The binomial symptom index for the evaluation of temporal association between cardiorespiratory symptoms and gastroesophageal reflux in neonates

Alejandro Barriga-Rivera, María José Moya, Manuel Lopez-Alonso *

Children’s Hospital Virgen del Rocío, Sevilla, Spain

Received 2 July 2015; accepted 7 September 2015
Available online 8 December 2015

KEYWORDS
Gastroesophageal reflux; Symptom association analysis; Binomial symptom index; Symptom association probability; Neonates; Cardiorespiratory; Anti-reflux surgery

Abstract
Introduction: The evaluation of symptom association between gastroesophageal reflux and cardiorespiratory events in preterm infants remains unclear. This paper describes a conservative approach to decision-making of anti-reflux surgery through symptom association analysis. Methods: Forty-three neonates with potentially reflux-related cardiorespiratory symptoms underwent synchronized esophageal impedance-pH and cardiorespiratory monitoring. Three indices were considered to evaluate symptom association, the symptom index (SI), the symptom sensitivity index (SSI) and the symptom association probability (SAP). A conservative strategy was adopted regarding the decision of anti-reflux surgery, and therefore, patients were scheduled for laparoscopic Nissen fundoplication if the three indices showed a positive assessment of symptom association. Retrospectively, these indices and the binomial symptom index (BSI) were contrasted against the decision of anti-reflux surgery using different windows of association. Results: Thirteen patients showed positive symptom association but only two underwent anti-reflux surgery. The SI and the SSI showed an increasing trend with the width of the window of association. The SAP was affected randomly by slightly altering the windowing parameters. The BSI showed the best performance with the two-minute window ($k = 0.78$).
Introduction

Gastroesophageal reflux (GER) can be described as the retrograde movement of the gastric content into the esophagus, having or not regurgitation or vomiting. It is a physiological condition and occurs in infants, children and adults several times a day. Most episodes are short and asymptomatic and occur during transient lower esophageal sphincter relaxations. However, when GER produces symptoms or complications one can refer to GER disease (GERD). These symptoms are occasionally nonspecific such as persistent crying, irritability, sleeping disorders, growth delays and respiratory complications, and may potentially cause neonatal death.

According to Tsoukali and Sifrim, GERD is the most costly gastrointestinal disease with a prevalence ranging between 10% and 20% in the western world. Combined multiple intraluminal impedance-pH (MII-pH) monitoring allows for a more accurate detection of GER episodes and facilitates the study of symptom association (SA). Particularly in preterm infants, cardiorespiratory (CR) episodes such as apnea, bradycardia or oxygen desaturation, may be temporally related to GER if they occur within a given time interval around the onset of a GER episode. It has been a subject of extensive research since the mid-1980s. Nevertheless, in the second decade of the 21st century, this discussion remains open.

Different statistical indices have been proposed in order to quantify this relationship: the symptom index (SI), the symptom sensitivity index (SSI) and the symptom association probability (SAP). However, none of them seem to have solved this problem and therefore surgical treatments may be inappropriately recommended as the

Conclusions: The pathology under study is known to improve with maturity. However, the severity of cardiorespiratory symptoms may threaten the neonate’s life and therefore, on some occasions, invasive treatments must be considered to protect life. The BSI provides a good prediction of a combination of positive SI, SSI and SAP, which may improve clinical decisions. However, further clinical studies are required to prove the BSI as an optimal predictor of clinical outcomes.

© 2015 Asociación Española de Pediatría. Published by Elsevier España, S.L.U. All rights reserved.
result of an overinterpretation of the current metrics. On the other hand, the binomial symptom index (BSI)\(^2\) fell by the wayside due to its apparent mathematical complexity.\(^1\) However, a recent computational analysis\(^10\) has shown that an approximation of the BSI it can provide an optimal measurement of SA. This paper highlights the shortcomings of SA in GER and presents the results of a clinical study of the evaluation of the temporal association between CR symptoms and GER in pediatric patients.

