



## OPEN Estimating trauma prevalence from incomplete human skeletal remains

Judith Beier<sup>1,2,✉</sup>, Matteo Santon<sup>3</sup> & Hannes Rathmann<sup>1,2,4,✉</sup>

Traumatic lesions on human skeletal remains are widely used for reconstructing past accidents or violent encounters and for comparing trauma prevalence across samples over time and space. However, uncertainties in trauma prevalence estimates increase proportionally with decreasing skeletal completeness, as once-present trauma might have gone missing. To account for this bias, samples are typically restricted to skeletal remains meeting a predefined minimum completeness threshold. However, the effect of this common practice on resulting estimates remains unexplored. Here, we test the performance of the conventional frequency approach, which considers only specimens with  $\geq 75\%$  completeness, against a recent alternative based on generalized linear models (GLMs), integrating specimen completeness as a covariate. Using a simulation framework grounded on empirical forensic, clinical, and archaeological data, we evaluated how closely frequency- and GLM-based estimates conform to the known trauma prevalence of once-complete cranial samples after introducing increasing levels of missing values. We show that GLM-based estimates were consistently more precise than frequencies across all levels of incompleteness and regardless of sample size. Unlike GLMs, frequencies increasingly produced incorrect relative patterns between samples and occasionally failed to produce estimates as incompleteness increased, particularly in smaller samples. Consequently, we generally recommend using GLMs and their extensions over frequencies, although neither approach is fully reliable when applied to largely incomplete samples.

**Keywords** Trauma prevalence, Crude frequencies, Generalized linear (mixed) models, Skeletal preservation, Simulations

Trauma analyses are an integral part of bioarchaeological investigations. Traumatic lesions result from accidents or violent encounters, and their examination represents a powerful tool to gain insights into past humans' fates and behaviors. The level and patterns of trauma present in a population allow for reconstructions of group-specific social, cultural, and environmental life realities and their comparisons across time and space<sup>1–10</sup>.

Trauma is commonly recorded as a binary variable (present/absent) for each bone or individual. Trauma prevalence in a sample is typically given as a percentage, representing the proportion of individuals or skeletal parts with injuries, and compared between samples using statistical tests such as chi-squared or Fisher's exact test. Such crude trauma frequencies (hereafter, CFs) are directly affected by skeletal preservation in both qualitative and quantitative means (hereafter, completeness), because once-present traumatic lesions might be missing from an incompletely preserved skeletal record<sup>11–13</sup>. Hence, uncertainties in trauma prevalence estimates increase proportionally with decreasing completeness of the remains and may result in false negative findings in incomplete samples where once-present trauma might have gone missing<sup>11,12,14–16</sup>. To account for this potential bias in estimates of trauma prevalence and their statistical analysis, samples are commonly restricted to skeletal remains in a pre-defined minimum completeness state, while neglecting less complete individuals or bones. However, the threshold at which completeness is deemed sufficient for inclusion varies among studies, and various thresholds have been employed, such as 25%<sup>17</sup>, 40%<sup>18</sup>, 50%<sup>18–21</sup>, 66.7%<sup>22,23</sup>, 80%<sup>12,24</sup>, 90%<sup>25</sup>, 100%<sup>13</sup>, or most commonly, 75%<sup>11,12,14,15,19,24,26–31</sup>. Other methodological differences exist, for example, in handling trauma observations; while some studies exclude trauma findings on remains smaller than the defined threshold<sup>11,13,15,26,28,29,31</sup>, others include them to maintain a wider range of analyzable variation<sup>12,18,24,25</sup>.

Employing CFs along with completeness thresholds enhances data quality and comparability between samples but, at the same time, severely reduces sample sizes if samples are overall poorly preserved. This issue especially applies to the human fossil record, which comprises large portions of partial remains. The use of

<sup>1</sup>DFG Center for Advanced Studies 'Words, Bones, Genes, Tools', University of Tübingen, 72070 Tübingen, Germany.

<sup>2</sup>Paleoanthropology, Department of Geosciences, Institute for Archaeological Sciences, University of Tübingen, 72070 Tübingen, Germany. <sup>3</sup>Ecology of Vision Group, School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, UK. <sup>4</sup>Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, 72070 Tübingen, Germany. ✉email: judith.beier@uni-tuebingen.de; hannes.rathmann@uni-tuebingen.de

alternative approaches, such as generalized linear models (hereafter, GLMs), to estimate and compare trauma rates within and between samples can help overcome the trade-off between sample sizes and comparability that occurs in the CF approach when dealing with incompletely preserved samples. GLMs are an extension of linear models, which describe the relationship between one response variable as a function of one or multiple predictors<sup>32</sup>. However, instead of assuming a Gaussian distribution family, GLMs are implemented with other distributions to analyse skewed, count, proportional, or, in our case, binary response variables such as trauma scores<sup>32–35</sup>. Moreover, GLMs allow for the simultaneous assessment of multiple predictor variables and their interactions, such as sex, age at death, or archaeological period. Importantly, this approach allows the inclusion of all available skeletal fragments for analysis, regardless of their size and without the need for a priori exclusion, achieved by assigning a completeness score to each fragment and incorporating completeness as a predictor. By adopting such a modeling approach, Beier et al.<sup>36,37</sup> investigated cranial trauma rates in highly fragmented Late Pleistocene human fossils from Eurasia, including even the smallest skeletal elements preserved, while factoring in variation in trauma prevalence owing to differential completeness. Given the advantages of maximizing sample sizes and evaluating multiple predictor variables simultaneously, they recommended utilizing GLMs and related techniques instead of CFs to estimate trauma prevalence in samples of fossilized or otherwise incomplete skeletal remains. However, this recommendation has not been analytically corroborated, prompting them to advise an evaluation of the performance of the two approaches in a designated study in the future.

