Voting with Their Feet (VWF) Endpoint: A Meta-Analysis of an Alternative Endpoint in Clinical Trials, Using 5-ASA Induction Studies in Ulcerative Colitis

Sujal C. Rangwalla, DO,* Akbar K. Waljee, MD,† and Peter D.R. Higgins, MD, PhD, MSc†

Background: Strict clinical remission endpoints in ulcerative colitis (UC) trials produce low remission rates and do not reflect the good outcomes of UC therapy. We proposed the use of the VWF (Voting With their Feet) endpoint, the percentage of subjects leaving a randomized controlled trial (RCT) arm for lack of efficacy. The aims were 1) to determine if the VWF endpoint can be extracted from 5-aminosalicylate (5-ASA) RCTs in UC; 2) to perform meta-analyses of VWF and clinical remission (CR) endpoints; and 3) to determine the statistical power of the VWF endpoint.

Methods: Fixed effects meta-analysis and power calculations were used.

Results: In 5 studies, including 1048 subjects, 9.5% of patients left 5-ASA study arms for lack of efficacy, versus 28.3% leaving placebo. The rate of failure to achieve CR was 68.2% with 5-ASA, versus 86.9% with placebo. The relative risk (RR) of treatment failure for 5-ASA using the VWF endpoint was 0.33 (95% confidence interval [CI] 0.24–0.44), which was significantly smaller than with the CR endpoint (RR 0.81, 95% CI 0.76–0.88). The statistical power of VWF was slightly greater than CR.

Conclusions: VWF is inexpensive, intuitive, and has similar statistical power to CR. The VWF endpoint can confirm the validity of outcome measures in clinical trials, and estimate real-world clinical efficacy.

(Inflamm Bowel Dis 2009;15:422–428)

Key Words: disease activity index, inflammatory bowel disease, endpoints, clinical trials, ulcerative colitis

Ulcerative colitis (UC) is a chronic disease that is commonly diagnosed in the second or third decade of life and affects most patients for 50 years or more.1,2 The chronic inflammation and relapsing course of UC require maintenance therapy that can suppress inflammation effectively over many years, and thereby reduce the risk of both colectomy3 and colon cancer.4 However, studies of clinical trial endpoints in UC suggest that the most statistical power can be obtained in short clinical trials with very strict definitions of clinical success, which maximize the differences between treatment and placebo arms.5

While the current approach of short clinical trials with physician-centered endpoints and strict definitions of remission lowers the placebo remission rate, the sample size, and the cost of trials, the results do not reflect the success rates in clinical practice. In fact, a majority of patients with UC initiated on 5-aminosalicylate (5-ASA) medications will obtain remission and remain on the medication. A large pharmacy database study6 by Kane and Shaya found that 55% of patients with UC and new 5-ASA prescriptions will continue to reliably refill their 5-ASA through 1 year. However, because of strict clinical endpoints randomized, placebo-controlled clinical trial data suggest that less than one-third of patients will obtain remission with 5-ASA medications.7

This disparity is problematic, as it is difficult to predict real-world clinical efficacy from the results of randomized controlled trials (RCTs). A similar disparity is seen in Crohn’s disease, where the 54-week remission rate in RCTs of infliximab is 39%–45%,8 but the maintenance of remission (MOR) rate in clinical practice at 30 months is 69%.9 It would be very helpful to have a clinical trial endpoint that would more accurately reflect clinical practice and maintenance of clinical benefit.

Ideally, this endpoint would: 1) be readily obtained during clinical trials; 2) be useful in long maintenance studies, not just short induction studies; 3) have a falling placebo rate with time; 4) allow comparisons and meta-analysis across trials; 5) reflect clinical practice outcomes; and 6) provide reasonable statistical power.

In this article we propose and test a new clinical trial endpoint, the VWF (Voting With their Feet) endpoint. This
endpoint is defined as the percentage of subjects in each arm of a clinical trial who drop out due to lack of efficacy of their blinded therapy. In theory, subjects should drop out more frequently from the treatment arm that is less efficacious. We aim to test whether: 1) the VWF endpoint can be extracted retrospectively from RCTs of 5-ASA medications; 2) the VWF endpoint outcomes parallel the clinical response endpoints in direction and magnitude of drug effect; 3) the VWF endpoint more accurately reflects the good results seen in clinical practice with 5-ASAs in UC than standard clinical endpoints; and 4) whether the VWF endpoint would provide adequate statistical power for clinical trials in comparison with standard clinical trial endpoints.

