Ulcerative Colitis: Clinical Trial Endpoints

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to http://www.regulations.gov. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact Dr. Anil Rajpal at 301-796-2120.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

August 2016
Clinical/Medical

Ulcerative Colitis: Clinical Trial Endpoints

Guidance for Industry

Additional copies are available from:

Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353; Email: druginfo@fda.hhs.gov
http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

August 2016
Clinical/Medical

TABLE OF CONTENTS

Ulcerative Colitis: Clinical Trial Endpoints Guidance for Industry[1]

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible for this guidance as listed on the title page.

I. INTRODUCTION

The purpose of this guidance is to assist sponsors in the clinical development of drugs for the treatment of ulcerative colitis (UC) in adult and pediatric patients.[2] Specifically, this guidance addresses the Food and Drug Administration's (FDA's) current thinking regarding efficacy endpoints for UC clinical trials. This draft guidance is intended to serve as a focus for continued discussions among the Division of Gastroenterology and Inborn Errors Products (DGIEP), pharmaceutical sponsors, the academic community, and the public.[3]

This guidance does not address the treatment or prevention of long-term complications of UC; for example, this guidance is not intended to discuss endpoints for prevention or reduction in risk of colorectal cancer.

This guidance also does not contain discussion of the general issues of statistical analysis or clinical trial design. Those topics are addressed in the ICH guidances for industry *E9 Statistical Principles for Clinical Trials* and *E10 Choice of Control Group and Related Issues in Clinical Trials*, respectively.[4]

[1] This guidance has been prepared by the Division of Gastroenterology and Inborn Errors Products in the Center for Drug Evaluation and Research at the Food and Drug Administration.

[2] For the purposes of this guidance, all references to *drugs* include both human drugs and therapeutic biological products unless otherwise specified.

[3] In addition to consulting guidances, sponsors are encouraged to contact the DGIEP to discuss specific issues that arise during the development of a given drug.

[4] We update guidances periodically. To make sure you have the most recent version of a guidance, check the FDA Drugs guidance Web page at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. BACKGROUND

A. Ulcerative Colitis

UC is a chronic, relapsing disease characterized by diffuse mucosal inflammation of the colon (Langan, Gotsch, et al. 2007). The precise etiology of UC is unknown; however, it is thought to be caused by an inappropriate inflammatory response to the gut contents in genetically predisposed individuals (Talley, Abreu, et al. 2011; Abraham and Cho 2009). The estimated U.S. incidence of UC is 9 to 12 cases per 100,000 persons per year, and the estimated prevalence is 205 to 240 cases per 100,000 persons (Danese and Fiocchi 2011). There is considerable variability in incidence and prevalence of UC around the world. Approximately 20 percent of patients with UC present before age 20 (Kelsen and Baldassano 2008). In the pediatric population (ages 0 to 17 years), the estimated U.S. incidence of UC is 3.2 cases per 100,000 persons per year, and the estimated prevalence is 19.5 cases per 100,000 persons (Abramson, Durant, et al. 2010).

UC involves the rectum and it may extend proximally in a contiguous pattern to affect part of the colon or the entire colon. Clinical manifestations of active disease include bloody diarrhea (with or without mucus), urgency, tenesmus, abdominal pain, weight loss, fever, and malaise. In patients with extensive or severe inflammation, acute complications such as severe bleeding and toxic megacolon, which can lead to perforation, may occur (Danese and Fiocchi 2011). There is an increased risk of colorectal cancer in UC patients compared to the general population; risk factors include long duration of disease, extensive colonic involvement, severe inflammation and epithelial dysplasia, and childhood-onset disease (Danese and Fiocchi 2011). The signs and symptoms of UC in adults and children are similar;[5] however, abdominal pain, disease involving the entire colon, extra-intestinal manifestations, proctitis (among girls), and disease severity necessitating colectomy are more common in children (Kelsen and Baldassano 2008; Malaty, Abraham, et al. 2013).

[5] A *sign* is defined as any objective visual or measured evidence of a disease, or health condition. Signs, which are observed or measured, are distinct from symptoms or clinical outcomes or treatment-related effect. Signs are usually observed and interpreted by the clinician but may be noticed and reported by the patient. A *symptom* is defined as any subjective evidence of a disease or health condition or treatment-related effect that can be noticed and known only by the patient. See the guidance for industry *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims* (patient-reported outcome guidance).

B. Treatment Goals

The short-term treatment goal of an active disease flare is to provide relief to the patient by decreasing the severity of and achieving resolution of the signs and symptoms of active disease. After this has been achieved, the long-term treatment goal is to decrease the frequency of subsequent disease flares. In both treatment phases (treatment of active disease flare and long-term treatment), a related goal of treatment is to affect the disease process itself (by decreasing the mucosal inflammation of the colon).

