Surrogate Endpoints for the Treatment of Venous Leg Ulcers

Joel M. Gelfand,*† Ole Hoffstad,† and David J. Margolis*†

*Department of Dermatology and the †Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA, U.S.A.

Surrogate markers are endpoints that occur early in the course of treatment and are intended to predict the true, meaningful clinical endpoint. Surrogate markers have been used to study treatments for a wide range of diseases in which the true outcome is delayed. The evaluation of therapies for venous leg ulcers is challenged by the prolonged observation period necessary to reach the endpoint of healing. We have performed a large cohort study to examine wound healing characteristics as candidate surrogate markers of venous leg ulcer healing using the Curative Health Services population. A total of 58,038 wounds met our definition of venous leg ulcer; however, 1550 wounds were excluded based on size, depth, site, and/or involvement of tendon or bone, leaving 56,488 wounds in 29,189 patients for analysis. The median wound size was 189 mm^2, with a median wound duration of 3 mo. Using a large cohort of diverse venous leg ulcer patients, we demonstrate that after only 4 wk of treatment the wound parameters log healing rate, log wound area ratio, and percentage change in wound area can be valid surrogate markers of healing at 12 or 24 wk of care. Based on the area under the receiver operator characteristic curve log rate, log area ratio, and percentage change in area can discriminate which patients will heal at 12 or 24 wk of care (receiver operator characteristic 0.72–0.80). These surrogates were further validated by demonstrating that established risk factors for not healing such as wound size and wound duration are also important risk factors for not achieving the surrogate endpoint. These surrogate markers for venous leg ulcer healing may allow for early clinical trials to be more efficient, and can allow clinicians to identify patients unlikely to heal early in the course of treatment in order to expedite referral to specialty centers or for the selection of stepped treatment algorithms. ***Key words: epidemiologic/leg ulcer/outcome assessment/randomized controlled trials/varicose ulcer. J Invest Dermatol 119:1420–1425, 2002***

Surrogate endpoints have been defined by Prentice (1989) as "a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint". This rigorous definition demonstrates the statistical principles underlying the interpretation of surrogates. A surrogate endpoint as defined by Temple (1999) is "a laboratory measurement or a physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions or survives". This definition demonstrates the clinical principles underlying the interpretation of surrogates. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint. A surrogate endpoint must not be simply a correlate of the true clinical outcome; it must fully capture the net effect of the treatment on the clinical outcome. Therefore, a valid surrogate endpoint is related to the outcome of interest, and is affected by the treatment of interest to the same degree and in a manner that accurately reflects the effect of the treatment on the true outcome. For example, lower blood pressure has been used as a surrogate endpoint to predict fewer cardiovascular events.

Although many surrogate endpoints have been used in clinical trials, few have been extensively studied. Failure to study extensively a surrogate endpoint can result in invalid study results. For example, some anti-arrhythmia drugs have been approved based on their ability to suppress ventricular arrhythmia (i.e., the surrogate endpoint), but were later found to increase sudden death (i.e., the true endpoint) (Fleming and DeMets, 1996; Pstay *et al*, 1999).

Surrogate endpoints occur sooner than the primary endpoint, and help to make clinical trials more efficient (e.g., less follow-up time, smaller sample size required for adequate power). Valid surrogate endpoints therefore help speed the development of new effective medications and minimize the exposure of patients to ineffective developmental therapies. Furthermore, surrogate endpoints may allow a clinician to assess a patient's response to therapy during treatment. These concepts were recently reiterated in a recent NIH workshop on surrogate endpoints in clinical trials (De Gruttola *et al*, 2001). To be complete some authors differentiate surrogate markers from intermediate endpoints, in that an intermediate endpoint like a surrogate occurs before the true outcome but unlike a surrogate it also represents a clinical state that is of benefit to the patient (Freedman *et al*, 1992; Temple, 1999; Ellenberg, 2001).

The validity of a surrogate endpoint is supported by the biologic plausibility of the relationship, the demonstration in epidemiologic studies of the prognostic value of the surrogate endpoint for the clinical outcome, and evidence from clinical trials that the treatment effects on the surrogate endpoint

Manuscript received June 13, 2002; revised September 2, 2002; accepted for publication September 8, 2002

Reprint requests to: Joel M. Gelfand, 3600 Spruce Street, 2 Maloney Building, Philadelphia, PA 19104, U.S.A. Email: Joel.gelfand@uphs.upenn.edu

Abbreviations: AROC, area under the receiver operating characteristic curve; CHS, Curative Health Services; ROC, receiver operating characteristic curve.

correspond to effects on the outcome (Fleming and DeMets, 1996; Pstay *et al*, 1999). Surrogate markers have been accepted by the Food and Drug Administration as proof of drug efficacy provided that they are studied in large numbers of patients with appropriate follow-up periods (FDA ICH E9, 1998; Temple, 1999). A surrogate, however, must always be carefully evaluated because the effect of a novel treatment on the surrogate does not always translate into a similar effect on the real outcome.

