**Caustic ingestion: development and validation of a prognostic score**

**Authors**
Joan Tosca1, Rosana Villagrasa1, Ana Sanahuja1, Ana Sanchez1, Galo Antonio Trejo1, Belen Herreros2, Isabel Pascual1, Pilar Mas1, Andres Peña1, Miguel Minguez1

**Institutions**
1 Department of Digestive Medicine, Hospital Clínic Universitari, Universitat de Valencia, Valencia, Spain
2 Department of Digestive Medicine, Hospital de la Marina Baixa de la Vila Joiosa, Alicante, Spain

**submitted** 18.5.2020
**accepted after revision** 23.10.2020
**published online** 23.10.2020

**Bibliography**
Endoscopy 2021; 53: 784–791
DOI 10.1055/a-1297-0333
ISSN 0013-726X
© 2020. Thieme. All rights reserved.
Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

**ABSTRACT**

**Background** Caustic ingestion is a potentially severe condition and early identification of poor outcome is essential to improve management; however, prediction based on endoscopy alone can overestimate severity. This study aimed to develop and validate a prognostic score.

**Methods** A prospective cohort study was designed to include all consecutive patients aged >15 years who presented with caustic ingestion between 1995 and 2017. Adverse outcome was defined by intensive care unit admission, urgent surgery, or death. The predictive value of clinical, analytical, and endoscopic variables was assessed in the first cohort (derivation cohort) and a prognostic score based on the resulting risk factors was developed by logistic regression. Internal validation (bootstrapping) was performed and then external validation was checked in an independent sample of patients (validation cohort).

**Results** 469 cases of caustic ingestion were included, 265 in the derivation cohort and 204 in the validation cohort. Ingestion of acidic substances (odds ratio [OR] 3.13, 95% confidence interval [CI] 2.33–4.21), neutrophil count (OR 1.05, 95%CI 1.04–1.06), metabolic acidosis (bicarbonate value, OR 0.82, 95%CI 0.78–0.85), and endoscopic injury (OR 3.81, 95%CI 3.35–4.34) were independent risk factors for poor outcome. The prognostic score based on these variables provided better accuracy than endoscopy alone ($P=0.04$), with high sensitivity, specificity, positive and negative predictive values (93.3%, 92.7%, 72.7%, 98.5%, respectively), and area under the curve (0.976, 95%CI 0.973–0.979; $P<0.001$).

**Conclusions** This score allowed a reliable prognosis of caustic ingestion and was more accurate than endoscopy-based evaluation.

**Introduction**
Ingestion of corrosive substances is a medical problem characterized by wide clinical expression. In the acute setting, some patients are asymptomatic, while other episodes lead to life-threatening situations, intensive care unit (ICU) admission, urgent surgery, and even death. Early detection of patients with a poor prognosis is key to accurately managing these cases; however, recognizing an adverse outcome is not always straightforward and reliable prognostic scores are lacking.

Almost all published articles studying prognostic factors are retrospective and assess only certain features of the disease.
Moreover, data on the predictive value of the symptoms are inconclusive, with some studies questioning the association between presence of symptoms and outcome of caustic ingestion [1–4], while others attribute a predictive value to the existence of some symptoms [5–14]. Endoscopic evaluation has been the mainstay of prognosis since the publication of the Zargar series [1, 2, 15], although some weaknesses are inherent to this approach. Isolated endoscopic evaluation may overestimate severity by failing to consider either the depth of the injury in the gastroesophageal wall or systemic involvement [16]. Moreover, endoscopy may sometimes be contraindicated, usually in the most severe cases where a precise prognosis is most necessary.

We hypothesized that the combination of clinical, endoscopic, and analytical data could improve predictive ability and enable a reliable prognosis during the acute period.

Our study aims were: 1) to identify clinical, analytical, and/or endoscopic parameters that determine poor clinical outcome; 2) to develop and validate a prognostic model that enables early detection of cases with poor prognosis; and 3) to compare the accuracy of the model with that of endoscopy alone.