III. IDEAL PRIMARY EFFICACY ASSESSMENT TOOL

A *test*, *tool*, or *instrument* is defined as "an assessment system comprising three essential components: 1) materials for measurement; 2) an assay for obtaining the measurement; and 3) method and/or criteria for interpreting those measurements."[6]

There are three clinical outcome assessment types relevant to the measurement of UC signs and symptoms:[7]

(1) **Patient-reported outcome:** A measurement based on a report that comes directly from the patient (i.e., study subject) about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else. A patient-reported outcome can be measured by self-report or by interview provided that the interviewer records only the patient's response. Symptoms or other unobservable concepts known only to the patient can only be measured by patient-reported outcome measures. Patient-reported outcomes can also assess the patient perspective on functioning or activities that may also be observable by others.

(2) **Observer-reported outcome:** A measurement based on a report of observable signs, events, or behaviors related to a patient's health condition by someone other than the patient or a health professional. Generally, observer-reported outcomes are reported by a parent, caregiver, or someone who observes the patient in daily life and are particularly useful for patients who cannot report for themselves (e.g., infants or individuals who are cognitively impaired). An observer-reported outcome measure does not include medical judgment or interpretation.

(3) **Clinician-reported outcome:** A measurement based on a report that comes from a trained health care professional after observation of a patient's health condition. Most clinician-reported outcome measures involve a clinical judgment or interpretation of the

[6] FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Maryland: Food and Drug Administration (US)/National Institutes of Health (US); 2016 Jan 28 (http://www.ncbi.nlm.nih.gov/books/NBK338448/?report=reader).

[7] The source for all three clinical outcome assessment definitions is the BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet] (FDA-NIH Biomarker Working Group, 2016, Jan 28 (http://www.ncbi.nlm.nih.gov/books/NBK338448/?report=reader)).

observable signs, behaviors, or other manifestations related to a disease or condition. Clinician-reported outcome measures cannot directly assess symptoms that are known only to the patient.

FDA believes that the ideal primary efficacy assessment tool used in clinical trials to support marketing approval for the treatment of UC would consist of the following scales:[8]

- A signs and symptoms assessment scale, best measured by a patient-reported outcome instrument (or, in the case of young children (5 to 6 years old) or those unable to provide valid and reliable self-report, an observer-reported outcome instrument (Matza, Patrick, et al. 2013); and

- An endoscopic and histological assessment scale, best measured by a clinician-reported outcome instrument. A claim of "mucosal healing" would not be supported by endoscopy findings that provide only an assessment of the visual appearance of the mucosa; any claim related to findings on endoscopy, in the absence of a validated histological assessment of the mucosa, would be limited to the "endoscopic appearance of the mucosa."

Sponsors are encouraged to develop such primary efficacy assessment tools.[9]

The intent of this dual measurement is to ensure that an observed improvement in signs and symptoms is related to an effect on underlying inflammation, and vice versa. Onset of achievement of the targeted improvement in these two components would not have to occur concurrently. One can lag behind the other; however, the improvement in whichever component occurs first must be sustained through achievement of the remaining component.

Ideally, a signs and symptoms assessment scale should consist of a well-defined and reliable patient-reported outcome instrument or observer-reported outcome instrument that measures the clinically important signs and symptoms of UC, including stool frequency and rectal bleeding. Identification of relevant signs and symptoms should be confirmed by qualitative research with patients and caregivers. Instrument items should be clear and easily interpretable. Additional tools, such as the Bristol Stool scale (Lewis and Heaton 1997), may be useful for visual evaluation of the diarrhea and can be incorporated into the signs and symptoms assessment for adults; the pediatric version of the Bristol Stool scale should be considered for studies in children. Qualitative research with patients and caregivers, as appropriate, also should be conducted to document understanding of the instrument, including its instructions, items, and

[8] *Scale* is defined as the system of numbers or verbal anchors by which a value or score is derived for an item. Examples include visual analogue scales, Likert scales, and numeric rating scales (see the patient-reported outcome guidance).

[9] Because instrument development can be time and resource intensive, instrument developers, sponsors, and other interested parties should consider collaboration to develop publicly available clinical outcome assessments in a precompetitive manner using the process described in the guidance for industry and FDA staff *Qualification Process for Drug Development Tools*.

response options. Use of event logs and items with a short recall period generally are recommended when signs and symptoms of a condition vary frequently.

Although patient-reported outcome instruments should capture patient-reported signs and symptoms, an observer assessment of signs may be needed for children who are not able to reliably report their signs and symptoms (e.g., young children);[10] sponsors can consider a separate age-appropriate instrument with a score that captures both patient reports of symptoms and observer reports of signs for use in children who cannot self-report on the full range of signs and symptoms. Any assessment where an observer is asked to rate unobservable aspects of the child's condition (e.g., abdominal pain severity) should be avoided. Instead, observers should be asked to rate only signs and behaviors that are observable.

A well-defined and reliable clinician rating scale that measures the level of mucosal inflammation would be the ideal endoscopic and histological assessment tool. Endoscopy should be used in conjunction with histology for an assessment of mucosal healing. Endoscopy alone (without histology) only provides an assessment of the visual appearance of the mucosa. Ideally, clinicians performing the rating would have agreement on the definitions of the scale components, any procedures that are necessary when performing the rating (e.g., procedures for video recordings/equipment in the endoscopic assessment, and procedures for biopsy in the histological assessment), and any specific criteria to guide their judgments when selecting their response on the rating scale.

