

1 **Policy:** **Evidence Selection and Evaluation**

2

3 **Date of Implementation:** **June 18, 2020**

4

5 **Product:** **All Products**

6

7

8

9

10

11

12

<p>Related Policies:</p> <ul style="list-style-type: none"> • QM 32: Evidence Based Health Information Evaluation/Technology Assessment
--

13

14

15

16

American Specialty Health (ASH) lines of business include an array of services which impact the health of members in a variety of ways. These lines of business include American Specialty Health – Specialty, American Specialty Health Digital, and American Specialty Health – Fitness.

17

18

19

20

21

22

23

24

25

26

- The Clinical Network benefits administration manages member benefits for access to healthcare professionals (i.e., contracted in the ASH system and out of network healthcare professional such as chiropractors, physical therapists, and others).
- The Fitness programs provide members with access to fitness facilities, on-line digital resources, and health improvement information.
- The American Specialty Health Digital services provide members with digital format health information on the web and apps and, when eligible, access to coaching relative to healthy lifestyle (e.g., diet, exercise, tobacco use) chronic illness and musculoskeletal (MSK) conditions.

27

28

29

30

31

32

For each of these lines of business, there is a relevant body of scientific literature that informs ASH policies and guidelines. These bodies of literature are overlapping, but not identical. For each of these lines of business the breadth and depth of that literature varies. And for each of these lines of business there is a different set of clinical/scientific questions that must be answered.

33

34

35

36

37

ASH has created policies and procedures that dictate how this scientific literature is identified and evaluated. Broadly speaking, these policies and procedures conform to what has become known as “Evidence Based Medicine” (EBM). The rules for EBM do not change from one scenario to another. And this holds true within ASH. These different lines of business all respect the principles of EBM equally.

1 What does differ is the breadth and depth of available published evidence, the quality of
 2 the published evidence, how the evidence will and can be used by different delivery
 3 systems (e.g., network providers; coaches; digital content).

4
 5 Thus, a single set of principles and definitions, as defined below, are required to describe
 6 how ASH interprets scientific literature. These principles are also used to guide the work
 7 of the ASH Internal Evidence Evaluation Committee (IEEC) and the External Evidence
 8 Evaluation Committee (EEEC).

9
 10 **Sources of Evidence**

11 Acceptable sources of evidence for Clinical Practice Guidelines (CPGs), Coaching
 12 Guidelines (CGs), and Health Information Resources (HIRs), and subsequently the
 13 information contained in training manuals and consumer content may include:

- 14 • Peer-reviewed journals;
- 15 • Peer-reviewed textbooks;
- 16 • Professional clinical or health-related organizations;
- 17 • Core curriculum material from accredited educational institutions and programs;
- 18 • Expert and consensus opinion documents by professional organizations;
- 19 • Applicable State and Federal regulatory oversight agencies; and
- 20 • Applicable accreditation organizations (e.g., NCQA, URAC).

21
 22 The standards for evaluation may include, but may not be limited to:

- 23 • Randomization
- 24 • Sample size
- 25 • Pre/post measurement features
- 26 • Valid/reliable measures of outcomes
- 27 • Inclusion of a control group
- 28 • Effect size
- 29 • Impartial assessment of study outcome
- 30 • Well defined inclusion/exclusion criteria
- 31 • Assessment/measurement of co-interventions
- 32 • Appropriate statistical analyses
- 33 • Appropriate accounting of all study participants: drop-outs, compliance with
- 34 intervention, loss to follow-up
- 35 • Potential effects of conflicts of interest (e.g., study sponsorship or funding).

36
 37 A variety of study designs may include a sufficient number of the above characteristics to
 38 be used for the evaluation of clinical effectiveness. In addition, non-experimental designs

1 such as prospective cohort studies with non-randomized control groups will be considered
 2 as well as variations on this basic study design. Case reports and case series studies will
 3 not be used to inform the question of treatment effectiveness; however, these may be
 4 reviewed for informational purposes in the presence of other acceptable studies.

