Cureus
Part of SPRINGER NATURE

Open Access Original Article

Sting-to-Care Interval and Recorded Cardiorespiratory Complications After Scorpion Stings in Brazil, 2008-2022: A Nationwide Cross-Sectional Study of Surveillance Notifications

Review began 10/27/2026
Review ended 11/10/2026
Published 11/12/2026

© Copyright 2026
Carilo et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.119540

Guilherme A. Carilo [1], Jaqueline E. Beirith [2], Rafaella Ottaviani [3], Gabriel P. Andreotti [4], Ana Clara Aragão Fernandes [5], Pietra S. Wladimirski [6]

1. Faculty of Medicine, Universidade Anhembi Morumbi (UAM) - Campus Piracicaba, Piracicaba, BRA 2. General Practice, Faculdades Pequeno Príncipe (FPP), Curitiba, BRA 3. Faculty of Medicine, Universidade Nove de Julho (UNINOVE), São Paulo, BRA 4. Faculty of Medicine, Universidade de São Paulo, São Paulo, BRA 5. Internal Medicine, Hospital Universitário dos Servidores do Estado, Rio de Janeiro, BRA 6. Faculty of Medicine, Universidade de Santo Amaro, São Paulo, BRA

Corresponding author: Ana Clara Aragão Fernandes, claraaragao.provas@gmail.com

Abstract

Background

Cardiorespiratory complications after scorpion stings can be studied in surveillance data, but their measurement is heterogeneous. We assessed the association between the recorded sting-to-care interval and these complications in Brazil, including an analysis of recorded shock in Pará state.

Methodology

We analyzed scorpion sting notifications from the Brazilian Notifiable Diseases Information System (SINAN), 2008-2022. The exposure comprised six interval categories; the primary outcome was a composite of recorded shock or respiratory complication. Logistic regression was adjusted for age, sex, year, and federative unit, with municipality-clustered robust variance, and estimated standardized risk ratios (RRs).

Results

Of the 1,563,983 notifications, 1,296,606 entered the primary model, with 4,547 events. The interval was globally associated with the composite (Wald $p = 3.47 \times 10^{-14}$); RRs for 1-3, 3-6, 6-12, 12-24, and ≥24 hours versus 0-1 hours were 1.40, 1.70, 2.37, 3.30, and 1.98, respectively. All four secondary outcomes retained global associations after Holm correction. The association persisted without Pará, but a score bootstrap clustered by federative unit did not reach a p-value <0.05 (p = 0.0735). Pará accounted for 2,087 of 3,218 recorded shocks; its adjusted RR for shock fell from 79.46 under the broad definition to 0.97 (95% confidence interval = 0.39-2.38; eight events in Pará) under the strict definition.

Conclusions

In this exploratory analysis of surveillance records, the recorded interval was associated with recorded cardiorespiratory complications, with non-monotonic estimates; evidence varied with the inferential procedure and outcome definition. The heterogeneity of shock records in Pará calls for clinical validation and improved recording; it does not demonstrate a causal effect of delay, a shock mechanism, or individual recording error.

Categories: Environmental Health, Emergency Medicine, Epidemiology/Public Health
Keywords: brazil, public health surveillance, respiratory insufficiency, scorpion stings, shock, time to treatment

Introduction

Scorpionism is a major public health problem in Brazil because of the frequency of stings and their potential to progress to severe complications and death. In a bulletin published in 2025, the Ministry of Health reported 202,714 scorpion sting notifications for 2023 in the Notifiable Diseases Information System (SINAN), 131 of which recorded death from the sting. Case fatality, the proportion of deaths among notified cases, was approximately 0.06% nationally and 0.20% among children aged up to nine years. These figures underline the need to examine factors associated with unfavorable outcomes across age groups [1].

Recent national studies have described rising notification counts, a heterogeneous geographical distribution, and associations between sting occurrence and indicators of social vulnerability. Because they focus on territorial patterns, these ecological analyses help identify priority areas, but they do not assess the association between time to care and cardiorespiratory complications at the notification level [2,3]. In a

national case-control study of children aged up to 10 years, an interval of three hours or more to treatment was associated with death; this evidence, however, refers to the pediatric population and to the fatal outcome [4]. A national analysis covering all age groups can therefore complement these findings by examining recorded cardiorespiratory complications specifically, together with their regional heterogeneity.

The clinical relevance of this question lies in the capacity of severe envenoming to cause cardiovascular dysfunction, shock, respiratory failure, and pulmonary edema. National guidelines emphasize early recognition of systemic manifestations and timely care [5]. Surveillance data, however, have measurement limits: the sting-to-care interval does not capture the time to antivenom administration, the shock field does not specify its hemodynamic mechanism, and a single field combines respiratory failure and acute pulmonary edema. These features justify assessing the internal coherence of the records when interpreting the associations observed, without equating them with individually confirmed clinical diagnoses [6].

The primary objective of this study was to analyze the adjusted association between the sting-to-care interval and recorded cardiorespiratory complications in scorpion sting notifications in Brazil from 2008 to 2022, across all age groups. The secondary objectives, addressed in exploratory analyses, were to assess the regional heterogeneity of these associations and the internal coherence of shock records, with a specific analysis of Pará state.

Materials And Methods

Study design

This was a nationwide, retrospective, observational cross-sectional study of secondary data from compulsory notification. Reporting followed the STROBE and RECORD recommendations [7,8].

Setting and data source

We used de-identified scorpion sting notifications recorded in SINAN, obtained from the preserved national annual files classified as final for 2008-2022 and accessed on September 3, 2026 [6,9-11]. The Mortality Information System (SIM) was used only for aggregate comparison of deaths, without individual linkage between the two databases [1,12].

Study population

The population comprised people treated for a scorpion sting in Brazil and notified to SINAN between 2008 and 2022. The unit of analysis was the notification. We included notifications with a valid age, sex recorded as female or male, a valid notification location, and an evaluable exposure and outcome. Identical records were retained because there was no way to establish that they represented duplicate episodes.

Variables

The exposure was the time elapsed between the sting and care, categorized as 0-1, 1-3, 3-6, 6-12, 12-24, and ≥24 hours, with 0-1 hour as the reference and no assumption of a linear trend across categories.

The primary outcome was a composite that required a recorded systemic complication together with at least one of two components: shock or respiratory complication. When the systemic complication field was positive, any positive component defined an event, even if the other component was absent or invalid. Non-events were records with a positive systemic complication field and both components explicitly negative, or with a negative systemic complication field and both components blank. All other patterns were classified as not evaluable. The respiratory field combines respiratory failure and acute pulmonary edema; the shock field does not identify the hemodynamic mechanism.

Secondary outcomes, analyzed separately with their own denominators, were respiratory complication, shock, severe final classification, and death attributed to the sting versus recovery. Deaths from other causes were excluded from the mortality denominator (Appendix, Table S10).

The adjustment covariates were age group (nine categories), sex, notification year, and federative unit (the 26 states and the Federal District). Sex was the female/male field recorded in the notification form. Initial severity was not available in the system. Final severity classification and antivenom use were not included in the adjustment.

Quantitative variables

In the main analysis, time to care and age were analyzed as categorical variables (six interval categories and nine age groups), and year as a categorical variable, without assuming a linear relationship with the outcome.

Data sources and measurement

All variables were extracted from the fields of the SINAN notification form. The complete rules used to classify the exposure, the composite outcome, and the non-evaluable patterns are provided in the Appendices. Outcomes were not imputed.

Bias control

To reduce confounding by regional and temporal differences, the model included fixed effects for the 27 federative units and for notification year. Possible bias from incomplete records was assessed by weighting for the probability of observation. In the main models, standard errors accounted for the clustering of notifications within municipalities.

Study size

All eligible notifications in the period were included, without a prior sample size calculation.

Statistical methods

In brief, the analysis had a primary component and secondary, exploratory components. The primary analysis estimated how the probability of the recorded composite outcome differed across interval categories after accounting for age, sex, calendar year, and federative unit. Logistic regression was used because the outcome is binary, and its estimates were converted by standardization into adjusted probabilities, risk ratios (RRs), and risk differences (RDs), which express the results directly on the probability scale. Because notifications from the same municipality may share reporting practices and care resources, standard errors were made robust to clustering by municipality. Secondary outcomes were analyzed with the same approach or, when the logistic model could not be estimated, with weighting. Sensitivity analyses repeated the primary analysis under alternative assumptions to show how much the findings depended on particular states, model choices, missing data, and the level of clustering. The Pará-specific comparison of shock definitions and the interaction analyses were secondary and exploratory.

The main analysis used logistic regression with the sting-to-care interval, age group, sex, year, and federative unit as explanatory variables. Standard errors accounted for clustering by municipality, with the CR1 small sample correction. From the model, marginal standardization over the covariate distribution of the analyzed population (g-standardization) yielded adjusted probabilities, RRs, and RDs for each category relative to 0-1 hour, with nominal 95% confidence intervals (CIs).

The global association between interval and composite outcome was assessed with the Wald test. The five RR contrasts relative to 0-1 hour were corrected with the Holm procedure; the four global tests for secondary outcomes formed a separate family. This correction controls for multiple comparisons but does not allow the associations to be interpreted as causal.

Respiratory complication and severe classification were analyzed by logistic regression. For shock and death, the absence of events in some federative units produced separation in the logistic model with federative-unit fixed effects; we therefore used inverse probability weighting for the interval category, keeping all federative units, incorporating the uncertainty of the weight model parameters into the variance, and applying CR1 correction by municipality.

Sensitivity analyses assessed geographical exclusions, alternative model specifications, missing data, and dependence between observations. They included weighting for the probability of observation, conditional logistic regression by municipality, and a score test with a bootstrap clustered by federative unit, which differs from the main Wald test in both the statistic and the clustering level. The age × interval and Pará × interval interactions formed a third, exploratory family, also corrected by Holm. For the age × interval interaction only, we used five age groups (<5, 5-9, 10-19, 20-59 and ⩾60 years) and two interval groups: codes 1-2 versus 3-6, labeled ⩽3 and >3 hours. The Pará × interval interaction kept the six interval categories.

In a secondary, exploratory Pará-specific analysis, six partially nested definitions of shock were compared in a common denominator, adjusted for age, sex, interval, and year, with clustering by municipality. Definition D1 used hierarchical shock; D2 to D5 combined clinical markers; and D6 required shock, severe classification, systemic manifestation, and at least one of respiratory complication or attributed death. Systemic manifestation and systemic complication are distinct fields of the form. These definitions assessed convergence among themselves, without diagnostic validation or a demonstrated hierarchy of specificity.

The analytical protocol was drafted after exploration of the same data. A subsequent reproduction was carried out in R 4.6.1 (data.table *1*.18.6.1; survival 3.8.6 for the exact conditional estimator), using the same preserved sources, without prospective pre-registration or an independent sample. Operational details, equations, model diagnostics, and protocol amendments are provided in the Appendices. The literature search was targeted and non-systematic, and no claim of absolute novelty is made.

Ethical aspects

Because the study used secondary, public, and de-identified data, it was exempt from review by a Research Ethics Committee under Resolution No. 510/2016 of the Brazilian National Health Council [13].

Data availability

The source data are public. SINAN national final files on venomous animal accidents were obtained from the DATASUS file transfer service (https://datasus.saude.gov.br/transferencia-de-arquivos/), and SIM national annual mortality files from the Ministry of Health open data catalogue (https://dadosabertos.saude.gov.br/dataset/sim), both accessed on September 3, 2026. Only public, de-identified data were used, and no new data were generated; all derived estimates are reported in the article and the Appendices. The Appendices document the integrity checks of the preserved source files (SHA-256 signatures; Section S1), the complete outcome coding rules (Section S2), and the software and package versions used, identifying which analyses come from the original implementation and which from the later R reproduction (Section S9). The R scripts used for data extraction and analysis are publicly available in the Open Science Framework (OSF) repository (https://doi.org/10.17605/OSF.IO/67UXK).