An instrument user manual is recommended for clinician-reported, patient-reported, and observer-reported outcome instruments to standardize instrument administration and provide any recommended procedures and definitions for use.

IV. INTERIM APPROACHES TO EFFICACY ASSESSMENTS

Until well-defined and reliable clinician-reported, patient-reported, and observer-reported outcome instruments become available for use in clinical trials, sponsors should consider the strategies discussed in the following sections when designing UC clinical trials. Traditionally, the Mayo Score and Ulcerative Colitis Disease Activity Index (UCDAI) have been the most commonly used tools to support registration trials in UC (see Appendixes 1 and 2). Both the Mayo Score and the UCDAI incorporate scoring of stool frequency, rectal bleeding, endoscopic findings, and the physician's assessment of disease activity.

The Mayo Score and UCDAI share limitations. A key limitation of both the UCDAI and the Mayo Score is the physician's assessment of disease activity and the Physician's Global Assessment (PGA) subscores, respectively.[11] A single general item cannot adequately capture

[10] See Matza, Patrick, et al. 2013.

[11] The PGA is a clinician-reported assessment that reflects "the patient's recorded symptoms, the proctoscopic appearance of the rectosigmoid mucosa, and other pertinent clinical indexes, such as physical findings and the patient's performance status" (Schroeder, Tremaine, et al. 1987).

whether benefit is achieved in all, or only some, of the important signs and symptoms. Additionally, as previously discussed, a signs and symptoms assessment scale is best measured by a patient-reported outcome instrument as opposed to a clinician-reported outcome instrument. Use of the PGA subscore/physician's assessment of disease activity subscore or the full Mayo Score or the full UCDAI (which incorporate the PGA subscore/physician's assessment of disease activity subscore) are not recommended as endpoint measures to support a marketing application. Although we acknowledge the limitations of the Stool Frequency and Rectal Bleeding subscores in these tools,[12,13] we recommend their use in addition to the Endoscopy subscore as endpoint measures for clinical trials until well-defined and reliable endpoint measures become available.

Data collection procedures for the assessment of the Rectal Bleeding, Stool Frequency, and Endoscopy subscores should be submitted for review and concurrence by the DGIEP. The goal is to ensure that the following issues are addressed: (1) the patient-reported assessments of stool frequency and rectal bleeding are standardized across patients; (2) the effect of inter-observer variability on the endoscopic assessment is minimized; and (3) the limitations from the PGA subscore is removed (see Table 1). Strategies to improve the reliability and precision of the components are discussed below. The following modifications recommended for the Mayo Score also should be applied to the UCDAI.

A. Stool Frequency and Rectal Bleeding Subscores

Patients should be provided standardized instructions for recording the number of stools and their worst rectal bleeding over a 24-hour period; investigational sites should be provided instructions for calculation of the Stool Frequency and Rectal Bleeding subscores.

1. Standardized Instructions for Recording Number of Stools and Worst Rectal Bleeding

An example of standardized instructions (for the Mayo Score) is summarized in Table 1. However, we are open to considering alternative approaches as long as they are well-defined.[14] Standardized instructions should be drafted early in drug development and shared with FDA for comment.

[12] The Stool Frequency subscore may not be well-defined primarily because it relies on a reference "normal" number of stools per day (based on the number of stools when the patient is not experiencing a flare) rather than the reported absolute number of stools per day.

[13] The Rectal Bleeding subscore may not be well-defined primarily because it requires patients to report the answer to a *double-barreled* question (it asks patients to choose streaks of blood with stool less than half the time versus obvious blood with stool most of the time) (Higgins 2012).

[14] Because instrument development can be time and resource intensive, instrument developers, sponsors, and other interested parties should consider collaboration to develop publicly available clinical outcome assessments in a precompetitive manner using the process described in the guidance for industry and FDA staff *Qualification Process for Drug Development Tools*.

Table 1. Example of Standardized Instructions for Recording Number of Stools and Worst Rectal Bleeding (for the Mayo Score) (Each Over a 24-Hour Period)

Category of Instructions	Specific Instructions to Patients
Definition of *Stool*	• Patients should be instructed that a stool is defined as a trip to the toilet when the patient has either a bowel movement, or passes blood alone, blood and mucus, or mucus only
Reference Remission Stool Frequency (over 24 hours)	• The patient should be asked to identify at the screening visit how many stools he or she had in a 24-hour period when in remission from UC • If the patient does not report that he or she has achieved remission, then the patient should be asked to identify the number of stools he or she had per day before initial onset of signs and symptoms of UC – Sponsors should record if the reference remission stool frequency is based on reported stool frequency when the patient was in remission or reported stool frequency before initial onset of signs and symptoms of UC. – Both the remission and pre-UC stool frequency should be collected at baseline. This allows exploration of the natural history of pre-diagnosis stool frequency versus remission stool frequency.
Most Severe Category of Rectal Bleeding (in a given 24-hour period)	• Patients should be instructed to indicate the most severe category that describes the amount of blood they had in their stools for a given day • Categories of rectal bleeding should be defined as follows: – No blood seen – Streaks of blood with stool less than half the time – Obvious blood (more than just streaks) or streaks of blood with stool most of the time – Blood alone passed • Patients should be instructed to select "No Blood Seen" in the rectal bleeding section if they do not have stool during a given day
Completion of Event Log or Diary*	• Patients should be trained on the completion of the event log or diary • The instructions for completion of the stool frequency and rectal bleeding assessments should be incorporated into the event log or diary for ready reference by the patient
Recording of Rectal Bleeding and Stool Frequency Assessments	• Patients should be directed to capture their rectal bleeding and stool frequency assessments in event logs or daily diaries* for 1 week before each visit