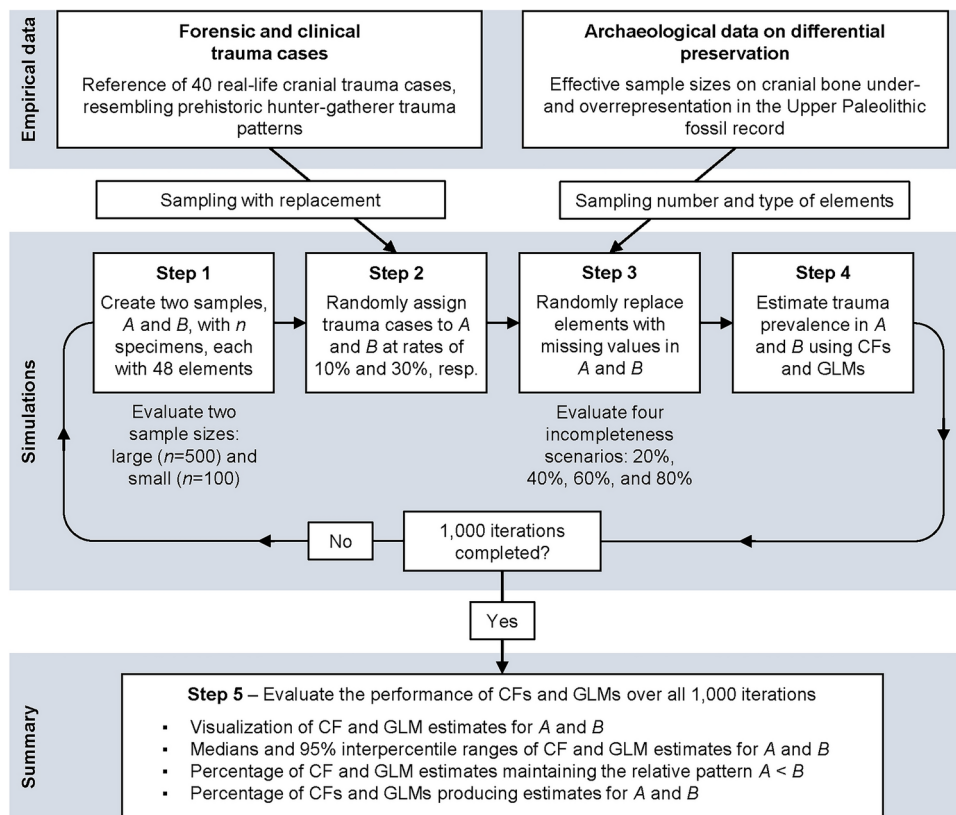
In this study, we employ a simulation framework based on empirical forensic, clinical, and archaeological data to explicitly test how closely CF- and GLM-based trauma prevalence estimates align with the known prevalence of once-complete skeletal samples after introducing increasing amounts of missing values. Our focus is solely on exploring the performance of these two common approaches in capturing the true trauma rate in incomplete samples. We do not address the statistical comparison between samples, which may also be part of bioarchaeological trauma investigations, such as when testing hypotheses about inter- and intra-sample trauma rates, including differences between sexes, sites, or age cohorts.

## Methods

Figure 1 summarizes all key stages of our simulation framework. In each simulation run, we first created an artificial dataset with two samples, A and B, each comprising  $n$  specimens. For simplicity, and in line with our previous research, our simulation approach focuses solely on the skull. Each specimen was represented by a vector of cranial elements, which served both as a scheme for recording trauma and for allocating missing values across the skull. We subdivided the skull into 14 major bones, namely, the frontal and occipital bones, as well as the left and right elements of the parietal, temporal, maxilla, mandible, zygomatic and nasal bones. Each cranial bone was further divided into four roughly equal-sized regions, oriented, for example, on anatomical structures or midlines. Due to their small size, the left and right nasal and zygomatic bones were divided into just two regions. The partitioning resulted in a total of 48 cranial elements per skull. While other procedures for assessing skeletal completeness exist<sup>14,38–41</sup>, we followed the cranial bone partitioning and trauma recording protocol previously employed by one of us<sup>36,37</sup>, allowing us to incorporate existing empirical data on cranial completeness into our simulations (see below). We performed our simulations once with large initial (i.e., before introducing missing values) sample sizes for A and B ( $n=500$ ) and once with small initial sample sizes ( $n=100$ ).

In the second step, we defined 10% of the specimens in sample A to be affected by traumatic injuries, and 30% in sample B. These arbitrary numbers reflect an average cranial trauma rate on one hand, and a slightly elevated rate on the other. For example, trauma rates around 10% are commonly found in Neolithic cranial samples<sup>42</sup>, whereas highly traumatized samples such as those resulting from violent massacres show a cranial trauma prevalence of up to 50%, or even more<sup>9,42,43</sup>. Trauma on any of the 48 cranial elements was recorded as either absent (i.e., '0') or present (i.e., '1'). Because traumatic fractures are not arbitrarily distributed across the skull but follow distinct biomechanical breakage patterns<sup>44–48</sup>, we used real-life forensic and clinical trauma cases as a reference. We compiled a set of 40 trauma cases from publications and online resources, selecting only injury scenarios that are realistic for prehistoric hunter-gatherer populations. We focused on blunt force trauma resulting from interpersonal violence (without modern-day weapons or tools), falls from a height, animal encounters, as well as projectile trauma from arrow shots. We excluded sharp force trauma, high-velocity trauma such as motor vehicle accidents or plane crashes, explosives, and gunshot trauma, as these represent injury causes with distinct trauma patterns that were not present in the deep past, such as during the Paleolithic. Being composed of both forensic and clinical cases, the severity of trauma in the selected cases ranges from minor (with only a few '1s' distributed over the 48 cranial elements) to extensive and fatal (with multiple '1s' distributed over the 48 cranial elements). Details on the fractures, injury situations, and associated references are provided in Table 1. Further information about the proportion of individual elements affected by trauma within the pool of real-life trauma cases is provided in the Supplementary Information Figure S1. From the reference pool of 40 trauma cases, we randomly assigned cases to samples A and B at rates of 10% and 30%, respectively, using sampling with replacement, meaning each case always had the same chance of being selected in every draw. Consequently, some trauma cases could appear once or multiple times in either sample, while others might not be selected at all. This sampling procedure ensures that our simulations are not systematically biased towards a possible over- or underrepresentation of minor or extensive trauma cases in either sample. This is crucial because extensive trauma cases involving multiple skeletal elements of a skull are more likely to persist in a sample when skeletal preservation decreases.

In the third step, we introduced missing values per specimen in samples A and B by deleting trauma absence/presence scores recorded on any of the 48 cranial elements (i.e., replacing '0' or '1' for a given cranial element with a missing value). This procedure may remove portions of a previously larger fracture, i.e., when one or more elements affected by a fracture extending over multiple elements are removed, but not all. This is intentional, as larger fractures may be partially preserved in the archaeological record if crania are incompletely preserved.



**Fig. 1.** Workflow summarizing all key stages of our simulation framework. In each simulation run, we first created an artificial dataset with two samples, A and B, each comprising  $n$  specimens. Each specimen was represented by a vector of 48 cranial elements, serving both as a scheme for recording trauma and for allocating missing values. We evaluated large ( $n = 500$ ) and small ( $n = 100$ ) initial sample sizes. Second, we defined 10% of the specimens in sample A to be affected by trauma, and 30% in sample B. Trauma cases were randomly sampled from a pool of 40 real-life forensic and clinical trauma cases, approximating injury scenarios and patterns that are realistic for prehistoric hunter-gatherer populations. Third, missing values were randomly added to the specimens in samples A and B, allowing some specimens to have more missing data than others, mimicking variation in preservation in the archaeological skeletal record. Drawing on empirical data from the Upper Paleolithic fossil record<sup>37</sup>, we gave cranial elements typically underrepresented in the skeletal record a higher likelihood of being assigned missing values, while those commonly well-represented were less likely to be removed. We evaluated four incompleteness scenarios with increasing levels of data missingness (on average, 20%, 40%, 60%, and 80% incomplete). Fourth, the incomplete data was used to generate CF- and GLM-based trauma prevalence estimates for samples A and B. In the CF approach, trauma frequency was estimated only over specimens with  $\geq 75\%$  completeness, while excluding all less complete remains with and without trauma. In the GLM approach, we generated model-derived point estimates of trauma prevalence prediction, using specimen completeness as a continuous covariate. For each scenario, we ran 1000 simulation iterations. Fifth, we quantified from the resulting distributions of trauma prevalence estimates: (1) the medians to assess the two approaches' overall accuracy in relation to the benchmarks; (2) the 95% interpercentile ranges to assess the approaches' precision; (3) the percentage of CF and GLM estimates that maintained the correct relative prevalence pattern between samples A and B; and (4) the percentage of how often the two approaches were able to produce prevalence estimates for either sample.