**Methods**

**Patients**

Forty-three preterm neonates of gestational age varying between 24.0 and 36.0 weeks (median 29.0 weeks) and weighing 0.73–3.18 kg (median 1.31 kg) were considered for this study between 2004 and 2008. At the time of recruitment, their postnatal age varied between 13 and 119 days (median 36.5 days). All showed recurrent CR symptoms, potentially related to GER, and were included in the study as recommended by Vandenplas and co-workers.\(^1\) Twenty-five of them (11 females and 14 males) developed symptoms during the study and were the subjects under research, whereas the remaining 18 did not present any CR symptoms during and after the study. Therefore these patients were not considered under this methodology. This research was approved by the Ethics Committee of the Children’s Hospital Virgen del Rocio (CHVR), Seville, Spain. Parents or legal guardians signed an informed consent form prior to enrolling. Patients were hospitalized at the neonatal intensive care unit of the CHVR under oxygen-enriched atmosphere (21–30%). One hour of nasogastric feeding was followed by a 2-h fasting period cyclically during hospitalization.

**Instrumentation**

All patients underwent 24-h MII-pH and CR monitoring. Six-french transnasal catheters were used to monitor MII-pH (Femu, Aachen, Germany).\(^3\) The probe had seven impedance electrodes separated 10 mm from each other and two pH electrodes separated 5 cm from each other. The pressure inversion point (PIP) was determined manometrically. Thus, the proximal pH electrode was placed at 20 mm from the PIP, with the distal electrode located inside the stomach. The distal impedance electrode was located at the position of the PIP. On the other hand, heart rate (HR), respiratory rate (RR) and oxygen saturation (SpO\(_2\)) were recorded using a bedside patient monitor. MII-pH and CR signals were acquired and digitalized at 50 Hz per channel with a National Instruments DAQCard-700. Synchronization was carried out by simultaneously sampling all channels every 20 ms. GER events were identified according to the recommendations provided by the ESPGHAN and the NASPGHAN,\(^1\) that is, a drop in the pH value below 4 and/or a drop of 50% of the impedance base line in two consecutive channels in the retrograde direction. Then, pathologic GER was determined when the reflux index was greater than 7% or when the number of reflux episodes exceeded 70 episodes in 24 h.\(^3\) Similarly, CR episodes were considered when the SpO\(_2\) fell below 80%, when breathing was suspended for at least 20 s or when the heart rate dropped below 80 systolic cycles per minute. SA was analyzed computationally using numerical analysis software (Matlab 2008).

**Clinical methodology**

A total of 25 patients were evaluated for SA. In doing so, the SI, the SSI and the SAP were calculated using a 2-minute window. Those patients having none of these indices with a positive value (SA−) were studied for reflux only and excluded from SA analysis. Patients with a positive value in at least one of the indices (SA+) were carefully assessed, adopting a conservative strategy at all times. Patients were scheduled for laparoscopic Nissen fundoplication when the three estimators were positive, as evidence of SA was proven. However, when only one or two indices were positive, the patient underwent medical treatment and was controlled carefully. All subjects with SA+ were followed up during five years as shown in Fig. 1. At the time of the study, the BSI was compared with the previous methodology recalculating all the indices using both a one-minute and a two-minute window of association.

**Statistical analysis**

The three major metrics for SA were evaluated to determine the existence of a temporal relationship between GER and CR events: the SI, the SSI and the SAP. Additionally, the BSI\(^10\) was added and compared retrospectively against anti-reflux surgery. The indices were calculated varying the length of the window of association between 1 and 300 s. In addition, different temporal offsets were added to the start of the windowing to study the effect time shifts. Although the current recommended width of the window of association is 120 s,\(^22,25\) a one-minute window of association was also studied. Thus, the Pearson’s correlation coefficient (r) was applied to contrast agreement between 60 and 120 s windows with a level of confidence of 95%. The Cohen’s kappa (κ) coefficient was used to evaluate agreement between the indices and the decision of anti-reflux surgery based on the combination of positive SI, SSI and SAP.