Clinical trials and evaluations of the therapeutic response in individual patients with venous ulcers are limited by the prolonged time period needed (e.g., 24 wk), and the corresponding expense incurred to measure the outcome of complete wound healing. A valid surrogate marker of complete wound healing would make early clinical trials and trials being used to screen potential therapeutic agents more efficient in terms of cost, time, and number of subjects required. A valid surrogate marker may also allow for the more rapid screening of potential therapies and therefore may aid in the discovery of novel treatments for venous leg ulcers. Finally, a valid surrogate marker may allow for the identification of patients who are not likely to heal by standard methods early in the patient's course of treatment, thereby allowing for expedited referral to specialty centers or the expedited initiation of stepped treatment algorithms.

The initial healing rate of venous ulcers and the percentage change in the venous ulcer area over the initial 4 wk of therapy have been demonstrated in small studies to predict ulcer healing (Pecoraro *et al*, 1991; Gross *et al*, 1993; Margolis *et al*, 1993; Tallman *et al*, 1997; Kantor and Margolis, 1999, 1998, 2000a). The initial rate of healing and the percentage change in the ulcer area are promising surrogate markers for ulcer healing; however, surrogate markers require extensive epidemiologic testing to ensure validity. For example, recent statements from the FDA have indicated that in order for a surrogate marker to be acceptable, it must be shown to predict healing with adequate follow-up on a large sample of patients (FDA ICH E9, 1998). We hypothesize that wounds that have a faster healing rate or a greater absolute change, or those that achieve a greater percentage reduction in wound size, will be more likely to heal. The goal of this study is to determine if these surrogates correctly predict who will heal after 12 or 24 wk of care.

MATERIALS AND METHODS

Patient population Curative Health Services (CHS) has been directly involved in the care of individuals with chronic wounds since 1988 (Doucette *et al*, 1989; Fylling and Knighton, 1989; Knighton *et al*, 1989, 1990a,b; Hotta and Holohan, 1992; Margolis *et al*, 2001). They are a disease management firm. Their involvement has been centered on marketing, health-care provider education, treatment algorithms, and dispensing an autologous product that they believed improved the probability that a chronic wound would heal. At the time of this study, CHS had managed more than 150 distinct wound care facilities in 38 states in the U.S.A. As part of their association with the medical centers, CHS maintained an administrative and patient record database for each patient seen in a wound care center since 1988. We have previously described this database in detail (Kantor and Margolis, 2000b; Margolis *et al*, 2001). All subjects for this investigation had been treated at a CHS center between 1988 and 2000. To avoid including subjects who were one-time specialty center consultations, any individual who did not have a second office visit or documentation of a surgical procedure within 6 wk of the first office visit was excluded.

Ascertaining disease and outcome In order to ascertain that a person has a venous leg ulcer we used the CHS coding system for this disorder and excluded patients that had a diagnosis code consistent with lower limb arterial disease, diabetic etiology (e.g., neuropathic), or a code consistent with a pressure etiology. Venous leg ulcer was defined as a chronic wound of the lower extremity located in the gaiter area (from just below the ankle or malleolus up to but not including the knee flexural crease). The wounds could not have a depth greater than 2 cm or involve underlying structures such as tendon, ligament, or bone, or be larger than 150 cm^2. Wounds greater than 150 cm^2 were excluded because these wounds are unlikely to heal, and in our experience measurements in wounds of this size are unstable and subject to increased error (Kantor and Margolis, 1998). There was no minimum size criterion for inclusion in this study; however, patients who healed within 6 wk of care were excluded. Wound size was determined by measuring the maximum length of the wound and multiplying it by the maximum width, a method that closely approximates planimetric techniques (Kantor and Margolis, 1998). Wounds located on the head, face, neck, chest, sternum, arms, elbow, hands, shoulders, scapulae, ears, nose, chin, mouth, stomach, abdomen, buttocks, ischium, iliac crest, sacrum, hip, thigh, plantar surface of the foot, and toe were excluded.

We previously used a similar coding algorithm to ascertain individuals with diabetic neuropathic foot ulcers and we were able to validate this diagnosis scheme using patient chart review (Kantor and Margolis, 2000b; Margolis *et al*, 2001). Owing to recent and evolving changes in regulations that govern access to patients' charts, we were not able to abstract patients' charts for this study. In lieu of direct access chart validation we investigated the validity of the database diagnosis of venous leg ulcer using an algorithm approach. It was determined a priori that at least 90% of the ulcers should be in an anatomic location (e.g., calf, malleolus, leg, ankle, and shin) that is consistent with the diagnosis of venous leg ulcer. A random sample of 100 wounds identified by the inclusion/exclusion criteria described above was examined for anatomic location.