5 **The Clinical Evidence Hierarchy**

6 In implementing EBM, it is necessary to understand that all evidence is not the same. Each
 7 type of evidence provides different levels of quality, insights into the proof or disproof or
 8 a hypothesis, and confidence in the outcomes and results communicated in the document.
 9 Certain types of evidence are more persuasive than others. There is general agreement over
 10 this hierarchy of evidence; the higher the hierarchy a methodology is ranked (#1 below
 11 considered the highest), the more robust and closer to objective proof of fact it is assumed
 12 to be. In descending order, the hierarchy of clinical evidence is:

- 13
14
15 1. **Systematic reviews and meta-analyses.** A systematic review is a method of
 16 identifying, reviewing, and analyzing a body of data that uses explicit methods to locate
 17 primary studies and explicit criteria to assess their quality. A meta-analysis is a
 18 statistical analysis that combines or integrates the results of several independent clinical
 19 trials considered by the analyst to be "combinable" usually to the level of re-analyzing
 20 the original data. The advantage of a meta-analysis is that it may be possible, by
 21 combining study results, to more definitively identify and quantify treatment effects.
 22 Both systematic reviews and meta-analyses are typically predicated on the existence of
 23 a sufficient body of evidence, particularly of randomized controlled trials.
 24
- 25 2. **Randomized controlled trials (RCT).** Individuals are randomly allocated to a control
 26 group and a group who receive one or more specific intervention. When assigned
 27 randomly, the groups are usually very similar on other significant variables, such as
 28 age, sex, etc. They are followed up for specific end points and evaluated using specific
 29 outcome measures. RCTs are considered the essential method for establishing clinical
 30 effectiveness.
 31
- 32 3. **Non-randomized controlled trials.** These are similar to RCTs, with the exception that
 33 patients either self-select to a particular treatment group or are assigned (non-randomly)
 34 to a treatment group. There is a considerable reduction in methodological quality from
 35 RCT to Non-RCT. Without randomization it is much harder to determine if patients in
 36 each treatment group were comparable, especially on variables that are not measured.

- 1 4. **Prospective clinical series or pre-post study design.** A series of patients are
 2 systematically evaluated as they are being managed with a single treatment regimen.
 3 There is no randomization and no comparison group.
 4
- 5 5. **Case reports.** A case report is a systematic analysis and published report based on a
 6 single patient or subject. A case report is best utilized as a teaching tool rather than as
 7 providing evidence of effectiveness (or ineffectiveness). Sometimes a series of case
 8 reports are collected and reported together. It must be emphasized that simply adding
 9 more case reports (even if they show consistent results) does not elevate the hierarchical
 10 status of case reports.
 11
- 12 6. **Expert opinion.** An expert opinion is a consensus of experience and opinion from
 13 leaders or experts in a field. Although useful for certain purposes when done
 14 systematically, simply invoking expert opinion does not always carry significant
 15 methodological weight. Expert opinion with the same conclusions from different
 16 organizations and experts across disciplines adds strength to expert opinion but the
 17 conclusions must be used with caution as they remain opinion even if multiple experts
 18 opine similarly.
 19
- 20 7. **Clinical anecdote.** Personal anecdotes and experience (either the health care
 21 practitioner's or the patient's) are meaningful only to the individual involved and by
 22 itself should not be generalized to inform clinical decisions about others.
 23

24 While there are no rigid standards, it is generally understood that establishing the existence
 25 of clinical effectiveness requires some evidence indicating effectiveness in categories #1
 26 and #2 above. In the absence of such evidence, it is necessary to refer to other principles
 27 of EBM to determine appropriateness.
 28

29 ASH does not/will not perform its own formal systematic review of the literature where
 30 that review already exists and is authored by institutions like the Cochrane Collaboration
 31 and the Agency for Healthcare Research and Quality (AHRQ). These organizations are the
 32 first two sources that ASH committees' access in their evidence evaluation process. As
 33 appropriate, other systematic reviews are used as are specific RCTs that represent core
 34 findings for the intervention or question at hand.

1 **DEFINITIONS USED BY ASH EVIDENCE EVALUATION COMMITTEES**

2 The ASH evidence evaluation committees have approved the following definitions as
3 references to use during their review of research and evidence:

4
5 **Clinical effectiveness:** The capacity of a therapeutic intervention or coaching
6 methodology to more likely than not result in or contribute to a favorable outcome for the
7 condition and population to which it is applied.

8
9 **Safety:** The terms “safe” and “safety,” are used only with specific reference to the absence
10 of direct harm. Direct harm would include any injury to a patient caused by the mechanical,
11 thermal, biological, chemical, pharmacological, electrical, electromagnetic, or psycho-
12 dynamic properties of a diagnostic or therapeutic procedure, and as such, the procedure
13 would be considered unsafe.