Methods
Study design
A prospective and longitudinal cohort study was designed, and two cohorts of patients were sequentially recruited to develop and validate a prognostic score. The hypothesis and the study variables were defined before the start of the study, at the study design stage before patient recruitment. Data collection was subsequently conducted according to this initial protocol in both patient cohorts. The first cohort (derivation cohort) included patients who presented with caustic injury between January 1995 and September 2009. The second cohort (validation cohort) comprised patients with caustic ingestion from October 2009 to May 2017. The study was performed at the Hospital Clinic Universitari de Valencia, a tertiary care center in Valencia, Spain.

The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki and its later amendments, as reflected in approval by the Clinical Research Ethics Committee of the Hospital Clinic Universitari de Valencia.

Patients
Inclusion criteria for the study population were patients aged > 15 years who had accidentally or voluntarily ingested any kind of substance with caustic properties. The exclusion criteria were nonsignificant intake of a very low volume, specifically in situations where the caustic substance made contact with the oral cavity but was ejected without being swallowed, and patients who declined to participate in the study. These criteria applied to both cohorts.

Patients who met these criteria were identified from our institutional registry, in which every case of caustic ingestion presenting at our hospital has been prospectively recorded since 1995, including patients who attended the emergency service of our hospital as well as those referred from primary care or another tertiary center. The recruitment was systematic and all patients who met the selection criteria were consecutively included in the derivation or validation cohorts. Baseline features of both cohorts were compared in order to detect any selection bias.

As some patients were exposed to caustic injury more than once, we considered the number of cases of caustic ingestion in our analysis rather than the number of patients. As there were no major advances in the management of caustic ingestion over the study period, the same endoscopy-based treatment was applied in both cohorts.

Variables
The main outcome of the study was adverse evolution, defined as ICU admission, urgent surgery, or death, during the acute period of caustic injury. Chronic complications such as esophageal or gastric strictures and esophageal cancer were not considered in the study.

The indications for ICU surveillance were persistent hemodynamic failure despite fluid replacement, need for central venous pressure monitoring, respiratory failure despite oxygen administration, and need for invasive ventilation. Urgent surgery was performed if perforation was proven in endoscopic or radiological examinations or in the event of clinical suspicion (progressive clinical worsening) despite the absence of endoscopic or radiological confirmation.

Surviving patients were monitored for at least 3 months after discharge, to exclude any later complication related to the caustic injury. The main outcome was ascertained after this 3-month follow-up period in both cohorts by a different researcher from the one who collected data for potential predictors.

Sex, age, and history of gastroenterological and psychiatric disorders were considered as potential predictors. The characteristics of caustic ingestion (i.e. type of substance, volume, dilution, and reason for ingestion) were also evaluated. The remaining clinical variables examined were presence of symptoms, along with evaluation of oral, pharyngeal, and laryngeal injury by an otolaryngologist. We collected different analytical data such as hemoglobin (g/dL), leukocytes (×10⁹/L), platelets (×10⁹/L), urea (mg/dL), creatinine (mg/dL), electrolytes (mmol/L), and arterial (if oxygen saturation was <95%) or venous blood gas with pH and bicarbonate (mEq/L) in order to test these factors as potential prognostic markers. According to the study protocol, urgent endoscopy was proposed to all patients, regardless of the severity of caustic ingestion. Therefore, endoscopy was performed except in cases of patient refusal or contraindication, specifically severe otolaryngologic injury, persistent hemodynamic or respiratory failure despite treatment or suspected perforation; endoscopic injury was graded according to Zargar’s classification [Fig. 1] [15]. All these variables were gathered during emergency room attendance in both cohorts, according to a data collection protocol designed before the start of the study.
Sample size

The study size was calculated on the basis of the following data: the expected specificity of the prognostic score was 86.4%, that is 15% higher than Zargar’s grade [15]; the selected confidence level was 95% and the margin of error was 5%; the estimated ratio between favorable evolution and adverse outcome was 6 (this ratio was based on the included cases after the first 2 years of recruitment). According to Epidat software (version 4.2; Galician Health Service [Xunta de Galicia, Spain], PAHO-WHO, and CES University [Medellín, Colombia]), 181 cases of good clinical evolution and 31 cases of poor outcome were expected. We also expected that endoscopy would not be performed in 15% of cases. Thus, 213 cases of favorable evolution and 37 cases of adverse outcome were needed; when the required numbers were reached, recruitment for the derivation cohort was stopped and recruitment for the validation cohort was started.