Results

Of the 1,563,983 scorpion sting records notified between 2008 and 2022, 1,296,606 (82.90%) met the inclusion criteria. Of these, 4,547 presented the composite outcome, defined as a systemic complication accompanied by shock or respiratory complication. Notifications were concentrated in the 0-1 and 1-3-hour care categories. Record completeness varied across federative units. Excluded records differed from included records mainly in federative unit, missing race/color information, and anatomical site of the sting (Figure *1*, Tables *1*, *2*; Appendix, Tables S8, S9).

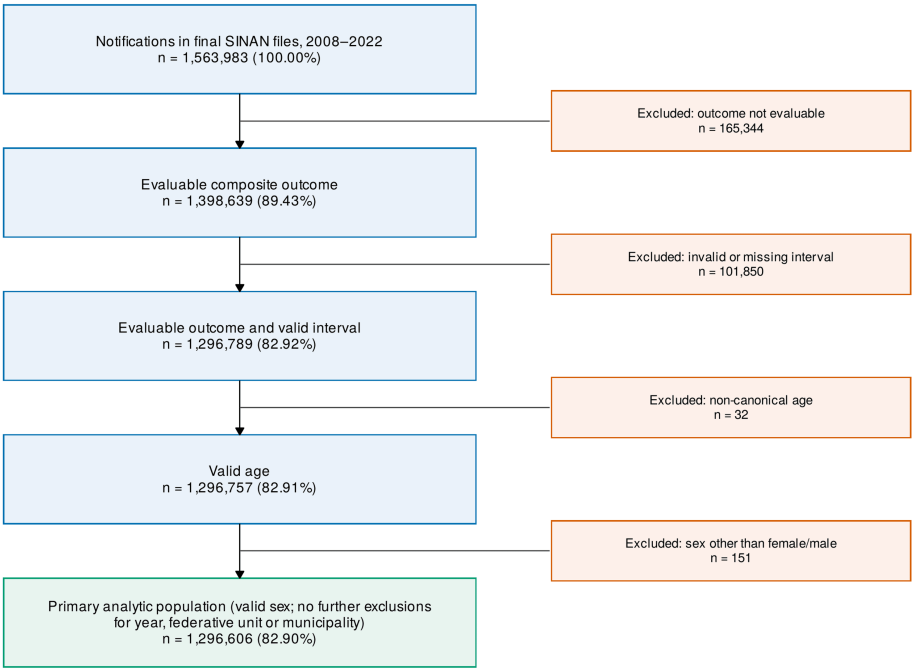


FIGURE 1: Sequential selection flow for the primary analytic population among scorpion sting notifications in the final SINAN files, Brazil, 2008-2022.

Exclusions at each arrow are sequential. Marginal outcome and interval missingness overlap and therefore should not be added.

SINAN: Brazilian Notifiable Diseases Information System

Characteristic	0–1 hours	1–3 hours	3–6 hours	6–12 hours	12–24 hours	≥24 hours	Total
Primary analytic population	807,190	321,697	85,803	38,592	25,839	17,485	1,296,606
Age group, n (%)							
<1 year	9,892 (1.2)	4,236 (1.3)	1,170 (1.4)	418 (1.1)	301 (1.2)	219 (1.3)	16,236 (1.3)
1–4 years	39,437 (4.9)	15,642 (4.9)	3,649 (4.3)	974 (2.5)	657 (2.5)	450 (2.6)	60,809 (4.7)
5–9 years	50,822 (6.3)	19,413 (6.0)	4,107 (4.8)	1,172 (3.0)	843 (3.3)	572 (3.3)	76,929 (5.9)
10–14 years	55,300 (6.9)	20,925 (6.5)	4,361 (5.1)	1,523 (3.9)	1,118 (4.3)	736 (4.2)	83,963 (6.5)
15–19 years	65,267 (8.1)	23,705 (7.4)	5,614 (6.5)	2,310 (6.0)	1,668 (6.5)	1,208 (6.9)	99,772 (7.7)
20–39 years	259,055 (32.1)	95,739 (29.8)	24,979 (29.1)	11,976 (31.0)	7,963 (30.8)	5,617 (32.1)	405,329 (31.3)
40–59 years	210,720 (26.1)	88,646 (27.6)	25,730 (30.0)	12,415 (32.2)	8,204 (31.8)	5,282 (30.2)	350,997 (27.1)
60–79 years	103,821 (12.9)	47,272 (14.7)	14,258 (16.6)	6,937 (18.0)	4,476 (17.3)	3,000 (17.2)	179,764 (13.9)
≥80 years	12,876 (1.6)	6,119 (1.9)	1,935 (2.3)	867 (2.2)	609 (2.4)	401 (2.3)	22,807 (1.8)
Sex, n (%)							
Female	407,170 (50.4)	158,532 (49.3)	39,996 (46.6)	18,986 (49.2)	13,618 (52.7)	8,614 (49.3)	646,916 (49.9)
Male	400,020 (49.6)	163,165 (50.7)	45,807 (53.4)	19,606 (50.8)	12,221 (47.3)	8,871 (50.7)	649,690 (50.1)
Calendar period, n (%)							
2008–2012	113,977 (14.1)	60,580 (18.8)	17,790 (20.7)	8,034 (20.8)	5,417 (21.0)	3,254 (18.6)	209,052 (16.1)
2013–2017	232,740 (28.8)	98,768 (30.7)	27,012 (31.5)	12,337 (32.0)	8,380 (32.4)	5,437 (31.1)	384,674 (29.7)
2018–2022	460,473 (57.0)	162,349 (50.5)	41,001 (47.8)	18,221 (47.2)	12,042 (46.6)	8,794 (50.3)	702,880 (54.2)
Geographic group, n (%)							
Pará	8,896 (1.1)	7,410 (2.3)	4,001 (4.7)	1,650 (4.3)	823 (3.2)	471 (2.7)	23,251 (1.8)
Other federative units	798,294 (98.9)	314,287 (97.7)	81,802 (95.3)	36,942 (95.7)	25,016 (96.8)	17,014 (97.3)	1,273,355 (98.2)
Observed events by outcome-specific denominator, n/N (%)							
Primary cardiorespiratory composite	1,756/807,190 (0.2175)	1,399/321,697 (0.4349)	664/85,803 (0.7739)	362/38,592 (0.9380)	266/25,839 (1.0295)	100/17,485 (0.5719)	4,547/1,296,606 (0.3507)
Recorded respiratory component	931/807,171 (0.1153)	600/321,685 (0.1865)	171/85,798 (0.1993)	120/38,590 (0.3110)	160/25,838 (0.6192)	48/17,484 (0.2745)	2,030/1,296,566 (0.1566)
Severe classification	7,701/856,962 (0.8986)	4,788/343,602 (1.3935)	1,530/91,869 (1.6654)	508/41,238 (1.2319)	259/27,545 (0.9403)	216/18,752 (1.1519)	15,002/1,379,968 (1.0871)
Recorded shock	1,004/807,158 (0.1244)	974/321,674 (0.3028)	541/85,795 (0.6306)	299/38,589 (0.7748)	183/25,837 (0.7083)	67/17,483 (0.3832)	3,068/1,296,536 (0.2366)
Attributed death	541/832,987 (0.0649)	402/331,939 (0.1211)	130/88,374 (0.1471)	54/39,744 (0.1359)	39/26,551 (0.1469)	34/18,031 (0.1886)	1,200/1,337,626 (0.0897)

TABLE 1: Characteristics and events by sting-to-care interval category in the primary analytic population.

Characteristics are derived exclusively from the verified final cells of the primary model; percentages use the column total. The five outcomes use outcome-specific complete denominators and should not be compared as though they shared the same population.

Criterion	Valid, n	Missing/Invalid, n	Completeness, %
Evaluable composite outcome	1,398,639	165,344	89.43
Valid interval (marginal)	1,426,329	137,654	91.20
Evaluable outcome ∩ valid interval	1,296,789	267,194	82.92
Canonical age (marginal)	1,563,940	43	100.00
Female/male sex (marginal)	1,563,653	330	99.98
Complete primary analytic set	1,296,606	267,377	82.90

TABLE 2: Evaluability and missing data in the source population of 1,563,983 notifications.

Marginal missingness counts overlap: 35,804 records had both a non-evaluable outcome and an invalid interval; the union of these two domains was 267,194. Only 183 additional sequential exclusions occurred after that intersection. Race/skin color, anatomical site, and month of sting are not shown by interval because no final audited cross-tabulation exists for this table; they were not recomputed from microdata.

Time to care was associated with the composite outcome (global Wald test, $p = 3.47 \times 10^{-14}$). After adjustment for age, sex, year, and federative unit, the recorded categories from 1-3 to ⩾24 hours had a higher risk of the composite than the 0-1-hour category. All five comparisons remained significant after Holm correction. The largest estimated difference occurred in the 12-24-hour category, with a risk of 0.88% versus 0.27% at 0-1 hours (RR = 3.30, 95% CI = 1.62-6.75; RD = 0.61 percentage points, 95% CI = 0.002-1.22). Point estimates did not increase continuously across categories: they were higher at 12-24 hours than at ⩾24 hours, without a direct test between these two categories (Table *3*, Figure *2*).

Interval	N; events	Crude risk, %	Standardized risk, % (95% CI)	Standardized RR (95% CI)	RD, pp (95% CI)	Nominal p	Holm p
0–1 hours	807,190; 1,756	0.2175	0.2659 (0.2130–0.3320)	1.00 (reference)	0.0000 (reference)	—	—
1–3 hours	321,697; 1,399	0.4349	0.3731 (0.3018–0.4611)	1.40 (1.23–1.60)	+0.1071 (0.0604–0.1538)	7.85×10^{-7}	2.35×10^{-6}
3–6 hours	85,803; 664	0.7739	0.4508 (0.3563–0.5703)	1.70 (1.41–2.04)	+0.1849 (0.1049–0.2648)	3.32×10^{-8}	1.33×10^{-7}
6–12 hours	38,592; 362	0.9380	0.6311 (0.4829–0.8243)	2.37 (1.76–3.19)	+0.3652 (0.2024–0.5279)	1.17×10^{-8}	5.85×10^{-8}
12–24 hours	25,839; 266	1.0295	0.8785 (0.4367–1.7592)	3.30 (1.62–6.75)	+0.6125 (0.0019–1.2231)	0.0011	0.0011
≥24 hours	17,485; 100	0.5719	0.5274 (0.4223–0.6586)	1.98 (1.50–2.63)	+0.2615 (0.1409–0.3820)	1.94×10^{-6}	3.88×10^{-6}

TABLE 3: Sting-to-care interval and the recorded cardiorespiratory composite.

Robust global Wald test for the five time terms: $\chi^2(5) = 72.261$; $p = 3.47 \times 10^{-14}$. N = 1,296,606; 4,547 events. Logistic model with age group, sex, categorical year, and federative-unit fixed effects; municipality-clustered CR1 sandwich; risks standardized by g-computation. Crude risk = observed events/n. RD is in percentage points. Category-specific 95% CIs and p-values are nominal; Holm was applied to the five contrasts only after the global test. Findings are associational; point estimates are non-monotonic, and the difference between 12–24 hours and ≥24 hours was not tested directly. Marginal-probability 95% CIs were calculated using the delta method on the logit scale and back-transformed to the probability scale. RR CIs use the log scale; RD CIs use the identity scale. Crude proportions: Wilson CIs where indicated. Because RR and RD intervals are computed on different scales (log and identity), an RR interval that excludes 1 can coexist with an RD interval close to or including 0.

