittee *a priori*, to ensure clarity about the engagement from the outset.
3. Methods to promote high quality narrative reviews	Four key approaches (Items) were adopted with the intent of supporting delivery of high-quality narrative reviews for the eight topic areas: 1. Video presentation to define best practice for narrative reviews (recommended viewing) 2. CrowdCARE online evidence appraisal training for clinical research studies (optional) 3. Standardized electronic database search strategies (optional) 4. Curation of reliable and relevant systematic reviews for each topic area report (compulsory use; verified by the Evidence Quality Subcommittee)	• Items 1&2: Given that Workshop contributors often undertake a spectrum of research, spanning preclinical to implementation science, baseline familiarity with clinical study designs, evidence hierarchies and evidence appraisal may be heterogeneous. Ensuring a consistent level of understanding of these concepts could be facilitated by requiring the undertaking of Items 1 and 2 as a pre-requisite for contribution to the Workshop. • Item 3: Whether this step can be instituted may relate to several practical factors, including the breadth of the topic area(s) and familiarity of the contributors with structured electronic database search methods. Whilst desirable, complete implementation of this item is likely context dependant, and should not to be expected for narrative reviews. • Item 4: Whilst this is a substantial undertaking, mechanisms to support citation of high quality, relevant systematic review evidence aligns with the methods used in international clinical practice guideline development and adds a dimension of rigor to the overall consideration of the body of evidence.
4. Incorporation of a systematic review	• As the remit of the Evidence Quality Subcommittee was developed, adapted, and ultimately finalized during the early stages of the Workshop, the decision to undertake a systematic review of a focussed research question was confirmed only once the narrative review outlines were close to finalization. • The process for deciding on the 'priority' question was informal, involving a discussion amongst each topic area subcommittee, led by the topic area subcommittee Chair. • There was a call for at least two volunteers within each topic area subcommittee to contribute to the systematic review; the assignment of writing tasks was variable across subcommittees, as decided by the topic area Chair (i.e., sometimes these individuals also contributed sections to the narrative review, whereas in other cases they primarily contributed to the systematic review component). • Standardized, robust methods were adopted, to the extent possible, across all systematic reviews.	• From the outset, define the overall structure of each topic area report, including the potential for narrative and systematic review subcomponents with a consistent framework to define the integration of the sub-sections. • Consider using a structured approach (e.g., a Delphi method) that engages clinicians and/or patients to define the 'priority' research question. • During subcommittee member selection processes, ask applicants to indicate if they have systematic review expertise (although not a pre-requisite) and whether they would like to contribute to this component of the Workshop specifically. This will assist with attracting appropriate expertise and capacity, and identify relevant mentoring opportunities for those interested in learning these methods. • Create, and potentially publish, an overarching protocol that serves as a framework to define the expected minimum requirements for each systematic review.
5. Post-Workshop evaluation	• Not formally assessed.	• Consider instituting a formal evaluation of the process with members and end-users of the products resulting from the Workshop.

Note: The level of engagement of the full Workshop membership with Items 1 to 3 was not assessed, and so is unclear.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jtos.2023.04.009.

References

[1] Craig JP, Nelson JD, Azar DT, Belmonte C, Bron AJ, Chauhan SK, et al. TFOS DEWS II report executive summary. Ocul Surf 2017;15:802–12.

[2] Nichols JJ, Willcox MD, Bron AJ, Belmonte C, Ciolino JB, Craig JP, et al. The TFOS international Workshop on contact lens discomfort: executive summary. Invest Ophthalmol Vis Sci 2013;54. Tfos7-tfos13.

[3] Nichols KK, Foulks GN, Bron AJ, Glasgow BJ, Dogru M, Tsubota K, et al. The international Workshop on meibomian gland dysfunction: executive summary. Invest Ophthalmol Vis Sci 2011;52:1922–9.

[4] The definition and classification of dry eye disease: report of the definition and classification subcommittee of the international dry eye WorkShop. Ocul Surf 2007;5:75–92. 2007.

[5] Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71–2.

[6] Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2: e124.

[7] Pianta MJ, Makrai E, Verspoor KM, Cohn TA, Downie LE. Crowdsourcing critical appraisal of research evidence (CrowdCARE) was found to be a valid approach to assessing clinical research quality. J Clin Epidemiol 2018;104:8–14.

[8] Downie LE, Keller PR. Nutrition and age-related macular degeneration: research evidence in practice. Optom Vis Sci 2014;91:821–31.

[9] (NHMRC) National Health and Medical Research Council. NHMRC levels of evidence and grades for recommendations for developers of guidelines. 2009.

[10] Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Base Med 2016;21:125.

[11] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. Amstar 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008.

[12] Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016;69:225–34.

[13] Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898.

[14] Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919.

[15] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.

[16] Wells GA, Wells G, Shea B, Shea B, O'Connell D, Peterson J, et al. The newcastle-ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2014.