The outcome, a healed wound by the 12th or 24th week of care, for all subjects was determined using a previously validated algorithm (Kantor and Margolis, 2000b; Margolis *et al*, 2001). These time periods were selected as they are consistent with most venous leg ulcer clinical trials. Patients who were lost to follow-up after 6 wk of care were treated as not healed for the 12 and 24 wk endpoints.

Analysis

Surrogate formulas Because a surrogate should be measured at a time period much sooner than the outcome, we created and evaluated surrogates that occurred at weeks 2, 4, and 6 for the 12th week of care outcome and weeks 2, 4, 6, and 8 for the 24th week of care outcome. All the surrogates that we evaluated were based on changes in the size of the wound. We used the following seven formulas to generate our surrogates:

Absolute change in area	$Area_0 - Area_t$
Healing rate	$(Area_0 - Area_t)/t$
Area ratio	$Area_t/Area_0$
Percentage change area	$((Area_0 - Area_t)/Area_0)*100$
Log absolute change in area	$Ln(Area_0) - Ln(Area_t)$
Log healing rate	$(Ln(Area_0) - Ln(Area_t))/t$
Ratio of log areas	$Ln(Area_0)/Ln(Area_t)$

$Area_t$ is the wound area at weeks 2, 4, 6, or 8. $Area_0$ is the baseline wound area. All wound areas are in mm^2 and t was measured in days. Natural logarithm transformation of wound area was performed due to previous experience with the non-normality of this parameter. These formulas were evaluated for all of the wounds on the patient and for the first wound on the patient treated by the wound care center.

Descriptive analyses All continuous variables were described using means with standard deviations or medians with 25% and 75% percentiles. Comparisons were made between those that healed and those that did not heal for normally distributed continuous variables using a t test and ranksum test for those continuous variables that were not normally distributed.

Surrogate discrimination Discrimination or the ability of a test to differentiate between two individuals, one who healed and one that did not was measured by the area under the receiver operating characteristic curve (AROC and ROC *sans* area under). An ROC curve is a graphical representation of sensitivity (true positives) on the y-axis *vs* 1-specificity (false positives) on the x-axis over all possible cut-points for a test (see **Fig 1**). Therefore, the ROC curve provides a measure of the trade-off between the true positive rate *vs* the false positive rate over all possible dichotomous cut points for a test. The better the test the closer the AROC is to 1.0. An AROC of 0.5 suggests that the test is no better than chance, whereas an AROC of >0.7 is acceptable, >0.8 is good, and 0.9 is considered excellent (Hosmer and Lemeshow, 2000). ROC curves were generated for the candidate surrogate markers at weeks 2, 4, and 6 compared with the healed outcome at the 12th week of care and at weeks 2, 4, 6, and 8 compared with the healed outcome at the 24th week of care. Candidate markers were therefore screened based on the highest AROC. This analysis was performed based on the wound healing characteristics of the first venous leg ulcer that the patient had treated by CHS (e.g., primary wound analysis). A secondary analysis, including subsequent venous leg ulcers, that a patient may have been treated by CHS (e.g., all wounds analysis) was also performed.

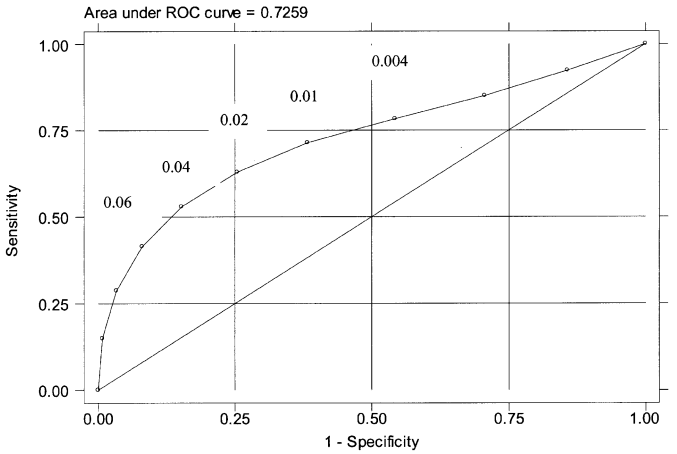

Figure 1. Receiver Operator Characteristic curve for the surrogate log rate of healing at 4 wk *vs* healing status at 24 wk. The sensitivity and specificity of potential cut-points are represented graphically, with sensitivity (true positive rate) on the y-axis, and the false positive rate (1-Specificity) on the x-axis. Note that as the rate of healing required to test positive increases the sensitivity decreases and the specificity increases.