**Symptom index**

The SI was first introduced by Ward et al.\(^20\) as the ratio between the number of symptoms that are related to a reflux episode (N\(_{RBS}\)) and the total number of symptoms (N\(_s\)) during time of study, as in Eq. (1). A positive association can be considered if SI ≥ 50%, according to studies conducted in adults.\(^17\)

\[
SI = \frac{N_{RBS}}{N_s} \times 100
\]  

(1)

**Symptom sensitivity index**

The SSI represents the positive predictive value taking into account the total number of reflux episodes,\(^17,21\) Eq. (2) defines this index as the ratio of those symptom-related reflux episodes (N\(_{RSS}\)) to the total number of reflux occurrences (N\(_R\)). It is considered as positive if SSI ≥ 10%.\(^21\)

\[
SSI = \frac{N_{RSS}}{N_R} \times 100
\]  

(2)
The binomial symptom index in neonates

Figure 1  The figure shows the algorithm followed in this study. Patients underwent synchronized multiple intraluminal impedance-pH and cardiorespiratory (MII-pH + CR) monitoring. Symptom association (SA) was evaluated and considered positive if at least one of the indices was positive. Otherwise, patients underwent reflux study only. Those having three indices with positive values were scheduled for anti-reflux surgery. All patients with SA+ were followed up for five years.

Symptom association probability
This index provides a measure of probability of association. The study is divided into temporal windows to form a 2×2 contingency array, and then the Fischer’s exact test is applied to obtain the probability of reflux and symptoms to occur by chance (p). Thus, Eq. (3) represents the degree of temporal association, as a measure of probability, in terms of the SAP:

\[ SAP = (1 - p) \times 100 \]  

(3)

This index denotes a statistical temporal association when \( SAP \geq 95\% \).

Binomial symptom index
The BSI was first described by Ghillebert et al. to determine SA in esophageal pH studies and has been recently studied using computational methods. The BSI can be calculated according to Eq. (4), where \( N_s \) represents the total number of symptoms, \( N_e \) is the total number of ebs, \( T \) is the monitoring time, \( w \) is the width of the window of association and \( N_a \) represents the number of reflux-related symptoms. The BSI was suggested to provide a positive assessment, in a 24-h monitoring study, when \( BSI \geq 99\% \).

\[ BSI = \left[ 1 - \sum_{i=N_a}^{N_s} \left( \frac{N_s}{i} \right) \left( \frac{N_R}{T} \right)^i \left( 1 - \frac{N_R}{T} \right)^{N_s-i} \right] \times 100 \]  

(4)

Results
The average monitoring time was 21.8±1.1h. The total number of reflux events ranged between 19 and 146 (mean 77.5) whereas the total number of symptom episodes varied between 1 and 46 (mean 15.0). In the first stage, at least one of the three main indices under study determined a positive SA in 13/25 patients as in Fig. 1. All the indices were positive in 3/25 subjects. These patients were scheduled for anti-reflux surgery. However, one of them stopped having CR symptoms spontaneously a few days before surgery and therefore a conservative approach was adopted. This patient had a good evolution and did not require surgery. On the other hand, 10/25 patients had one or two indices with a positive value. They were carefully studied and underwent eight weeks medical treatment, including proton pump inhibitors and prokinetic drugs to treat reflux, and theophylline and caffeine for the CR pathology. Afterwards, these patients were reevaluated for SA: five of them were studied for reflux only for not having shown CR symptoms again, whereas the other 5/25 had a negative SA but still presented some CR episodes. Therefore, these subjects were included in the five year follow-up period. After this time, all patients were free from CR symptoms. All subjects were operated by the same surgical team. Detailed results on SA are provided in the supplementary table.

Dissimilarities were observed between the SI, the SSI, the SAP and the BSI, as shown in the contingency matrix (Table 1). This fact was previously reported by Lühold et al. Note that Table 1 represents the number of positive values that occurred simultaneously for each two entries of the

Table 1
Figure 2 The scatter plots contrast the results of using one-minute and two-minute windows of association. Panels A–D are referred to the SI, the SSI, the SAP and the BSI respectively. Note that the numeric subscripts represent the length of the window of association in seconds. The dashed line depicts the linear regression model.

table. These data were used to contrast the performance of all the indices to suggest anti-reflux therapy, and in particular, laparoscopic Nissen fundoplication.

