Defining Criteria Used to Evaluate Response to Treatment of Acne Vulgaris

James Q. Del Rosso, DO

Results from controlled studies form the basis of overall perceptions regarding the efficacy and safety of specific treatments. In acne vulgaris, determining statistical significance related to mean percentage reduction in inflammatory and noninflammatory lesion counts, investigator global assessment, and patient (subject) global assessment have formed the basis of most studies. Results may be impacted by several mitigating factors related to inclusion and exclusion criteria and variations in study "power."

Recently, standards for evaluation of response to acne treatment have been reconsidered, with new methodologies suggested throughout the approval process. For example, the standard of "complete clearance" has been introduced. How the new methodologies compare with previous standards, and how new criteria will impact the reporting and interpretation of trial results are reviewed in this article. Specific study outcomes, including those reported in more recent trials with topical adapalene, are utilized as illustrative examples.

Cutis. 2006;78:117-121.

Evaluating Treatment of Acne Vulgaris

When a new product is introduced, clinical trials form the primary foundation for clinical decision making. The highest level of evidence is the double-blinded, randomized, vehicle- or placebo-controlled trial that is adequately powered and appropriately analyzed by accepted and established statistical methodologies. Although materials and methods are delineated, both investigators and clinicians reading the outcomes of a study after its completion may interpret results based on their own perceptions of what is being described. For example, specific grading systems and definitions of levels of therapeutic response may not be adequate to achieve a consistent picture of what is being described. Difficulty occurs when one attempts to translate what the data from a trial are demonstrating to what clinicians actually observe and evaluate in clinical practice in terms of grading severity and evaluating response to therapy. Essentially, words are being utilized to describe results that are based on visual analysis that usually is inclusive of multiple clinical investigators.

Translating Research Results to Clinical Practice

I became intrigued with the potential for inaccuracy or inconsistency associated with clinical evaluation of acne vulgaris after first learning that at least 25 scales have been described for evaluating global acne severity and after attending workshops in which clinical dermatologists and research investigators were asked to grade severity of individual cases of acne vulgaris and rosacea. The variability of responses among a group of experienced dermatologists was surprising.[1] Lesion count variability also was observed. The Dermatologic and Ophthalmic Drugs Advisory Committee (DODAC) to the US Food and Drug Administration (FDA) in November 2002 recognized the difficulty related to standardizing the evaluation of acne vulgaris in clinical trials.[2] In 1990, a consensus conference of the American Academy of Dermatology discussed the difficulty of developing a universal scale for trials, primarily because of the pleomorphic nature of acne vulgaris.[3] Factors confounding analysis include multiple lesion types, variability of inflammatory and noninflammatory lesion types, involvement of facial and extrafacial sites, and the changing nature of acne lesions as they progress through their normal life cycle (natural course of the disease). Clinical researchers and government officials involved in the drug approval process continue to strive for a system

Accepted for publication September 28, 2005.

From the Department of Dermatology, University of Nevada School of Medicine, Las Vegas.

Dr. Del Rosso is a consultant and speaker for Allergan, Inc; CollaGenex Pharmaceuticals, Inc; Connetics Corporation; Coria Laboratories, Ltd; Doak Dermatologics, a subsidiary of Bradley Pharmaceuticals, Inc; Galderma Laboratories, LP; Intendis; Medicis Pharmaceutical Corporation; Novartis Pharmaceuticals Corporation; OrthoNeutrogena; QLT Inc; SkinMedica, Inc; Stiefel Laboratories, Inc; and Warner Chilcott.

Reprints not available from the author.

Table 1.