Dichotomization and test characteristics The candidate markers with the best AROC were further investigated. In order to simplify the use of the surrogate markers they were dichotomized. A cut-point for the candidate surrogate marker was generated such that it maximized the correct classification of the outcome (complete wound healing) in the population studied. Once a cut-point for the surrogate marker was identified based on correct classification (probability that the test result (e.g., positive or negative) matches the outcome result (e.g., outcome present or outcome not present, respectively), the sensitivity (the probability of testing positive if the outcome is truly present), specificity (the probability of testing negative if the outcome is truly absent), positive predictive value (the probability of having the outcome given a positive test), and negative predictive value (the probability of not having the outcome given a negative test) for the population studied were calculated. The performance of the dichotomized cut point for each of the surrogate markers was further investigated by examining the effect of different prevalences of healing on the correct classification, positive predictive value, and negative predictive value of the test.

Test validation To be consistent with published opinions on surrogates, we investigated whether exposure known to be associated with a wound healing by the 24th week of care was also associated with the surrogate endpoint (Prentice, 1989; Temple, 1999; Buyse *et al*, 2000; Ellenberg, 2001). Wound size and wound duration have been previously demonstrated to be key risk factors associated with a wound healing by the 24th week of care (Margolis *et al*, 1999a,b). The association between these risk factors and the dichotomous surrogate were estimated as odds ratios using logistic regression.

Internal validation To assess the internal validity of the surrogate markers we performed bootstrap analysis with 1000 replications (Steyerberg *et al*, 2001). Using this technique, correct classification was estimated for the surrogate markers by sampling the original data set with replacement. These bootstrap samples were ultimately the same size as the original sample but as the original sample was sampled with replacement an individual might not be part of the bootstrapped sample or could even appear more than once. For our study this process was repeated 1000 times. Ultimately, this analysis yields an estimate of bias (the difference between the original correct classification and that 1000 observed bootstrap samples) and a confidence interval for the correct classification for the original correct classification. The estimate bias can then be used to adjust the original estimate of correct classification (Steyerberg *et al*, 2001).

All statistical analyses were performed using STATA 7.0 (College Station, TX) for a PC.

RESULTS

The CHS database contains data on 153,270 wounds. Initially, 95,232 wounds were excluded because of diagnosis codes consistent with arterial, diabetic, or pressure etiologies. A total 58,038 wounds had a code consistent with a venous ulcer and no evidence of arterial, pressure, or diabetic (neuropathic) etiology. Next, 177 wounds were excluded for depth >20 mm, 827 wounds were excluded for size greater than 150 cm^2, 125 were excluded based on site, and 421 were excluded because the wound included tendon or bone. All of these findings are unusual for a venous leg ulcer. After these exclusions we were left 56,488 wounds consistent with our definition of a venous leg ulcer in 29,189 patients. A random sample of 100 wounds ascertained by this algorithm demonstrated that 95% were located in the gaiter area. Review of the diagnostic codes for these 100 wounds demonstrated no significant inclusion of alternative etiologies such as malignancy, post-traumatic origin, or collagen vascular disease. Finally, 1597 (5.5%) patients were lost to follow-up within the 24th week of care. For the purpose of our primary analyses these individuals were coded as having not healed.

Table I. Patient characteristics and week 4 surrogate characteristics for those that healed and those that did not by the 24th week of care

Variable	Healed	Unhealed
No. of individuals (%)	7529 (65.6)	3943 (34.4)
No. of wounds (%)	14,037 (55.25)	11,369 (44.75)
Mean age (25th, 75th percentile)	68.31 (57,80)	68.66 (57,80)
Median wound size (mm^2) (25th, 75th percentile)	84 (14,320)	642 (199, 1805)
Median wound duration (mo) (25th, 75th percentile)	3 (1,6)	5 (2,13)
Median log wound size mm^2 (mean)	4.87 (1.69)	6.35 (1.64)
Median log wound duration (mo; mean)	1.15 (1.29)	1.75 (1.50)
Log rate mm^2 per day (SD) ($p<0.0001$)	0.0316 (0.073)	0.0024 (0.037)
Percent change (SD) ($p<0.0001$)	58.68 (220)	6.62 (341)
Log area ratio (SD) ($p<0.0001$)	0.84 (0.42)	0.99 (0.22)

The total number of individuals who had wound measurements (i.e., an office visit) on a given week varied. This occurred either because a patient healed before a given date or their patient visit occurred before or after our surrogate date. For example, 11,369 patients with 25,406 wounds were seen at week 4. By the 24th week of care 65.6% of their primary wounds had healed and 55.3 of all of their wounds had healed (**Table I**). By the 12th week of care 45.2% of patients had healed. Those that healed had at baseline smaller wounds, and wounds of shorter durations as compared with those that did not heal (all p-values $<$ 0.001). With respect to our surrogates, patients that healed had a larger log wound rate, a larger percent change and a smaller log area ratio at all surrogate time points (for week 4 see **Table I**).

