Dissecting Lewis score under the light of fecal calprotectin; an analysis of correlation of score components with calprotectin levels in capsule endoscopy

Anastasios Koulaouzidis, Artur Nemeth, Gabriele Wurm Johansson, Ervin Toth
The Royal Infirmary of Edinburgh, Scotland, UK; Skåne University Hospital, Malmö, Sweden

Abstract

Background Lewis Score (LS) is an inflammatory score in small-bowel capsule endoscopy (SBCE). Fecal calprotectin (FC) is considered the non-invasive, ‘gold standard’ marker of gastrointestinal (GI) inflammation. Recently, we reported that LS shows only a moderate correlation with FC. In this study, we aim to evaluate which LS parameters have greater correlation with FC.

Methods A retrospective, two-center study; 74 patients who underwent SBCE within 7 (median 1.5) days from a FC measurement. LS was calculated; univariate and multivariate analyses were performed, investigating LS correlation with FC, and which LS parameters had stronger correlation coefficient (rs) with FC.

Results 74 patients had an FC measurement within 7 days of their SBCE examination (median 22 time-interval: 1.5 days; IQR: 5). Coefficient rs between LS and FC was moderate (0.454). In univariate analysis, the variables that gave the strongest association with FC were: the higher tertile subscore for ulcer, the summative ulcer subscore, the higher tertile ulcer score (only with descriptors of ulcer size and number), the summative ulcer score (only with descriptors of ulcer size and number), and subscores including various combinations of the stenosis descriptors. In multivariate analysis, the only positive predictor for FC was the higher tertile ulcer subscore (only with descriptors of ulcer size and number).

Conclusion LS shows only moderate correlation to FC. This is due to a) an inherent limitation of LS, and b) the notion of correlating the 2 parameters, and consideration should be given to development of a new, simplified (or composite) inflammation score/index for SBCE.

Keywords Capsule endoscopy, calprotectin, small-bowel, Lewis score, inflammation

Introduction

Crohn’s disease (CD) is a chronic form of inflammatory bowel disease of the gastrointestinal (GI) tract [1,2]. The diagnosis of small-bowel (SB) CD often presents a significant challenge. To that end, capsule endoscopy (CE) offers an enhanced diagnostic yield (DY) [3]. However, CE ‘suffers’ from lack of specificity as many of the detected SB mucosal changes are generic changes of inflammation [4,5]. Furthermore, minor SB abnormalities can be observed in up to 1 in 5 asymptomatic individuals [3,6].

The use of CE in monitoring the extent and activity of SB inflammation was – up until recently – defined by a lack of standardization in the reported SB mucosal inflammation [7]. In 2008, two scoring indices were developed. Niv et al developed and later validated the CE CD activity index (CECDAI) [8,9]. This index evaluates three parameters of SB pathology namely inflammation, extent of disease and presence of a stricture. Almost concurrently, Gralnek et al developed another scoring system – ever since known as the Lewis score (LS) – based on evaluation of the following parameters: villous edema, ulceration and stenosis (all weighed based on the extent and severity) [10]. Recently, we showed that LS performs better than the CECDAI at describing SB inflammation [11].
Fecal calprotectin (FC), a protein complex of the S-100 family [12], is considered an accurate and non-invasive marker of intestinal inflammation [12,13] with homogenous distribution in feces [14]. Therefore, spot samples of <5 g of fecal material are extremely reliable [15]. FC shows excellent correlation with excretion of 111-indium-labeled leukocytes in feces and can predict abnormal SB radiology [16], hence it is considered to be the ‘gold standard’ for defining gut inflammation [12,14,15].

Recently, it was shown that the measurement of FC – prior to referral for CE – is a useful clinical selection tool. A FC >100 μg/g is good predictor of positive CE findings, while FC >200 μg/g is associated with a higher CE positive DY (65%) [17]. However, that study had a significant methodological limitation in that we had allowed a 30-day interval between the CE study and FC measurement [7]; while relatively short, this time-gap may be significant due to subtle and spontaneous changes in the degree of active SB inflammation [18]. Furthermore, it has been suggested that components pertinent to the LS (weighing of parameters and/or final score calculation based only on the most severely affected tertile) may also “interfere” with attempts for a “desirable” correlation with a surrogate marker of SB inflammation, i.e. FC [7]. The primary endpoints of this retrospective study was to assess the correlation between LS, its elements (subscores), and FC. As secondary endpoint we were interested in checking possible difference between LS/FC correlation between 2 different SBCE systems.