Acne Severity Grading Scales*[2,4]

Scale	Comments
Leeds (Cunliffe) scale	A 10-point scale ranging from 0 (no acne lesions) to 10 (very severe acne); due to subdivisions within several grades, becomes a 26-point scale; especially cumbersome for clinicians
Cook scale	Five grading definitions; due to ability to assign additional grades that may be arbitrary, becomes a 9-point scale; completely clear status not identified
Dynamic scale	Memory-dependent; requiring correlation with baseline status; lacking clinical descriptors; describing levels of improvement (eg, moderate improvement) does not translate to a definitive clinical picture
Newer proposals	Static definitions used to evaluate current status and treatment response; not compared with baseline; current acne status defined at point in time of evaluation; limited number of levels for ease of use; defines completely clear and almost clear responses, which may be more practical for clinicians; usefulness dependent on precise and concise definitions of acne severity included in the trial and on level descriptors that correlate accurately with the true clinical picture observed by the investigator or clinician; if used as end points by FDA for approval, the break point for approval needs to correlate with reasonable expectations; lesion counts and standard global assessments still valuable as study end points (eg, secondary)

*FDA indicates US Food and Drug Administration.

of evaluation for acne vulgaris inclusive of a universally accepted severity grading scale and measures to evaluate treatment response.

Conventional Methods Used to Evaluate Acne Therapy

Grading Acne Severity—At the DODAC meeting in November 2002, it was explained that "a number of different scales have been published in the literature and a number of different scales have been proposed by sponsors for use at the agency [FDA]."[2] Problems also have existed in reaching a consensus for what should be assessed during investigator global evaluation. For example, some authorities indicate that greater emphasis should be placed on inflammatory lesions. Others argue that patient assessments may be misleading because therapy that exhibits dramatic reduction in acne lesion counts may still be graded as poor because the only acceptable response for some patients is complete absence of acne lesions. Still, others have questioned the need for specification regarding the anatomic sites included in the clinical trials (eg, facial vs extrafacial sites). Examples of acne severity grading scales and their potential limitations are listed in Table 1.[2,4]

Since the establishment of DODAC in 1996, products evaluated for mild to moderate acne have received FDA approval based on the following guidelines[2]:

- Two properly designed clinical trials that compare active product as monotherapy versus topical vehicle (if active product is a topical agent) or oral placebo (if active product is an oral agent)
- Assessment based on a physician global severity scale that demonstrates statistically significant superiority of active product versus vehicle (topical) or placebo (oral)
- Reduction in 2 lesion counts (usually calculated as mean percentage reduction) inclusive of total lesion count, inflammatory lesion count, and/or noninflammatory (comedonal) lesion count that documents statistically significant superiority of active product versus vehicle (topical) or placebo (oral). Many studies include data on all 3 lesion counts

Table 2.

Acceptable Global Evaluation Scale for Acne Vulgaris Trials[2]

0=healthy clear skin with no evidence of acne vulgaris

1=almost clear; rare noninflammatory lesions present; rare noninflamed resolving papules (may be hyperpigmented but not pink-red)

2=some noninflammatory lesions present; few inflammatory lesions (papules/pustules only; no nodulocystic lesions)

3=noninflammatory lesions predominate; multiple inflammatory lesions present; several to many comedones and papules/pustules; one small nodulocystic lesion

4=inflammatory lesions predominate; many comedones and papules/pustules; may or may not be a few nodulocystic lesions

5=highly inflammatory lesions predominate; variable number of comedones; many papules/pustules and nodulocystic lesions

Although these approval standards remain intact as important end points for evaluation of acne therapies, new proposals like these are being introduced as standards of drug evaluation and approval for several disorders, such as acne vulgaris, rosacea, and actinic keratosis.