MATERIALS AND METHODS

Literature Search
A literature search was performed using the Medline and the Cochrane databases from January 1, 1966, through December 31, 2007. The search used the terms: (“aminosalicylic acid” OR “5-ASA” OR “sulfasalazine” “mesalazine” OR “olsalazine” OR “asacol” OR “pentasa” OR “salofalk” OR “rowasa” OR “asamax” OR “canasa” OR “SPD476” OR “lialda” OR “mesavant” OR “mesasa” OR “claversal” OR “dipentum” OR “colazide” OR “colazal” OR “balsalazine”) AND “ulcerative colitis” AND “placebo-controlled.” Limits were placed to restrict the search to human studies, English language studies, and RCTs. Hand-searching the bibliographies of individual articles led to the identification of additional relevant studies.

Study Selection
Two investigators (S.R. and A.W.) independently performed the computerized bibliographic search and independently reviewed the titles, abstracts, contents, and references of the articles. Once the relevant articles were identified, selection criteria were applied. The inclusion criteria included: human studies; placebo-controlled RCTs; studies of induction of remission by 5-ASAs; included blinding and contained a placebo arm; published as a full article; reported relevant data including how many patients were in each arm of the study, what number of subjects left each arm due to lack of medication efficacy, and clearly defined primary trial endpoints; studies using at least full doses (≥2 g/d) of 5-ASAs (as a wide range of doses might produce statistical heterogeneity); and data not duplicated in other articles. In addition, articles studying olsalazine, in which a prominent side effect of the drug (diarrhea) could not readily be differentiated from lack of treatment effect, were excluded.

Data Extraction
Each article was critically evaluated and the following data were extracted: the medication used for the study; year of the study; duration of the usage of the drug in the trial; dose used in the trial (data was collected for all study arms using at least 2 g/d of 5-ASA); the number of patients in each arm of the study; the results of the primary endpoints of the trial; the number or percentage of patients who achieved the primary clinical trial endpoint using the intention-to-treat population; and the number of patients in the placebo and the experimental group that dropped out of the study due to lack of efficacy of the drug or nonimprovement of UC symptoms.

There was greater than 95% agreement between the 2 data extractors and discrepancies were resolved by consensus. Discordant data were also discussed with the senior author (P.D.R.H.), who reviewed the consensus decisions about these data points.

Statistical Analysis
Meta-analysis was performed using Stata 10.1 (Stata-Corp, Galveston, TX). Meta-analysis was performed using both fixed and random effects models. Tests for heterogeneity were performed. The presence of publication bias was evaluated using a funnel plot. An omission plot was also performed to determine whether there were any studies that were significant outliers.

Sample Size Analysis
The sample size required for a clinical trial comparing 5-ASA to placebo was calculated with the sampsi command in Stata, using both the clinical trial endpoint success rates and the VWF success rates from the clinical trials included in the meta-analysis. The sample sizes were compared to determine whether the VWF endpoint would offer reasonable statistical power in comparison with standard clinical trial endpoints.

RESULTS
The bibliographic search was performed using Medline and the Cochrane Database and 55 studies were identified. Of the 55 studies, only 10 studies were studies of induction that were placebo-controlled. These 10 studies were reviewed in detail and our a priori criteria were applied.7,10-18 (See the QUOROM diagram in Fig. 1.)

Five studies10,11,13,17,18 were excluded because they were either unavailable in full text format, used a low (<2 g/d) dose of 5-ASA, or they did not report the data required to perform the analysis. The Mesalamine Study Group trial11 was excluded because the dosage of mesalamine used for the study was less than 2 g/day, in order to maximize the homogeneity of the studies in the meta-analyses. Meyers et al19 did not include specific information about how many subjects left the study for failure of the experimental drug versus placebo. Feurle et al10 and Hetzel et al13 used olsalazine and did not clearly differen-
tiate between diarrhea caused by olsalazine versus lack of treatment effect. Sutherland et al. did not specify exactly how many patients left the study for lack of efficacy in each arm.