[17] NIH National Heart, Lung and Blood Institute. Study quality assessment tools. 2021. Available at: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools. 14 August 2022.

[18] Daly J, Willis K, Small R, Green J, Welch N, Kealy M, et al. A hierarchy of evidence for assessing qualitative health research. J Clin Epidemiol 2007;60:43–9.

[19] Egger M, Smith GD. Meta-Analysis. Potentials and promise. BMJ 1997;315:1371–4.

[20] Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol 2011;64:380–2.

[21] Treadwell JR, Singh S, Talati R, McPheeters ML, Reston JT. AHRQ methods for effective health care. A framework for "best evidence" approaches in systematic reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2011.

[22] Treadwell JR, Singh S, Talati R, McPheeters ML, Reston JT. A framework for best evidence approaches can improve the transparency of systematic reviews. J Clin Epidemiol 2012;65:1159–62.

[23] Sullivan DA, da Costa XA, Del Duca E, Doll T, Grupcheva CN, Lazreg S, et al. TFOS Lifestyle: impact of cosmetics on the ocular surface. Ocul Surf 2023. In press.

[24] Markoulli M, Arcot J, Ahmad S, Arita R, Benitez-del-Castillo J, Caffery B, et al. TFOS Lifestyle: impacts of nutrition on the ocular surface. Ocul Surf 2023. In press.

[25] Downie LE, Ng SM, Lindsley KB, Akpek EK. Omega-3 and omega-6 polyunsaturated fatty acids for dry eye disease. Cochrane Database Syst Rev 2019;12:Cd011016.

[26] Ma L-L, Wang Y-Y, Yang Z-H, Huang D, Weng H, Zeng X-T. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res 2020;7:7.

[27] Pussegoda K, Turner L, Garritty C, Mayhew A, Skidmore B, Stevens A, et al. Systematic review adherence to methodological or reporting quality. Syst Rev 2017;6:131.

[28] Lavis JN. How can we support the use of systematic reviews in policymaking? PLoS Med 2009;6:e1000141.

[29] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264–9. w64.

[30] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.

[31] Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. PROSPERO at one year: an evaluation of its utility. Syst Rev 2013;2:4.

[32] Cullis PS, Gudlaugsdottir K, Andrews J. A systematic review of the quality of conduct and reporting of systematic reviews and meta-analyses in paediatric surgery. PLoS One 2017;12:e0175213.

[33] Kelly KD, Travers A, Dorgan M, Slater L, Rowe BH. Evaluating the quality of systematic reviews in the emergency medicine literature. Emerg Med 2001;38:518–26.

[34] van der Pol CB, McInnes MD, Petrcich W, Tunis AS, Hanna R. Is quality and completeness of reporting of systematic reviews and meta-analyses published in high impact radiology journals associated with citation rates? PLoS One 2015;10: e0119892.

[35] Popovich I, Windsor B, Jordan V, Showell M, Shea B, Farquhar CM. Methodological quality of systematic reviews in subfertility: a comparison of two different approaches. PLoS One 2012;7:e50403.

[36] Sun X, Wang D, Wang M, Li H, Liu B. The reporting and methodological quality of systematic reviews and meta-analyses of nursing interventions for chronic obstructive pulmonary disease - a systematic review. Nurs Open 2021;8:1489–500.

[37] George PP, DeCastro Molina JA, Heng BH. The methodological quality of systematic reviews comparing intravitreal bevacizumab and alternates for neovascular age related macular degeneration: a systematic review of reviews. Indian J Ophthalmol 2014;62:761–7.

[38] Downie LE, Makrai E, Bonggotgetsakul Y, Dirito LJ, Kristo K, Pham M-AN, et al. Appraising the quality of systematic reviews for age-related macular degeneration interventions: a systematic review. JAMA Ophthalmol 2018;136:1051–61.

[39] Le JT, Qureshi R, Twose C, Rosman L, Han G, Fapohunda K, et al. Evaluation of systematic reviews of interventions for retina and vitreous conditions. JAMA Ophthalmol 2019;137:1399–405.

[40] Qureshi R, Azuara-Blanco A, Michelessi M, Virgili G, Barbosa Breda J, Cutolo CA, et al. What do we really know about the effectiveness of glaucoma interventions?: an overview of systematic reviews. Ophthalmol Glaucoma 2021;4:454–62.

[41] Golozar A, Chen Y, Lindsley K, Rouse B, Musch DC, Lum F, et al. Identification and description of reliable evidence for 2016 American Academy of Ophthalmology preferred practice pattern guidelines for cataract in the adult eye. JAMA Ophthalmol 2018;136:514–23.

[42] Saldanha IJ, Lindsley KB, Lum F, Dickersin K, Li T. Reliability of the evidence addressing treatment of corneal diseases: a summary of systematic reviews. JAMA Ophthalmol 2019;137:775–85.