* Sponsors are encouraged to propose an electronic data collection method (e.g., voice response system, electronic diary, or Web-based system) as an alternative to pen and paper data collection. If an electronic data collection method is proposed, sponsors should provide instructions for training in electronic methods.

2. Calculation of Stool Frequency and Rectal Bleeding Subscores

Generally, sponsors have used an approach of capturing stool frequency and rectal bleeding data from the most recent 3-day consecutive period within the week before the visit to calculate the Stool Frequency and Rectal Bleeding subscores. Either the average or the worst of the most recent 3-day consecutive period has been used for the calculation of the respective subscore. Missing data may be an issue if a consecutive 3-day period is not available (see section VII.A., Missing Data). Alternative proposals can be considered such as the collection of data from at least 3 days (including nonconsecutive days) in the week before the visit. Instructions to

investigators for calculation of Stool Frequency and Rectal Bleeding subscores should be drafted early in drug development and shared with FDA for comment.

B. Endoscopic Assessment

Endoscopic assessment should be documented by the endoscopist performing the procedure, and ideally by blinded central readers reviewing video recordings of the procedure. Charters that standardize procedures, video recordings/equipment, and assessment of endoscopy should be drafted early in drug development and shared with FDA for comment. The protocol should specify how discrepancies between the assessment of the endoscopist and the central reader will be handled in the efficacy analyses. Particularly important is the interpretation and methodology underlying the scoring of endoscopic characteristics that may have subjective elements (e.g., friability).

The Endoscopy subscore of the Mayo Score should be modified so that a value of 1 does not include friability. This is because the presence of friability (even if considered to be mild by the endoscopist/central reader) is not consistent with *clinical remission* (the recommended definition of this endpoint includes a Mayo Endoscopy subscore of 0 or 1) (see section V.A., Clinical Remission and Clinical Response, and Appendix 1).

C. PGA Assessment/Physician's Assessment of Disease Activity

The utility of the PGA subscore (in the full Mayo Score)/Physician's Assessment of Disease Activity subscore (in the full UCDAI) is questionable because the concept it purports to measure that is distinct from the other components is not clear. Use of these subscores as part of the Mayo or UCDAI and the resulting full scores that incorporate them are not recommended as an endpoint measure to support a marketing application. Possible use of these or modified global scores as an exploratory endpoint (a physician's global) could be considered.

V. EFFICACY ENDPOINT DEFINITIONS

This section focuses on endpoint definitions for phase 3 clinical trials, and the associated claims that would be supported.

Trial design issues such as the assessment of disease severity (on entry) and the *induction* and *maintenance* paradigm (used in prior UC drug approvals),[15] and possible alternatives to this paradigm are beyond the scope of this guidance.

[15] The paradigm of induction and maintenance trials has been used for prior UC drug approvals (i.e., induction trials are conducted to demonstrate efficacy after a short duration such as 1 to 3 months; and a maintenance trial is conducted to demonstrate durability of treatment effect over a longer period, such as 6 months or 1 year).

A. Clinical Remission and Clinical Response

Clinical remission (responder definition based on an absolute total Mayo Score and absolute Mayo endoscopy, stool frequency, rectal bleeding, and PGA subscores) and *clinical response* (responder definition based on a reduction in total Mayo Score and reduction in rectal bleeding subscore) have been accepted as primary endpoints in clinical trials that have supported prior approvals of treatments of UC.[16] We currently recommend a primary endpoint of clinical remission (responder definition based on Stool Frequency, Rectal Bleeding, and Endoscopy scores) (see section IV., Interim Approaches to Efficacy Assessments). Until a valid patient-reported outcome instrument for UC signs and symptoms and a valid clinician rating scale for mucosal inflammation in UC become available, a modified Mayo or modified UCDAI score omitting the physician's global or disease activity ratings, as described in section IV, can be used as an endpoint measure.

The following definition of clinical remission is recommended:

- Stool Frequency subscore = 0
- Rectal Bleeding subscore = 0
- Endoscopy subscore = 0 or 1 (modified) on Mayo Score; or 0 on UCDAI

An alternative to a Stool Frequency subscore of 0 in the definition of clinical remission is the following:

> At least one point decrease in Stool Frequency subscore from Baseline (start of trial) to achieve a Stool Frequency subscore = 0 or 1

Inclusion of a maximum score of 1 for the Stool Frequency subscore in the definition of clinical remission would not support a labeling claim that includes normalization of stool frequency.

Sponsors should discuss their proposed endpoints for phase 3 trials at the end-of-phase 2 meeting. See also section VII., Statistical Considerations. Sponsors should begin discussions of endpoint measures with FDA as early as possible during drug development.