Description of the Studies

The clinical response data (primary clinical endpoints) from each of the 5 studies are presented in Table 1 and the VWF endpoint data are presented for comparison in Table 2.

**Schroeder 1987**

This placebo-controlled, double-blind, randomized trial enrolled 76 subjects over 17 months. Subjects ranged in age from 15 to 70, with mild to moderate UC. Asacol 400 mg tablets or matching placebo were used as 12 tablets daily for 6 weeks. Evaluation occurred at 3 and 6 weeks. Clinical response, described as “complete,” “partial,” or “no response,” was determined on the basis of stool frequency, amount of rectal bleeding, and physician’s global assessment (which included sigmoidoscopic appearance) on 4-point scales compared to baseline data. In the mesalamine arm, 9/38 achieved a complete response, while only 2/38 in the placebo arm achieved a complete response. For the VWF endpoint, 1 of 38 subjects left the mesalamine arm of the study due to lack of efficacy (LOE) of their therapy. In contrast, in the placebo arm 13 of 38 subjects left the study due to LOE.

**Sninsky 1991**

This multicenter, double-blind, placebo-controlled, randomized trial enrolled 87 subjects at 10 sites. Eighty-seven subjects aged 18–75 years with mildly to moderately active UC were enrolled. Mesalamine (1.6 or 2.4 g/d) or matching placebo tablets containing microcellulose were administered. Clinical grading was based on stool frequency, rectal bleeding, sigmoidoscopic findings, and patient’s functional assessment, each on a 4-point scale, which together gave the physician’s global assessment, also on a 4-point scale. The change in this clinical grade was indicated by classifying each patient as being “in remission,” “improved,” “maintained,” or “worsened.” In terms of the clinical remission endpoint, 6/43 subjects in the 2.4 g/d mesalamine arm achieved remission, while 2/44 in the placebo arm achieved remission. For the VWF endpoint, in the mesalamine arm 4/43 subjects left the study due to drug LOE, and in the placebo group 15/44 subjects left the study due to LOE.

### TABLE 1. Primary (Clinical Response) Endpoint Results

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Remission</th>
<th>Nonremission</th>
<th>Total Treatment</th>
<th>Placebo Remission</th>
<th>Placebo Nonremission</th>
<th>Placebo Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schroeder</td>
<td>1987</td>
<td>9</td>
<td>29</td>
<td>38</td>
<td>2</td>
<td>36</td>
<td>38</td>
</tr>
<tr>
<td>Sninsky</td>
<td>1991</td>
<td>6</td>
<td>37</td>
<td>43</td>
<td>2</td>
<td>42</td>
<td>44</td>
</tr>
<tr>
<td>Hanauer</td>
<td>1993</td>
<td>56</td>
<td>136</td>
<td>192</td>
<td>11</td>
<td>79</td>
<td>90</td>
</tr>
<tr>
<td>Lichtenstein</td>
<td>2007</td>
<td>56</td>
<td>121</td>
<td>177</td>
<td>11</td>
<td>74</td>
<td>85</td>
</tr>
<tr>
<td>Kamm</td>
<td>2007</td>
<td>97</td>
<td>158</td>
<td>255</td>
<td>19</td>
<td>67</td>
<td>86</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>224</td>
<td>481</td>
<td>705</td>
<td>45</td>
<td>298</td>
<td>343</td>
</tr>
</tbody>
</table>
Hanauer 1993

This multicenter, double-blind, placebo-controlled, randomized, dose-response trial was conducted at 20 sites. In addition to daily patient diaries, clinical assessments and sigmoidoscopy were performed at weeks 1, 4, 8 or upon withdrawal. The subjects were required to be at least 18 years old, with mild to moderate active UC confirmed by clinical and colonoscopic evidence, and a score of 5 or greater on a 15-point index. Patients were stratified according to extent of disease. Mesalamine or placebo was provided in capsules in identical blister cards for 8 weeks. The primary clinical endpoint was achieved if subjects achieved physician global assessment score (based on 6 categories) and divided into treatment success, treatment benefit or treatment failure, sigmoidoscopic index (based on erythema, friability, granularity/ulceration, mucopus, mucosal vascular pattern), and biopsy score (4-point scale used by pathologists). Data were pooled from the 2- and 4-g/d groups. In terms of the clinical remission endpoint, 56/192 subjects in the mesalamine arm achieved remission, while 11/90 in the placebo arm achieved remission. In the placebo arm, 21/90 subjects dropped out due to LOE, while in the pooled mesalamine group 12/192 subjects dropped out due to LOE.