The surrogates percentage change in area, log healing rate, and log area ratio discriminated well at all time points with respect to differentiating between those that healed and those that did not heal by the 12th or 24th week of care (**Table II**) (Mayfield *et al*, 1996; Hosmer and Lemeshow, 2000). A representative example of the ROC curve for the surrogate log rate of healing at week 4 *vs* healing status at week 24 is given in **Fig 1**. The ability of the surrogates to discriminate did not vary whether the primary wound analysis was evaluated (e.g., only first venous wound that a patient had treated by CHS was included) or all wounds analysis was evaluated (e.g., inclusion of all venous ulcers for patients with multiple venous wounds) (**Table II**). The surrogate's absolute change in area, healing rate, and area ratio demonstrated poor discrimination at all time points with respect to differentiating between those that healed and those that did not heal by the 12th or 24th week of care, as demonstrated by an AROC of less than 0.60 (data not shown).

The surrogates were dichotomized at a point that optimized the ability of the surrogate to classify correctly a wound as healed or unhealed by the 24th week of care. An

Table II. The receiver-operating-characteristic-curve area for surrogate endpoints measured after 2 wk (W2), 4 wk (W4), 6 wk (W6), or 8 wk (W8) for a 12 wk or 24 wk healed wound outcome

Surrogate	Wound	Endpoint	ROC-W2	ROC-W4	ROC-W6	ROC-W8
Log rate (mm^2 per day)	Primary	Week 12	0.74	0.80	0.84	NA
Percent change	Primary	Week 12	0.74	0.80	0.84	NA
Log ratio	Primary	Week 12	0.74	0.81	0.85	NA
Log rate (mm^2 per day)	All	Week 12	0.73	0.79	0.83	NA
Percent change	All	Week 12	0.73	0.79	0.83	NA
Log ratio	All	Week 12	0.73	0.79	0.84	NA
Log rate (mm^2 per day)	Primary	Week 24	0.69	0.73	0.75	0.77
Percent change	Primary	Week 24	0.69	0.73	0.75	0.77
Log ratio	Primary	Week 24	0.69	0.73	0.76	0.78
Log rate (mm^2 per day)	All	Week 24	0.67	0.72	0.74	0.73
Percent change	All	Week 24	0.68	0.72	0.74	0.73
Log ratio	All	Week 24	0.68	0.72	0.74	0.74

Table III. Correct classification maximized dichotomous surrogate endpoints at 4 wk for a 24 wk healed outcome. Ninety-five percent confidence intervals generated from bootstrap analysis

Surrogate	Wound	Cut-point	Sensitivity	Specificity	PPV	NPV	CC (95% confidence interval)
Log rate (mm^2 per day)	All	0.021	0.60	0.74	0.74	0.60	0.66 (0.65, 0.67)
Log rate (mm^2 per day)	Primary	0.012	0.67	0.69	0.81	0.53	0.68 (0.68, 0.69)
Percent change	All	43.76	0.60	0.74	0.74	0.60	0.66 (0.65, 0.67)
Percent change	Primary	28.79	0.67	0.69	0.80	0.52	0.68 (0.67, 0.69)
Log ratio	All	0.95	0.61	0.71	0.63	0.69	0.67 (0.66, 0.67)
Log ratio	Primary	0.98	0.74	0.74	0.80	0.67	0.69 (0.68, 0.69)

PPV, positive predictive value; NPV, negative predictive value; CC, correct classification.

example of the effect of varying the log rate of healing necessary to test positive for the surrogate at week 4 compared with the test characteristics of predicting healing status at week 24 is shown in **Fig 1**. This analysis was focused on a 24 wk outcome to be consistent with the most frequent endpoint for a venous leg ulcer clinical trial. The selected cut-points and test characteristics (sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratio, and correct classification) for the surrogates percent change in area, log rate of healing, and log area ratio are listed in **Table III**. Note that the bootstrap analysis yielded narrow confidence intervals around the correct classification observed in the original data set, demonstrating the precision of our estimate. Similarly, bias estimates using the bootstrap technique for our surrogates were all near 0 (none to four decimal places) and therefore indicate the robustness of our estimates within our population. Because positive predictive value, negative predictive value, and correct classification are dependent on the prevalence of the outcome (i.e., a healed wound), we adjusted these test characteristics for varying healing rates (**Table IV**). There is very little change in correct classification within the expected range for likely healing outcomes normally encountered in venous leg ulcer wound healing studies (**Table IV**).