CI: confidence interval; CR1: cluster-robust covariance with finite-sample correction; IPTW: inverse probability of treatment weighting; pp: percentage points; RD: risk difference; RR: risk ratio

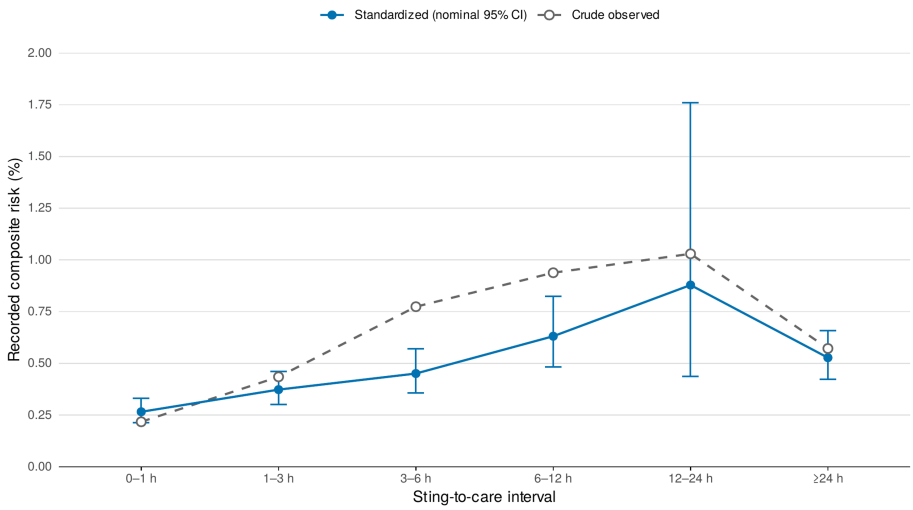


FIGURE 2: Crude observed and standardized risks of the recorded cardiorespiratory composite across the six sting-to-care interval categories.

Error bars show nominal 95% CIs for standardized marginal risks, calculated on the logit scale using the delta method and back-transformed to probabilities. Crude points are descriptive. Standardized risks can lie below crude risks in intermediate categories because standardization holds the covariate distribution, including federative unit, fixed across categories. The point-estimate pattern is non-monotonic, and the comparison between the final two categories was not directly tested; categories are intervals and do not support a causal per-hour interpretation.

CI: confidence interval

Time to care was also associated with each secondary outcome in the Holm-adjusted global tests: respiratory complication, severe classification, shock, and death attributed to the sting. For all of them, the 1-3 to ⩾24-hour categories had higher point estimates than 0-1 hours; the category-specific 95% CIs are nominal (Tables *4*-*6*). The relationship between time to care and the composite outcome varied with age (adjusted p = 0.000572). In children younger than five years, the >3 versus ⩽3-hour contrast yielded an RR of 0.98 (95% CI = 0.74-1.30), showing no evidence of a difference in this contrast and without demonstrating equivalence (Appendix, Table S4).

Outcome	N; events	Method	Wald $\chi^2(5)$	Raw p	Holm p (family of 4)
Recorded respiratory component	1,296,566; 2,030	Federative-unit fixed-effects logit, g-computation, and municipality CR1	82.681	2.30×10^{-16}	6.91×10^{-16}
Severe classification	1,379,968; 15,002	Federative-unit fixed-effects logit, g-computation, and municipality CR1	119.100	4.87×10^{-24}	1.95×10^{-23}
Recorded shock	1,296,536; 3,068	Multinomial time-category IPTW, Hájek means, and municipality-CR1 stacked sandwich	32.194	5.44×10^{-6}	5.44×10^{-6}
Attributed death	1,337,626; 1,200	Multinomial time-category IPTW, Hájek means, and municipality-CR1 stacked sandwich	72.550	3.02×10^{-14}	6.03×10^{-14}

TABLE 4: Secondary outcomes by sting-to-care interval: denominators, methods, and global tests.

Category-specific estimates, confidence interval conventions, and abbreviations for this outcome set are described in the note to Table 6.

Outcome	Interval	N; events	Crude risk, %	Adjusted risk, % (95% CI)	RR (95% CI)	RD, pp (95% CI)
Recorded respiratory component	0–1 hours	807,171; 931	0.1153	0.1156 (0.0928–0.1440)	1.00 (ref.)	0.0000 (ref.)
	1–3 hours	321,685; 600	0.1865	0.1753 (0.1501–0.2046)	1.52 (1.22–1.88)	+0.0597 (0.0303–0.0890)
	3–6 hours	85,798; 171	0.1993	0.1920 (0.1431–0.2578)	1.66 (1.17–2.36)	+0.0764 (0.0168–0.1360)
	6–12 hours	38,590; 120	0.3110	0.4045 (0.2514–0.6502)	3.50 (2.22–5.51)	+0.2889 (0.1034–0.4744)
	12–24 hours	25,838; 160	0.6192	0.8963 (0.3125–2.5428)	7.75 (2.82–21.32)	+0.7806 (−0.1538–1.7151)
	≥24 hours	17,484; 48	0.2745	0.4025 (0.2896–0.5590)	3.48 (2.30–5.26)	+0.2868 (0.1497–0.4239)
Severe classification	0–1 hours	856,962; 7,701	0.8986	0.8930 (0.8345–0.9556)	1.00 (ref.)	0.0000 (ref.)
	1–3 hours	343,602; 4,788	1.3935	1.3428 (1.1599–1.5540)	1.50 (1.31–1.72)	+0.4498 (0.2688–0.6308)
	3–6 hours	91,869; 1,530	1.6654	1.6394 (1.4399–1.8659)	1.84 (1.62–2.08)	+0.7464 (0.5471–0.9456)
	6–12 hours	41,238; 508	1.2319	1.5390 (1.2713–1.8620)	1.72 (1.44–2.07)	+0.6460 (0.3641–0.9279)
	12–24 hours	27,545; 259	0.9403	1.2861 (1.0482–1.5771)	1.44 (1.19–1.75)	+0.3931 (0.1441–0.6422)
	≥24 hours	18,752; 216	1.1519	1.5514 (1.3007–1.8496)	1.74 (1.46–2.07)	+0.6584 (0.3909–0.9259)
Recorded shock	0–1 hours	807,158; 1,004	0.1244	0.1788 (0.1139–0.2806)	1.00 (ref.)	0.0000 (ref.)
	1–3 hours	321,674; 974	0.3028	0.2606 (0.1725–0.3933)	1.46 (1.26–1.68)	+0.0818 (0.0417–0.1218)
	3–6 hours	85,795; 541	0.6306	0.3255 (0.2234–0.4741)	1.82 (1.48–2.25)	+0.1467 (0.0836–0.2099)
	6–12 hours	38,589; 299	0.7748	0.5073 (0.3464–0.7422)	2.84 (2.06–3.91)	+0.3285 (0.1817–0.4752)
	12–24 hours	25,837; 183	0.7083	0.4578 (0.2466–0.8481)	2.56 (1.50–4.37)	+0.2790 (0.0299–0.5281)
	≥24 hours	17,483; 67	0.3832	0.3585 (0.2275–0.5645)	2.01 (1.34–2.99)	+0.1797 (0.0493–0.3101)
Attributed death	0–1 hours	832,987; 541	0.0649	0.0705 (0.0611–0.0813)	1.00 (ref.)	0.0000 (ref.)
	1–3 hours	331,939; 402	0.1211	0.1123 (0.0961–0.1313)	1.59 (1.38–1.84)	+0.0418 (0.0269–0.0568)
	3–6 hours	88,374; 130	0.1471	0.1638 (0.1288–0.2083)	2.32 (1.83–2.95)	+0.0934 (0.0558–0.1309)
	6–12 hours	39,744; 54	0.1359	0.2011 (0.1413–0.2861)	2.85 (1.99–4.09)	+0.1306 (0.0604–0.2008)
	12–24 hours	26,551; 39	0.1469	0.1935 (0.1305–0.2868)	2.75 (1.84–4.09)	+0.1230 (0.0477–0.1983)
	≥24 hours	18,031; 34	0.1886	0.1881 (0.1296–0.2730)	2.67 (1.81–3.94)	+0.1176 (0.0474–0.1878)

TABLE 5: Secondary outcomes by sting-to-care interval: category-specific estimates.

Category-specific estimates, confidence interval conventions, and abbreviations for this outcome set are described in the note to Table 6.

CI: confidence interval; RD: risk difference; RR: risk ratio.

Specification	RR 6–12 hours (95% CI)	RR ≥24 hours (95% CI)
Main IPTW, 27 federative units	2.85 (1.99–4.09)	2.67 (1.81–3.94)
p1/p99-truncated IPTW, 27 federative units	2.35 (1.73–3.20)	2.66 (1.81–3.90)
Region fixed-effects logit, 27 federative units	2.49 (1.87–3.30)	3.55 (2.46–5.14)
Federative-unit FE logit, 25 informative units	2.33 (1.76–3.09)	3.49 (2.44–4.99)

TABLE 6: Secondary outcomes by sting-to-care interval: specification heterogeneity for death in later categories.

All category-specific 95% CIs are nominal and multiplicity-unadjusted; family-wise inference derives from the four Holm-adjusted global tests. Marginal-risk CIs use a logit transformation with the delta method; risk-difference CIs remain on the identity scale, where negative values are valid. Death estimates in later categories depend on specification/estimand: the main IPTW was higher at 6–12 hours, whereas the region and 25-federative-unit logits were higher at ≥24 hours; this is between-model heterogeneity, not uniformly demonstrated robustness. Marginal-probability 95% CIs were calculated using the delta method on the logit scale and back-transformed to the probability scale. RR CIs use the log scale; RD CIs use the identity scale. Crude proportions: Wilson CIs where indicated. Because RR and RD intervals are computed on different scales (log and identity), an RR interval that excludes 1 can coexist with an RD interval close to or including 0.

CI: confidence interval; CR1: cluster-robust covariance with finite-sample correction; IPTW: inverse probability of treatment weighting; pp: percentage points; RD: risk difference; RR: risk ratio

The association between time to care and the composite outcome was maintained when specific federative units were excluded, when adjustment variables were added, when the modified Poisson model was used, and when weighting for the probability of observation was applied (Appendix, Tables S2, S11). This weighting brought the distribution of observed characteristics in the analyzed records closer to that of the eligible target population. When Pará was excluded, the highest point estimate of composite risk shifted to the ⩾24-hour category (RR = 3.06, 95% CI = 2.10-4.47, relative to 0-1 hours) (Figure *3*; Appendix, Table S1). In the score test with a bootstrap clustered by federative unit, which differs from the main Wald test in both the statistic and the clustering, the association between time to care and the composite outcome was not significant at 5% (p = 0.0735). The analysis restricted to 2017-2022 did not produce finite adjusted estimates because Rondônia had no events in the eligible set. In conditional logistic regression, which compares notifications within the same municipality, 710 informative municipalities contributed data; its measures are conditional odds ratios (ORs), not marginal RRs (12-24 versus 0-1 hour: OR = 3.80, 95% CI = 1.78-8.11; Appendix, Table S12).

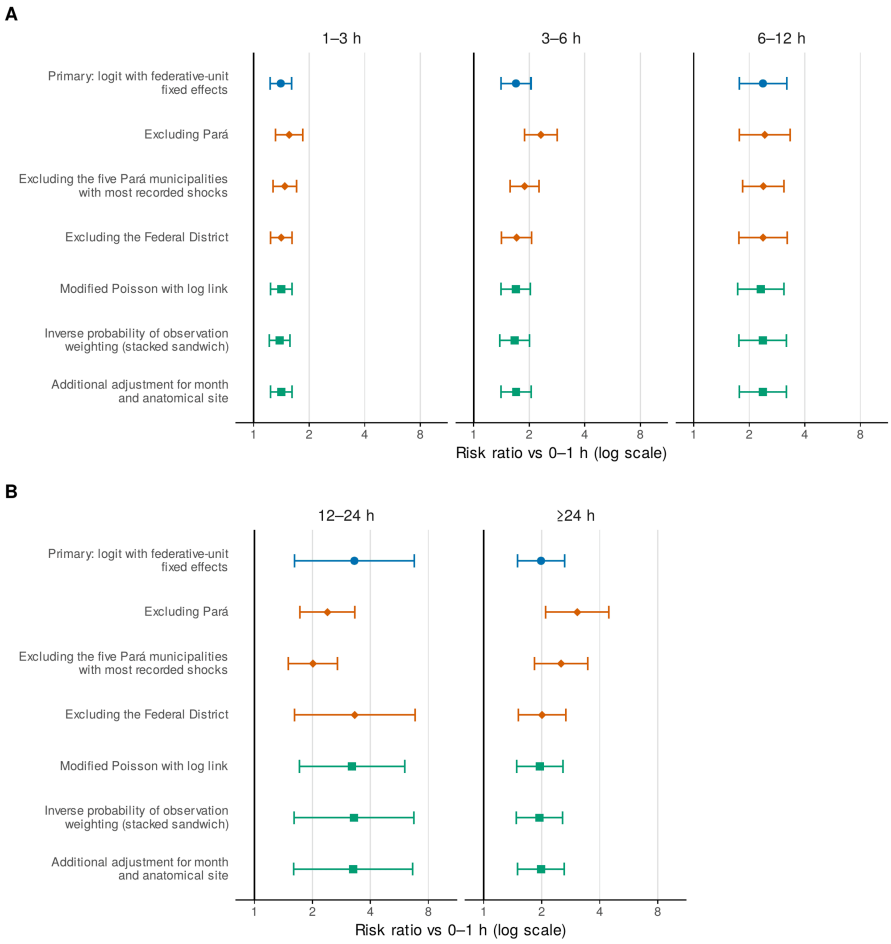


FIGURE 3: Forest plot of risk ratios from the primary model and six estimable sensitivity scenarios for the five non-reference sting-to-care interval categories.