Lichtenstein 2007

This randomized, double-blind, parallel-group, placebo-controlled, multicenter study enrolled 262 subjects with mild to moderately active UC. Subjects included adults with newly diagnosed or relapsing mild to moderately active UC (score of 4–10 on a modified version of the Sutherland UC-disease activity index [UCDAI], with a sigmoidoscopic score ≥1 and a Physician’s Global Assessment [PGA] score <2) with compatible histology. Patients were randomized equally between MMX mesalamine 2.4 g/day given once daily (QD), MMX mesalamine 4.8 g/day given QD, mesalamine 2.4 g/day given in 3 divided doses, and placebo. Treatment lasted for 8 weeks. The primary clinical trial endpoint was the number of patients in clinical (UCDAI score of <1 with a score of 0 for rectal bleeding and stool frequency and at least a 1-point reduction from baseline) and endoscopic remission after 8 weeks of treatment. For the primary clinical response, 56/177 subjects achieved CR in the pooled mesalamine arms, and 11/85 subjects achieved CR in the placebo arm. For the VWF endpoint, in the pooled MMX arms 18/177 patients left the study due to LOE of therapy, and in the placebo group 24/85 patients left the study due to LOE.

Kamm 2007

This double-blind, multicenter study enrolled 335 subjects in a trial comparing MMX mesalamine versus placebo for the treatment of active UC. Subjects included adults with newly diagnosed or relapsing (relapsed 6 weeks prior to baseline) active, mild-to-moderate UC (score of 4–10 on a modified UCDAI, with a sigmoidoscopic score ≥1 and a PGA score <2) with compatible histology. Patients were randomized equally between MMX mesalamine 2.4 g/day given once daily (QD), MMX mesalamine 4.8 g/day given QD, mesalamine 2.4 g/day given in 3 divided doses, and placebo. Treatment lasted for 8 weeks. The primary clinical trial endpoint was the number of patients in clinical (UCDAI score of <1 with a score of 0 for rectal bleeding and stool frequency and at least a 1-point reduction from baseline) and endoscopic remission after 8 weeks of treatment. In the pooled 5-ASA arms, 97/255 subjects achieved CR, while 19/86 achieved CR in the placebo arm. For the VWF endpoint, in the pooled MMX arms 18/177 patients left the study due to LOE, and in the placebo group 24/86 patients left the study due to LOE.

Meta-analysis

Meta-analysis was performed using both fixed and random effects models for the data from both the primary study endpoints and for the VWF endpoint. The 2 models produced
very similar results, and the fixed effect model results are presented below.

Figure 2 presents the forest plot of the fixed effects model meta-analysis using the clinical response data from each study. Mesalamine therapy is associated with a lower risk of failure to achieve CR (RR = 0.81; 95% confidence interval [CI] 0.76–0.86). There was low heterogeneity, with a chi squared statistic of 2.69 (df = 4, P = 0.61) suggesting consistent mesalamine efficacy across the trials.

Figure 3 presents the forest plot of the fixed effects model meta-analysis using the VWF endpoint. Mesalamine therapy is associated with a lower risk of leaving the study due to therapeutic LOE, with a RR using the fixed effect model of 0.33 (95% CI 0.24 – 0.44). There was low heterogeneity, with a chi squared statistic of 4.35 (df = 4, P = 0.36) suggesting consistent mesalamine efficacy across the trials for this endpoint.

In order to evaluate for publication bias, funnel plots were constructed for both endpoints, and no evidence of significant publication bias was found, although the limited sample size limits the power of this approach. Egger’s bias plot was also constructed for each endpoint, with nonsignificant P values for bias (0.98 for the CR endpoint, and 0.1 for the VWF endpoint). An omission plot was also performed to evaluate the influence of each study on the data and detect outliers. There was no study which was an outlier or significantly influenced the summary risk ratio when excluded.