14
15 **Diagnostic utility:** The capacity for a diagnostic test or procedure to both:

- 16 • Provide reliable and valid information about a specific clinical condition; and
- 17 • Influence the course of management of a specific condition in a manner expected
18 to improve prognosis and/or clinical outcomes.

19
20 **Preponderance of evidence:** The conclusion reached on the quality and quantity of
21 considered sources of information regarding the safety, and clinical effectiveness or
22 diagnostic utility. The conclusion is one of the following: Positive, Negative, or Equivocal.

23
24 **Benefit versus risk profile:** The relative effectiveness or utility of a therapeutic
25 intervention, coaching methodology, or diagnostic procedure versus its potential for direct
26 harm. The conclusion is one of the following:

- 27 • Positive (benefits outweigh risks);
- 28 • Negative (risks outweigh benefits); or
- 29 • Equivocal (available information is inconclusive).

30
31 **Scientific plausibility:** A belief, theory, or mechanism of health and disease that can be
32 explained within the existing framework of scientific methods, reasoning, and available
33 knowledge. A treatment intervention or diagnostic procedure is said to be implausible if it
34 requires the existence of forces, mechanisms, or biological processes that are not known to
35 exist within the current framework of scientific methods, reasoning, and available
36 knowledge.

1 **Substitution harm:** Compromised clinical outcomes caused by:

- 2 • Utilizing a specific diagnostic or therapeutic procedure or coaching methodology
- 3 whose safety, clinical effectiveness, or diagnostic utility is either unknown or is
- 4 known to be unsafe, ineffective, or of no diagnostic utility, if implemented *in place*
- 5 *of or instead of* a diagnostic or therapeutic procedure or coaching methodology
- 6 known to be safe, be clinically effective, or to have diagnostic utility; or
- 7 • The utilization of a diagnostic or therapeutic procedure or coaching methodology
- 8 that is substantially less effective or safe than another procedure with established
- 9 safety, and clinical effectiveness or utility.

10
11 **Labeling effects:** The harm that results from communicating to a patient that they have a
12 condition or a finding that requires care/services or has an effect on their health when that
13 condition or finding is actually incidental, not clinically significant, or not clinically valid
14 (e.g., a diagnosis, physical or mental or functional, is defined but has no support in the
15 evidence).

16
17 Safety of Information Provided to Participants/Members/Patients – Significant
18 consideration is given to the process of creating consumer-directed health content to ensure
19 the information provided does not:

- 20 • Constitute the practice of medicine, psychology/psychiatry, nursing, dietetics, or
- 21 other health professions; or
- 22 • Prescribe or provide medical, pharmacy, physical therapy or any other clinical
- 23 service; or
- 24 • Interfere with the recommendations of the participant’s health care practitioner; or
- 25 • Interfere with the practitioner/patient relationship; or
- 26 • Provide information that can easily be misunderstood or misinterpreted by the
- 27 coach, participant/member/patient, or health care practitioner.

28
29 Both the External Evidence Evaluation Committee (EEEC) and the Internal Evidence
30 Evaluation Committee (IEEC) utilize a clinical decision-making algorithm along with the
31 above considerations to determine the applicability and validity of the information being
32 evaluated and applied in ASH program management. Both EEEEC and IEEEC provide
33 clinical policy recommendations, as applicable, to the appropriate committee(s) for
34 approval. The algorithm determines the appropriate classification (Established, Not
35 Established, or Unacceptable Health/Safety Risk) for each technique/procedure evaluated
36 by either the EEEEC or the IEEEC. The decision at each point in the algorithm is driven by
37 the supporting information and criteria described above.

1 The clinical techniques/procedures are categorized as follows:

2
3 Established – The diagnostic and/or therapeutic procedure:

- 4 • Is established as safe and is established as effective; or
- 5 • Is not established as safe; however, the benefit: risk profile is favorable.

6
7 Not Established – The diagnostic and/or therapeutic procedure:

- 8 • Is established as safe, is not established as effective, and poses no substantial
- 9 substitution and/or labeling risk; or
- 10 • Is not established as safe, the benefit:risk profile is not established as favorable, the
- 11 benefit:risk profile is not established as unfavorable and poses no substantial
- 12 substitution risk and/or labeling risk and/or the procedure is not determined to be
- 13 scientifically implausible.