As can be noted in **Table IV**, about 68% of the time the dichotomous 4 wk surrogate correctly classified the primary wound as healed by the 24th week of care. These rates are very conservative in that all patients who dropped out were coded as not healed and patients who healed before the fourth week of care were dropped from the data base. In all cases, the surrogate would appear to function better if healed patients were included in the final assessment and even better if those who dropped out were excluded. For example, if for the 4 wk log rate dichotomous surrogate patients who healed before the fourth week of care were included (e.g., as they healed they must have achieved the surrogate healed status at some point prior to healing) then the surrogate correctly classified 71% of patients, and if those that healed before the fourth week are included and those that dropped out are excluded then the surrogate correctly predicted 81% of the subjects.

Table IV. Effect of changing the prevalence of a wound healing using the dichotomous 4 wk log rate surrogate for a 24 wk outcome (typical rates of healing by the 24th week of treatment shown in bold)

Prevalence	PPV	NPV	CC
0.10	0.19	0.95	0.69
0.20	0.35	0.89	0.68
0.40	**0.59**	**0.76**	**0.68**
0.60	**0.76**	**0.58**	**0.68**
0.80	0.90	0.35	0.68

PPV, positive predictive value; NPV, negative predictive value; CC, correct classification.

Two risk factors known to be associated with an ulcer healing by the 24th week of care had similar associations with the 4 wk log rate surrogate. Using multivariable logistic regression, initial wound log area was associated with a wound healing by the 24th week by an odds ratio of 0.63 (95% confidence interval 0.61, 0.65) and with log rate by an odds ratio of 0.86 (95% confidence interval 0.84, 0.88). Wound log duration was associated with a wound healing by the 24th week of care by an odds ratio of 0.74 (95% confidence interval 0.71, 0.76) and with the log rate surrogate by an odds ratio of 0.85 (95% confidence interval 0.83, 0.88). As expected, because the log rate surrogate is area based, the odds ratio of association between the log rate dichotomous surrogate and a healed wound at 24 wk was confounded by wound log area [unadjusted 4.55 (95% confidence interval 4.19, 4.94) and adjusted by log area 1.74 (95% confidence interval 1.58, 1.93)], but not by wound duration.

DISCUSSION

For this study we described several surrogate endpoints for a venous leg ulcer wound healing study. These surrogates were all based on wound measurements, which are commonly measured in leg ulcer studies. Initially, using ROC analysis, the continuous surrogates percent change in area, log healing rate, and log area ratio at weeks 2, 4, and 6 were shown to discriminate between a wound that healed by 12 wk of care and one that did not. This was also true for wounds followed for 24 wk (albeit we used an 8 wk not a 6 wk surrogate). We ultimately chose to focus on a 4 wk surrogate, because this time period seemed to maximize accuracy and minimize the time to surrogate. Dichotomization of the surrogate markers at week 4 demonstrated that a wound's healing status at 24 wk can be correctly classified at a rate of 66–69% depending on the marker utilized. It should be noted that the CHS database is quite large and includes data from more than 150 centers. Therefore, these surrogates have demonstrated the ability to properly predict who will heal by the 24th week of care in a large and diverse population of venous leg ulcer patients. In addition, to demonstrate that the surrogates follow the dictates of Prentice (1989), Temple (1999), and others who have published on this topic, we used regression models to demonstrate that risk factors for wound healing, such as wound duration and wound size, were also important risk factors for the dichotomized surrogate markers (Freedman *et al*, 1992; Margolis *et al*, 1999a, b; Temple, 1999; Buyse *et al*, 2000; Ellenberg, 2001). Finally, we have conservatively called these endpoints surrogate endpoints and not intermediate endpoints because we are not certain that any of these surrogates truly represent an improved beneficial clinical state and because FDA guidance has consistently represented that only a healed wound is of true benefit to a patient (FDA Wound Healing Clinical Focus Group, 2001; Lindblad, 2001). Nevertheless, a strong case can be made to consider our surrogates as intermediate endpoints (Robson *et al*, 1999; Robson *et al*, 2000).

