The continuing story of peptic ulcer bleeding
van Leerdam, M.E.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 7

Interobserver agreement for the Forrest classification in peptic ulcer bleeding and the relation with endoscopic Doppler assessment

Monique E. van Leerdam, MD\textsuperscript{1}, Erik A.J. Rauws, MD PhD\textsuperscript{1}, Herman J. Adèr, PhD\textsuperscript{2},
Guido N.J. Tytgat, MD, PhD\textsuperscript{1}

\textsuperscript{1}Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam,
\textsuperscript{2}Department of Clinical Epidemiology and Biostatistics, Medical Center Vrije Universiteit,
Amsterdam, The Netherlands

Submitted for publication
Abstract

**Background:** Stigmata of recent hemorrhage in peptic ulcer bleeding are an important prognostic factor. This study evaluated interobserver agreement in Forrest classification among endoscopists. Furthermore differences between Forrest classification and endoscopic Doppler assessment were evaluated.

**Methods:** Video recordings from emergency endoscopies were shown to internists and gastroenterologists. A digital answering system was used for the registration of the Forrest classification.

**Results:** Forrest classifications of 19 video-fragments of gastroduodenal ulcers were obtained from 56 endoscopists. The intra-class correlation was 0.47 for the internists and 0.52 for the gastroenterologists. Using a dichotomous scale after dividing ulcers into 1: high-risk ulcer for rebleeding or 2: low risk ulcers for rebleeding, the overall Kappa statistic was 0.23. Agreement was fair ($\kappa$ 0.32) among gastroenterologists and poor ($\kappa$ 0.17) among internists. Out of 14 ulcers with a positive endoscopic Doppler signal, which theoretically need endoscopic treatment, a mean of 25% would not have been treated based on the Forrest classification.

**Conclusions:** There is a high variation among endoscopists in classification of stigmata of recent hemorrhage in gastroduodenal ulcers. There is lack of agreement between the visual interpretation and the endoscopic Doppler assessment of the ulcer base, which means that endoscopic therapy might be wrongly withheld or unnecessarily given.
Introduction

Acute upper gastrointestinal bleeding is a common medical emergency with an incidence of 62-172 cases per 100,000 persons per year. Gastroduodenal ulcers are responsible for almost half of all cases. Despite advances in diagnosis and treatment in the past years, mortality is still about 6-14% and rebleeding occurs in about 10-30% of those successfully treated. Especially rebleeding is a strong risk factor for mortality, with a mortality rate between 30-37%. Endoscopic appearance, including stigmata of recent hemorrhage (SRH), location and size of the ulcer, dictates management in terms of therapy, level of hospital care and length of hospitalization. The (modified) Forrest classification is one of the most accepted classifications of SRH, and is used as a prognostic indicator. Furthermore this classification is used to select patients for inclusion in therapeutic trials. However, classification on the basis of ulcer appearance is limited by subjectivity.

Recently endoscopic echo-Doppler has been introduced to assess the ulcer base for the presence of non-occluded superficial vessels. Based on the results of endoscopic Doppler investigation decisions for endoscopic treatment might be made.

In this study, we evaluated interobserver agreement of the Forrest classification among internists with endoscopic experience and gastroenterologists, using endoscopic video fragments of patients who presented with peptic ulcer bleeding (PUB). Furthermore, we evaluated agreement between the Forrest classification and endoscopic Doppler assessment of the ulcer base.

Patients and Methods

Video-recordings of ulcers at emergency endoscopy in PUB patients were used to evaluate interobserver agreement. All patients participated in an ongoing trial evaluating the value of endoscopic Doppler investigation. A montage of these video-recordings was shown during a National Endoscopy Meeting in the Academic Medical Center in Amsterdam, The Netherlands. Gastroenterologists and internists, all with endoscopic experience, working
either in academic or regional hospitals throughout The Netherlands, attended the meeting. Demographic questions were asked using an electronic multiple answering device regarding specialty (gastroenterologists or internist) and familiarity with the Forrest classification. The video fragments lasted 22 seconds on average. After each video fragment the audience was asked to classify the ulcer according the modified Forrest classification: 1: spurring bleeding (Ia), 2: oozing bleeding (Ib), 3: visible vessel (IIa), 4: adherent clot (IIb), 5: hematin stained ulcer base (IIc) and 6: clean ulcer base (III). Electronic answering devices were used and all answers were collected in an Excess database. Data about the presentation of the patient, hemodynamic parameters, or hemoglobin level were not given to prevent bias.