14
15 Health and Safety Risk – The diagnostic and/or therapeutic procedure:

- 16 • Is established as safe, is not established as effective, and poses substantial
- 17 substitution and/or labeling risk; or
- 18 • Is not established as safe, the benefit:risk profile is not established as favorable, the
- 19 benefit:risk profile is established as unfavorable; or
- 20 • Is not established as safe, the benefit:risk profile is not established as favorable, the
- 21 benefit:risk profile is not established as unfavorable; and poses substantial
- 22 substitution risk and/or labeling risk and/or the procedure is determined to be
- 23 scientifically implausible.

24
25 **The Bradford-Hill Criteria**

26 A set of criteria has been established for the purpose of evaluating and analyzing all the
27 disparate evidence that might exist on a specific clinical question. These “Bradford-Hill
28 Criteria” (named after Sir Austin Bradford-Hill, the British epidemiologist who first
29 identified the causal relationship between cigarette smoking and lung cancer) were first
30 developed and used to evaluate the relationship between certain risk factors (e.g., smoking)
31 and specific diseases (e.g., lung cancer), but these have been extended to cover most
32 clinical science questions: safety, effectiveness, diagnostic utility, etc.

33
34 At the core of these criteria is the question of causality—is the risk factor, treatment, etc.,
35 causally related to the clinical outcome of interest? This clinical outcome might be the
36 development of disease, the prevention of disease, or a resolution or improvement in the
37 disease process. These criteria were explicitly established under the premise that the
38 evidence is always flawed, incomplete, and even contradictory and that the idea of “proof”
39 is nonsensical. These criteria assume that clinical decision-making is always done under a

1 cloud of uncertainty. Thus, the criteria are particularly useful when applied to practical
 2 questions (Is procedure X an appropriate modality?). A discussion of the criteria is
 3 provided below:

- 4
- 5 1. **Strength of association.** The principle here is that large effects (either good or bad)
 6 are easily seen and that these large effects (if real) are unlikely to be produced by
 7 spurious effects. For example, the strength of the association between cigarette
 8 smoking and lung cancer is a twenty-fold increase in the likelihood of cancer among
 9 smokers. An effect of this magnitude is impossible to explain other than by a causal
 10 relationship. By contrast, many clinical studies find very small effects. Small effects
 11 (e.g., treatment A is slightly better than a placebo) are consistent both with a causal
 12 relationship or with no relationship whatever. Small and marginal effect can be
 13 explained by many means other than causality.
 - 14
 - 15 2. **Consistency.** The reason that small effects are suspect is that there are always many
 16 confounding variables that might have been responsible for the results. However, the
 17 set of confounding variables are not likely to be the same in different populations,
 18 different studies, etc. Thus, if different studies, with different methodologies, in
 19 different populations, produce similar results, it strongly suggests a causal relationship.
 20 If on the other hand, as is often the case, study results are inconsistent among different
 21 populations and different studies, it suggests either an absence of causality, or very
 22 small effect sizes.
 - 23
 - 24 3. **Specificity.** The principle of specificity is one cause, one effect. That is, it is presumed
 25 to be biologically improbable that a single risk factor is a cause of multiple diseases.
 26 So, when it is reported or suspected that a particular pollutant is causing MS, breast
 27 cancer, arthritis, autism, etc., this evidence is given very little credence. This principle
 28 is also extended to the idea that a single treatment is unlikely to be effective for a broad
 29 range of conditions. One can think of exceptions to the specificity rule—smoking does
 30 cause a variety of problems and exercise is effective in treating/preventing a broad
 31 range of diseases-but in general the specificity rule is reliable.
 - 32
 - 33 4. **Temporality.** This criterion is applied exclusively to questions of risk factors. The
 34 problem arises in cross-sectional population studies of risk factors when it is not clear
 35 whether the risk variable was in fact present before the onset of a disease process. This
 36 issue is rarely a problem relative to evaluating treatment effectiveness.
 - 37
 - 38 5. **Dose-response.** The principle is that if a causal relationship exists, there is likely to be
 39 a linear relationship (within limits) between the causal agent and the effect. This is
 40 certainly true relative to cigarette smoking—one pack a day is much worse than five

1 cigarettes a day. It is also generally true with regard to treatment effectiveness.
 2 Assuming a treatment is effective, it should be possible to increase or decrease the
 3 treatment dose and achieve different levels of effectiveness. The absence of a dose-
 4 response finding suggests that the presumed treatment effectiveness may be spurious.