We evaluated four incompleteness scenarios with increasing levels of overall data missingness: on average, 20%, 40%, 60%, and 80% incomplete. The frequency of missing values introduced per specimen in each of the four incompleteness scenarios was not fixed. Instead, it was randomly picked from a normal distribution with a mean corresponding to either 20%, 40%, 60%, or 80%, and a standard deviation of 20. If the sampled value fell below 0%, it was set to 0%. Similarly, if the sampled value exceeded 100%, it was capped at 100%. For instance, in the 40% incomplete scenario, most specimens have around 40% missing values, with some having more, perhaps up to 100%, while others having less, perhaps 0%. This procedure allows for variability, ensuring that, in each overall incompleteness scenario, all specimens had similar but slightly different missing data amounts, mimicking preservation-related variation in the archaeological skeletal record.

The randomly selected number of missing values per specimen was then allocated to the 48 cranial elements again in a random manner. Previous research<sup>37,76</sup> has shown that some cranial elements, such as the mandible, frontal, and parietal bones, are more likely represented in the archaeological record than others, such as the

Trauma category	Age	Sex	Situation	Trauma type	Nr. of injured elements	Case ID	References
Sports accident	35	Male	Headbutt from opposition player in football game	Depressed fracture on right frontal	1	97910	<a href="#">49</a>
Sports accident	25	–	Football tackle	Fractures on the left mandibular body and the right mandibular subcondylar area	3	89829	<a href="#">50</a>
Animal encounter	Adult	Male	Kicked in the head by horse	Depressed fracture on right parietal	4	–	<a href="#">51</a>
Animal encounter	28	Male	Tamer attacked by three tigers during evening show, bite to the head	Penetrating fractures from tigers' canines and associated linear fractures on left parietal, temporal and occipital	6	–	<a href="#">52</a>
Animal encounter	35	Male	Attacked by bear in the wild, hit with paw	Multiple fractures of the right zygomatic arch, zygoma, and maxilla	6	Case 1	<a href="#">53</a>
Animal encounter	18	Female	Attacked by bear in the wild from behind, first hit with paw	Comminuted fractures of the left zygoma, zygomatic arch, and maxilla	5	Case 2	<a href="#">53</a>
Animal encounter	16	Male	Attacked by tiger in the wild, single hit with the paw	Multiple fractures of the left mandibula (posterior body and subcondylar area), left zygomatic arch, left zygoma, and left and right maxillae	9	Case 3	<a href="#">53</a>
Animal encounter	50	Female	Kicked and trampled by a horse	Comminuted fractures of the right craniofacial region involving the frontal, maxilla, nasal bone, zygoma, and zygomatic arch	9	44849	<a href="#">54</a>
Fall from a height	80	Male	Fell from a roof	Occipital condyle fracture	1	87674	<a href="#">55</a>
Fall from a height	50	Male	Fall	Fractures of the right mandibular body and subcondylar area	2	56134	<a href="#">56</a>
Fall from a height	Young adult	Female	Jumped from 7 <sup>th</sup> floor of a building (approx. 21 m), landed on lower balcony	Extensive comminuted fractures across the cranial vault, face, and mandible	34	FT-1	<a href="#">57</a>
Fall from a height	Young adult	Female	Jumped from a bridge (approx. 58 m), landed in a river	Fractures of the left mandibular body and the right mandibular angle	3	FT-2	<a href="#">57</a>
Fall from a height	Young adult	Male	Jumped from a bridge (approx. 58 m), landed in a river	Linear fracture of the left maxilla	1	FT-3	<a href="#">57</a>
Fall from a height	Middle aged adult	Female	Jumped off a cliff (approx. 45 m), landed on sand, possible impacts with the cliff edge during fall	Extensive comminuted fractures across the face, mandible, and cranial vault, particularly on the left side	19	FT-4	<a href="#">57</a>
Fall from a height	Elderly adult	Male	Fell from the roof of a house (approx. 3 m), through a plastic covering, landed on concrete floor	Linear fractures of the left and right temporal and parietal bones and the occipital	8	FT-12	<a href="#">57</a>
Fall from a height	Adolescent	Male	Fell from a bridge (approx. 15 m) onto a footpath	Linear fractures of the right temporal and parietal bones and the occipital	4	FT-14	<a href="#">57</a>
Fall from a height	Elderly adult	Male	Jumped headfirst from ladder, fell approx. 2 m, landed on concrete floor	Linear fractures of the left frontal and maxilla	4	FT-8	<a href="#">57</a>
Fall from a height	Elderly adult	Male	Suffered from natural heart disease event causing him to fall approx. 3 m from a building framework without attempt to break the fall	Multiple fractures across the frontal, occipital, left parietal and maxilla	8	FT-16	<a href="#">57</a>
Fall from a height	30	Male	Fell headlong from 2 m	Linear fracture of the frontal and right parietal	4	23979	<a href="#">58</a>
Miscellaneous clinical	15	Female	Facial trauma 8 years back, healed	Complete bone ankylosis of the right temporomandibular joint	3	88120	<a href="#">59</a>
Miscellaneous clinical	25	Male	Nasal bone fracture	Left and right nasal bone fracture	2	91256	<a href="#">60</a>
Miscellaneous clinical	30	Male	Blunt frontal trauma	Depressed frontal bone fracture involving the frontal sinus	2	90655	<a href="#">61</a>
Miscellaneous clinical	40	Male	Nasal bone fracture	Left and right nasal bone fracture	2	68197	<a href="#">62</a>
Miscellaneous clinical	Adult	–	Depressed skull fracture	Depressed fracture of left parietal	1	13084	<a href="#">63</a>
Projectile/arrow shot	25	Male	Arrow injury to the face	Penetrating fracture through the left mandibular ramus	1	73267	<a href="#">64</a>
Projectile/arrow shot	35	Male	Arrow shot through the forehead	Penetrating fracture through the frontal bone at the right frontal sinus	1	72101	<a href="#">65</a>
Violence	Adolescent	Male	Punch to head with fist causing fall to the ground	Linear fracture of the right parietal	2	BFT-2	<a href="#">66</a>
Violence	Young adult	Male	Punch to head with fist causing fall backward onto concrete pavement	Fractures of the occipital bone at the foramen magnum and the left petrous bone	4	BFT-3	<a href="#">66</a>
Violence	Middle aged adult	Male	Punched with fist, kicked with feet (no shoes), struck with statue and chair	Multiple fractures of the face involving the frontal and nasal bones, zygomas, zygomatic arches, maxillae, and mandible	19	BFT-8	<a href="#">66</a>
Continued							