Materials and methods

This was a 2-center, retrospective study. Participating centers were the Royal Infirmary of Edinburgh (RIE, Edinburgh Scotland, UK) and Skåne University Hospital (Sneh, Malmö, Sweden). University hospitals and tertiary referral centers for small bowel disease in Sweden), University hospitals and tertiary referral centers for the CE for the south east of Scotland and Skåne prefecture, respectively. A combined computerized search of the SBCE and biochemistry databases (March 2005 to December 2010) was performed in order to identify patients who had an FC measurement within 7 days around their CE examination. Part of the RIE small-bowel capsule endoscopy cohort has been used in a previous publication [11].

Table 1 The parameters and descriptors of Lewis score (LS) used to grade inflammation in small-bowel capsule endoscopy

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small bowel segmentation</td>
<td></td>
</tr>
<tr>
<td>Villi</td>
<td>Normal/edematous, Short/long/whole tertile, Single/patchy/diffuse</td>
</tr>
<tr>
<td>Ulcers</td>
<td>Non/single/few/multiple, Short/long/whole tertile, &lt;25%, 25%-50%, &gt;50%</td>
</tr>
<tr>
<td>Stricture</td>
<td>Non/single/few/multiple, Ulcerated/non-ulcerated, Traversed/non-traversed</td>
</tr>
</tbody>
</table>

Further selection was performed by checking the electronic records for previous endoscopic work-up. Only those subjected to an (ileo)-colonoscopy to exclude active inflammation in another part of the GI tract (with either unremarkable findings or irrelevant findings/pathology e.g. polyps that were dealt before obtaining the stool sample for FC measurement) in the 12-month period prior to the SBCE were included for further analysis. This was done to eliminate any concerns that an abnormal FC result could be the effect of a GI condition other than SB inflammation.

CE

SBCE was performed with the PillCam®SB1/SB2 (Given Imaging Ltd, Yoqneam, Israel) and the MiroCam® (IntroMedic Co., Seoul, South Korea) capsule endoscopes using the predefined – for each unit – procedure’s protocol (simethicone ± prokinetic) [19]. Certified gastroenterologists (AK, ET) with extensive experience in SBCE interpretation and LS calculation, blinded to the FC results, reviewed the CE sequences for LS calculation. The sequence review was performed using the RAPID® 7 software and the MiroView™ (2nd version) software. The SBCE video sequences were not de-identified.

LS

For LS calculation the SB transit time is divided into tertiles. A short segment is defined as occupying <10% of the tertile; a long segment is defined as occupying between 10–50% of the tertile; and a whole segment is defined as occupying >50% of the tertile, (Table 1) [10]. A total LS is created with the following mathematical calculation:

Maximum tertile score \(\{((villous \text{ parameter } x \text{ extent } x \text{ descriptor}) + (ulcer \text{ parameter } x \text{ extent } x \text{ size)})\} + (stenosis \text{ number } x \text{ ulcerated } x \text{ traversed})\) + (ulcer parameter x extent x descriptor) for tertile or (villous parameter x extent x descriptor) + (ulcer parameter x extent x descriptor) or (villous parameter x extent x descriptor) + (ulcer parameter x extent x descriptor) + (stenosis number x ulcerated x traversed). Hence, LS was calculated (using the necessary quantitative and qualitative descriptors) with the help of the LS screen incorporated into the RAPID® 7 software.

Score of LS components (parameters) per tertile i.e. villous score, ulcer (total and only number/size descriptors) score and stenosis (total, only ulcer and only number/traversed) score were registered in an Excel (Microsoft Office® 2014) spreadsheet. Thereafter, the correlation of the total LS and the aforementioned subscores with FC was calculated.

FC

FC, in both centers, was measured using CALPRO Calprotectin ELISA (Enzyme-Linked Monoclonal – polyclonal antibody combination immunosorbent assay) test in accordance to the manufacturer’s (CALPRO AS, Lysaker, Norway) instructions. The normal range for FC has been well defined as <50 μg/g; levels <20 μg/g are consistent with non-
detectable calprotectin in stools. To assist with analysis, by convention we replaced all $\text{FC}<20\ \mu\text{g/g}$ with 0.

**Statistical analysis**

Individual characteristics of the patients were recorded. Continuous data are presented as mean ± standard deviation (SD) and/or median; interquartile range (IQR). The Student’s $t$-test is used to compare the means of two samples. Multivariate logistic regression analysis was used to identify subscores’ association with FC. Independent parameters of LS were first analyzed by univariate analysis using the log-rank test in the Kaplan-Maier model (setting the FC variable as “dependent”). All variables from the univariate analysis with a $P<0.05$ were included in a Cox proportional hazards regression using the stepwise selection method. Spearman’s rank correlation coefficient ($r_s$) was used to assess the correlation between LS or the simplified LS and FC. Therefore, the strength of correlation was defined as follows: $r_s$ values ≤0.1 were considered no correlation; 0.1–0.3 weak to modest; 0.3–0.49 moderate; 0.5–0.79 strong; and ≥0.8 very strong correlation [20]. A two-tailed probability $P$ value <0.05 was considered to be statistically significant. Statistical analyses were carried out with a statistical package program for Windows (StatsDirect® version 2.7.8 Software, StatsDirect Ltd., Altrincham, Cheshire, UK).