The 0–1-hour category is the reference. Panel A shows the 1–3, 3–6, and 6–12-hour intervals. Panel B shows the 12–24 and ≥24-hour intervals; the same scenarios are repeated for readability. Bars show nominal 95% CIs. Blue circles identify the primary model; orange diamonds, geographical restrictions; green squares, alternative estimator, weighting, or additional adjustment. The 2017-2022 scenario (N = 805,309; 1,826 events) did not produce finite estimates because of separation; no federative unit was removed to resolve it. Municipality conditional logistic regression, initially incomplete, was subsequently completed in R with the same exact estimator and is reported separately in Table S12 (Appendix) as ORs, not RRs (710 informative municipalities). The panel also records the modified Poisson probability-bound check (18/1,296,606 predictions above 1, i.e. 0.0014% of records; maximum 1.853) and the federative-unit Webb bootstrap (original Python run: p = 0.0752 and control p = 0.0759; R re-execution: p = 0.0735 and control p = 0.0704; 9,999 replicates per run). The bootstrap uses a null-restricted score and federative-unit clustering; it differs from the primary Wald test in statistic as well as clustering. Sensitivity analyses do not replace the primary municipality-CR1 test. The 27 leave-one-unit-out analyses, the age-by-time interaction, and the unknown-time diagnostic are reported in Tables S3-S5 (Appendix). Numbers refer to notifications, not deduplicated persons.

CI: confidence interval; CR1: cluster-robust covariance with finite-sample correction; IPW: inverse probability of observation weighting; OR: odds ratio; RR: risk ratio

Among all notifications in the period, Pará accounted for 2,087 (64.9%) of the 3,218 shock records in the country under the hierarchical rule, before covariate requirements (the outcome-specific and common panel denominators contain 3,068 and 2,728 shock records, respectively), and five municipalities accounted for 65.1% of the shock records in the state. Among shock records with evaluable information for each marker, those from Pará less often had systemic manifestation (705/2,031, 34.7%, versus 772/1,104, 69.9%), severe classification (160/2,022, 7.91%, versus 681/1,099, 62.0%), and recorded antivenom therapy (1,488/2,057, 72.3%, versus 964/1,110, 86.8%) than those from other federative units (Tables *7*, *8*). Table *9* expresses the same numerators over all 2,087 broad shock records in Pará. The evolution of records over the years varied among municipalities in Pará.

Definition	Events in common N	Pará: n/N; crude risk, % (Wilson 95% CI)	Other federative units: n/N; crude risk, % (Wilson 95% CI)	Crude RR (95% CI)	Crude RD, pp (95% CI)
D1 — Recorded shock under the hierarchical flow	2,728/1,217,626	1,802/21,192; 8.5032 (8.1351–8.8863)	926/1,196,434; 0.0774 (0.0726–0.0825)	109.87 (101.61–118.79)	+8.4258 (8.0502–8.8014)
D2 — Shock with systemic manifestations	1,268/1,217,626	623/21,192; 2.9398 (2.7208–3.1759)	645/1,196,434; 0.0539 (0.0499–0.0582)	54.53 (48.89–60.83)	+2.8859 (2.6584–3.1133)
D3 — Shock with ≥1 high-severity marker	874/1,217,626	252/21,192; 1.1891 (1.0518–1.3442)	622/1,196,434; 0.0520 (0.0481–0.0562)	22.87 (19.77–26.46)	+1.1371 (0.9911–1.2831)
D4 — Shock with a recorded respiratory complication	507/1,217,626	118/21,192; 0.5568 (0.4652–0.6664)	389/1,196,434; 0.0325 (0.0294–0.0359)	17.13 (13.94–21.03)	+0.5243 (0.4241–0.6245)
D5 — Shock with severe classification	713/1,217,626	139/21,192; 0.6559 (0.5558–0.7739)	574/1,196,434; 0.0480 (0.0442–0.0521)	13.67 (11.37–16.45)	+0.6079 (0.4992–0.7167)
D6 — Strict convergent shock definition	341/1,217,626	8/21,192; 0.0378 (0.0191–0.0745)	333/1,196,434; 0.0278 (0.0250–0.0310)	1.36 (0.67–2.73)	+0.0099 (−0.0164–0.0362)

TABLE 7: Pará panel, six partially nested shock definitions: crude measures in the common denominator.

Common denominator, model specification, definitions, and abbreviations are described in the note to Table 9.

Definition	Standardized risk in Pará, % (95% CI)	Standardized risk in other federative units, % (95% CI)	Adjusted RR (95% CI)	Adjusted RD, pp (95% CI)
D1 — Recorded shock under the hierarchical flow	6.1965 (3.7825–9.9913)	0.0780 (0.0612–0.0994)	79.46 (45.83–137.77)	+6.1185 (3.1011–9.1360)
D2 — Shock with systemic manifestations	1.9174 (1.2312–2.9745)	0.0545 (0.0427–0.0694)	35.20 (21.27–58.25)	+1.8629 (1.0167–2.7092)
D3 — Shock with ≥1 high-severity marker	0.7904 (0.4578–1.3613)	0.0525 (0.0406–0.0679)	15.06 (8.31–27.30)	+0.7379 (0.3073–1.1685)
D4 — Shock with a recorded respiratory complication	0.3493 (0.1175–1.0334)	0.0329 (0.0250–0.0432)	10.62 (3.52–32.09)	+0.3164 (−0.0631–0.6959)
D5 — Shock with severe classification	0.4759 (0.3442–0.6575)	0.0483 (0.0373–0.0625)	9.85 (6.47–14.99)	+0.4276 (0.2727–0.5825)
D6 — Strict convergent shock definition	0.0271 (0.0116–0.0635)	0.0280 (0.0216–0.0364)	0.97 (0.39–2.38)	−0.0009 (−0.0254–0.0235)

TABLE 8: Pará panel, six partially nested shock definitions: adjusted and standardized measures.

Common denominator, model specification, definitions, and abbreviations are described in the note to Table 9.

Definition or municipal concentration	Pará shock records satisfying the definition or notified in the municipalities, n (% of 2,087 broad shock records)
D1 — Recorded shock under the hierarchical flow	2,087 (100.0)
D2 — Shock with systemic manifestations	705 (33.8)
D3 — Shock with ≥1 high-severity marker	285 (13.7)
D4 — Shock with a recorded respiratory complication	134 (6.4)
D5 — Shock with severe classification	160 (7.7)
D6 — Strict convergent shock definition	11 (0.5)
Shocks in the top 5 notifying municipalities	1,358 (65.1)
Shocks in the top 10 notifying municipalities	1,781 (85.3)
Shocks in the top 20 notifying municipalities	2,063 (98.9)

TABLE 9: Pará panel, six partially nested shock definitions: internal consistency and municipal concentration.

Common N = 1,217,626 for all six definitions (Pará: 21,192; other federative units: 1,196,434). Models are adjusted for age, sex, time category, and year, using municipality CR1 and standardization over a shared joint covariate distribution. Adjusted 95% CIs are nominal; marginal-risk CIs use a logit transformation with the delta method, and risk-difference CIs use the identity scale where negative values are valid. Definitions are partially nested, unvalidated, and non-ordinal. Table 9 uses all 2,087 broad recorded shock records in Pará rather than the common denominator. Municipal concentration reflects notification location rather than mechanism, causality, species, or recording error. D3 requires shock plus at least one of severe classification, recorded respiratory complication, or attributed death. D6 requires shock, severe classification, systemic manifestation, and either recorded respiratory complication or attributed death (noting that systemic manifestation [MCLI_SIST] and systemic complication [COM_SISTEM] are distinct fields). In Pará, D6 is met by 8 records in the common denominator and 11 of the 2,087 broad shock records. Table 9 percentages use all 2,087 broad shock records, whereas Results report numerators over evaluable denominators. Recorded antivenom therapy among shock records with evaluable fields: Pará, 1,488/2,057 (72.3%); other federative units, 964/1,110 (86.8%). Marginal-probability 95% CIs use the delta method on the logit scale (back-transformed to probability); RR CIs use the log scale; RD CIs use the identity scale. Crude proportions use Wilson CIs where indicated.

CI: confidence interval; CR1: cluster-robust covariance with finite-sample correction; IPTW: inverse probability of treatment weighting; pp: percentage points; RD: risk difference; RR: risk ratio

Six partially nested definitions of shock were compared, including D1 (shock according to the hierarchical logic of the fields) and D6 (shock accompanied by severe classification, systemic manifestation, and at least one of respiratory complication or attributed death). Among 1,217,626 notifications, the adjusted risk of shock in Pará was 79.46 times that of the other federative units under D1. The contrast was strongly attenuated under D6 (RR = 0.97, 95% CI = 0.39-2.38), with eight events in Pará and 333 in the other units; this imprecision does not demonstrate regional equivalence. The higher risk of shock in Pará under D1 persisted when only records whose sting and notification municipalities coincided were analyzed (RR = 97.69, 95% CI = 55.97-170.51). The Pará × interval interaction did not reach significance at 5% (adjusted p = 0.0757), without demonstrating equality of regional patterns (Table *8*, Figure *4*; Appendix, Table S6).

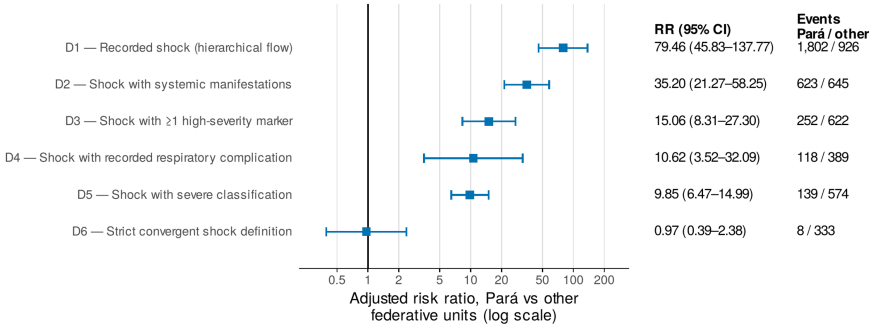


FIGURE 4: Heterogeneity of recorded shock in Pará: adjusted risk ratios comparing Pará with other federative units for six partially nested shock definitions.

Common denominator N = 1,217,626 (Pará, 21,192; other federative units, 1,196,434); models adjusted for age, sex, interval category and year, with municipality CR1 covariance. Squares and bars show RRs and nominal 95% CIs on the log scale; values and event counts (Pará/other) are shown on the right and correspond to Table 8. D3 requires shock and at least one of severe classification, recorded respiratory complication, or attributed death. D6 requires shock, severe classification, systemic manifestation, and either recorded respiratory complication or attributed death; it is not a gold standard, has eight Pará events in the common denominator, and its wide CI does not establish equivalence. The six definitions are partially nested and do not form an ordinal scale. Numbers refer to notifications, not deduplicated persons.