Trauma category	Age	Sex	Situation	Trauma type	Nr. of injured elements	Case ID	References
Violence	Middle aged adult	Female	Assault: blunt-force head injury and compressive neck injury inflicted with hands	Fractures of the left zygoma, zygomatic arch and maxilla	5	BFT-10	<sup>66</sup>
Violence	Elderly adult	Male	Assault: fell, dragged, and disposed in creek, beaten with tree branch	Multiple fractures of the left frontal and temporal bones, nasal bones, mandible, and maxillae	19	BFT-15	<sup>66</sup>
Violence	45	Male	Assault	Subcondylar mandibular fracture on the right side	1	155152	<sup>67</sup>
Violence	20	Male	Punched in the jaw	Fracture of the left mandibular body	1	82151	<sup>68</sup>
Violence	20	Male	Alleged assault	Fracture of the left mandibular body	2	57289	<sup>69</sup>
Violence	25	Male	Punched in the jaw	Fractures of the right mandibular body and the left angle	4	54433	<sup>70</sup>
Violence	25	Male	Altercation, punched in the face	Fracture of the left zygomatic arch	2	48597	<sup>71</sup>
Violence	20	Male	Multiple punches to the face	Fracture of the right zygomatic arch	1	47820	<sup>72</sup>
Violence	16	Male	Hit with a stick	Depressed fracture of the right anterior maxillary sinus wall	1	97910	<sup>73</sup>
Violence	15	Male	Fight at school	Fractures of left and right nasal bones and left maxillary frontal process	4	89829	<sup>74</sup>
Violence	Young adult	Female	Assault	Fractures of the right temporal and parietal	3	-	<sup>75</sup>

**Table 1.** Clinical and forensic trauma cases used as a reference. Individual fracture patterns were recorded across the 48 cranial elements from published images and online CT representations. Diastatic fractures were disregarded due to their limited recognizability in a fragmentary archaeological skeletal record. Column “Nr. of injured elements” indicates how many of the 48 cranial elements were affected by trauma. Raw data are given in the R code (Supplementary Information Code S1).

nasal bones and zygoma, owing to various taphonomic factors, including differential bone structure properties and decomposition or animal and human activity<sup>39,76</sup>. To account for this variability in skeletal representation, we incorporated probability weights in the random allocation of missing values to the 48 cranial elements per specimen. Weights were obtained from published skeletal element completeness counts of an Upper Paleolithic cranial sample (~40,000–10,000 years before present) from Beier et al.<sup>37</sup> (see their Table 1; see also Beier<sup>77</sup> for raw data). From this data, we calculated ‘effective sample sizes’ for each of the 14 cranial bones by adding up fractional elements to ‘complete’ bones (i.e., the number of elements preserved per completeness category multiplied by their completeness scores, i.e., 0.25, 0.5, 0.75, 1). This approach has been described by Walker<sup>78</sup> to quantify how much a bone fragment contributes to the total sample, though for individuals instead of single cranial bones as used here. The obtained effective sample sizes used for weighing are given in the code (Supplementary Information Code S1) and are illustrated in Supplementary Information Figure S1. The weighting procedure ensures that commonly poorly represented cranial bones, such as mid-facial bones<sup>37</sup> (Fig. S1), have a higher likelihood of being allocated a missing value for one or several of their elements, while commonly well-represented cranial bones and their elements, such as vault bones and the mandible<sup>37</sup> (Fig. S1), are less likely to be removed.

In the fourth step, we recorded for each specimen: (1) the completeness, defined as the percentage of remaining cranial elements out of 48 (ranging from 0 to 100% complete), and (2) whether any of the remaining elements exhibit a trauma or not (recorded as either ‘0’ or ‘1’). This summary data per specimen was then used to estimate trauma prevalence with the two different approaches: CFs and GLMs. In the CF approach, the trauma frequency in samples A and B was estimated only over specimens with  $\geq 75\%$  completeness, while excluding all less complete remains with and without trauma, following common practice in bioarchaeological trauma studies<sup>11,15,26,28,29,31</sup>. In the GLM approach, the response variable (absence or presence of trauma) was modelled using a binomial distribution (link=log). We used the categorial predictor ‘sample’ (A or B) and the continuous covariate ‘completeness’ as fixed predictors in the models. We implemented the GLMs with the *glmmTMB* R-package<sup>79</sup>, following a custom-written guided linear modelling R-routine<sup>35</sup>. Model assessment involved the inspection of the distribution of randomized quantile residuals<sup>80</sup> and posterior predictive checks to assess overall model fit<sup>35</sup>. From the posterior distributions of fitted values obtained from 10,000 sets of model parameters, we predicted trauma prevalence for samples A and B for the 75% to 100% completeness range (to ensure comparability between GLM- and CF-based estimates) and averaged those estimates to arrive at point estimates of trauma prevalence prediction.

For both large ( $n = 500$ ) and small ( $n = 100$ ) initial sample sizes, and for each of the four incompleteness scenarios (on average, 20%, 40%, 60%, and 80% missing data), we ran 1000 simulation iterations. In the fifth and final step, we evaluated the performance of CFs and GLMs over all 1000 iterations and (1) visually displayed the shape of the resulting distributions of trauma prevalence estimates using sina plots and error bars; (2) quantified the medians to assess the overall accuracy of CF- and GLM-based trauma prevalence estimates in relation to the benchmarks; (3) estimated the 95% interpercentile ranges to assess their precision; (4) calculated the percentage with which each approach maintained the correct relative prevalence pattern between samples A and B; and (5) quantified the percentage of how often each approach was able to produce prevalence estimates for either sample.

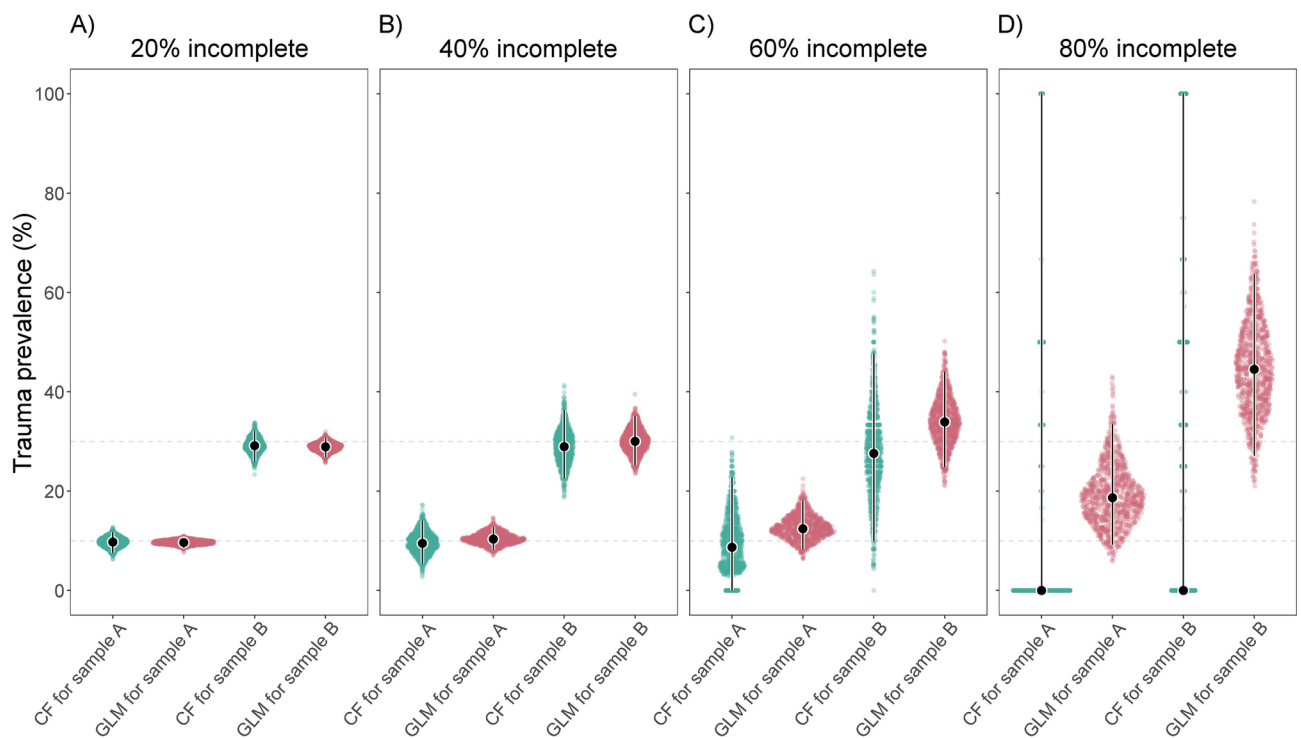
All analyses were performed in the programming language R<sup>81</sup>. Graphics were created using the package ggplot2<sup>82</sup>.