**Power and Sample Size**

The effect difference for the clinical response endpoint was 18.6% (31.8 CR with mesalamine versus 13.1% with placebo), and the effect difference for the VWF endpoint was 22% (89.7% did not leave for LOE with mesalamine versus 67.7% with placebo). To perform a clinical trial with the VWF endpoint, comparing 5-ASA to placebo, with 80% power and a 2-sided alpha of 0.05, would require 80 subjects per arm. To perform an identical clinical trial with the standard clinical trial endpoints would require 88 subjects per arm (Table 3).

**DISCUSSION**

We performed meta-analyses using both standard clinical response endpoints and the VWF endpoint with data extracted from RCTs of induction of remission in UC with 5-ASA. We were able to extract VWF data retrospectively from published placebo-controlled RCTs, particularly if they had a CONSORT diagram. We found that the VWF endpoint did provide an estimate of clinical success of >50% (90% with VWF) with 5-ASAs in UC, which reflects clinical practice better than the estimate of short-term clinical success from the primary clinical trial endpoints (32%). The measured effect difference between drug and placebo with the VWF endpoint favored 5-ASA, with a magnitude similar to the standard clinical trial endpoints. From our experience using 5-ASA drugs, the short-term effectiveness of 5-ASAs is closer to the results obtained from our VWF data extraction rather than the remission rates using the primary endpoint.

Remarkably, the VWF endpoint in the clinical trials analyzed here provides slightly more statistical power and allows for a smaller sample size than the standard clinical trial endpoints. This low sample size is due in part to the fact that the short-term rate of staying on 5-ASAs is quite high in these trials. Sample sizes in studies with dichotomous outcomes become smaller as the success rates are closer to the extremes of 0% and 100%. This is 1 reason why standard clinical endpoints are defined strictly, to obtain placebo rates as close to 0% as possible. However, this artificially lowers the drug success rate to a point that clinical trial results no longer reflect the real-world efficacy of therapy. In contrast, the VWF endpoint has a high drug success rate that provides good statistical power. Using the VWF endpoint to assess the outcome of a
clinical trial may be a simple yet effective way of estimating the real-world effectiveness of medications.

The VWF endpoint does have several important limitations. First, the success rate is likely to be influenced by the clinical trial setting, as large incentives or personable study coordinators will tend to keep subjects in a study longer. Second, the VWF endpoint is impossible to use when a prominent side effect of a drug (i.e., diarrhea with olsalazine) is very similar to lack of efficacy. Third, the VWF endpoint can only be used in trials in which the subjects and study staff are both blinded to the treatment assignment, to avoid possible bias against placebo. The VWF endpoint may also tend to have a high placebo rate in short trials, as many subjects will be willing to “hold on” a bit longer, hoping for later success.

The strengths of the VWF endpoint include: 1) the placebo rate by definition will fall over time as patients drop out, in contrast to our current clinical trial endpoints, which benefit from short trials; 2) it is relatively simple to obtain the data and calculate the results; 3) it is a patient-centered endpoint, in which the patient, rather than the investigator, makes the judgment of whether the therapy is working well enough to continue on it; and 4) it may better reflect the real-world efficacy of 5-ASAs.

Not all studies reported the number of subjects in each arm that left each arm for lack of efficacy. In fact, we had to drop 2 studies from the meta-analysis due to the failure of these studies to report these data. In more recent studies, especially the ones using mesalamine MMX the CONSORT guidelines were followed, and these data were consistently provided. We support the use of the CONSORT guidelines in reporting randomized clinical trials, and in particular, the reporting of the number of patients that left each arm for lack of efficacy of the treatment.

VWF has never been tested prospectively as the primary endpoint in any clinical trial, and it may be subject to externally introduced bias in clinical trials. Prominent drug-specific side effects (i.e., highly reproducible injection-site reactions) could effectively break the blinding and make VWF an invalid endpoint. However, in double-blind trials, where neither the subjects nor the investigators and study staff know the treatment assignment, it should be difficult to alter the differential between the drug and comparator arms. At this point we would recommend using VWF only as a secondary, “reality-check” endpoint, particularly when new or modified clinical trial endpoints are being used. If the VWF endpoint produces results that are similar in direction and magnitude of effect, this would support the validity of new endpoints, and the VWF endpoint may provide a good estimate of real-world maintenance of clinical effect.

REFERENCES