CI: confidence interval

Regarding mortality, SINAN recorded 1,342 deaths attributed to scorpion stings, counted by notification year; of these, 1,200 had a valid interval and entered the death analysis, and 142 had an unknown interval (Appendix, Table S10). In SIM, counted by year of death, there were 527 non-fetal deaths with X22 or T63.2 as the underlying cause and 610 with these codes as the underlying cause or in the audited current mention fields. The total of 610 includes the 527 deaths with the underlying cause (Appendix, Table S7).

Discussion

The sting-to-care interval was associated with recorded cardiorespiratory complications: all categories from 1-3 to ⩾24 hours had a higher adjusted probability than the 0-1-hour category. The largest point estimate occurred at 12-24 hours (RR = 3.30, 95% CI = 1.62-6.75). In absolute terms, the probabilities were 0.88% and 0.27%, respectively: approximately six additional recorded events per 1,000 notifications in the adjusted comparison. This absolute difference helps to size the association for complications that are potentially severe, although infrequent in the records (Table *3*).

The pattern was not one of continuous increase: at ⩾24 hours, the estimated probability was 0.53%. Because this category was not compared directly with 12-24 hours, a statistically significant reduction between them was not demonstrated. This finding does not define a safe waiting period. The SINAN interval also does not precisely identify first care or antivenom administration [6].

Respiratory complication, severe classification, shock, and death also had estimates above the 0-1-hour reference in the categories from 1-3 hours onwards. In the joint assessment of the interval categories, the association with each outcome was statistically significant, even after correction for multiple testing (Holm). The largest estimates did not occur in the same categories: 12-24 hours for the respiratory component (RR = 7.75, 95% CI = 2.82-21.32), 6-12 hours for shock and death, and 3-6 hours for severe classification. These maxima are descriptive, as no direct comparison between categories was made to demonstrate differences between them. The analyses had their own populations and methods, and the category-specific CIs were not corrected for multiple comparisons (Tables *4*-*6*). Studies in Jequié and Londrina also associated longer intervals with moderate/severe cases and pediatric hospitalization, respectively, but used different outcomes and definitions of time of care [14,15].

In children younger than five years, the exploratory comparison between grouped categories of >3 and ⩽3 hours yielded an RR of 0.98 (95% CI = 0.74-1.30). The CI is compatible with both lower and higher risk and does not demonstrate equivalence or the safety of waiting. Almeida et al. associated an interval of ⩾3 hours to treatment with death in children aged up to 10 years [4]. Differences in population, outcome, and time categories prevent these results from being treated as contradictory. The recommendation of timely care remains appropriate [5] (Appendix, Table S4).

The association with the composite persisted under geographical exclusions, additional adjustment, weighting for incomplete records, and the Poisson model. Without Pará, however, the largest point estimate shifted to ⩾24 hours (RR = 3.06, 95% CI = 2.10-4.47), indicating a regional influence on the observed pattern. Statistical confirmation also depended on the procedure: the alternative test that clustered notifications by federative unit (bootstrap) did not reach significance (p = 0.0735). This test changed both the statistic and the clustering, from municipal to state level; the divergence therefore cannot be attributed to the state scale alone. The Poisson model, used as a sensitivity analysis, also generated some individual predicted probabilities above 100%, an additional limitation (Appendix, Tables S1, S6).

In the 2017-2022 subset, the absence of eligible events in Rondônia prevented the model from producing valid adjusted estimates while keeping all federative units; this does not demonstrate the absence of an association. The conditional analysis, which compared notifications within 710 informative municipalities, maintained positive associations: at 12-24 versus 0-1 hour, the OR was 3.80 (95% CI = 1.78-8.11). It helps to control for stable characteristics of each municipality, but estimates ORs in a restricted population rather than the national RRs; it therefore complements, without replacing, the main analysis (Appendix, Table S12).

Pará accounted for 64.9% of national shock records under the hierarchical rule (2,087/3,218), and five notifying municipalities accounted for 65.1% of the shocks in the state. Among records with each item evaluable, systemic manifestation, severe classification, and antivenom therapy were less frequent in Pará than in the other federative units: 34.7% versus 69.9%, 7.91% versus 62.0%, and 72.3% versus 86.8%, respectively. The territorial concentration and the low frequency of systemic manifestations and severe classification among shock records justify clinical investigation. These data do not, however, prove undertreatment: confirmation of severity, indication, and timeliness of treatment are lacking (Tables *7*-*9*; Appendices).

The regional difference depended strongly on the definition of shock. In the same set of 1,217,626 notifications, the RR of shock in Pará relative to the other federative units was 79.46 (95% CI = 45.83-137.77) under D1 and 0.97 (95% CI = 0.39-2.38) under D6, which required combined clinical markers. This was a strong attenuation, not a demonstration of equivalence: D6 included only eight events in Pará and 333 in the other units. The definitions are not all progressively more restrictive than one another. They may select different presentations and share recording errors; they do not constitute a validated scale of diagnostic specificity (Tables *7*, *8*).

The excess under D1 (shock recorded in SINAN) persisted when the recorded sting and notification municipalities coincided (RR = 97.69, 95% CI = 55.97-170.51). This coincidence, however, does not exclude transfers or differences between services. Nor does the non-significant Pará × interval interaction (Holm-adjusted p = 0.0757) prove equality of temporal associations between regions. These findings limit single explanations for the Pará effect (Tables *7*-*9*, Figure *4*; Appendix, Table S6).

Clinical series from Pará describe an electric shock sensation and neuromuscular manifestations in envenoming attributed to *Tityus obscurus* [16-18]. The national protocol acknowledges that the usual classifications do not adequately capture this Amazonian presentation [5]. This justifies investigating a possible ambiguity of the term “choque” (shock) in the form, without proving incorrect completion of SINAN. Without individual species identification or review of medical records, biological, care-related, and recording causes cannot be distinguished. Clinical validation, with the appropriate authorizations, remains necessary.

Differences in mortality between systems also call for caution. SINAN recorded 1,342 deaths attributed to the sting, by notification year, and SIM recorded 527 with X22 or T63.2 as the underlying cause, by year of death. When the audited current fields of the death certificate were also included, SIM totaled 610 deaths, already including the 527. Distinct definitions, calendars, and flows, without individual linkage, prevent adding the series, quantifying person-level discordance, establishing superior capture, or inferring the speed of fatal progression (Appendix, Table S7).

Strengths and limitations

The national coverage, the explicit rules for assessing records, and the computational reproducibility are strengths, but they do not validate the diagnoses. Underreporting, differential completion, the absence of information on comorbidities, the impossibility of ensuring unique individuals, and limitations in matching municipalities over the years persist. Weighting considers only observed characteristics and does not recover unreported cases. In addition, planning followed exploration of the same data, without an independent confirmatory sample.

These limitations are not reduced by the size of the dataset. More than one million notifications increase precision but cannot correct systematic recording bias or establish whether complications began before or after care, because the form does not record when complications started. The strong dependence of the Pará estimates on the shock definition and the absence of statistical significance in the bootstrap clustered by

federative unit show that the findings depend on how outcomes were recorded and on inferential assumptions. Generalization to individual clinical care, or to settings with different recording practices, should therefore be cautious.

Interpretation is further limited by the absence of information on severity before care, onset of complications, and the care pathway. These gaps prevent a possible effect of time from being distinguished from other explanations, such as initial differences in severity, faster care of patients who were already severe, and selection related to survival and notification. These explanations remain hypotheses. The study did not estimate the effect of earlier care or earlier antivenom administration; this limitation does not demonstrate the absence of benefit from these interventions.

Conclusions

In Brazil, between 2008 and 2022, the recorded sting-to-care interval was associated with cardiorespiratory complications in the main analysis, with higher probabilities in the 1-3-hour to ⩾24-hour categories than at 0-1 hour. The estimates did not, however, increase continuously across categories. Although the association persisted in several sensitivity analyses, statistical significance varied with the procedure used. Because the design is cross-sectional and relies on recorded surveillance data, these findings are exploratory: they do not establish whether complications preceded or followed care, and they depend on recording quality and analytical assumptions. In Pará, the concentration of shock records and the strong attenuation of the regional difference under combined clinical criteria indicate the need for validation against medical records. These findings do not prove individual recording error, the involvement of a specific species, or equivalence between regions. Prospective studies documenting care times and clinical course may clarify these differences. The current results, on their own, do not support changes in treatment.

Appendices

Numbers refer to notifications, not deduplicated persons. The appendices contain the supplementary methods (Sections S1-S9) and Tables S1-S12 cited in the main text.

Supplementary methods

S1. Data Sources and Integrity

SINAN scorpion sting notifications for 2008-2022 were imported into R directly from the preserved national annual DBC files classified as final, keeping the annual schemas and missing values. The 15 annual files contain 1,563,983 scorpion sting notifications. The integrity of each source file was checked by SHA-256 signature, size, header, and format descriptors; a matching signature identifies byte preservation, not clinical accuracy. A numeric field filled with NUL characters was kept distinct from a negative clinical answer, and missing values were never converted to zero. Notification years outside the study window (2007 and 2023-2026) were processed separately and were not added to the analytic population.

S2. Composite Outcome and Its Components

The systemic complication field (COM_SISTEM, P) is the parent field; it is distinct from the systemic manifestation marker (MCLI_SIST). The shock field (COM_CHOQUE, H) and the respiratory field (COM_EDEMA, R; respiratory failure/acute pulmonary edema, a single field) are its components. Code 1 means yes, and code 2 means no. The composite outcome Y followed a three-valued hierarchical rule:

Y = 1 if P = 1 and (H = 1 or R = 1);

Y = 0 if P = 1 and H = 2 and R = 2;

Y = 0 if P = 2 and H and R are both blank;

Y = not evaluable for all other combinations.

When the parent field was positive, the composite could be positive even if the other component was not evaluable. For each component analyzed alone, a positive parent required the component to be coded 1 or 2; a negative parent produced a negative only when both components were blank. An unknown or incompatible parent was never recoded as negative. The respiratory field does not separate respiratory failure from pulmonary edema, and recorded shock is not necessarily cardiogenic.

S3. Age, Interval, and Eligibility

Age (NU_IDADE_N) required four canonical digits: units 1, 2, and 3 represent hours (0-23), days (0-30), and months (0-11); unit 4 accepts years (0-120). Invalid codes were not padded into valid ages. The nine age groups were <1, 1-4, 5-9, 10-14, 15-19, 20-39, 40-59, 60-79 and ⩾80 years. The interval has six recorded

categories (0-1, 1-3, 3-6, 6-12, 12-24 and ⩾24 hours); these are recorded intervals, not exact times, and no mean in hours can be derived from them without assumptions.

Sequential selection: Scorpion sting universe 2008-2022, 1,563,983; evaluable composite, 1,398,639 (165,344 excluded); valid interval among these, 1,296,789 (101,850 excluded); valid age, 1,296,757 (32 excluded); valid sex, 1,296,606 (151 excluded); valid year, federative unit and municipality, 1,296,606 (no additional exclusion). The primary population has 4,547 events and 4,778 notifying municipalities. Physical duplicates: 254 identical lines correspond to 127 surplus lines, 107 of which would be eligible for the primary analysis; none of the surplus lines had a positive composite, shock, respiratory component, or death. They were not removed automatically, because identical lines do not prove the same person or episode.

S4. Main Logistic Model and Marginal Standardization

The main model contains indicators for interval, age group, sex, file year, and notifying federative unit (references: 0-1 hours, 20-39 years, female, 2008 and São Paulo), totaling 55 parameters. Severity, antivenom therapy, and outcome were not added as post-sting adjustments. With $\eta = x^{\top}\beta$ and $p = \text{expit}(\eta)$, the grouped binomial log-likelihood $\ell(\beta) = \Sigma_j \{y_j \log p_j + (n_j - y_j) \log(1 - p_j)\}$ was maximized by iteratively reweighted least squares (tolerance 10^{-12}, maximum 100 iterations); 416,888 municipal cells were collapsed into 29,816 identical patterns for fitting only, and scores were rebuilt at the municipal level. The fit converged in 10 iterations without warnings (maximum absolute score 1.13×10^{-10}).