Global Severity Scales—Unlike lesion counts, which also may be subject to variability, a global severity scale that is universally accepted may serve to equalize the playing field among both investigators and practicing clinicians. The problem is developing a universal global severity scale. The "ideal" global scale for evaluation of acne vulgaris has been suggested as follows[2]:

- A user-friendly and practical scale with a limited number of evaluation levels. The evaluation scale must demonstrate a high degree of correlation with lesion counts
- Evaluation levels that are appropriately described to accurately correlate with the true clinical picture of acne status and to significantly limit both intraobserver and interobserver variability
- Levels that indicate clear or almost clear, defining when therapy may be discontinued or when a maintenance regimen may be introduced
- Static measures that evaluate acne status as a designated point in time and eliminate memory dependence or correlation with baseline status by the examiner
- An evaluation scale that is universally accepted for both investigational and real-world clinical use, and a scale that is utilized in all phase 3 and most phase 4 studies, except those that are specially designed to evaluate other designated or specific end points

Table 2 is an example of an acceptable proposed global assessment scale for acne vulgaris as discussed at the DODAC meeting in November 2002.[2] Evaluation of this and other suggested ideal scales will undoubtedly uncover individual grading scenarios that are not adequately addressed, resulting in arbitrary grading and a greater potential for interinvestigator variability.

The development of a universally acceptable investigator global evaluation scale is a dynamic and evolving process. In the meantime, methodologies that are used must be clearly defined with continued education provided to both study investigators and practicing clinicians so that study results may be accurately interpreted and appropriately applied to clinical practice.

Recent Studies Evaluating Adapalene Gel 0.3%

Two recent trials evaluating the efficacy of adapalene gel 0.3% versus both the adapalene gel 0.1% formulation and vehicle provide a comparison of different methodologies used to evaluate response to acne therapy.[5,6]

Conventional Evaluation—The first trial described is a multicenter, randomized, investigator-blinded, parallel comparison of adapalene gel 0.3% versus adapalene gel 0.1% versus vehicle administered over 12 weeks for treatment of acne vulgaris.[6] Table 2 describes the conventional evaluation of therapeutic response determined by this study based on mean percentage lesion count reductions, a "language" that is familiar to both investigators and clinicians because of a long track record of use. Despite certain limitations, lesion count evaluations remain as important end point parameters to be used in clinical trials.

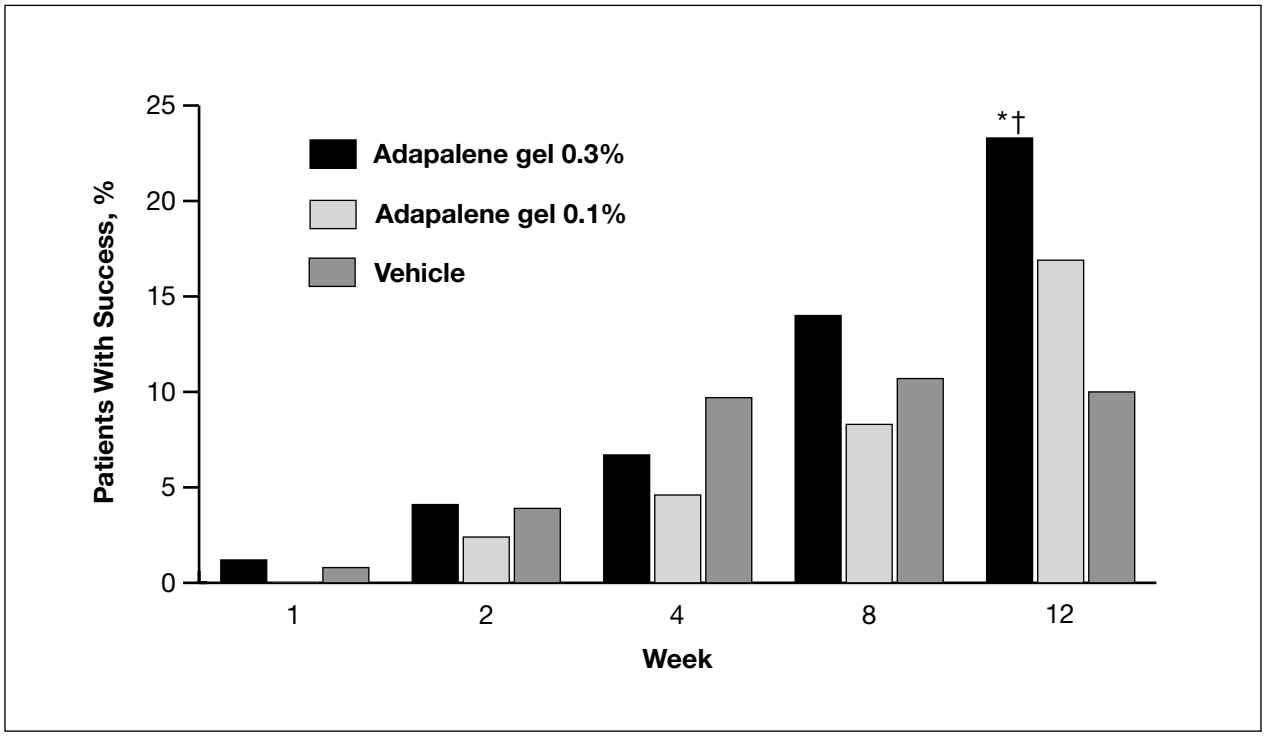
New Evaluation Methods—The second trial described is a multicenter, randomized, double-blinded, vehicle-controlled, parallel comparison of adapalene gel 0.3% versus adapalene gel 0.1% versus vehicle administered over 12 weeks for acne vulgaris.[5] This study assessed efficacy using both new static