The results of this large cohort study extend findings from previous small studies that the rate of healing and percentage change in area can predict healing at week 24. Previous small studies have found that the initial rate of healing can predict healing status at 24 wk without the need for log transformation of the data or the need to determine percentage change in area (Tallman *et al*, 1997; Kantor and Margolis, 1998, 1999). In these studies, the potential effect of different wound shapes and sizes on predicting healing was corrected for by wound perimeter as described by Gilman (1990). Wound perimeter, which may be difficult to measure without special equipment (e.g., computer-based planimetry), was not available in the CHS database. Of note, our surrogate markers (e.g., absolute change in area and healing rate) that were not log transformed provided poor discrimination with respect to complete healing at 12 or 24 wk. The log transformation of the variables as well as standardizing each wounds' change relative to its self (e.g., percentage change in area), however, resulted in good discrimination of healing status. This finding may suggest that like Gilman's correction, the effect of different wound sizes or shapes on the ability to predict healing can be adequately corrected for by log transformation or by adopting percentage change in area. Of note, our estimate of wound size, which was ultimately used in all of our surrogate formulas, was based on measuring the maximum length of the wound and multiplying it by the maximum width. We have previously shown that this method for determining wound size is similar to planimetric techniques, which were frequently used in the previously published studies (Kantor and Margolis, 1998). Furthermore, the technique we used is much more readily available for clinicians than planimetric approaches.

There are important potential limitations to the validity of these surrogate markers. Venous ulcers in this study were identified and defined by using diagnostic codes and validated using a predictive value approach based on the anatomic site of the wound. Ideally, the diagnosis would have been validated using a chart review method; however, this was unavailable due to regulatory issues related to patient confidentiality. As CHS coding has demonstrated as valid for identifying other chronic wounds (diabetic ulcers) by chart review studies, there is no reason to suspect that the coding of venous leg ulcers would not be accurate. The surrogate markers may not have captured the full effect of treatment (standard therapy with limb compression) on the true endpoint (e.g., complete wound healing) as the correct classification was not 100%. We know of no surrogate, however, that is always correct and in fact many of the accepted surrogates are far worse than what we report in this study (Ellenberg, 2001; Lonn, 2001). Thus, use of these surrogates in randomized controlled trials will result in some misclassification of outcome. We would therefore recommend using these surrogates in phase II studies or other studies being conducted to screen initially new wound healing agents. Additionally, the identified surrogate markers have not been validated in a population external to the one from which they were derived. As stated above, however, there is no larger and more diverse a database for chronic wounds than the one we used. Moreover, the size of this study and results of the bootstrap analysis strongly support that the results are applicable to the CHS population (e.g., internal validity), which is a substantial number of individuals with venous leg ulcers. Finally, how well the surrogates capture the effect of treatments other than the standard therapies used at the wound care centers has not been investigated. In fact, many authors feel that it is essential to investigate the use of a surrogate for each new therapy (Freedman *et al*, 1992; Buyse *et al*, 2000).

We have demonstrated in a large diverse patient population that the percent change in area, log healing rate, and log area ratio at the fourth week of care can serve as important surrogate markers of complete wound healing at 12 or 24 wk of care. These surrogate markers can be easily applied by clinicians as they do not require specialized equipment or training to implement, and involve negligible expense. The validity of these surrogate markers in a large and diverse patient population suggests that they will be valid in other populations, such as patients cared for outside of CHS and patients in clinical trials. The ability to identify correctly patients who will heal or not heal by the fourth week of care has important implications for screening new wound healing agents in early clinical trials, and for the management of venous leg ulcer patients with novel treatments beyond standard compression therapy.

This work was funded by an unrestricted grant from Pfizer to the Trustees of the University of Pennsylvania, and NIH grants AR44695, AR02212, and AR48100.

REFERENCES

Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H: Statistical validation of surrogate endpoints: Problems and proposals. *Drug Info J* 34:447–454, 2000

De Gruttola VG, Clax P, DeMets DL, *et al*: Considerations in the evaluation of surrogate endpoints in clinical trials: Summary of a National Institutes of Health workshop. *Control Clin Trials* 22:485–502, 2001

Doucette MM, Fylling C, Knighton DR: Amputation prevention in a high-risk population through comprehensive wound-healing protocol. *Arch Phys Med Rehabil* 70:780–785, 1989

Ellenberg SS: Surrogate endpoints: the debate goes on. *Pharmacoepidemiol Drug Saf* 10:493–496, 2001

FDA ICH E9: E9 Statistical Principles for Clinical Trials. *www. FDA.gov* September, 1998

FDA Wound Healing Clinical Focus Group: Guidance for industry. Chronic cutaneous ulcer and burn wounds-developing products for treatment (Draft-Not for implementation). *Wound Repair Regeneration* 9:258–268, 2001

Fleming TR, DeMets DL: Surrogate endpoints in clinical trials: Are we being misled? *Ann Intern Med* 125:605–613, 1996

Freedman LS, Graubard BI, Schatzkin A: Statistical validation of intermediate endpoints for chronic diseases. *Stat Med* 11:167–178, 1992