Statistical analysis

Quantitative variables were reported as mean and standard deviation for normally distributed data, and as median and interquartile range for non-normally distributed data. Kolmogorov–Smirnov was applied as the normality test for a sample size $\geq 50$, and Shapiro–Wilk test was selected for a smaller sample.

Bivariate analysis was performed with qualitative variables by means of Yates’ chi-squared test or Fisher’s exact test when more than 20% of expected values were $\leq 5$; multiple comparisons of variables were adjusted using the Holm–Bonferroni method. Ordinal variables were analyzed by linear-by-linear association and quantitative variables were compared using the Student’s $t$ test for independent samples or the Mann–Whitney $U$ test if normal distribution could not be assumed. For quantitative variables, strength of association was calculated by simple binary logistic regression for binary outcomes, and was reported as odds ratio (OR) with 95% confidence interval (CI).

Bivariate analysis was performed in cases belonging to the derivation cohort in order to identify the potential determinants of poor prognosis. Variables resulting from this first step were considered for multivariate analysis. A multiple binary logistic regression model (Wald’s test, complete-case analysis) was manually constructed with the data from the same derivation cohort, to identify the risk factors for adverse outcome. Selected variables were those that: 1) were considered clinically...
more important; 2) enabled the building of a model that best explained the outcome with the fewest variables; and 3) reduced multicollinearity. The presence of multicollinearity was suspected if the correlation coefficients among covariates were > 0.7 (ordinal variables were considered continuous to this purpose). Multiple linear regression was performed with the different independent variables to calculate their tolerance and their variance inflation factor. Tolerance values < 0.2 and variance inflation factor values > 3 were considered indicative of possible multicollinearity.

The logistic regression model was evaluated according to statistics of goodness of fit (deviance and Hosmer–Lemeshow test) and measures of predictive power, such as R-square and area under the receiver operating characteristic (ROC) curve (AUC).

Missing data were assumed to occur at random depending on the severity of ingestion (e. g., asymptomatic patients with low-volume intake of a weak caustic substance may lack blood tests or endoscopy). For multivariate analysis, multiple imputation was performed in predictors that were recorded in < 95% of cases. Missing data were estimated by means of logistic regression and linear regression for dichotomous and quantitative variables, respectively. Multiple imputation was based on an iterative Markov chain Monte Carlo method and Rubin’s rules were used to combine the estimates of each imputed dataset into one overall estimate. The variables included in the multiple imputation model were age, type of caustic substance, neutrophil count, pH, grade of endoscopic injury, and main outcome. A total of 20 different imputed data sets were created and combined to obtain an overall estimate of each regression coefficient and a model performance measure.

As the main goal of the study was prognostic rather than explanatory, quantifiable values were managed as continuous variables for the bivariate analysis, but for greater applicability they were also dichotomized for logistic regression analysis. Cutoff points were chosen depending on the result of the ROC curve analysis, to provide the best combination of sensitivity and specificity (Youden’s index).

Depending on the regression coefficient values derived from the logistic regression analysis performed with these categorical variables, a score was assigned to each predictor to obtain the final prognostic index. The prognostic model was validated twice: first, internal validation was accomplished by a resampling method (bootstrapping, with 2000 samples) with data from the derivation cohort. Once internal validation was confirmed, the prognostic index was applied in the validation cohort and predicted outcomes were compared with observed outcomes in order to assess external validation. Accuracy was measured by sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and AUC analysis. The diagnostic accuracy of the prognostic index and the endoscopic lesion alone (same endoscopic cutoff, IIb in esophagus and III in stomach) were compared by McNemar’s test; the comparison of both ROC curves was performed by nonparametric test for related samples. All P values were two sided and the global significance level was set at P ≤ 0.05. Statistical analysis was performed using IBM SPSS Statistics, version 26.0 (IBM Corp, Armonk, New York, USA). Results were reported following the TRIPOD statement [17, 18].