Of all ulcers, endoscopic Doppler assessment of the ulcer base and ulcer margin was obtained during the same initial emergency endoscopy. A pulsed wave Multi-Range 16 MHz Doppler system was used, manufactured by DWL (Neurosoft, Sipplingen, Germany). The Doppler probe has a diameter of 1.8 mm, and can be passed down the working channel of the endoscope. During the Doppler investigation, the probe was placed on multiple points of the surface of the ulcer base, including the region around and on any SRH. The whole ulcer base was assessed systematically. The Doppler system simultaneously scanned at eight different depths, ranging from 0.4 mm up to 2.5 mm. Gain was set at 80%. These variables were kept constant. A positive Doppler ultrasound signal was defined as a reproducible waveform of at least four consecutive cycle durations (indicating pulsating blood flow and minimizing the chance of a false-positive signal) up to 1-mm depth. The waveform was both visible and audible to the endoscopists and was recorded on the videotape. A waveform originating deeper than 1 mm was considered irrelevant for predicting rebleeding risk. Results of the endoscopic Doppler investigation were compared with the Forrest classification.

Statistics
Data were collected in Excel spreadsheet and were submitted for statistical testing using the Statistical Package for the Social Sciences (SPSS), version 10.1. The overall agreement was determined, calculating intra-class correlation coefficients (ICC) for gastroenterologists and internists. A value of 1.00 represents perfect agreement between observers, whereas 0.00 represents no agreement whatever.
When using a dichotomous scoring scale: 1; high-risk ulcer for rebleeding that should be treated endoscopically (Forrest Ia, Ib, Ila and IIb)\textsuperscript{9,11,19,22,23} and 2; low risk ulcer for rebleeding that do not need treatment (Forrest IIc and III), interobserver agreement using Kappa statistics was calculated. In medical research Kappa statistics are most often the first choice statistics for interobserver agreement.\textsuperscript{11} The customary interpretation of the Kappa findings is as follows: $\kappa = 0.81-1.00$ very good agreement; $\kappa = 0.61-0.80$ good agreement; $\kappa = 0.41-0.60$ moderate agreement; $\kappa = 0.21-0.40$ fair agreement; $\kappa = <0.20$ poor agreement.\textsuperscript{24} Differences between groups were assessed using Student’s t-test to compare means or the Chi-square test for evaluation of qualitative values. Two-sided p-values less than 0.05 were considered statistically significant.

**Results**

During the meeting, classifications were obtained from 30 gastroenterologists and 26 internists. From the 56 endoscopists, 3.7% was not familiar with the Forrest classification; 55.6% was familiar with the classification, but did not use it in daily practice, and 40.7% used the classification during daily practice. Significantly more gastroenterologists than internists used the Forrest classification (55% vs 24%, $p = 0.02$).

**Agreement**

On average, answers were obtained from 51–56 endoscopists per video fragment. When considering the gastroenterologists, the ICC was 0.52, whereas the ICC of the internists was 0.47.

When using a dichotomous scoring scale: high-risk ulcers for rebleeding that should be treated endoscopically (Forrest Ia to IIb) and ulcers that do not need treatment (IIC and III)\textsuperscript{9,11,19,22,23}, overall agreement was only fair. Kappa 0.23. Agreement among gastroenterologist was fair ($\kappa = 0.32$), among internists agreement was poor ($\kappa = 0.17$).

From the 19 ulcers with video-recordings, 14 had a positive endoscopic Doppler signal. Of the 14 ulcers with a positive Doppler signal, which theoretically might need endoscopic treatment, 75% or 8% would have been treated based on the Forrest classification (Ia. Ib. Ila...
and IIb). Of the ulcers without a superficial Doppler flow signal, a Forrest classification Ia to IIb was given in 41% ± 18%. So 41% of the Doppler negative ulcers would have been treated based on the endoscopic appearance (figure 1).

**Forrest classification**

![Graph showing Forrest classification](image)

**Discussion**

The appearance of the ulcer base is probably the best available predictor of patient outcome. However, it is clear that the Forrest classification is subjective; there is an important interobserver variation, overall ICC was only 0.52 and 0.47 for gastroenterologists and internists respectively. It is clinically important to estimate the risk of rebleeding, to determine whether endoscopic therapy will be beneficial, and to decide whether early discharge is acceptable. But even if we divide the ulcers in high risk (Forrest Ia to IIb) and low risk (IIc and III) ulcers for rebleeding, agreement is only fair (Kappa 0.23). Lau et al evaluated the interobserver agreement of SRH among 14 international experts, using video fragments. For ulcers with spurting bleeding, good agreement was found (kappa 0.66). Moderate agreement was obtained for ulcers with oozing bleeding and adherent clots (kappa 0.42 and 0.43 respectively). For non-bleeding visible vessels, flat-pigmented spots and clean-based ulcers, agreement was fair (kappa 0.34, 0.39, and 0.37 respectively). Other studies, either using slides or video fragments, also showed lack of agreement.