None of the patients had a positive association for all the indices considering both window sizes. There was no perfect agreement between surgery and SA analysis. However, the BSI, using the two-minute window of association, showed the best agreement in terms of the Cohen’s kappa (0.78), as shown in Table 2.

On the other hand, there was a good linear correlation between the one-minute and the two-minute windows regarding the SI ($r=0.91$), the SSI ($r=0.94$) and the BSI ($r=0.96$). However, it was poor in the case of the SAP ($r=0.67$) as shown in Fig. 2.

Fig. 3 shows an example of the evolution of each index with the window length. Clearly, the wider the window of association the higher the value of the SI and the SSI. Nevertheless, the SAP showed chaotic behavior. The SI and SSI curves exhibit an increasing non-linear profile when plotted against the window length whereas the SAP showed a random variability. On the other hand, the BSI was reported to show a maximum value when the window length matches the characteristic symptom-reflux lag. This can be appreciated in Fig. 3, although the typical value of this temporal lag remains under discussion, and therefore the

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Number of positive association cases in common between the indices, including surgery. The subscript denotes the width of the window of association for each index.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI$_{60}$</td>
<td>SSI$_{60}$</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>SURGERY</td>
</tr>
</tbody>
</table>

| SI$_{120}$ | SSI$_{120}$ | SAP$_{120}$ | BSI$_{120}$ | SURGERY |
| 10       | 3       | 6       | 3       | 2       | SI$_{120}$ |
| 6       | 6       | 3       | SSI$_{120}$ |
| 9       | 3       | SAP$_{120}$ |
| 3       | BSI$_{120}$ |
| 2       | SURGERY |

| SI$_{60}$ | SSI$_{60}$ | SAP$_{60}$ | BSI$_{60}$ | SURGERY |
| 8       | 2       | 3       | 4       | 2       | SI$_{120}$ |
| 1       | 5       | 3       | 2       | SSI$_{120}$ |
| 4       | 4       | 5       | SAP$_{120}$ |
| 1       | 2       | 1       | BSI$_{120}$ |
| 1       | 2       | 1       | SURGERY |

<table>
<thead>
<tr>
<th>Table 2</th>
<th>The table shows the agreement with the decision of anti-reflux surgery using 60-s and 120-s windows of association.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen’s kappa</td>
<td>60 s</td>
</tr>
<tr>
<td>SI</td>
<td>0.08</td>
</tr>
<tr>
<td>SSI</td>
<td>0.63</td>
</tr>
<tr>
<td>SAP</td>
<td>0.15</td>
</tr>
<tr>
<td>BSI</td>
<td>0.63</td>
</tr>
</tbody>
</table>
two-minute window is still considered as the optimal window of association.\textsuperscript{22,25}

Note that the effect of windowing the recording for the calculation of the SAP affects the results of the estimator.\textsuperscript{27} Therefore, the value of the SAP depends, at random, on the starting time of the monitoring as shown in Fig. 4. It is an example that represents the value of the SAP using a two-minute window when different time shifts were added to the onset of the windowing. The standard deviation of the SAP index, when varying the onset from zero to the width of the window, ranged between 0\% and 46.7\%. On the contrary, the other indices did not vary by altering the onset of the windowing.