Cluster-robust CR1 covariance: $H = \Sigma_j n_j p_j(1 - p_j) x_j x_j^{\top}$; $U_g = \Sigma_{j \in g} x_j(y_j - n_j p_j)$; $c = \{G/(G - 1)\} \cdot \{(N - 1)/(N - k)\}$; $V = c\, H^{-1}(\Sigma_g U_g U_g^{\top}) H^{-1}$. In the main model, G = 4,778, N = 1,296,606, k = 55, and c = 1.00025. The Federal District forms a single municipal cluster, which produces one degenerate direction in the score covariance (rank = 54); the block of five interval coefficients remains full rank, the Federal District was not removed, and no generalized inverse was used to force a test of all parameters.

Marginal standardization used the fixed empirical covariate distribution of the 1,296,606 eligible notifications (6,947 patterns). For each interval category t, only the interval was changed in the prediction vector, keeping the same composition of age, sex, year, and federative unit: $P_t = \Sigma_j a_j \text{expit}\{x_j(t)^{\top}\beta\}$, with $a_j = n_j/N$ and gradient $g_t = \Sigma_j a_j p_j(t)\{1 - p_j(t)\} x_j(t)$. $\text{Var}(P_t) = g_t^{\top} V g_t$ and $\text{Cov}(P_t, P_s) = g_t^{\top} V g_s$. Gradients were checked against centered finite differences (maximum discrepancy 3.10×10^{-14}). The 95% CI of each probability was computed on the logit scale, $\text{expit}\{\text{logit}(P_t) \pm 1.96 \cdot SE(P_t)/(P_t(1 - P_t))\}$, as specified in the current protocol amendment; it is not a symmetric probability band truncated at 0 or 1. $RR_t = P_t/P_0$ with a log-scale CI based on $h_t = g_t/P_t - g_0/P_0$; $RD_t = P_t - P_0$ with an identity-scale CI based on $d_t = g_t - g_0$. The covariance between each category and the reference enters both calculations, and negative RD limits were retained. Exponentiated logistic coefficients are ORs, not the RRs reported here.

S5. Global Tests, Multiplicity, and the Federative Unit Bootstrap

The Wald test of the five interval coefficients was $W = (C\theta)^{\top}(CVC^{\top})^{-1}(C\theta)$ ($\chi^2(5) = 72.261$, $p = 3.47 \times 10^{-14}$). The Holm procedure was applied within predefined families: the five interval contrasts of the main model; the four global tests of the secondary outcomes; and the two interactions (age × interval and Pará × interval). Holm-adjusted p-values do not make the reported 95% CIs simultaneous.

The restricted score bootstrap used federative units as clusters. Under H_0, a logit without the five interval indicators was fitted (intercept, age, sex, year, and federative unit). With Z the design of the 50 nuisance parameters, T the five tested indicators, and $W_0 = \text{diag}\{n_j p_{0j}(1 - p_{0j})\}$, the projected design was $R_T = T - Z(Z^{\top}W_0 Z)^{-1} Z^{\top} W_0 T$ and the cluster scores $s_g = \Sigma_{j \in g} R_T{}_j(y_j - n_j p_{0j})$. $\Omega = c_{UF} \Sigma_g s_g s_g^{\top}$, $S = \Sigma_g s_g$ and $Q_{obs} = S^{\top}\Omega^{-1}S$, with $c_{UF} = (27/26) \cdot \{(N - 1)/(N - 55)\} = 1.0385$; $Q_{obs} = 7.994$. Replicates used six-point Webb weights $v_g \in \{-\sqrt{1.5}, -1, -\sqrt{0.5}, \sqrt{0.5}, 1, \sqrt{1.5}\}$ with equal probabilities, $S = \Sigma_g v_g s_g$, $Q = S^{\top}\Omega^{-1}S$, and $p = \{1 + \#(Q^* \geqslant Q_{obs})\}/(B + 1)$ with B = 9,999; Ω was kept fixed. In R, seed 20260904 (inferential run) gave 734 exceedances (p = 0.0735; Monte Carlo SE 0.0026), and seed 20260905 (check) gave 703 (p = 0.0704). The earlier Python runs gave p = 0.0752 and 0.0759; seeds with the same number do not guarantee identical draws across programs, and no seed was chosen by its p-value.

S6. Sensitivity Analyses and Interactions

Thirty-three estimable scenarios were run: exclusion of Pará, of five prespecified Pará municipalities (Rurópolis, Trairão, Medicilândia, Santarém, and Novo Progresso) and of the Federal District; modified Poisson regression; 27 leave-one-federative-unit-out models; unknown interval as a diagnostic category; and additional adjustment for month and anatomical site. Each scenario was refitted with its own N, G, k, and target population. The modified Poisson model ($\mu = \exp(x^{\top}\beta)$, CR1 covariance, standardization with derivative e^η) produced 18 individual predictions above 1, which were kept as a diagnostic and not truncated; marginal means remained between 0 and 1. The month/site model had N = 1,296,606, 4,547 events, 76 parameters, and an interval Wald statistic of 72.715. The unknown time diagnostic used N = 1,398,426 and 4,803 events; it is an extended evaluable population, not imputation of hours. The 2017-2022 restriction showed separation because Rondônia had zero events under the prescribed federative-unit logit;

no penalization or silent removal of the state was used to force finite estimates.

Age × interval interaction: five age groups (<5, 5-9, 10-19, 20-59, and ⩾60 years) and the interval dichotomized according to the form categories into ⩽3 hours (codes 1-2) and >3 hours (codes 3-6); logit with age, interval, sex, year, federative unit, and four interaction terms. The joint test of the four interaction terms gave $\chi^2(4) = 21.222$, nominal p = 0.000286, and Holm p = 0.000572. Standardized RRs (>3 hours versus ⩽3 hours) were 0.98 (95% CI = 0.74-1.30) for <5 years, 1.41 (1.08-1.83) for 5-9 years, 1.53 (1.16-2.02) for 10-19 years, 2.29 (1.61-3.26) for 20-59 years, and 1.94 (1.44-2.62) for ⩾60 years. An RR close to 1 in children younger than five years does not show that delay is safe. The Pará × interval interaction kept the six interval categories ($\chi^2(5) = 9.982$; nominal and Holm p = 0.0757).

S7. Conditional Logistic Regression and Secondary Outcome Estimators

The exact conditional likelihood conditions on the number of events in each municipality; only municipalities with both events and non-events are informative (841,654 notifications, 4,546 events, 710 municipalities, 28 coefficients for interval, age, sex, and year; federative unit is absorbed by the municipal stratum). The default exact call had returned −Inf and missing coefficients; the same likelihood was stabilized by within-stratum constant shifts, which cancel in the numerator and denominator, and converged in three iterations. No Efron or Breslow approximation, penalization, or additional exclusion was used. An independent Poisson-binomial recursion reproduced the log-likelihood (−17,779.554; difference between implementations 7.28×10^{-12}) and scores, and all 784 elements of the information matrix were checked by finite differences. The CR1 Wald test for the interval terms was $\chi^2(5) = 119.168$ ($p = 4.71 \times 10^{-24}$). Results are conditional ORs (Table S12), obtained in the later R reproduction; no baseline or marginal probabilities were derived from them.

Secondary outcomes each had their own evaluability. Death compared EVOLUCAO = 2 (death attributed to the sting) with EVOLUCAO = 1 (recovery); death from other causes was not counted as a non-event. Respiratory component and severe classification used the main logit with federative-unit fixed effects. For shock and death, a multinomial model of the six interval categories on age, sex, year, and federative unit (without severity, outcome, or antivenom therapy as predictors) produced stabilized weights $w_i = \rho_A/\pi_A(z_i)$; Hájek means were estimated jointly with the propensity model, and the stacked municipal sandwich included the derivative of the means with respect to the estimated propensities. In the 27-unit scenario, there were 250 multinomial coefficients and six means; the test of the five mean differences remained full rank. For untruncated shock weights, the range was 0.189-6.586, and the largest absolute standardized difference after weighting was 0.0158 (0.0145 for death). Standardized differences used the fixed binary standard deviation of the target population as the denominator; this corrects a divergent sentence in an earlier internal manual (documentary erratum; no numbers changed). Good balance of the covariates used does not exclude unmeasured confounding.

S8. Missing Data and Weighting for the Probability of Observation

The baseline set included 1,563,610 notifications with eligible basic covariates (373 fewer than the descriptive universe); within it, R = 1 identified the 1,296,606 records with jointly evaluable interval and composite, and R = 0 identified 267,004. The observation model $q_i = \Pr(R_i = 1 \mid z_i) = \text{expit}(z_i^T\gamma)$ included age, sex, year, federative unit, month, anatomical site, and an indicator of missing race/color (72 parameters). Weights were $w^*_i = c/q_i$ for R = 1, with c = 1,296,606/1,563,610 = 0.8292, truncated at the type-1 quantiles a = 0.8808 and b = 1.6823 (1.9997% truncated; effective sample size 1,269,710). The outcome model kept the 55 parameters of the main model, and the stacked covariance (127 parameters, CR1 factor 1.00029) propagated the uncertainty of the observation model. Margins were standardized to the full baseline target of 1,563,610 records. The weighted global test gave $\chi^2(5) = 75.085$ ($p = 8.93 \times 10^{-15}$). The largest absolute standardized difference against the target fell from 0.0952 to 0.0238 (Table S11). Missingness explained only by observed variables is an untested assumption; weights do not recover unreported stings or correct outcome misclassification.

Comparison of included and excluded records used standardized differences with pooled variance. The global standardized difference was 0.617 for federative unit, 0.393 for anatomical site, and 0.501 for missing race/color; these describe selection and are not tests of comparability. Thirty dates with a year before 2000 or after the file year were flagged without automatic exclusion.

S9. Pará Shock Definitions, SIM Selection, and Computing Environment

With H = hierarchical shock, R = hierarchical respiratory component, S = systemic manifestation (MCLI_SIST), G = severe classification and D = death in the panel (EVOLUCAO 2 versus 1/3), the six definitions were D1 = H; D2 = H ∧ S; D3 = H ∧ (G ∨ R ∨ D); D4 = H ∧ R; D5 = H ∧ G; D6 = H ∧ G ∧ S ∧ (R ∨ D), using three-valued logic (false ∧ unknown = false; true ∨ unknown = true). The common panel required all components to be evaluable (N = 1,217,626). D1-D6 are not necessarily nested in sequence (for example, D2 and D3 impose different conditions) and do not form a validated scale of clinical specificity; requiring markers associated with the outcome itself is an internal-coherence analysis, not an external gold

standard. In the hierarchical universe of shock records, before covariate requirements, there were 2,087 records in Pará and 1,131 in the other units (64.85%).

SIM: non-fetal deaths (TIPOBITO = 2) were selected when a cause token started with X22 or equaled T63.2 (normalized T632), with alphanumeric boundaries so that strings such as AX22 or T6329 were not accepted. B = selection in the underlying cause (CAUSABAS); M = selection in the underlying cause or current mention fields (LINHAA-LINHAD, LINHAII and ATESTADO). Year was derived from the date of death. For 2008-2022, ΣB = 527 and ΣM = 610; because B is contained in M, the totals are not additive, and the difference of 83 reflects a documentary criterion, not an estimate of underreporting. No individual SIM-SINAN linkage was performed.

The original analyses were run in Python 3.12.14 (NumPy, pandas, SciPy, and statsmodels). The reproduction used R 4.6.1 on Windows 11, data.table *1*.18.6.1 and explicit routines for scores, information, robust covariances, and standardization; the exact conditional estimator used the computational core of survival 3.8.6 with likelihood-preserving numerical stabilization. Scripts, outputs, sessions and file signatures are preserved by module; the R scripts are publicly available in the Open Science Framework (OSF) repository (https://doi.org/10.17605/OSF.IO/67UXK). Protocol amendments: the analytical protocol was written after exploration of the same data; the current amendment specifies the logit-delta interval for marginal probabilities (Section S4), and the conditional model and the R bootstrap are later additions reported as such.

Supplementary tables

The primary logistic model remains the main inferential analysis. The analyses below assess robustness, influence, missingness, and heterogeneity; none replaces or "rescues" the primary test. All estimate-level 95% CIs are nominal.