## Results

### Initial sample sizes of 500 specimens

Figure 2 displays the performance of CFs and GLMs in estimating cranial trauma prevalence for samples A and B as the level of missing data increases, with each sample consisting of 500 specimens initially. The corresponding values are presented in Table 2a. When the amount of missing data is low to moderate (20% and 40% incomplete), CFs and GLMs produced similarly accurate estimates for both samples (i.e., the medians of the spread of both CF and GLM estimates were similarly close to the benchmarks of samples A and B). However, with high to excessive amounts of missing data (60% and 80% incomplete), CFs tended to increasingly underestimate the benchmarks for both samples, while GLMs tended to overestimate them. Notably, GLM estimates were consistently more precise than CF estimates across all levels of incompleteness for both samples (i.e., the 95% interpercentile ranges of the spread of GLM estimates were narrower compared to those of CF estimates). This difference in precision becomes most pronounced in the 80% missing data scenario, where CF-based prevalence estimates span the entire range from 0 to 100% in both samples A and B. Moreover, note that CF-based estimates become increasingly skewed towards 0% as the amount of missing data increases, starting at 60% incomplete, particularly for sample A. For instance, in the 80% missing data scenario, CFs yielded a prevalence of 0% in 82.8% of simulations for sample A and in 53% for sample B, although the simulated raw data with missing values never exhibited a prevalence of 0%. In contrast, GLMs always produced prevalence estimates > 0%, in compliance with the simulated raw data.

GLMs consistently maintained the correct relative trauma prevalence pattern between samples A and B (with  $A < B$ ) across all levels of incompleteness (Table 2b). In contrast, CFs only maintained the correct relative trauma prevalence pattern when the amount of missing data was low to moderate (20% and 40% incomplete). With high to excessive amounts of missing data (60% and 80% incomplete), CFs maintained the correct relative pattern in only 94.7% and 29% of simulations, respectively, while in the remaining cases they yielded an incorrect, reversed pattern (with  $A > B$ ). An examination of the simulated raw datasets revealed that the  $A < B$  pattern was always maintained, despite the increasing amounts of missing values. While GLM estimates consistently reproduced this pattern given the raw data at hand, the CF estimates failed increasingly to do so, likely due to the excessive exclusion of specimens that were less than 75% complete.



**Fig. 2.** Sina plots showing the performance of CFs and GLMs in estimating cranial trauma prevalence for samples A and B (initially each  $n = 500$ ) with increasing levels of data missingness, assessed with 1000 simulated datasets each: (A) 20% missing data; (B) 40% missing data; (C) 60% missing data; (D) 80% missing data. Color-coding denotes CF- (green) and GLM-derived (red) estimates. Dotted horizontal lines illustrate the prevalence benchmarks for samples A and B (10% and 30%, respectively). Error bars are superimposed on the distributions to display medians (dots; indicating accuracy in relation to the benchmarks) and 95% interpercentile ranges (bars; indicating precision).

(a) Median (and 95% interpercentile range) of trauma prevalence estimates				
Scenario	CF		GLM	
	Sample A	Sample B	Sample A	Sample B
20%	9.748 (7.616–11.726)	29.132 (25.937–32.144)	9.656 (8.659–10.517)	28.916 (26.896–30.690)
40%	9.483 (5.453–14.050)	28.947 (22.608–36.363)	10.343 (8.213–12.749)	30.020 (25.269–35.059)
60%	8.696 (0.000–22.736)	27.586 (10.000–47.624)	12.426 (8.392–18.197)	33.925 (24.881–43.979)
80%	0.000 (0.000–100.000)	0.000 (0.000–100.000)	18.665 (9.261–33.494)	44.518 (27.173–63.727)
(b) Percentage of maintaining the correct relative prevalence pattern between samples A and B				
Scenario	CF		GLM	
20%	100%		100%	
40%	100%		100%	
60%	94.7%		100%	
80%	29%		100%	
(c) Percentage of producing a trauma prevalence estimate for samples A and B				
Scenario	CF		GLM	
	Sample A	Sample B	Sample A	Sample B
20%	100%	100%	100%	100%
40%	100%	100%	100%	100%
60%	100%	100%	100%	100%
80%	81%	83.2%	100%	100%

**Table 2.** Summary statistics on the performance of CFs and GLMs in estimating cranial trauma prevalence for samples A and B (benchmarks: 10% and 30%, respectively) with increasing levels of data missingness (on average 20%, 40%, 60%, and 80% incomplete), each of the four incompleteness scenarios assessed with 1000 simulated datasets. Initial sample sizes for samples A and B are 500.