Fylling CP, Knighton DR: Amputation in the diabetic population. incidence, causes, cost, treatment, and prevention. *J Enterostomal Ther* 16:247–255, 1989

Gilman TH: Parameter for the measurement of wound closure. *Wounds* 3:95–101, 1990

Gross EA, Wood CR, Lazarus GS, Margolis DJ: Venous leg ulcers. an analysis of underlying disease. *Br J Dermatol* 129:270–274, 1993

Hosmer DW, Lemeshow S: *Appl Logistic Regression* 2nd edition. John Wiley and Sons, Inc: Newyork. 2000

Hotta SS, Holohan TV: Health technology review. Procuren-a platelet derived wound healing formula. 1–4, 1992

Kantor J, Margolis DJ: From clinical trials to clinical practice: the efficacy and prognostic value of simple wound measurements. *Arch Dermatol* 134:1571–1574, 1998

Kantor J, Margolis DJ: Efficacy and prognostic value of simple wound measurements. *Arch Dermatol* 134:1571–1574, 1999

Kantor J, Margolis DJ: A multicentre study of percentage change in venous leg ulcer area as a prognostic index of healing at 24 weeks. *Br J Dermatol* 142:960–964, 2000a

Kantor J, Margolis DJ: The accuracy of using a wound care specialty clinic database to study diabetic neuropathic foot ulcers. *Wound Repair Regen* 8:169–173, 2000

Knighton DR, Fiegel VD, Doucette MM, Fylling CP, Cerra F: The use of topically applied platelet growth factors in chronic nonhealing wounds: a review. *Wounds* 1:71–78, 1989

Knighton DR, Fylling CP, Fiegel VD, Cerra F: Amputation prevention in an independently reviewed at-risk diabetic population using a comprehensive wound care protocol. *Am J Surg* 160:466–471, 1990a

Knighton DR, Citresi K, Flegel VD, Austin LL, Butler EL: Stimulation of repair in chronic, nonhealing cutaneous ulcers using platelet-derived wound healing formula. *Surg Gynecol Obstet* 170:56–60, 1990b

Lindblad WJ: Welcome (?) guidance from the Food and Drug Administration. *Wound Repair Regeneration* 9:257, 2001

Lonn E: The use of surrogate endpoints in clinical trials: focus on clinical trials in cardiovascular diseases. *Pharmacoepidemiol Drug Saf* 10:497–508, 2001

Margolis D, Gross E, Wood CR, Lazarus GS: Planimetric rate of healing in venous ulcers of the leg treated with pressure bandage and hydrocolloid dressing. *J Am Acad Dermatol* 28:418–421, 1993

Margolis DJ, Kantor J, Berlin JA: Healing of diabetic neuropathic foot ulcers receiving standard treatment: a meta analysis. *Diabetes Care* 22:692–695, 1999a

Margolis DJ, Berlin JA, Strom BL: Risk factors associated with the failure of a venous leg ulcer to heal. *Arch Dermatol* 135:920–926, 1999b

Margolis DJ, Kantor J, Santanna J, Strom BL, Berlin JA: Effectiveness of platelet releasate for the treatment of diabetic neuropathic foot ulcers. *Diabetes Care* 24:483–488, 2001

Mayfield JA, Reiber GE, Nelson RG, Greene T: A foot risk classification system to predict diabetic amputation in Pima Indians. *Diabetes Care* 19:704–709, 1996

Pecoraro RE, Ahroni JH, Boyko EJ, Stensel VL: Chronology and determinants of tissue repair in diabetic lower-extremity ulcers. *Diabetes* 40:1305–1313, 1991

Prentice RL: Surrogate endpoints in clinical trials. definitions and operational criteria. *Stat Med* 8:431–440, 1989

Pstay BM, Weiss NS, Furberg CD, *et al*: Surrogate endpoints, health outcomes, and the drug-approval process for the treatment of risk factors for cardiovascular disease. *JAMA* 282:786–795, 1999

Robson MC, Maggi SP, Smith PD, Wassermann RJ, Mosiello GC, Hill DP, Cooper DM: Original articles .ease of wound closure as an endpoint of treatment efficacy. *Wound Repair Regeneration* 7:90–96, 1999

Robson MC, Hill DP, Woodske ME, Steed DL: Wound healing trajectories as predictors of effectiveness of therapeutic agents. *Arch Surg* 135:773–777, 2000

Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema DF: Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 54:774–781, 2001

Tallman P, Muscare E, Carson P, Eaglstein WH, Falanga V: Initial rate of healing predicts complete healing of venous ulcers. *Arch Dermatol* 133: 1231–1234, 1997

Temple R: Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA* 282:790–795, 1999