Scenario	N; events	Interval	Standardized risk, % (95% CI)	RR (95% CI)	RD, pp (95% CI)
Primary: logit with federative-unit fixed effects	1,296,606; 4,547	1–3 hours	0.3731 (0.3018–0.4611)	1.40 (1.23–1.60)	+0.1071 (0.0604–0.1538)
		3–6 hours	0.4508 (0.3563–0.5703)	1.70 (1.41–2.04)	+0.1849 (0.1049–0.2648)
		6–12 hours	0.6311 (0.4829–0.8243)	2.37 (1.76–3.19)	+0.3652 (0.2024–0.5279)
		12–24 hours	0.8785 (0.4367–1.7592)	3.30 (1.62–6.75)	+0.6125 (0.0019–1.2231)
		≥24 hours	0.5274 (0.4223–0.6586)	1.98 (1.50–2.63)	+0.2615 (0.1409–0.3820)
Excluding Pará	1,273,355; 2,431	1–3 hours	0.2291 (0.1975–0.2658)	1.56 (1.32–1.85)	+0.0821 (0.0512–0.1129)
		3–6 hours	0.3401 (0.2851–0.4058)	2.31 (1.89–2.83)	+0.1931 (0.1376–0.2486)
		6–12 hours	0.3572 (0.2675–0.4767)	2.43 (1.77–3.34)	+0.2102 (0.1077–0.3126)
		12–24 hours	0.3519 (0.2599–0.4764)	2.39 (1.72–3.32)	+0.2049 (0.0994–0.3103)
		≥24 hours	0.4498 (0.3249–0.6225)	3.06 (2.10–4.47)	+0.3028 (0.1536–0.4521)
Excluding the five Pará municipalities with the most recorded shocks	1,289,763; 3,169	1–3 hours	0.2815 (0.2453–0.3231)	1.48 (1.27–1.71)	+0.0907 (0.0567–0.1248)
		3–6 hours	0.3604 (0.3075–0.4225)	1.89 (1.58–2.27)	+0.1697 (0.1161–0.2232)
		6–12 hours	0.4550 (0.3602–0.5747)	2.39 (1.85–3.08)	+0.2643 (0.1604–0.3681)
		12–24 hours	0.3840 (0.2945–0.5004)	2.01 (1.50–2.70)	+0.1932 (0.0912–0.2952)

		≥24 hours	0.4808 (0.3654–0.6323)	2.52 (1.84–3.46)	+0.2900 (0.1556–0.4244)
Excluding the Federal District	1,286,742; 4,515	1–3 hours	0.3742 (0.3024–0.4629)	1.41 (1.23–1.62)	+0.1091 (0.0621–0.1560)
		3–6 hours	0.4531 (0.3581–0.5732)	1.71 (1.42–2.06)	+0.1880 (0.1077–0.2683)
		6–12 hours	0.6302 (0.4810–0.8252)	2.38 (1.76–3.21)	+0.3651 (0.2010–0.5291)
		12–24 hours	0.8789 (0.4346–1.7693)	3.32 (1.61–6.81)	+0.6138 (−0.0018–1.2294)
		≥24 hours	0.5321 (0.4263–0.6640)	2.01 (1.51–2.66)	+0.2670 (0.1457–0.3883)
Modified Poisson with log link	1,296,606; 4,547	1–3 hours	0.3732 (0.3022–0.4607)	1.41 (1.23–1.61)	+0.1086 (0.0623–0.1548)
		3–6 hours	0.4476 (0.3561–0.5626)	1.69 (1.41–2.03)	+0.1830 (0.1069–0.2592)
		6–12 hours	0.6127 (0.4756–0.7890)	2.32 (1.74–3.09)	+0.3481 (0.1981–0.4981)
		12–24 hours	0.8494 (0.4642–1.5492)	3.21 (1.71–6.02)	+0.5847 (0.0730–1.0965)
		≥24 hours	0.5165 (0.4161–0.6409)	1.95 (1.48–2.57)	+0.2519 (0.1371–0.3666)
Missingness IPW with stacked sandwich	1,563,610; 4,547	1–3 hours	0.3695 (0.2968–0.4599)	1.38 (1.21–1.58)	+0.1024 (0.0563–0.1485)
		3–6 hours	0.4454 (0.3498–0.5670)	1.67 (1.38–2.01)	+0.1783 (0.0986–0.2580)
		6–12 hours	0.6329 (0.4841–0.8271)	2.37 (1.76–3.19)	+0.3658 (0.2026–0.5290)
		12–24 hours	0.8789 (0.4374–1.7582)	3.29 (1.61–6.74)	+0.6118 (0.0011–1.2224)
		≥24 hours	0.5197 (0.4196–0.6436)	1.95 (1.48–2.56)	+0.2526 (0.1380–0.3673)
Additional adjustment for month and anatomical site	1,296,606; 4,547	1–3 hours	0.3744 (0.3027–0.4629)	1.41 (1.23–1.61)	+0.1087 (0.0618–0.1556)
		3–6 hours	0.4510 (0.3565–0.5704)	1.70 (1.41–2.05)	+0.1853 (0.1050–0.2657)
		6–12 hours	0.6298 (0.4839–0.8194)	2.37 (1.77–3.18)	+0.3642 (0.2039–0.5244)
		12–24 hours	0.8664 (0.4335–1.7243)	3.26 (1.60–6.64)	+0.6008 (0.0034–1.1982)
		≥24 hours	0.5263 (0.4224–0.6557)	1.98 (1.50–2.62)	+0.2607 (0.1415–0.3798)

TABLE 10: S1: Sensitivity analyses, federative-unit influence, and interactions: estimates displayed in Figure 3.

Notes and conventions for Tables S1-S6 are given below Table S6.

Estimable scenario	N; events	Global χ^2 test (df)	Global p	Role
Primary: logit with federative-unit fixed effects	1,296,606; 4,547	72.26 (5)	3.47×10^{-14}	Primary inference
Excluding Pará	1,273,355; 2,431	81.24 (5)	4.61×10^{-16}	Sensitivity analysis
Excluding the five Pará municipalities with the most recorded shocks	1,289,763; 3,169	63.66 (5)	2.13×10^{-12}	Sensitivity analysis
Excluding the Federal District	1,286,742; 4,515	72.91 (5)	2.53×10^{-14}	Sensitivity analysis
Modified Poisson with log link	1,296,606; 4,547	65.64 (5)	8.26×10^{-13}	Sensitivity analysis
Missingness IPW with stacked sandwich	1,563,610; 4,547	75.08 (5)	8.93×10^{-15}	Sensitivity analysis
Additional adjustment for month and anatomical site	1,296,606; 4,547	72.72 (5)	2.79×10^{-14}	Sensitivity analysis

TABLE 11: S2: Sensitivity analyses, federative-unit influence, and interactions: global tests for estimable scenarios

Notes and conventions for Tables S1-S6 are given below Table S6.

Interval	Models	RR range (excluded unit at minimum; at maximum)	Range of lower limits	Range of upper limits	95% CI excludes 1
1–3 hours	27	1.310 (SP)–1.558 (PA)	1.138–1.315	1.487–1.846	27/27
3–6 hours	27	1.565 (MG)–2.313 (PA)	1.266–1.890	1.912–2.831	27/27
6–12 hours	27	2.218 (SP)–2.480 (ES)	1.651–1.848	2.981–3.341	27/27
12–24 hours	27	2.394 (PA)–3.449 (ES)	1.484–1.724	3.323–6.989	27/27
≥24 hours	27	1.786 (MG)–3.060 (PA)	1.353–2.096	2.357–4.466	27/27

TABLE 12: S3: Sensitivity analyses, federative-unit influence, and interactions: leave-one-unit-out influence (27 models per interval).

Notes and conventions for Tables S1-S6 are given below Table S6.

Age group	≤3 hours: observed events/N; standardized risk, % (95% CI)	>3 hours: observed events/N; standardized risk, % (95% CI)	RR >3 hours vs ≤3 hours (95% CI)	RD, pp (95% CI)
<5 years	762/69,207; 1.2559 (1.0290–1.5320)	172/7,838; 1.2330 (0.9075–1.6734)	0.98 (0.74–1.30)	−0.0228 (−0.3690–0.3233)
5–9 years	513/70,235; 0.8572 (0.6994–1.0501)	136/6,694; 1.2052 (0.8987–1.6144)	1.41 (1.08–1.83)	+0.3480 (0.0397–0.6564)
10–19 years	455/165,197; 0.2823 (0.2157–0.3694)	155/18,538; 0.4313 (0.3370–0.5519)	1.53 (1.16–2.02)	+0.1490 (0.0475–0.2505)
20–59 years	1,174/654,160; 0.1908 (0.1356–0.2685)	779/102,166; 0.4371 (0.3197–0.5974)	2.29 (1.61–3.26)	+0.2463 (0.1222–0.3705)
≥60 years	251/170,088; 0.2114 (0.1494–0.2991)	150/32,483; 0.4108 (0.2882–0.5851)	1.94 (1.44–2.62)	+0.1994 (0.0858–0.3130)

TABLE 13: S4: Sensitivity analyses, federative-unit influence, and interactions: age-by-time interaction (>3 hours versus ≤3 hours).

Notes and conventions for Tables S1-S6 are given below Table S6.

Category	n; events	Standardized risk, % (95% CI)	RR (95% CI)	RD, pp (95% CI)	Interpretation
0–1 hours	807,190; 1,756	0.2597 (0.2079–0.3243)	1.00 (1.00–1.00)	+0.0000 (0.0000–0.0000)	Known exposure category
1–3 hours	321,697; 1,399	0.3663 (0.2966–0.4523)	1.41 (1.23–1.61)	+0.1066 (0.0612–0.1521)	Known exposure category
3–6 hours	85,803; 664	0.4450 (0.3523–0.5619)	1.71 (1.42–2.07)	+0.1853 (0.1069–0.2637)	Known exposure category
6–12 hours	38,592; 362	0.6252 (0.4769–0.8193)	2.41 (1.78–3.25)	+0.3655 (0.2018–0.5293)	Known exposure category
12–24 hours	25,839; 266	0.8698 (0.4298–1.7522)	3.35 (1.63–6.89)	+0.6101 (0.0000–1.2202)	Known exposure category
≥24 hours	17,485; 100	0.5204 (0.4169–0.6495)	2.00 (1.51–2.66)	+0.2608 (0.1415–0.3801)	Known exposure category
Unknown	101,820; 256	0.3376 (0.2729–0.4177)	1.30 (1.08–1.57)	+0.0779 (0.0211–0.1348)	Missingness diagnostic; not an exposure

TABLE 14: S5: Sensitivity analyses, federative-unit influence, and interactions: unknown-time category diagnostic.

Notes and conventions for Tables S1-S6 are given below Table S6.

Analysis	Result	Interpretation/Status
Federative-unit Webb bootstrap	Q = 7.994; p = 0.0752; 751 exceedances in 9,999 replicates; QA p = 0.0759	Locked seed; does not replace primary municipality CR1
Poisson: predicted probabilities >1	18/1,296,606 (0.0014%); maximum = 1.853	Probability-bound gate (0 to 1) failed; interpret only as a sensitivity analysis
Age-by-time interaction	$\chi^2(4) = 21.22$; $p = 2.86 \times 10^{-4}$; Holm = 5.72×10^{-4}	Exploratory; family shared with Pará-by-time
Pará-by-time interaction	$\chi^2(5) = 9.98$; p/Holm = 0.0757	Exploratory
2017–2022 period	1,826/805,309; RO = 0 events	Non-estimable under the STOP gate; RO retained; no fallback
Municipality conditional logistic model	710 informative municipalities; 841,654 records; 4,546 events	Original implementation encountered a recursion limit; exact R reimplementation completed; conditional ORs are reported in Table S12, not marginal RRs

TABLE 15: S6: Sensitivity analyses, federative-unit influence, and interactions: gates, bootstrap, and multiplicity.