B. Secondary Endpoints of Interest

Sponsors can propose to evaluate changes between the treatment arms of each of the subscores (Stool Frequency, Rectal Bleeding, and Endoscopy) and/or the total score (i.e., sum of the Stool Frequency, Rectal Bleeding, and Endoscopy subscores).

C. Assessment of Mucosal Healing vs. Endoscopic Appearance of the Mucosa

Mucosal healing (based on the Mayo Endoscopy subscore) has been included as a secondary endpoint in many clinical trials. In many clinical trials, mucosal healing has been defined as

[16] See the Glossary of Historical Endpoints for the definitions of *clinical remission* and *clinical response*.

Mayo Endoscopic subscore of less than or equal to 1 point. However, a claim of mucosal healing would not be supported through endoscopy that provides only an assessment of the visual appearance of the mucosa. Any claim related to findings on endoscopy, in the absence of validated histological assessment of the mucosa, would be limited to the endoscopic appearance of the mucosa.

There are currently limitations of histologic scoring systems and of community standards for definitions of histologic improvement; thus, there are currently no criteria for histological assessment of mucosal healing. Sponsors intending to pursue a claim of mucosal healing should discuss their proposed development plan with FDA. Histologic evaluations of biopsy specimens, when obtained, should be conducted centrally. Charters that standardize biopsy and histology procedures and assessments (e.g., scoring for histology) should be drafted early in drug development and shared with FDA for comment. Methods to address discrepancies in assessments between site and central readers should be prespecified within the protocols. Grading scales and scoring techniques should be discussed with FDA. Use of measures that are not validated are unlikely to support labeling claims.

Sponsors should evaluate histological effects in phase 2 trials to inform how to effectively incorporate these assessments in phase 3 trials.

D. Corticosteroid-Free Remission

Corticosteroid-free remission (often defined as clinical remission without concomitant corticosteroids at a particular time point in patients who were using corticosteroids at baseline) has been included as a secondary endpoint in many clinical trials.[17] To ensure clinical meaningfulness of this endpoint, its definition should be based on a minimum duration of time over which a patient is considered to be both corticosteroid-free and in clinical remission; adequate justification should be provided for the proposed minimum duration.

VI. TRIAL PROCEDURES AND TIMING OF ASSESSMENTS

At screening, it is important that all patients undergo endoscopy with biopsy to obtain histological confirmation of disease activity. Sponsors should propose a maximal time between conduct of the screening endoscopy and entry into the trial. Sponsors seeking a claim of mucosal healing should discuss with FDA their proposal for histological assessments of biopsy specimens obtained during endoscopy at both the screening and end-of-treatment visits. Table 2 summarizes the key assessments throughout the trial.

[17] See the Glossary of Historical Endpoints for the definition of *corticosteroid-free remission*.

Table 2. Key Assessments During Screening, On-Treatment, and End-of-Treatment Visits

Visit(s)	Key Assessments
Screening	• The following subscores of the Mayo Score:* – Stool Frequency and Rectal Bleeding subscores (calculated from patient diary cards§) – Endoscopy subscore • Histological assessment – Disease activity should be confirmed in all patients through histological examination of endoscopically obtained biopsy specimens – Sponsors seeking a claim for mucosal healing should discuss their proposal for histological assessments of biopsy specimens obtained during endoscopy
On-Treatment‡	• The following subscores of the Mayo Score: – Stool Frequency and Rectal Bleeding subscores (calculated from patient diary cards§) • Pharmacokinetic samples • Anti-drug antibody samples (for biologic products)
End-of-Treatment	• The following subscores of the Mayo Score:* – Stool Frequency and Rectal Bleeding subscores (calculated from patient diary cards§) – Endoscopy subscore • Ideally histological assessment should be obtained# – Sponsors seeking a claim for mucosal healing should discuss their proposal for histological assessments of biopsy specimens obtained during endoscopy

* See section IV., Interim Approaches to Efficacy Assessments, for further information.

§ Sponsors are encouraged to propose an electronic data collection method (e.g., voice response system, electronic diary, or Web-based system) as an alternative to the paper diary card.

‡ Pharmacokinetic samples and anti-drug antibody samples (for biologic products) also should be obtained during on-treatment visits.

See also section V.C., Assessment of Mucosal Healing vs. Endoscopic Appearance of the Mucosa, regarding limitations of histologic scoring systems and of community standards (for definitions of histologic improvement).

The Stool Frequency, Rectal Bleeding, and Endoscopy subscores of the Mayo score should be obtained at screening and end-of-treatment visits. The Stool Frequency and Rectal Bleeding subscores of the Mayo Score should also be obtained at on-treatment visits (see section V.B., Secondary Endpoints of Interest).

Appropriate assays for pharmacokinetic and anti-drug antibody assessments should be used; there should be a prospective plan for analyses of these data.