**Results**

During the recruitment period, 645 episodes of caustic ingestion were attended to at our hospital and assessed for eligibility; 469 met the selection criteria and were included in the study, 265 in the derivation cohort and 204 in the validation one (Fig. 2). Baseline characteristics of patients and distribution of predictors and outcomes were similar in the two cohorts (see Table 1 in the online-only supplementary material).

**Adverse outcome**

Overall, 61 cases (13.0%, 95% CI 10.0 – 16.0) progressed unfavorably over the study period, 39 (14.7%, 95% CI 10.4 – 19.0) in the derivation cohort and 22 in the validation cohort (10.8%, 95% CI 6.5 – 15.1). In the derivation cohort, 38 cases (14.3%, 95% CI 10.1 – 18.6) required ICU admission, 8 (3.0%, 95% CI 0.9 – 5.1) underwent surgery, and 14 (5.3%, 95% CI 2.6 – 8.0) died. Both cohorts had similar clinical characteristics (Table 1).

**Predictors of adverse outcome**

Age (OR 1.02, 95% CI 1.01 – 1.04) and ingested volume (OR 1.01, 95% CI 1.00 – 1.01) progressively increased the risk of poor evolution, but acidic ingestion was the strongest predictor
of adverse outcome (OR 22.6, 95% CI 9.6—52.9) in univariate analysis. Voluntary intake, neutralization, dilution, and psychiatric disorders were confounding factors in multivariate analysis. All demographic and clinical data are summarized in Table 2.

All symptoms were associated with a higher risk of complication (Table 3). Leukocytosis (OR 11.8, 95% CI 5.3—26.3 for leukocytes ≥15 × 10⁹/L), neutrophilia (OR 10.0, 95% CI 4.7—21.3 for neutrophil count ≥10 × 10⁹/L), and metabolic acidosis (OR 15.5, 95% CI 6.8—35 for pH < 7.35) were analytical markers of severity of caustic intake. Degree of metabolic acidosis was related to the nature and volume of the ingested caustic substance, but also to the severity of the endoscopic lesion (Fig. 1).

Endoscopy was performed in 239 cases in the derivation cohort. In the remaining 26 cases, endoscopy was not performed due to patient refusal (15 patients) or contraindications (11 cases). Contraindication to endoscopy implied an increased risk of poor outcome (OR 17.9, 95% CI 4.5—71.1).

The risk of complications increased with degree of endoscopic injury, but not in a linear way. The threshold of injury that increased the risk of an adverse outcome was determined by Zargar’s grade: IIb for esophagus and III for stomach and duodenum (Table 5).

Although results were not significant in the derivation cohort, major complications also depended on the extent of the gastric lesion when analysis was performed with the whole series (OR for pangastric vs. localized injury was 3.4, 95% CI 1.7—6.7).

Model development

The acidity of the caustic substance (OR 3.13, 95% CI 2.33—4.21), neutrophil count (OR 1.05, 95% CI 1.04—1.06), bicarbonate value (OR 0.82, 95% CI 0.78—0.85), and degree of endoscopic injury (OR 3.81, 95% CI 3.35—4.34) proved to be the model with the best predictive ability (Table 6). For all these independent variables, correlation coefficients were < 0.7, tolerance values were > 0.6, and variance inflation factors were < 1.7 (Table 7).

The internal validation of these variables was checked by bootstrapping with 2000 samples (Table 6); once their internal validity was verified, continuous variables were dichotomized and a score was assigned for each item depending on their regression coefficient, in order to convert the model into a prognostic index (Table 1, Table 2). In this resulting index, the scores of each predictor must be added to obtain the final result.