Lack of uniformity is also suggested by the wide range of values given in published studies for the prevalence and rebleeding rates of the same stigmata. For example, the prevalence of non-bleeding visible vessels varies between 4-35% and of adherent clot between 0-49%. The rate of further bleeding associated with these stigmata in control groups without endoscopic
therapy ranges from 0-81% for the visible vessel and from 14-36% for the adherent clot. Differences in study design, patient selection and lack of agreement regarding appearances of SRH will play a significant role in study results.

Chen et al evaluated the endoscopic picture and the pathologic findings from 110 patients with subtotal gastrectomies for PUB. They found that a vessel wall protruding above the ulcer surface on pathologic examination was usually described as a ‘pearl’ colored vessel in the endoscopic examination. In 20% of patients with a Forrest III classification on endoscopic examination, the pathological specimen did contain an eroded vessel. The clean base appearance may be caused by the pearl color of the visible vessel, which is similar to the color of the ulcer base.

Twenty-five percent of the ulcers classified as hematin stained ulcer base or clean ulcer base in our study, had a superficial Doppler flow signal, and thus, might theoretically be at risk of rebleeding. In 41% of the ulcers classified as Forrest Ia to IIb no superficial doppler flow signal was detectable and therefore endoscopic hemostatic therapy would not be indicated based on the Doppler examination. Our findings reflect those of Kohler et al who also reported discrepancy between the Forrest classification and Doppler findings in PUB patients. Doppler investigation might provide a better method to distinguish between a blood vessel, a clot, or necrotic tissue in the base of the ulcer. However, whether Doppler assessment gives us clinical relevant information in predicting the outcome of the patient should be further evaluated.

Maybe the endoscopic Doppler system might overcome the subjectivity of the visual classification of SRH. We are still waiting for a randomized controlled trial evaluating the clinical value of the Doppler system in predicting clinical outcome of the patient. Until we have an objective diagnostic tool, we should realize that the Forrest classification is subjective and furthermore that we do not know whether we see a patent vessel, an already thrombosed vessel, an adherent clot, or just a necrotic piece of mucosa. Education for improvement of interobserver agreement only overcomes one part of this problem. Recognition of a pearl colored vessel in a pearl colored ulcer base remains very difficult. Decisions about endoscopic therapy and early discharge of the patient should never be based solely on such a subjective classification. Other parameters, such as age, co-morbidity, use of drugs.
presentation with melena or hematemesis, hemoglobin level, hemodynamic parameters etc should be taken into account. Furthermore, study results should be properly judged, because of the interobserver variation in classifying the ulcer.

In this study video fragments were used, which, of course, never reflect the real situation. The fragments were rather short. However, agreement between the in vivo evaluation and video tape recordings in the study of Bour et al was good (kappa 0.60 ± 0.19), suggesting that bias due to video recording standardization was negligible. We did not inform the endoscopists about the clinical parameters of the patient (shock, hemoglobin level etc), which in the real situation undoubtedly will play an important role in ulcer classification and determining endoscopic management. The difference in agreement between internists and gastroenterologists in this study might be explained by the fact that in The Netherlands gastroenterologists are significantly more often dealing with patients presenting with peptic ulcer bleeding than internists. Differences in experience of endoscopists might partly explain lack of agreement.

Conclusions

There is a significant interobserver variation in classification of stigmata of recent hemorrhage in bleeding peptic ulcers. Also agreement in high or low risk ulcers for rebleeding is far from optimal. There is lack of agreement between the classification based on visual interpretation and the endoscopic Doppler assessment of the ulcer base. Theoretically, the Doppler investigation might be an objective diagnostic tool to assess the ulcer base. The clinical diagnostic value of the Doppler should be further evaluated before it can be used in daily practice.

Acknowledgment

We thank the endoscopists at the National Endoscopy Meeting for participating in this study.
The Forrest classification in peptic ulcer bleeding

References


Chapter 7