Sponsors are encouraged to prospectively collect data for non-endoscopic markers of inflammation (such as C-reactive protein level, fecal calprotectin and lactoferrin, or other putative biomarkers) throughout the trial; however, these markers may not support labeling claims. Sponsors are encouraged to investigate the use of proposed biomarker(s) for identifying patients with an increased likelihood of achieving clinical benefit. One approach is to characterize the biomarker(s) (i.e., to select the assay and cutoff values that are associated with clinical benefit) in early phase trials, and to validate the biomarker(s) in the phase 3 trials; sponsors should discuss their proposed approach with FDA. Sponsors are also encouraged to consult the Center for Devices and Radiological Health through the presubmission process for

development and validation of the assay(s) for the proposed biomarker(s) selected for study in phase 3 pivotal trials.[18]

VII. STATISTICAL CONSIDERATIONS

A. Missing Data

Sponsors should propose methods for handling missing data (e.g., one or more missing Mayo subscores). Different imputation methods should be considered before the study and the properties analyzed by simulation or other methods, perhaps leading to a conclusion that the patient was a nonresponder in some cases or absent in others. One possible method should be that if a patient is missing one or more Mayo subscores at the visit where the primary endpoint is being assessed, the patient would be considered a treatment failure, although this could also depend on the reason for discontinuation (see also section IV.A.2., Calculation of Stool Frequency and Rectal Bleeding Subscores).

Because there are a number of reasons why a patient may not have a protocol-mandated endoscopy performed, such assessments could introduce bias if there is nonrandomness of the missing data. With regard to analyses of endoscopic findings, the protocol should prespecify how the following patients will be handled with respect to defining endoscopy responders or nonresponders: patients who drop out,[19] are lost to follow-up, discontinue investigational drug, refuse endoscopy, or are otherwise deemed to be treatment failures. If there are imbalances between study arms in the proportion of patients who meet such criteria, it is likely that the endoscopic data will be biased and may result in invalid estimates of treatment benefit. Although this is not likely to be a significant problem in short-term trials, these concerns may limit the ability of trials of longer duration to provide meaningful endoscopic assessment data (see also section V.C., Assessment of Mucosal Healing vs. Endoscopic Appearance of the Mucosa).

B. Primary Analysis

We recommend that analyses of binary responses (e.g., clinical remission status) be based on a stratified Cochran-Mantel-Haenszel test of a difference in proportions. The primary analysis should be adjusted by the stratification factors used for randomization.

C. Secondary Analyses

Sponsors should calculate the change from baseline in Stool Frequency, Rectal Bleeding, and Endoscopy subscores, and/or total score and compare these changes between the treatment arms (see section V.B., Secondary Endpoints of Interest) as secondary analyses.

---

[18] See the guidance for industry and Food and Drug Administration staff *Requests for Feedback on Medical Device Submissions: The Pre-Submission Program and Meetings with Food and Drug Administration Staff*.

[19] The protocol should specify that patients who drop out of the trial will obtain an endoscopy upon exiting the trial.

VIII. PEDIATRIC DRUG DEVELOPMENT CONSIDERATIONS

A. Addressing Pediatric Research Equity Act Requirements

To comply with the Pediatric Research Equity Act (PREA) (21 U.S.C. 355c),[20] a new drug application or biologics license application must contain data that are adequate to assess the safety and effectiveness of the drug for the claimed indications to support dosing and administration in all relevant pediatric subpopulations unless the requirement is waived or deferred. In addition, an age-appropriate formulation must be developed for all relevant pediatric subpopulations. Sponsors should begin discussions about their pediatric formulation and clinical development plan early in development because they are required to submit pediatric study plans no later than 60 days after an end-of-phase 2 meeting. We recommend sponsors refer to PREA as amended by the Food and Drug Administration Safety and Innovation Act (FDASIA).[21,22]

The DGIEP has waived PREA requirements for pediatric studies of UC drugs in children younger than 2 years of age because studies would be impossible or highly impracticable in that age group. However, any request for waiver of studies in any specific pediatric age groups (e.g., infants and young children) must include appropriate justification(s).[23] Sponsors should consider enrollment of children of all ages affected by UC into pediatric study(ies).

Pediatric studies should begin as soon as there are adequate data to support safety and expected benefit in pediatric patients with UC. In some cases, a pediatric subpopulation (e.g., adolescents) can be enrolled in phase 3 trials for adults provided nonclinical studies do not raise specific safety concerns for this pediatric subpopulation.

All enrolled pediatric patients should have a documented diagnosis of UC that includes confirmation by both endoscopic and histological assessments. In addition, the presence of active inflammation at the time of entry into the clinical trial should be confirmed by visualization of the colon during the screening period. Sponsors should propose a maximal time between conduct of the screening endoscopy and entry into the trial.

---

[20] See section 505B(a)(2) of the Federal Food, Drug, and Cosmetic Act (FD&C Act) (pediatric assessments "shall contain data, gathered using appropriate formulations for each age group for which the assessment is required that are adequate (i) to assess the safety and effectiveness of the drug or the biological product for the claimed indications in all relevant pediatric subpopulation; and (ii) to support dosing and administration for each pediatric subpopulation for which the drug or the biological product is safe and effective)."

[21] See PREA (Public Law 108-155; section 505B of the FD&C Act; 21 U.S.C. 355B) as amended by FDASIA (Public Law 112-144).