Notes: 0–1 hour is the reference. Leave-one-unit-out analysis removes one federative unit at a time, preserving the full 27×6 CSV. The unknown time category has no exposure interpretation. The 2017-2022 scenario lacked finite estimates due to separation; a computational limitation was resolved in an exact R reimplementation (Table S12). For IPW, target N = 1,563,610 and R = 1 for 1,296,606 records; uncertainty was propagated, and weights truncated at the 1st/99th percentiles (ESS = 1,269,710). Month/site adjustment omitted only the unsupported “unknown month” column. In Table S4, fractions are observed counts, whereas percentages and 95% CIs are standardized estimates. The three-hour threshold represents grouping of recorded categories. Marginal-probability 95% CIs use the delta method on the logit scale; RR uses the log scale; RD uses the identity scale; crude proportions use Wilson CIs. In the R execution, the restricted federative-unit score test retained Q = 7.993900503. With 9,999 replicates per run (seeds 20260904 and 20260905), exceedances yielded p = 0.0735 and p = 0.0704; original Python values (p = 0.0752 and p = 0.0759) are preserved as distinct historical realizations. Cross-generator seed variations reflect Monte Carlo differences, not error corrections. None of the four realizations reached the 0.05 threshold.

Year	SINAN: death attributed to the notified condition, by notification year	SIM: X22 or T63.2 as underlying cause, by year of death	SIM: X22 or T63.2 as underlying cause or in any audited current mention field, by year of death
2008	87	43	51
2009	87	32	39
2010	70	26	29
2011	81	25	30
2012	84	30	33
2013	68	24	30
2014	71	33	40
2015	99	26	35
2016	117	42	47
2017	83	37	40
2018	81	50	53
2019	96	35	41
2020	119	48	52
2021	119	45	54
2022	80	31	36
Total	1,342	527	610

TABLE 16: S7: Aggregate triangulation of scorpion-related death records in SINAN and the Mortality Information System (SIM), Brazil, 2008-2022.

Notes. SINAN: scorpion-sting notifications with EVOLUCAO = 2, grouped by notification year. SIM: non-fetal deaths (TIPOBITO = 2) with X22, including subcategories, or normalized T63.2, grouped by year of death. The any-mention definition used only frozen current-cause fields available in each annual schema: CAUSABAS, LINHAA, LINHAB, LINHAC, LINHAD, LINHAII, and ATESTADO; CAUSABAS_O and CAUSAMAT were excluded. The three series use different units, calendars, definitions, and recording pathways. No individual-level linkage was performed; counts must not be summed, treated as person-level agreement, or used to infer time to death or superiority of either system.

COM_SISTEM	COM_EDEMA	COM_CHOQUE	Composite	Respiratory	Shock	Pará	Other units	Total
1	9	1	1	NA	1	9	35	44
1	2	9	NA	0	NA	0	4	4
1	1	9	1	1	NA	0	71	71
1	9	2	NA	NA	0	0	1	1

TABLE 17: S8: Discordant evaluability patterns.

Records with eligible covariates and interval. This table contains all patterns explaining differences in evaluability between the composite and its components, not all observed combinations. NA means non-evaluable, not zero. Code 9 = unknown. The 44 records with positive shock and unknown respiratory field allow a positive composite but not an evaluable respiratory component, and the four records with negative respiratory field and unknown shock do the opposite (44 − 4 = 40); for shock, the balance is 71 − 1 = 70. This reproduces the denominators 1,296,606 (composite), 1,296,566 (respiratory), and 1,296,536 (shock).

Definition	Negative	Positive	Non-evaluable	Total definition-specific N
D1	78,570	340	129,517	1,296,536
D2	166,423	131	41,873	1,384,180
D3	78,611	95	129,721	1,296,332
D4	78,826	44	129,557	1,296,496
D5	177,688	79	30,660	1,395,393
D6	195,276	31	13,120	1,412,933

TABLE 18: S9: Definitions in partially observed records.

The 208,427 records with at least one non-evaluable H/R/S/G/D marker belong to the 1,426,053 records with eligible covariates and time. The three state columns refer only to those 208,427 partially observed records. The final column is the total definition-specific denominator, adding determined positives and negatives in that group to the 1,217,626 records with every marker observed. Partial refers to the five-marker panel, not necessarily each definition's own requirements.

Interval	EVOL = 1	EVOL = 2	EVOL = 3	EVOL = 9	Blank
0–1 hours	832,446	541	47	16,192	35,676
1–3 hours	331,537	402	23	6,379	16,457
3–6 hours	88,244	130	9	1,848	5,015
6–12 hours	39,690	54	9	791	2,207
12–24 hours	26,512	39	5	603	1,540
≥24 hours	17,997	34	8	517	1,101
Unknown	114,753	142	21	8,053	14,588

TABLE 19: S10: Recorded outcome by interval before mortality selection.

Baseline-eligible records before exclusion of EVOLUCAO values unclassifiable for the specific outcome. EVOL = 1: recovery; 2: attributed death; 3: other-cause death; 9: unknown. No other codes occurred in this population. Across the six valid time categories, 88,427 records lay outside the death denominator: 101 other-cause deaths, 26,330 unknowns, and 61,996 blanks. The unknown time row is separate and is not included in that total or in the interval analysis.

Measure	Before	After
Largest absolute standardized difference	0.0952401	0.0238204
Unknown anatomical site among observed records, %	2.6962	3.8214
Unknown anatomical site in target, %	4.3049	4.3049

TABLE 20: S11: Observation-weighting diagnostic.

Comparison of R = 1 with the fixed empirical target. The level-specific standardized difference is observed proportion minus target proportion, divided by the square root of p_target(1 − p_target). This diagnostic uses 78 levels. It differs both from the unweighted included versus excluded comparison and from multinomial balance across time categories. Unknown anatomical site had the largest residual imbalance. Outcome model point estimates and inferential results were retained. This characterization does not establish missingness at random. In the first row, values closer to zero indicate better balance of observed marginal distributions.

Interval	Conditional OR	95% CI (CR1), lower	95% CI (CR1), upper
1–3 hours	1.3932	1.2111	1.6028
3–6 hours	1.6980	1.4056	2.0514
6–12 hours	2.6101	2.0046	3.3984
12–24 hours	3.8026	1.7823	8.1131
≥24 hours	2.5078	1.8590	3.3831

TABLE 21: S12: Conditional odds ratios by interval.

Reference: 0–1 hour. N = 841,654; events = 4,546; informative municipalities = 710; parameters = 28. Nominal global Wald $\chi^2(5) = 119.168$; $p = 4.71 \times 10^{-24}$. These are conditional odds ratios within informative municipalities, not the nationwide marginal RRs in the main tables. Baseline probabilities, marginal probabilities, and RDs were not derived from this conditional likelihood. CIs are nominal; this subsequent sensitivity analysis was not retrospectively added to a confirmatory testing family and does not replace Table 2.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Ana Clara Aragão Fernandes, Guilherme A. Carilo, Piettra S. Wladimirski

Acquisition, analysis, or interpretation of data: Ana Clara Aragão Fernandes, Guilherme A. Carilo, Jaqueline E. Beirith, Rafaella Ottaviani, Gabriel P. Andreotti, Piettra S. Wladimirski

Critical review of the manuscript for important intellectual content: Ana Clara Aragão Fernandes, Jaqueline E. Beirith, Rafaella Ottaviani, Gabriel P. Andreotti, Piettra S. Wladimirski

Supervision: Ana Clara Aragão Fernandes

Drafting of the manuscript: Guilherme A. Carilo, Rafaella Ottaviani, Gabriel P. Andreotti, Piettra S. Wladimirski

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Ministério da Saúde, Secretaria de Vigilância em Saúde e Ambiente: Epidemiologia dos acidentes escorpiônicos no Brasil em 2023. Bol Epidemiol. 2025, 56:1-10.
2. Lacerda AB, Cândido D, de Azevedo TS, Wen FH, Freitas GD, Dourado FS, Neto FC: Scorpionism in Brazil: space-time approach and risk areas in 2012 to 2024. PLoS Negl Trop Dis. 2025, 19:e0013602. 10.1371/journal.pntd.0013602
3. Siqueira TS, Dos Santos AJ, Santos JL, da Silva RR, Silva JR, Santos VS: Scorpion envenomation in Brazil and its relationship with the social determinants of health: a population-based ecological study. Acta Trop. 2024, 253:107165. 10.1016/j.actatropica.2024.107165
4. Almeida AC, Carvalho FM, Mise YF: Risk factors for fatal scorpion envenoming among Brazilian children: a case-control study. Trans R Soc Trop Med Hyg. 2021, 115:975-83. 10.1093/trstmh/trab120
5. Ministério da Saúde: Protocolo Clínico e Diretrizes Terapêuticas dos Acidentes Escorpiônicos . (2026). Accessed: September 25, 2026: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/a/animais-peconhentos/publicacoes/pcdt-dos-acidentes-escorpion....
6. Ministério da Saúde, Sistema de Informação de Agravos de Notificação: Ficha de investigação: acidentes por animais peçonhentos, versão 5. (2006). Accessed: September 25, 2026: https://portalsinan.saude.gov.br/images/documentos/Agravos/AAP/Animais_Peconhentos_v5.pdf.

7. Benchimol EI, Smeeth L, Guttmann A, et al.: The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015, 12:e1001885. 10.1371/journal.pmed.1001885
8. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med. 2007, 4:e296. 10.1371/journal.pmed.0040296
9. Ministério da Saúde, Sistema de Informação de Agravos de Notificação: Dicionário de dados: SINAN NET, agravo acidente animais peçonhentos, versão 5.0. (2010). Accessed: September 25, 2026: https://portalsinan.saude.gov.br/images/documentos/Agravos/AAP/DIC_DADOS_Animais_Pedonhentos_v5.pdf.
10. Ministério da Saúde, Departamento de Informática do SUS (DATASUS): Sistema de Informação de Agravos de Notificação (SINAN): acidentes por animais peçonhentos, national final annual files, 2008-2022 (dataset). (2026). Accessed: September 3, 2026: https://datasus.saude.gov.br/transferencia-de-arquivos/.
11. Ministério da Saúde: Como obter dados epidemiológicos: acidentes por animais peçonhentos . (2026). Accessed: September 25, 2026: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/a/animais-peconhentos/como-obter-dados-epidemiologicos/como-ob....
12. Ministério da Saúde: Sistema de Informações sobre Mortalidade (SIM): national annual files (dataset) . (2026). Accessed: September 3, 2026: https://dadosabertos.saude.gov.br/dataset/sim.
13. Conselho Nacional de Saúde: Resolução nº 510, de 7 de abril de 2016 . (2016). Accessed: September 25, 2026: https://www.gov.br/conselho-nacional-de-saude/pt-br/atos-normativos/resolucoes/2016/resolucao-no-510.pdf.
14. Carmo EA, Nery AA, Pereira R, Rios MA, Casotti CA: Factors associated with the severity of scorpio poisoning. Texto Contexto Enferm. 2019, 28:20170561. 10.1590/1980-265X-TCE-2017-0561
15. Tonin Dos Santos RH, González AD, Guidoni CM, Rufino JV, Tirolla RM, Girotto E: Characterization of scorpion envenoming in children care for a poisoning control center. Toxicon. 2025, 258:108305. 10.1016/j.toxicon.2025.108305
16. Borges RF, Dergan JN, Torrez PP, et al.: Scorpionism in Pará, Brazil: clinical assessment of neuromuscular manifestations. Rev Soc Bras Med Trop. 2025, 58:e00532025. 10.1590/0037-8682-0053-2025
17. Pardal PP, Ishikawa EA, Vieira JL, et al.: Clinical aspects of envenomation caused by Tityus obscurus (Gervais, 1843) in two distinct regions of Pará state, Brazilian Amazon basin: a prospective case series. J Venom Anim Toxins Incl Trop Dis. 2014, 20:3. 10.1186/1678-9199-20-3
18. Torrez PP, Quiroga MM, Abati PA, Mascheretti M, Costa WS, Campos LP, França FO: Acute cerebellar dysfunction with neuromuscular manifestations after scorpionism presumably caused by Tityus obscurus in Santarém, Pará / Brazil. Toxicon. 2015, 96:68-73. 10.1016/j.toxicon.2014.12.012