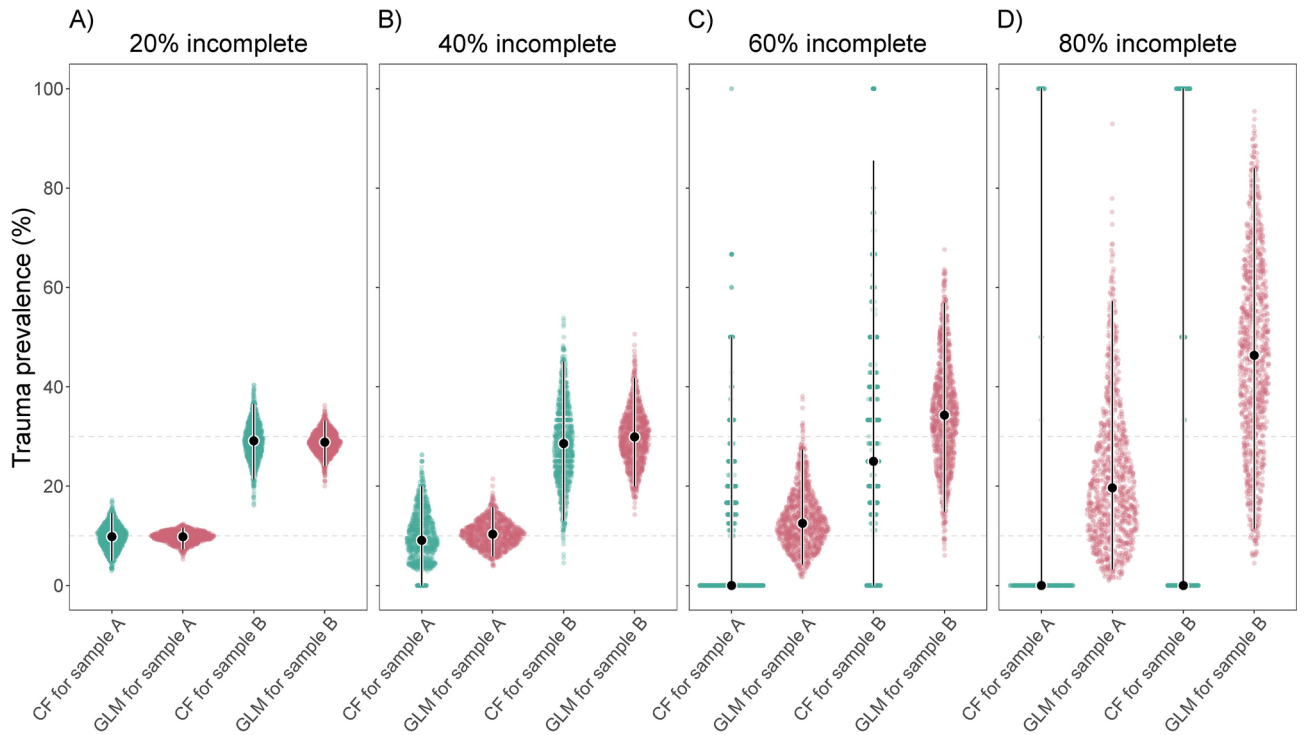
GLMs always produced trauma prevalence estimates for samples A and B across all levels of incompleteness (Table 2c). CFs only consistently produced trauma prevalence estimates when the amount of missing data ranged from low to high (20%, 40%, and 60% incomplete). With excessive amounts of missing data (80% incomplete), CFs failed in 19% of simulations in sample A and in 16.8% in sample B. Failure occurred because, after introducing missing values, none of the analyzed specimens met the predefined  $\geq 75\%$  completeness threshold.

### Initial sample sizes of 100 specimens

To assess the effect of overall smaller sample sizes, we performed our simulation analysis again using initial sample sizes of 100 instead of 500 specimens. The generated estimates are visualized in Fig. 3, with corresponding values presented in Table 3a. Again, we observe that, with low to moderate amounts of missing data (20% and 40% incomplete), both CFs and GLMs produced similarly accurate estimates for both samples. However, with high to excessive amounts of missing data (60% and 80% incomplete), CFs tended to increasingly underestimate the benchmarks for both samples, while GLMs tended to overestimate them. As before, GLM estimates were consistently more precise than CF estimates across all incompleteness scenarios and for both samples, although precision was lower overall given the now smaller sample sizes. The difference in precision between GLMs and CFs was again most pronounced in the 80% missing data scenario, where CF-based prevalence estimates spanned the entire range from 0 to 100% in both samples A and B. CF-based estimates increasingly skewed towards 0% as the amount of missing data increased, starting at 40% incomplete, especially for sample A. For example, in the 80% missing data scenario, the number of CFs yielding a trauma prevalence estimate of 0% increased to 87.3% for sample A and 63.5% for sample B, even though the simulated raw data with missing values again never exhibited a prevalence of 0%. In contrast, GLMs always produced prevalence estimates greater than 0%, in compliance with the simulated raw data.

As before, GLMs always maintained the correct relative trauma prevalence pattern between samples (with  $A < B$ ) across all levels of incompleteness (Table 3b). CFs only maintained the correct relative trauma prevalence pattern when the amount of missing data was low (20% incomplete). With moderate to excessive amounts of missing data (40%, 60%, and 80% incomplete), however, CFs yielded the correct relative pattern in only 96.7%, 61.5%, and 2.8% of simulations, respectively, and otherwise produced the incorrect, reverse pattern with  $A > B$ . An examination of the simulated raw datasets again confirmed that they always preserved the  $A < B$  pattern, regardless of the amounts of missing values. While GLM estimates consistently reflected this pattern based on the available raw data, CF estimates increasingly failed to do so due to the excessive exclusion of specimens that were less than 75% complete.

As also observed before, GLMs always produced trauma prevalence estimates for samples A and B across all incompleteness scenarios (Table 3c). CFs, on the other hand, only consistently produced trauma prevalence estimates when the amount of missing data ranged from low to moderate (20% and 40% incomplete). With high to excessive amounts of missing data (60% and 80% incomplete), CFs failed in 1.5% and 72.4% of simulations in



**Fig. 3.** Sina plots showing the performance of CFs and GLMs in estimating cranial trauma prevalence for samples A and B (initially each  $n=100$ ) with increasing levels of data missingness, assessed with 1000 simulated datasets each: (A) 20% missing data; (B) 40% missing data; (C) 60% missing data; (D) 80% missing data. Color-coding denotes CF- (green) and GLM-derived (red) estimates. Dotted horizontal lines illustrate the prevalence benchmarks for samples A and B (10% and 30%, respectively). Error bars are superimposed on the distributions to display medians (dots; indicating accuracy in relation to the benchmarks) and 95% interpercentile ranges (bars; indicating precision).

(a) Median (and 95% interpercentile range) of trauma prevalence estimates				
Scenario	CF		GLM	
	Sample A	Sample B	Sample A	Sample B
20%	9.836 (5.000–14.516)	29.129 (21.426–36.508)	9.820 (7.315–11.648)	28.850 (24.061–33.148)
40%	9.091 (0.000–20.000)	28.571 (13.043–45.000)	10.328 (5.904–15.760)	29.913 (19.997–41.732)
60%	0.000 (0.000–50.000)	25.000 (0.000–85.500)	12.521 (4.264–27.304)	34.298 (14.936–56.832)
80%	0.000 (0.000–100.000)	0.000 (0.000–100.000)	19.665 (3.237–57.237)	46.331 (11.459–83.908)
(b) Percentage of maintaining the correct relative prevalence pattern between samples A and B				
Scenario	CF		GLM	
20%	100%		100%	
40%	96.7%		100%	
60%	61.5%		100%	
80%	2.8%		100%	
(c) Percentage of producing a trauma prevalence estimate for samples A and B				
Scenario	CF		GLM	
	Sample A	Sample B	Sample A	Sample B
20%	100%	100%	100%	100%
40%	100%	100%	100%	100%
60%	98.5%	99%	100%	100%
80%	27.6%	28.2%	100%	100%

**Table 3.** Summary statistics on the performance of CFs and GLMs in estimating cranial trauma prevalence for samples A and B (benchmarks: 10% and 30%, respectively) with increasing levels of data missingness (on average 20%, 40%, 60%, and 80% incomplete), each of the four incompleteness scenarios assessed with 1000 simulated datasets. Initial sample sizes for samples A and B are 100.

sample A, and in 1% and 71.8% of simulations in sample B, respectively. Failure occurred because no specimens met the predefined  $\geq 75\%$  completeness threshold after missing values were introduced.