**Ethics**

This study was carried out in accordance with the World Medical Association Declaration of Helsinki, 1964 (incorporating all later amendments). After review by the local ethics committee of each participating center, further specific ethical review and approval were not required, as the study was considered a retrospective audit work using data obtained as part of regular patient care.

**Results**

The database searches identified 1,765 patients that had a SBCE performed in the two centers in the aforementioned period. Of those, 74 patients (55 RIE/19 MUH; 20M/54F) had an FC measurement within 7 days of their SBCE examination (median time-interval: 1.5 days; IQR: 5). Mean age of the patients was 42±18.19 years. Twenty six ($n=26$) SBCE were performed with MiroCam® and 48 with PillCam® SB1/SB2. Sixty-nine ($n=69$; 93.24%) SBCE were complete to cecum; only 2 capsules were delivered endoscopically in the duodenum. Median FC level was 127.5 μg/g; IQR: 280 μg/g and the median LS was 135; IQR: 450. The correlation coefficient $r_s$ between LS and FC was 0.454, denoting only modest correlation (Fig. 1).

When the SBCE system used in each exploration was taken into account, there was no statistical difference in LS and FC ($P=0.919$ and 0.771, respectively) between the 2 groups. Furthermore, no difference in LS/FC correlation was seen when the type of SBCE device (MiroCam® or PillCam® SB) was taken into account; $r_s$ 0.523 and 0.486, respectively ($P=0.67$).

To elucidate which LS subscores/parameters have the stronger association with FC, univariate and multivariate analyses were performed. Association univariate analysis results are described in detail in Table 2. In univariate analysis, the variables that gave the strongest association with FC were the higher tertile subscore for ulcer ($P=0.001$), the summative (tertiles 1, 2, and 3) ulcer subscore ($P=0.001$), the higher tertile ulcer score (only with descriptors of ulcer size and number) ($P<0.0001$), the summative (tertiles 1, 2, and 3) ulcer score (only with descriptors of ulcer size and number) ($P<0.0001$), and subscores including various combinations of the stenosis descriptors. In multivariate analysis, the only positive predictor for FC was the higher tertile ulcer subscore (only with descriptors of ulcer size and number), which outperformed the rest of the variables ($P=0.0003$). The correlation coefficient between the simplified version of LS (including only one sub-score) and FC was 0.41, once more denoting only modest correlation.

**Discussion**

CE has shown a superior ability to visualize small mucosal ulcers, often missed by other imaging modalities [6], and displays and incremental DY ranging between 23-40% [2]. However, up until a few years ago, there was a lack of standardization in describing the extent and severity of such lesions [3,7]. In 2004, the first scoring index was developed, maturing to the LS in 2008 [10]. Certain endoscopic variables such as edema, erythema, and nodularity are doubtful markers of inflammation severity. Hence, LS was developed as a tool to describe CE lesions and standardize/quantify mucosal damage reporting. An LS <135 denotes normal or clinically insignificant mucosal inflammatory change; 135 ≥ LS < 790 designates mild; and an LS ≥790 severe mucosal inflammation.

LS was initially field-tested using CE studies with inflammatory changes. The individual parameters and descriptors that synthesized it were subjectively ranked in
order of severity. Thereafter, values for each parameter were created using descent gradient methodology [10]. In LS, the SB transit time is divided in 3 equal parts, creating tertiles scored separately [10]. Due to concerns that adding tertile scores will lead to potential upgrading of mild cases, the official LS uses the score of the most severely involved tertile.

However, there are certain drawbacks to utilizing LS in CE. For instance, strictures are the most heavily weighted element of the LS (for example, a simple singe ulcer yields a LS score of 135, while a single traversed luminal stricture is scored 18-fold higher). Furthermore, there is scoring discrepancy as a patient who has involvement in all 3 tertiles may have LS equal to a patient who only has limited SB involvement [1,7].