global evaluation methodology and conventional determination of lesion count percentage reduction. The investigator's global (static) assessments, a primary end point for this trial, were based on a 6-point scale (0=clear, some residual hyperpigmentation and erythema may be present; 1=almost clear, patients may have a few scattered comedones and fewer than 5 small papules; 2=mild, acne is easily recognizable, but less than half the face is involved and there are multiple comedones, papules, and pustules; 3=moderate, more than half the face is involved and there are numerous comedones, papules, and pustules; 4=severe, the entire face is involved, covered with numerous comedones, papules, pustules, and a few nodules and cysts; 5=very severe, patients have highly inflammatory acne covering the entire face, with nodules and cysts also present). Using this system, success is described as clear and almost clear, with any other acne status described as failure at study end point.

Scores of 0 (clear) and 1 (almost clear) were considered treatment successes.[5] This criteria for treatment evaluation raises the bar when analyzing response to acne therapy; percentages reported for treatment success must not be confused with percentages reported using other more conventional methodologies (eg, mean percentage lesion reduction, median percentage lesion reduction, 1- or 2-grade reduction in global severity).

The Figure describes trial results using this static method of investigator global evaluation. At week 12, the subjects using adapalene gel 0.3% had a success rate of 23.3% versus 16.9% among those using adapalene gel 0.1% (P=.020) and 10% for those using vehicle (P=.005). The median percentage change in lesion count for inflammatory and noninflammatory lesions was highest in the group using adapalene gel 0.3%. The median percentage decrease for all lesions was –55.6% in the adapalene gel 0.3% group versus –48.2% in the adapalene gel 0.1% group (P=.020) versus –36.4% for those using vehicle (P<.001).

For all treatment groups, the greatest decreases were shown in inflammatory lesions. The group using adapalene gel 0.3% showed a median percentage decrease of –62.5% versus –57.8% in the group using adapalene gel 0.1% (P=.015) versus –47.2% in the vehicle group (P<.001). The median percentage change in noninflammatory lesions was –52.1% in the group using adapalene gel 0.3% versus –43.4% in the adapalene gel 0.1% group (P=.061) and –29.3% in the vehicle gel group (P<.001). The combination of both methodologies, assuming they are properly designed, clearly defined, and understood by the reader, provides additional information.[5]



Adapalene gel 0.3% trial investigator global assessment results. Asterisk indicates P=.020 for adapalene gel 0.3% versus adapalene gel 0.1%. Dagger indicates P=.005 for adapalene gel 0.3% versus vehicle. Data from Thiboutot et al.[5]