[22] See the draft guidance for industry *Pediatric Study Plans: Content of and Process for Submitting Initial Pediatric Study Plans and Amended Pediatric Study Plans*. When final, this guidance will represent the FDA's current thinking on this topic.

[23] See section 505B(a)(4)(A) of the FD&C Act.

B. Extrapolation of Efficacy

Based on sufficient similarities in the course of the disease and the effects of the drug in adults and pediatric patients with UC, FDA has concluded that partial extrapolation of efficacy is acceptable from adequate and well-controlled studies in adults for a systemically active drug if there are sufficient exposure-response data from adult and pediatric studies of that drug that support this approach, and the exposure-response relationships between the populations are similar.[24,25,26] It should be noted that because there is no established biomarker that can be relied upon to establish an exposure-response relationship in UC, trials designed with endpoints directly measuring clinical benefit (i.e., rectal bleeding, stool frequency, and endoscopy) should be used in pediatric studies to define the exposure-response relationship. It is important to adequately design the pediatric clinical trial to assess the dose-exposure-response relationship to ensure adequate dose selection in pediatrics. In addition, pharmacokinetic and efficacy comparison of pediatrics to adults should be used to support pediatric dose selection. FDA encourages the use of modeling and simulation to design pediatric trials by leveraging knowledge from the adult clinical program. Furthermore, the use of innovative model-based approaches is encouraged for analyzing the pediatric dose-exposure-response data.

C. Data Collection in Pediatric Patients

Depending on the age of the pediatric patient, observer-based reporting (e.g., reporting from a parent or caregiver) may be needed for the assessment of signs and symptoms. Sponsors should discuss their proposal to address age-specific data collection issues with FDA.

D. Other Pediatric Considerations

Growth parameter(s), including weight and height measured by a calibrated stadiometer (appropriate for patients 2 years of age and older), should be included as secondary endpoint(s) in pediatric trials.

[24] 21 CFR 314.55 and 21 CFR 601.27 state: "Where the course of the disease and the effects of the drug are sufficiently similar in adults and pediatric patients, FDA may conclude that pediatric effectiveness can be extrapolated from adequate and well-controlled studies in adults usually supplemented with other information obtained in pediatric patients, such as pharmacokinetic studies."

[25] See the Pediatric Study Decision Tree and discussion of extrapolation of efficacy in Dunne, Rodriguez, et al. 2011.

[26] This topic was discussed in a Gastrointestinal Drugs Advisory Committee meeting for an open label trial of a drug in pediatric patients that used the Mayo Score for its primary endpoint measure. See the meeting materials available at http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/GastrointestinalDrugsAdvisoryCommittee/ucm235600.htm.

GLOSSARY OF HISTORICAL ENDPOINTS[27]

Clinical remission: Mayo Score of less than or equal to 2 with no individual subscore greater than 1.

Clinical response: Reduction in Mayo Score of greater than or equal to 3 points and greater than or equal to 30 percent from baseline with an accompanying decrease in rectal bleeding subscore of greater than or equal to 1 point or absolute rectal bleeding subscore of less than or equal to 1 point.

Corticosteroid-free remission: Clinical remission in patients using oral corticosteroids at baseline who have discontinued corticosteroids and are in clinical remission at the end of the study.

[27] As defined in many previous clinical trials. See section V., Efficacy Endpoint Definitions, for current recommendations.

REFERENCES

Literature

Abraham C and Cho JH, 2009, Inflammatory Bowel Disease, *N Engl J Med.*, 361(21):2066-78.

Abramson O, Durant M, Mow W, et al., 2010, Incidence, Prevalence, and Time Trends of Pediatric Inflammatory Bowel Disease in Northern California, 1996 to 2006, *J Pediatr.*, 157:233.e1–239.e1.

Danese S and Fiocchi C, 2011, Ulcerative Colitis, *N Engl J Med.*, 365(18):1713-1725.

D'Haens G, Sandborn WJ, Feagan BG, Karel G, et al., 2007, A Review of Activity Indices and Efficacy End Points for Clinical Trials of Medical Therapy in Adults with Ulcerative Colitis, *Gastroenterology*, 132 (2):763-786.

Dunne J, Rodriguez WJ, et al., 2011, Extrapolation of Adult Data and Other Data in Pediatric Drug-Development Programs, *Pediatrics*, 128:e1242-1249.

Kelsen J and Baldassano RN, 2008, Inflammatory Bowel Disease: The Difference Between Children and Adults, *Inflamm Bowel Dis.*, 14 (Suppl. 2), S9-S11 (2008).

Kornbluth A, Sachar DB, and the Practice Parameters Committee of the ACG, 2010, Ulcerative Colitis Practice Guidelines in Adults: American College of Gastroenterology, Practice Parameters Committee, *Amer J Gastroenterol.*, 105:501-523.

Kugathasan S, Judd RH, et al., 2003, Epidemiologic and Clinical Characteristics of Children with Newly Diagnosed Inflammatory Bowel Disease in Wisconsin: A Statewide Population-Based Study, *J Pediatr.*, 143:525-531.