Calprotectin, a 36kDa protein, was first isolated in the 1980s from granulocytes; initially named L1 protein, it was later given its current name upon identification of its calcium/zinc-binding and antimicrobial properties. Calprotectin is a cytosolic protein and is released from neutrophils upon activation or death. It is resistant to bacterial degradation and stays stable in fecal material for up to 7 days [12,14]. FC concentrations have a narrow normal range [12,15]. FC is a valuable tool to differentiate patients with organic and functional bowel syndromes [15]. A cut-off threshold of 50 μg/mg of feces has been selected as the upper limit of normal [12,15]. Thus, both FC and LS are markers of intestinal inflammation and on their own are unable to be specific markers of SB CD. A previous study by our group has shown that LS performs well in describing the paucity of SB inflammation at levels of FC <100 μg/g, but at higher levels of FC there was little or no correlation [11]. However, there are potential explanations behind this mismatch; for instance, that possibility of a small active ulcer producing the same amount of FC to a sizeable chronic cobblestone mucosa patch or that of a fibrous stenosis/villous edema ability to produce/’leak’ FC should be considered. Furthermore, how accurate is the use of the tertile involvement for ulcers (as it is the individual mucosa break that yields FC), is there any other pathology/parameter left out of the current LS, such as mucosal erythema, capable of producing FC and likely the missing link and lastly, is the arbitrary use of frame quadrants to describe mucosal ulcer size enough to reflect their FC production capacity?

In the present study we attempted to re-address those issues. Therefore, we hypothesized that FC release is more likely to happen in areas of active mucosal inflammation and continuity breaks such as ulcers, cobblestone mucosa and/or erosions. Furthermore, we accept that FC remains the ‘gold standard’ of SB inflammation and that in order improve FC and LS correlation, certain LS components i.e., ulcers (number and size), and/or ulcerated stenoses should produce more FC than e.g., villous edema even if the whole tertile has edematous villi. Furthermore, on this occasion, we used a bigger group of patients and we were very rigorous in regards to the time frame between FC measurement and SBCE examination.

At first, the results of our analysis confirmed the fact that the correlation between the official LS and FC, considered as the gold standard of SB inflammation, are at the best modest (rs=0.454) [20]. Recently, Hoog et al recently showed that in 30 patients with known or suspected CD and inflammatory small changes (median FC 151 μg/g) the correlation between FC and LS was strong (rs=0.54) [22]. However, it should be noted that a different enzyme-linked immunosorbent, monoclonal antibody assay (Bühlmann Laboratories, Switzerland) was used and the interval between CE and FC was not reported (although we assume that the FC was performed at the time of inclusion in the study). Hence, in order to further investigate the correlation of separate elements/components of LS with FC, we undertook a univariate/multivariate regression analysis. In the multivariate analysis, only the higher tertile ulcer subscore (containing only descriptors of number and size) retained association with FC levels (rs=0.41). This is certainly supporting our hypothesis, based on the pathophysiology of FC product ion. Calprotectin constitutes ~60% of the soluble cytosol proteins of macrophages and upon activation and/or endothelial adhesion of monocytes is released in the gut lumen. Therefore, the number and surface size of mucosal breaks should be the main sites ‘leaking’ calprotectin in the intestinal lumen. However, the detection of a spot sample of FC cannot be considered as ideal because inflammation is a continuum or a spectrum of events and the FC might express a spot frame of this series of events.
This study has 2 main clinical applications:

a. The ulcer component of LS reflects equally well the official LS FC levels. When reporting examinations using CE software without integrated LS calculator, a ‘simplified LS’ (consisting only of the higher tertile ulcer score (number + size) is a reasonable substitute to the official LS; and,

b. The FC production sites are areas with breaks in mucosal continuity. By inference, any thoughts for using FC levels as a sole surrogate marker of SB patency and/or selection tool of patients for patency check with Agile capsule or cross-sectional imaging should be reconsidered. Chronic, fibrotic stenoses have limited capacity in FC production. Therefore, although in our cohort no retention was observed (2 capsules that did not reach the cecum by the end of the battery life were eventually excreted within the 2 weeks of ingestion, in both FC >250), FC cannot be recommended as surrogate patency marker.

We thought that it would be of great clinical interest if a modified/simplified LS is shown to retain, or even better, if it shows improved correlation with FC levels. This would lead to a simpler calculation and a wider application of LS in SBCE examinations read with other than the RAPID® software.

Despite our best efforts to ameliorate the conditions from previous studies this study also has limitations. Patients’ number, despite enrichment with patients from MUH, remains relatively small. Also, the fact that a single operator from each center performed the LS calculations may be seen as a relative weakness of our study. The use of 2 CE systems with different technical characteristics and specifications is another factor to consider. Lastly, although the same ELISA kit used in this study was the same in both centers, lab and temporal factors may produce further bias.

In conclusion, LS shows only moderate correlation to FC. This seems to be an inherent limitation of LS, as well as that of the notion of correlating the 2 parameters, and consideration should be given to the development of new, simplified inflammation score/index for SBCE.

References

17. Koulaouzidis A, Douglas S, Rogers MA, Arnott ID, Plevris JN. Fecal calprotectin: a select ion tool for small bowel capsule endoscopy in...
suspected IBD with prior negative bi-directional endoscopy. Scand J Gastroenterol 2011;46:561-566.