Potential Pitfalls of New Methodologies

New definitions of success (clear or almost clear) versus failure (any other clinical status) in acne vulgaris trials are subject to reasonable criticism. Based on our current knowledge of treatment response established by myriad clinical trials, it is not realistic to expect complete or near clearance with therapies used to treat diseases such as acne vulgaris within the confines of a standard 12-week trial, especially with monotherapy. As combination therapy is firmly entrenched as the standard of care in the majority of cases of acne therapy, what needs to be determined by monotherapy trials for FDA approval are the impact of a given agent in reducing acne lesion types (efficacy) versus vehicle/placebo, tolerability, and safety. Ultimately, success versus failure will be determined based on how a given agent establishes itself in real-world clinical practice and through phase 4 trials examining combination therapy.

Other Study Analysis Considerations

Additional considerations related to study analysis include understanding of *P* values, impact of study power, and interpretation of confidence intervals.[7] As stated by Bhardwaj et al[7]: "Dermatologists should not focus on *P* values alone to decide whether a treatment is clinically useful; it is essential to consider the magnitude of treatment differences and the power of the study." It is possible to demonstrate a statistically significant difference between 2 therapies based on *P* values alone where clinical significance is minimal or absent based on nominal evaluation. Method of analysis also is significant. For example, intent-to-treat analysis includes all enrolled patients in the denominator used for final efficacy determination regardless of study completion or the reason for not finishing the trial. This method often will serve to suppress the final efficacy result and is felt to accurately reflect the reality that clinicians will likely experience in real-world practice. The per protocol analysis method includes only those patients who completed the trial through study end point in the denominator for final efficacy determination. As a result, reported efficacy results may be inflated because of a lack of consideration of the reasons why some patients did not complete the trial.

Conclusion

Grading scales used to evaluate response to treatment in acne vulgaris trials continue to evolve. The goal is to achieve development of a convenient and accurate methodology that may be universally accepted and used by both investigators and clinicians. New methodologies for investigator global evaluation are static in nature and appear to be a better method of evaluation as compared with memory-dependent assessments that are compared with baseline. Definitions of success and failure used for global evaluation may be helpful in translating study results to a clinically useful picture of a patient's acne status; however, they are not as likely to correlate with what is realistically achievable over the course of a 12-week acne trial. Further work is needed in this area. Analysis of trial results is highly dependent on establishment of a clear frame of reference related to the parameters that are utilized. Without proper education regarding the differentiation between conventional and new methodologies used to evaluate therapies for acne vulgaris, there is potential for confusion among clinicians and investigators who may be less familiar with the nuances of study analysis.

REFERENCES

1. Lehmann HP, Robinson KA, Andrews JS, et al. Acne therapy: a methodologic review. *J Am Acad Dermatol.* 2002;47:231-240.
2. Executive Summary. Dermatologic and Ophthalmic Drugs Advisory Committee. Available at: http://www.fda.gov/OHRMS/DOCKETS/AC/02/transcripts/3904T1.htm. Accessed July 19, 2006.
3. Pochi PE, Shalita AR, Strauss JS, et al. Report of the consensus conference on acne classification. *J Am Acad Dermatol.* 1991;24:495-500.
4. O'Brien SC, Lewis JB, Cunliffe WJ. The Leeds revised acne grading system. *J Dermatol Treat.* 1998;9:215-220.
5. Thiboutot D, Pariser D, Egan N, et al. Adapalene gel 0.3% for the treatment of acne vulgaris: a multicenter, randomized, double-blind, controlled, phase III trial. *J Am Acad Dermatol.* 2006;54:242-250.
6. Pariser DM, Thiboutot DM, Clark SD, et al. The efficacy and safety of adapalene gel 0.3% in the treatment of acne vulgaris: a randomized, multicenter, investigator-blinded, controlled comparison study versus adapalene gel 0.1% and vehicle. *Cutis.* 2005;76:145-151.
7. Bhardwaj SS, Camacho F, Derrow A, et al. Statistical significance and clinical relevance: the importance of power in clinical trials in dermatology. *Arch Dermatol.* 2004;140:1520-1523.