Model performance

The goodness of fit of the predictive model was assessed by deviance (98.9) and Hosmer–Lemeshow test (5.69, P = 0.224); its predictive ability was measured with R-square (0.782) and the ROC curve (Fig. 3a–c), which defined an AUC of 0.976 (95% CI 0.973—0.979).

A cutoff of 4 points was chosen according to Youden’s index, so that a sum of ≥4 points implies an adverse outcome. This value provided both sensitivity and specificity of > 92% in the derivation cohort, with PPV and NPV of 75% and 98.2%, respectively.

The diagnostic ability of the prognostic model was then calculated in the validation cohort, where it performed similarly (95.2% sensitivity and 91.7% specificity), with a PPV of 69% and an NPV of 99% (Table 6).
The diagnostic performance of this prognostic score was higher than that of endoscopic injury alone. Combining endoscopic and analytical data improved detection of poor outcome cases. Considering only cases in which endoscopy was performed, for the same endoscopic cutoff points (IIb in the esophagus or III in the stomach), sensitivity, specificity, PPV, and NPV results from endoscopy alone were 89.8%, 90.3%, 56.4%, and 98.4%, while the diagnostic ability of the index was higher (93.8%, 93.1%, 71.4%, and 98.8%, respectively) in this setting. Overall, the accuracy of the prognostic index (93.1%, 95%CI 90.5–95.8%) was superior to that of endoscopy alone (90.5%, 95%CI 87.6–93.3%; \( P=0.04 \)). The AUC of the score calculated by the prognostic index (0.976, 95%CI 0.96–0.993) was also higher than the prediction based solely on the endoscopic damage (0.897, 95%CI 0.85–0.944) (\( P<0.001 \)) (▶Fig. 3d). In cases of contraindication to endoscopy, the prognostic index identified all cases of poor evolution (100% sensitivity).

Discussion

These results show that predicting the outcome of a caustic ingestion episode by simple parameters is feasible and safe, and that adding analytical data and the chemical nature of the caustic substance to the prognostic assessment improves on the traditional endoscopy-based approach.

One limitation of the study is the slightly low number of poor outcome cases, despite a long recruitment period. This limitation is inherent to the selection criteria, as all patients with caustic ingestion were consecutively included without any restriction based on severity of the ingestion; this condition reduced the prevalence of complications to 13%, lower than in most published series. This rate of poor evolution is nevertheless sufficient to ensure a ratio of 10 cases per predictor and may also better reflect the severity of caustic ingestion, possibly even enhancing the applicability of the predictive score.

Despite the prospective design, the occurrence of missing data could not be avoided, largely because patients referred from other centers in our healthcare area were included, but also because in the milder cases, not all planned complementary tests were performed. To prevent any bias from a complete case analysis, multiple imputation was performed, taking into account this latter condition, thus minimizing the potential impact of missing values in the prognostic model [17, 18].

The prognostic accuracy of the score in the validation cohort was similar to the performance in the derivation cohort. Although the swallowed volume was slightly lower and voluntary intake was less frequent in the validation cohort, these differences did not imply lower clinical severity, so both cohorts can be considered comparable. Although external validation confirms the predictive capacity of the model, the study was conducted in a Western population. The Asian series have a higher proportion of acidic caustic ingestion due to greater availability [15, 19–21]. Before applying this prognostic index in an Eastern population, it would be appropriate to verify its performance in a population in which acid ingestion predominates.

The high accuracy of the model is possibly due to the combination of endoscopic damage and other markers, such as the chemical nature of the caustic substance, the inflammatory response triggered by the injury (leukocytosis), or the degree of transmural involvement (acidosis). In essence, the model supplements the information provided by the mucosal lesion with its transmural and systemic involvement, making it easier to discern which lesions with severe mucosal involvement entail worse progression.