[43] Mayo-Wilson E, Ng SM, Chuck RS, Li T. The quality of systematic reviews about interventions for refractive error can be improved: a review of systematic reviews. BMC Ophthalmol 2017;17:164.

[44] Hoffmann F, Allers K, Rombey T, Helbach J, Hoffmann A, Mathes T, et al. Nearly 80 systematic reviews were published each day: observational study on trends in epidemiology and reporting over the years 2000-2019. J Clin Epidemiol 2021;138:1–11.

[45] Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. JAMA 1998;280:278–80.

[46] Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, et al. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. BMJ 2000;320:537–40.

[47] Campbell JM, Kavanagh S, Kurmis R, Munn Z. Systematic reviews in burns care: poor quality and getting worse. J Burn Care Res 2017;38:e552–67.

[48] Furuya-Kanamori L, Xu C, Hasan SS, Doi SA. Quality versus Risk-of-Bias assessment in clinical research. J Clin Epidemiol 2021;129:172–5.

[49] Banzi R, Cinquini M, Gonzalez-Lorenzo M, Pecoraro V, Capobussi M, Minozzi S. Quality assessment versus risk of bias in systematic reviews: AMSTAR and ROBIS had similar reliability but differed in their construct and applicability. J Clin Epidemiol 2018;99:24–32.

[50] Webster J, Watson RT. Analyzing the past to prepare for the future: writing a literature review. MIS Q 2002;26:xiii–xxiii.

[51] Snyder H. Literature review as a research methodology: an overview and guidelines. J Bus Res 2019;104:333–9.

[52] Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. Res Integr Peer Rev 2019;4:5.

[53] Mulrow CD. The medical review article: state of the science. Ann Intern Med 1987; 106:485–8.

[54] Jones L, Downie LE, Korb D, Benitez-Del-Castillo JM, Dana R, Deng SX, et al. TFOS DEWS II management and therapy report. Ocul Surf 2017;15:575–628.

[55] Cochrane Eyes and Vision US. CEV database of systematic reviews in eyes and vision. 2021. Available at: https://eyes.cochrane.org/resources/cev-database-systematic-reviews-eyes-and-vision.

[56] US Institute of Medicine. Finding what works in health care: standards for systematic reviews. Washington DC, USA: National Academies Press; 2011.

[57] Montori VM, Wilczynski NL, Morgan D, Haynes RB. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. BMJ 2005;330:68.

[58] Cochrane handbook for systematic reviews of interventions. Version 5.1.0 ed. London, UK: Cochrane Collaboration.

[59] Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, et al. Indirect comparisons of competing interventions. Health Technol Assess 2005;9:1–134 [iii-iv].

[60] Li T, Vedula SS, Scherer R, Dickersin K. What comparative effectiveness research is needed? A framework for using guidelines and systematic reviews to identify evidence gaps and research priorities. Ann Intern Med 2012;156:367–77.

[61] Gomes JAP, Azar DT, Bauduoin C, Bitton E, Chen W, Hafezi F, et al. TFOS Lifestyle: impact of elective medications and procedures on the ocular surface. Ocul Surf 2023. In press.

[62] Alves M, Asbell P, Dogru M, Giannaccare G, Grau A, Gregory D, et al. TFOS Lifestyle: environmental impacts on the ocular surface. Ocul Surf 2023. In press.

[63] Tran K, Cimon K, Severn M, Pessoa-Silva CL, Conly J. Aerosol generating procedures and risk of transmission of acute respiratory infections to healthcare workers: a systematic review. PLoS One 2012;7:e35797.

[64] O'Dwyer LC, Wafford QE. Addressing challenges with systematic review teams through effective communication: a case report. JMLA 2021;109:643–7.

[65] Jones L, Efron K, Bandamwar K, Barnett M, Jacobs DS, Jalbert I, et al. TFOS Lifestyle: impact of contact lenses on the ocular surface. Ocul Surf 2023. In press.

[66] Wolffsohn JS, Lingham G, Downie LE, Huntjens B, Inomata T, Jivraj S, et al. TFOS Lifestyle: impact of the digital environment on the ocular surface. Ocul Surf 2023. In press.

[67] Galor A, Britten-Jones AC, Feng Y, Ferrari D, Goldblum D, Gupta PK, et al. TFOS Lifestyle: impact of lifestyle challenges on the ocular surface. Ocul Surf 2023. In press.

[68] Stapleton F, Abad JC, Barabino S, Burnett A, Iyer G, Lekhanont K, et al. TFOS Lifestyle: impact of societal challenges on the ocular surface. Ocul Surf 2023. In press.

[69] Li T, Saldanha IJ, Jap J, Smith BT, Canner J, Hutfless SM, et al. A randomized trial provided new evidence on the accuracy and efficiency of traditional vs. electronically annotated abstraction approaches in systematic reviews. J Clin Epidemiol 2019;115:77–89.

[70] Glasziou P. Systematic reviews in health care: a practical guide. Cambridge: Cambridge University Press; 2001.