## Discussion

Estimating trauma prevalence from incomplete skeletal remains is hampered by the state of their completeness, as traumatic lesions that were once present may be absent and thus go undiagnosed<sup>11–13</sup>. This bias can result in false negative findings, increasing proportionally with decreasing completeness<sup>11,12,14–16</sup>, and can impede comparability between samples of varying completeness if not accounted for. Consequently, assessments of trauma prevalence based on CFs often restrict analyses to specimens meeting a minimum preservation criterion, such as  $\geq 75\%$  completeness<sup>11,12,14,15,19,24,26–31</sup>. While the systematic exclusion of partial remains below a predefined completeness threshold can help mitigate completeness bias, selecting an appropriate threshold remains challenging and may vary from study to study, as it involves a trade-off balancing reliable CFs against maintaining an adequate sample size, making it a potential source of bias itself. This underscores the need to integrate alternative data analysis routines, which can provide more generalized avenues for small and fragmented datasets commonly encountered in bioarchaeological investigations. GLMs are a promising solution to this problem as they can overcome the aforementioned trade-off by including even the smallest skeletal elements without requiring the prior exclusion of remains. The use of GLMs and related approaches for studying patterns of skeletal lesions has become increasingly popular among bioanthropologists in the last few decades, particularly to study enthesal changes and degenerative joint disease<sup>34,83–88</sup>. Recently, GLMs and their extensions have also gained traction for investigating trauma patterns<sup>20,36,37,89–92</sup>.

Here, we assessed how closely CF- and GLM-based trauma prevalence estimates align with the known prevalence of two once-complete cranial samples after introducing increasing amounts of missing data. This was done to explore the extent to which trauma rates in incomplete samples are adequately reflected by these two approaches. As expected, we found that accuracy and precision decreased for both CFs and GLMs as skeletal completeness decreased. This pattern was observed in both larger and smaller sample sizes but was more pronounced in the latter. When the amount of missing data was low to moderate (20% and 40% incomplete), both CFs and GLMs generated similarly accurate trauma prevalence estimates for both samples. However, with high to excessive amounts of missing data (60% and 80% incomplete), CFs tended to increasingly underestimate the benchmarks for both samples, while GLMs tended to overestimate them. Importantly, we found that GLM estimates were consistently more precise than CF estimates across all incompleteness scenarios and for both sample sizes. Alarming, as the amount of missing data increased, CF-based estimates skewed towards 0%, frequently yielded an incorrect reversed relative prevalence pattern across samples (with  $A > B$ ), and occasionally failed to produce any prevalence estimates, particularly with small sample sizes. The shortcomings of the CF approach result from the considerable reduction of sample sizes when a completeness threshold, such as  $\geq 75\%$  as employed in this study, is applied to increasingly incomplete skeletal samples. In these cases, resulting CF estimates are based on unrepresentative subsamples comprising only a few, or sometimes no, specimens that meet the completeness threshold from the original larger sample. Therefore, we caution against using CFs in bioarchaeological contexts where samples primarily comprise small fractions of skeletal specimens.

Although GLMs consistently outperformed CFs in terms of precision, the accuracy and precision of both approaches decreased as skeletal completeness decreased, with their absolute estimates often deviating substantially from the benchmarks when missing values were excessive. This underscores the most crucial prerequisite for obtaining reliable trauma prevalence estimates that applies to both approaches, namely, striving for well-designed studies with sufficiently large and representative samples. In bioarchaeological studies, however, sample size and representativity are usually dictated by the skeletal remains preserved and available from excavations, rather than by study design. In our view, it is therefore indispensable for any trauma prevalence study to quantify the completeness of all available skeletal remains in a sample prior to analysis to determine its representativity, which impacts the choice of analytical approach and the reliability and interpretation of results.

Even though our methodological exploration included simulated datasets with excessive amounts of missing data (80% incomplete) to exemplify preservation-related effects on trauma prevalence estimates, we generally consider such incomplete samples unsuitable for studying trauma rates, regardless of the approach used. In instances where critically incomplete samples are the only option available for analysis, it becomes even more crucial to quantify their completeness, report their state of preservation, and be cautious regarding the limits of interpretations. In this context, it is important to note that while CFs provide only a single point estimate of trauma prevalence, GLM-derived trauma prevalence predictions include two estimates: a point estimate of mean prevalence and a compatibility interval of the error around the mean, signaling uncertainty in the estimate, such as resulting from the underlying sample size. To ensure comparability among approaches in our simulation framework, we relied only on CF and GLM point estimates, without further considering the GLM compatibility intervals. In bioarchaeological applications, however, the GLM compatibility intervals can help evaluate the estimate and its uncertainty, potentially resulting from limited sample sizes—a feature absent in the CF approach.

In the 60% and 80% missing data scenarios, we observe that GLM estimates not only range widely around the benchmarks but also tend to overestimate trauma prevalence (Figs. 2C, D and 3C, D). By default, GLMs allow for the explicit specification of the completeness score for which trauma prevalence estimates are to be predicted. To ensure comparability in our simulation framework between the GLM and CF approaches, the latter employing a  $\geq 75\%$  completeness threshold, we predicted GLM trauma prevalence across the entire 75% to 100% completeness score range and averaged those estimates. We suspect that an overestimation of trauma prevalence might result from an underrepresentation of largely complete specimens in the 60% and especially in the 80% incomplete datasets, limiting the models' references for predicting prevalence among complete specimens, which appears to lead to an over-extrapolation from lower completeness scores. Indeed, when we predicted for the sequence of 50%–75% completeness instead, estimates were lower overall, while relative prevalence patterns

across samples remained constant (results not shown). Therefore, in cases of largely incomplete samples, we advise not to take the absolute GLM trauma predictions at face value but to focus on relative prevalence patterns between samples instead. This conclusion holds implications for previous studies that utilized such model-based predictions to infer cranial trauma prevalence from incomplete fossil remains of Late Pleistocene humans<sup>36,37</sup>. It cautions against overinterpreting predicted absolute trauma prevalence values, while reinforcing findings on relative trauma patterns in Neanderthals and Upper Paleolithic humans, including similar overall and sex-specific cranial trauma rates in the two taxa along with possible age-related differences<sup>36,37</sup>.

Another conceptual advantage of GLMs over CFs—not integrated into our study for the sake of direct comparability of the approaches—is their ability to simultaneously explore the effects of multiple predictor variables and their interactions on trauma prevalence (such as age at death, sex, or archaeological period). This stands in contrast to CFs, which can only assess a single predictor variable at a time. Moreover, in the form of generalized linear mixed models (hereafter, GLMMs), their utility is further enhanced by the ability to account for repeated observations and further dependency structures of the data through partial pooling using random effects<sup>35,93,94</sup>. This can help to avoid potential pseudoreplication, an issue that can arise in bioarchaeological trauma studies when linear models are implemented without specifying appropriate random effects—such as in cases where multiple cranial elements affected by the same fracture, or multiple traumas on the same individual, are recorded separately. The versatility of GL(M)Ms comes at the cost of higher computational complexity compared to CFs. However, a recently published step-by-step workflow in the programming language R, which requires minimal coding skills and is directed at users of different experience levels<sup>35</sup>, greatly facilitates the implementation of GL(M)Ms, alongside other best-practice and troubleshooting literature<sup>93–100</sup>.