Discussion

SA analysis is a mathematical technique to express temporal association between reflux and acute symptoms.\textsuperscript{17} The use of the SI, the SSI and the SAP is still questioned and the need for new methods is claimed by the scientific community.\textsuperscript{27,28,31} This study shows disagreement between the three main metrics and presents the BSI as an optimal estimation of SA in pediatric patients with CR symptoms. It is worth mentioning that the SI and the SSI can be calculated by hand whereas the SAP requires complex software.\textsuperscript{23} Although the BSI cannot be calculated by hand, parameters involved in the formulation of this index can be accounted easily. Then, common software tools can be used to compute the results of the index. This provides an alternative and more reliable method to calculate probability of association. Note that the temporal window was defined to study association only in the same way that the previous indices do. However, if one aims to study causality, latency between both events must be considered in one direction only as described in a previous study.\textsuperscript{30}
The existence of a temporal association between CR and GER is controversial. Both phenomena are common in preterm infants and are rarely related to each other. Thus, there are few recent scientific papers describing this problem<sup>24</sup> and include a scarce number of patients. This pathology is known to have a good spontaneous evolution as the patient ages.<sup>25</sup> The development of physiological barriers to GER with maturity explains these cases and supports the conservative approach adopted by the authors with regards to the use of anti-reflux surgery. However, due to the severity of these symptoms and the potential risk of neonatal death, anti-reflux surgery was, in some cases, the elected treatment to ensure the patient’s life. Approximately half of the infants with recurrent apnea have a positive diagnosis of GERD.<sup>26</sup> Further studies should consider the direction of the association to allow more reliable decisions in order to reduce hospital stays and to provide better therapies.

It is important to mention that the impedance-pH monitoring was carried out along with a nasogastric probe. This might have conditioned the results obtained in this study, especially those related to the diagnosis of GERD. Nevertheless, the control values in pediatric patients<sup>1</sup> were obtained with nasogastric probe as well, and therefore, both studies were conducted in similar conditions.

The complex nature of recruiting patients regarding this pathology makes research in this area a difficult task.<sup>27</sup> On the other hand, there is only one control group for pediatric gastroesophageal reflux reported in the scientific literature.<sup>2</sup> These two facts along with a poorly understood pathophysiology makes every contribution a valuable step. Note the number of patients that underwent fundoplication in this study (<i>n</i> = 2) cannot fully support conclusions regarding clinical outcomes. Anti-reflux surgery was decided based on a combination of a positive SI, SSI and SAP. That means the BSI predicts a combination of positive SI, SSI and SAP. Therefore, there is still a need for further studies to investigate the BSI as a predictor of clinical outcomes.<sup>26</sup>

This work supports the existence of an important inconsistency between the three indices under study.<sup>1,2</sup> Surgical treatments are often based on the evaluation of one, or a combination, of these indices (probability of association plus effect of size).

A very critical parameter in SA analysis is the time window in which symptom and GER are considered to be related. Lam et al.<sup>22</sup> stated that the optimal window length was 120 s. However, this has not been validated in neonates until a recent study by Omari et al.<sup>25</sup> It shows that two minutes is the optimal window length when calculating SAP in infants. However, the SAP was criticized on many occasions for being biased and affected by chance.<sup>26</sup> In this study, the 120-s window provided the best results for the BSI.

The study of the window of association concludes that the value of the SI and the SSI increase with the window size: the larger the window the greater the index; it occurs because the probability of a reflux to be symptom-related increases with the width of the window of association.<sup>22</sup> However, one cannot see the same behavior regarding the SAP, as shown in Fig. 2. Small variations of the factors involved in the windowing (onset and width) produce strong oscillations in the value of the SAP. On the contrary, the BSI has its maximum accuracy when the window of association matches the characteristic reflux-symptom delay.<sup>30</sup> The BSI was resembled and proved computationally to be an optimal estimator of symptom association.<sup>30</sup> Other authors<sup>27</sup> have considered the BSI as described by Ghillebert et al.<sup>29</sup> and concluded it is a good test for SA analysis. In this study the BSI seems to provide a good measurement of SA.

Conclusion

The evaluation of GER-CR association in infants remains an open problem. Current statistical indices, SI, SSI and SAP, do not provide a unique criterion for diagnosis: the SAP index presents multiple deficiencies whereas the BSI represents a good measurement of probability and size of effect. After having evaluated the BSI clinically against the decision of anti-reflux surgery, this study concludes the BSI simplifies the evaluation of SA in pediatric patients with CR symptoms. However, further investigations are required to prove the BSI as an optimal predictor of clinical outcomes.

Conflict of interest

The authors declare no conflict of interests.

Funding

This project (PI-0434-2010) was supported by the Department of Health of the Regional Government of Andalusia, Spain.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.anpede.2015.09.026.

References