Langan RC, Gotsch PB, Krafczyk MA, and Skillinge DD, 2007, Ulcerative Colitis: Diagnosis and Treatment, *American Family Physician*, 76:1323-30.

Lewis JD, Chuai S, Nessel L, et al., 2008, Use of the Non-Invasive Components of the Mayo Score to Assess Clinical Response in Ulcerative Colitis, *Inflamm Bowel Dis.*, 14 (12):1660-1666.

Lewis, SJ and Heaton KW, 1997, Stool Form Scale as a Useful Guide to Intestinal Transit Time, *Scand. J. Gastroenterol*, 32(9):920-4.

Malaty HM, Abraham BP, Mehta S, et al., 2013, The Natural History of Ulcerative Colitis in a Pediatric Population: A Follow-Up Population-Based Cohort Study, *Clinical and Experimental Gastroenterology*, 6:77-83.

Matza LS, Patrick DL, Riley AW, et al., 2013, Pediatric Patient-Reported Outcome Instruments for Research to Support Medical Product Labeling, *Value Health*, 16 (4):461–479.

Schroeder KW, Tremaine WJ, and Ilstrup DM, 1987, Coated Oral 5-Aminosalicylic Acid Therapy for Mildly to Moderately Active Ulcerative Colitis, *N Engl J Med*., 317(26):1625-9.

Sutherland LR, Martin F, Greer S, et al., 1987, 5-Aminosalicylic Acid Enema in the Treatment of Distal Ulcerative Colitis, Proctosigmoiditis, and Proctitis, *Gastroenterology*, 92:1894-8.

Talley NJ, Abreu MT, Achkar JP, et al., 2011, An Evidence-Based Systematic Review on Medical Therapies for Inflammatory Bowel Disease, *Am J Gastroenterol*., 106:S2–S25.

Presentations and Advisory Committee Meetings

Higgins, PDR, 2012, "Measuring Treatment Benefit in UC: Current Standard and Alternatives," Gastroenterology Regulatory Endpoints and the Advancement of Therapeutics (GREAT) Workshop, September 21. (Transcript available at http://www.regulations.gov/#!documentDetail;D=FDA-2012-N-0001-0093.)

Gastroenterology Regulatory Endpoints and the Advancement of Therapeutics (GREAT) Workshop, September 21, 2012. (Transcript available at http://www.regulations.gov/#!documentDetail;D=FDA-2012-N-0001-0093.)

Gastroenterology Regulatory Endpoints and the Advancement of Therapeutics 2 (GREAT 2) Workshop, October 22, 2013.

Guidances[28]

Draft guidance for industry *Pediatric Study Plans: Content of and Process for Submitting Initial Pediatric Study Plans and Amended Pediatric Study Plans*

Guidance for industry *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claim*

Guidance for industry *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products*

Guidance for industry and FDA staff *Qualification Process for Drug Development Tools*

Guidance for industry and Food and Drug Administration staff *Requests for Feedback on Medical Device Submissions: The Pre-Submission Program and Meetings with Food and Drug Administration Staff*

[28] Guidances can be found on the FDA Drugs guidance Web page at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

APPENDIX 1: MAYO SCORE

The following is taken from Schroeder, Tremaine, et al. 1987.

Table A. Mayo Score

Stool frequency*
- 0 = Normal no. of stools for this patient
- 1 = 1–2 stools more than normal
- 2 = 3–4 stools more than normal
- 3 = 5 or more stools more than normal

Rectal bleeding†
- 0 = No blood seen
- 1 = Streaks of blood with stool less than half the time
- 2 = Obvious blood with stool most of the time
- 3 = Blood alone passed

Findings of flexible proctosigmoidoscopy
- 0 = Normal or inactive disease
- 1 = Mild disease (erythema, decreased vascular pattern, mild friability)
- 2 = Moderate disease (marked erythema, absent vascular pattern, friability, erosions)
- 3 = Severe disease (spontaneous bleeding, ulceration)

Physician's global assessment‡
- 0 = Normal
- 1 = Mild disease
- 2 = Moderate disease
- 3 = Severe disease

---

*Each patient served as his or her own control to establish the degree of abnormality of the stool frequency.

†The daily bleeding score represented the most severe bleeding of the day.

‡The physician's global assessment acknowledged the three other criteria, the patient's daily record of abdominal discomfort and general sense of well-being, and other observations, such as physical findings and the patient's performance status.

APPENDIX 2: UCDAI

The following is taken from Sutherland, Martin, et al. 1987.

Table B. UCDAI

1. Stool frequency
   - 0 = Normal
   - 1 = 1–2 Stools/day >normal
   - 2 = 3–4 Stools/day >normal
   - 3 = >4 Stools/day >normal
2. Rectal bleeding
   - 0 = None
   - 1 = Streaks of blood
   - 2 = Obvious blood
   - 3 = Mostly blood
3. Mucosal appearance
   - 0 = Normal
   - 1 = Mild friability
   - 2 = Moderate friability
   - 3 = Exudation, spontaneous bleeding
4. Physician's rating of disease activity
   - 0 = Normal
   - 1 = Mild
   - 2 = Moderate
   - 3 = Severe

Maximum score = 12