In line with this idea, some studies have shown that involvement of the deepest layers of the digestive tract is a marker of severity. Although deep lesions detected by endoscopic ultra-

---

**Table 1** Prognostic model.

<table>
<thead>
<tr>
<th>Prognostic model with categorical variables</th>
<th>( \beta )</th>
<th>OR</th>
<th>95%CI</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid caustic</td>
<td>1.61</td>
<td>4.99</td>
<td>1.58–15.70</td>
<td>0.006</td>
</tr>
<tr>
<td>Neutrophil count ( \geq 75 % )</td>
<td>1.86</td>
<td>6.42</td>
<td>1.99–20.77</td>
<td>0.002</td>
</tr>
<tr>
<td>Bicarbonate ( \leq 22 \text{ mEq/L} )</td>
<td>2.13</td>
<td>8.44</td>
<td>2.84–25.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Esophageal injury ( \geq ) IIb or contraindicated endoscopy</td>
<td>1.34</td>
<td>3.81</td>
<td>1.19–12.17</td>
<td>0.024</td>
</tr>
<tr>
<td>Gastric injury ( \geq ) III or contraindicated endoscopy</td>
<td>2.63</td>
<td>13.83</td>
<td>4.27–44.79</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

\( \beta \), regression coefficient; OR, odds ratio; CI, confidence interval.

**Table 2** Prognostic score.

<table>
<thead>
<tr>
<th>Points</th>
<th>( \beta )</th>
<th>OR</th>
<th>95%CI</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intake of acidic caustic substance</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutrophil count ( \geq 75 % )</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bicarbonate ( \leq 22 \text{ mEq/L} )</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe endoscopic injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Esophagus ( \geq ) IIb</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stomach ( \geq ) III</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In case of contraindication for endoscopy, add both items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \geq 4 ) points imply an adverse outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
sound have been linked to a worse prognosis, this result has been more consistent in the long term for predicting stenosis rather than immediate complications [22, 23]. Nonetheless, radiological transmural necrosis evidenced by computed tomography (CT) has been associated more clearly with a worse evolution during the acute period and with higher rates of surgical intervention [24, 25]. According to our score, integrating mucosal assessment with the analytical expression of the transmural lesion improves the ability to identify cases with worse evolution and overcomes the limitations of an evaluation based on endoscopy alone. As leucocytosis and acidosis are very effective in identifying severe cases, we recommend performing these analyses whenever clinical complications are suspected (impaired clinical status, respiratory involvement, or severe endoscopic injury).

Several studies have evaluated the predisposing factors for poor evolution after caustic damage [26–28], the impact of symptoms on prognosis [1 – 11, 14, 29], and the predictive ability of endoscopy [2, 23, 30 – 34]. However, several of these studies contain methodologic limitations owing to their retrospective design, possible selection biases (some studies included only admitted patients or those who had undergone endoscopy), and the fact that most of their results have not been validated.

Another advantage of the score is that it quantifies and evaluates the burden of contraindication to endoscopy; to our knowledge, this issue has not been previously explored and yet it improves prediction performance and usefulness.

Some authors have proposed CT as the first examination to be performed after caustic ingestion [35]; however, the data are based on a series with a probable selection bias, with a large number of severe ingestions and critical patients who finally underwent surgery (20%). This strategy does not seem appropriate as a systematic method in real populations with a lower prevalence of surgery (3.2%), in which the PPV of CT is therefore reduced. In this scenario, our score could be helpful in identifying patients with a poor prognosis and who would benefit from a CT scan.

This score is simple and can be easily applied with limited data; furthermore, it improves accuracy over endoscopy alone and can even be performed in cases of contraindication to endoscopy. Its main utility is in establishing an early and reliable prognosis of the course of the episode. Nevertheless, given the heterogeneous evolution of caustic ingestions, more prognostic indicators are needed to select the best approach in each case and to establish evidence-based management of caustic ingestions.

Acknowledgment

We would like to dedicate this work to Adolfo Benages (1942 – 2012) – much more than an author.

Competing interests

The authors declare that they have no conflicts of interest.

References