5
 6 **6. Scientific plausibility.** The issue of scientific plausibility is both one of the more useful
 7 and one of the more contentious criteria. The question is whether the presumed
 8 mechanism of a treatment is or is not plausible. What is plausible to one person may
 9 not be plausible to another. But this can still be a powerful argument either for or
 10 against a particular procedure. Particularly when there is an absence of actual clinical
 11 outcomes studies it is entirely legitimate to ask: “Does this make sense? Is there a
 12 scientifically plausible explanation for the treatment’s effectiveness?” If a given
 13 procedure’s effectiveness, or diagnostic utility, is predicated upon mechanisms, forces,
 14 or biological processes that are not known to exist, the plausibility of that procedure is
 15 legitimately in doubt. And on the other hand, if there is a very strong (scientifically
 16 plausible) theory to explain an effect, it legitimately strengthens the case in favor of a
 17 procedure, practice, and/or lifestyle.

18
 19 **7. Coherence.** This is related to the above principle of scientific plausibility but is more
 20 general. The question is whether a particular approach, theory, or treatment and the
 21 cause and effect relationship that is hypothesized is consistent with the broad base of
 22 existing knowledge on the subject. The principle is that one should always choose the
 23 explanation that requires the fewest assumptions and one should avoid theories that
 24 require re-writing the textbooks. To say that a particular theory (or a procedure
 25 associated with that theory) is coherent is to say that it is consistent with the existing
 26 set of scientific literature on the subject. An incoherent approach is one that requires
 27 that we abandon most of what we know about a subject in order to accommodate this
 28 new approach.

29
 30 **8. Analogy.** The principle of analogy is a particularly useful concept when there is a
 31 complete absence of clinical literature. It permits us to observe that although we might
 32 not have any evidence of effectiveness of Procedure X on Condition B, we do have
 33 evidence of effectiveness of Procedure X on Conditions A & C, which are closely
 34 related to Condition B. The foremost example of the use of analogy would be:
 35 Procedure X = SMT; Condition B = Thoracic spine pain; Conditions A & C = Low
 36 back and cervical spine pain, respectively. There is in fact no evidence of effectiveness
 37 of Spinal Manipulative Therapy (SMT) (or any treatment, for that matter) for thoracic
 38 spine pain. So, we can observe that even though we have zero random controlled trials
 39 (RCT) evidence that SMT is effective for thoracic spine pain, we do have evidence for

1 the lumbar and cervical spine, and these are analogous situations. The question with
 2 any analogy is, how closely related are the analogous situations?
 3

- 4 **9. Experimental evidence.** There is a variety of experimental evidence that can be
 5 brought to bear on a clinical question. Experiments on the mechanical, physical,
 6 thermal, electromagnetic, or physiological properties of a treatment are all relevant and
 7 may either strengthen the case or weaken the case for a particular procedure. It should
 8 always be kept in mind, however, that this type of evidence is more useful in defining
 9 how a procedure might work, rather than if it works. The latter question is best
 10 addressed directly through clinical science methods.
 11

12 It is important to note that it is not presumed that each of these Bradford-Hill criteria must
 13 be met to establish a plausible cause/effect relationship. Indeed, that is very improbable.
 14 Rather, the point is to apply each of these criteria to a question and then weigh the totality
 15 of the evidence. The application of the Bradford-Hill criteria does not automatically result
 16 in a definitive conclusion, but it narrows considerably the range of possible conclusions.
 17 Probably the best instruction as to how and when to use the Bradford-Hill criteria comes
 18 from Sir Austin Bradford-Hill himself. He concluded that:
 19

20 *“All scientific work is incomplete – whether it be observational or experimental. All*
 21 *scientific work is liable to be upset or modified by advancing knowledge. That does not*
 22 *confer upon us a freedom to ignore the knowledge we already have, or to postpone the*
 23 *action that it appears to demand at a given time.”*
 24

25 ASH clinical/health evidence evaluation, interpretation, and application to business
 26 processes follows the principles outlined in this document. ASH committees may make
 27 changes to this policy as the knowledge and evidence management principles change
 28 within the research community. The governance and processes of policy and guideline
 29 review and adoption is further outlined in other ASH clinical policy documents.