In conclusion, we generally recommend GL(M)Ms for estimating trauma prevalence from incomplete skeletal samples and caution against using CFs, especially when dealing with largely incomplete and small samples. In addition to their overall superior performance demonstrated in this study, GL(M)Ms offer conceptual advantages that are typically pertinent to most bioarchaeological investigations. These include the ability to accommodate complex data structures, such as hierarchical designs, and to assess multiple predictor variables simultaneously in both the estimation of trauma prevalence and their statistical comparison between samples. We demonstrated that neither approach is fully reliable when dealing with largely incomplete samples. Nevertheless, in such cases, GL(M)Ms can at least help to assess relative prevalence patterns between samples. In this context, it is important to note that, while powerful, GL(M)Ms are only statistical models and should thus be used with caution. In particular, if the goal is causal inference, their formulation should be based on appropriate causal models<sup>101</sup>.

Lastly, it is important to emphasize that our simulation study is grounded on empirical data relevant to archaeological contexts of the Last Glacial, incorporating real-life cases of blunt force cranial trauma which are realistic for hunter-gatherers (resulting from interpersonal violence, animal encounters, falls from a height, and projectile trauma from arrow shots) and information about the completeness of cranial remains from the Upper Paleolithic period. Consequently, our findings may not necessarily apply to other archaeological contexts, such as medieval or early modern periods, where markedly different injury scenarios may have existed (including sharp force trauma, explosives, and gunshot trauma), and to taphonomic contexts where the relative representation patterns of skeletal remains might be different. Moreover, our study was restricted to cranial remains and focused exclusively on the impact of data missingness and sample sizes on trauma prevalence estimates using the two approaches of CFs and GLMs. Future research could, for example, expand such methodological explorations of sample incompleteness on trauma prevalence estimates to the post-cranial skeleton, explore more complex research designs involving multiple predictor variables and their interactions, and compare the accuracy of different GL(M)M algorithms and software packages for predicting trauma prevalence.

## Data availability

All data and code needed for reproducing or extending our analyses are presented or referenced in the main text and supplementary materials.

Received: 14 June 2024; Accepted: 11 October 2024

Published online: 12 November 2024

## References

- Allen, M. W., Bettinger, R. L., Codding, B. F., Jones, T. L. & Schmitz, A. W. Resource scarcity drives lethal aggression among prehistoric hunter-gatherers in central California. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12120–12125 (2016).
- Larsen, C. S. *Bioarchaeology: Interpreting Behavior from the Human Skeleton* (Cambridge University Press, 2015). <https://doi.org/10.1017/CBO9781139020398>.
- Martin, D. L., Harrod, R. P. & Fields, M. Beaten down and worked to the bone: Bioarchaeological investigations of women and violence in the ancient Southwest. *Landsc. Violence* **1**, 1–19 (2010).
- Martin, D. L. & Harrod, R. P. Bioarchaeological contributions to the study of violence. *Am. J. Phys. Anthropol.* **156**(Suppl 59), 116–145 (2015).
- Martin, D. L. & Harrod, R. P. The bioarchaeology of pain and suffering: Human adaptation and survival during troubled times. *Archeol. Pap. Am. Anthropol. Assoc.* **27**, 161–174 (2016).
- Mirazón Lahr, M. *et al.* Inter-group violence among early Holocene hunter-gatherers of West Turkana, Kenya. *Nature* **529**, 394–8 (2016).
- Redfern, R. C. *Injury and Trauma in Bioarchaeology: Interpreting Violence in Past Lives* (Cambridge University Press, 2017).
- Redfern, R. C. Identifying and interpreting domestic violence in archaeological human remains: A critical review of the evidence. *Int. J. Osteoarchaeol.* **27**, 13–34 (2017).
- Wahl, J. & Trautmann, I. The Neolithic massacre at Talheim: A pivotal find in conflict archaeology. In *Sticks, stones, and broken bones: Neolithic violence in a European perspective* (eds. Schulting, R. J. & Fibiger, L.) 77–100 (Oxford University Press, 2012).
- Wilkinson, R. G. & van Wagenen, K. M. Violence against women: Prehistoric skeletal evidence from Michigan. *Midcont. J. Archaeol.* **18**, 190–216 (1993).

11. Judd, M. A. Comparison of long bone trauma recording methods. *J. Archaeol. Sci.* **29**, 1255–1265 (2002).
12. Judd, M. A. Trauma in the city of Kerma: Ancient versus modern injury patterns. *Int. J. Osteoarchaeol.* **14**, 34–51 (2004).
13. Lovejoy, C. O. & Heiple, K. G. The analysis of fractures in skeletal populations with an example from the Libben site, Ottawa County, Ohio. *Am. J. Phys. Anthropol.* **55**, 529–541 (1981).
14. Buikstra, J. & Ubelaker, D. H. Standards for Data Collection from Human Skeletal Remains: Proceedings of a Seminar at the Field Museum of Natural History, vol. 44 (Arkansas Archeological Survey, Fayetteville, Ark., 1994).
15. Fibiger, L., Ahlström, T., Bennike, P. & Schulting, R. J. Patterns of violence-related skull trauma in Neolithic southern Scandinavia. *Am. J. Phys. Anthropol.* **150**, 190–202 (2013).
16. Pankowská, A., Galeta, P., Uhlík Spěváčková, P. & Nováček, K. Violence in European medieval monasteries: Skeletal trauma in Teplá monastery (Czech Republic). *Int. J. Osteoarchaeol.* **24**, 305 (2019).
17. Dittmar, J. M., Mitchell, P. D., Cessford, C., Inskip, S. A. & Robb, J. E. Fancy shoes and painful feet: Hallux valgus and fracture risk in medieval Cambridge, England. *Int. J. Paleopathol.* **35**, 90–100 (2021).
18. Scott, R. M. & Buckley, H. R. Exploring prehistoric violence in Tonga: Understanding skeletal trauma from a biocultural perspective. *Curr. Anthropol.* **55**, 335–347 (2014).
19. Alvrus, A. Fracture patterns among the Nubians of Semna South, Sudanese Nubia. *Int. J. Osteoarchaeol.* **9**, 417–429 (1999).
20. McCool, W. C., Tung, T. A., Coltrain, J. B., Accinelli-Obando, A. J. & Kennett, D. J. The character of conflict: A bioarchaeological study of violence in the Nasca highlands of Peru during the Late Intermediate Period (950–1450 C.E.). *Am. J. Phys. Anthropol.* **174**, 614–630 (2021).
21. Robb, J. Violence and gender in early Italy. In *Troubled Times: Violence and Warfare in the Past* (eds. Martin, D. L. & Frayer, D. W.) 111–144 (Gordon & Breach, 1997).
22. Krakowka, K. Patterns and prevalence of violence-related skull trauma in medieval London. *Am. J. Phys. Anthropol.* **164**, 488–504 (2017).
23. Šlaus, M., Novak, M., Bedić, Ž & Strinović, D. Bone fractures as indicators of intentional violence in the eastern Adriatic from the Antique to the late medieval period (2nd–16th century AD). *Am. J. Phys. Anthropol.* **149**, 26–38 (2012).
24. Judd, M. A. Continuity of interpersonal violence between Nubian communities. *Am. J. Phys. Anthropol.* **131**, 324–333 (2006).
25. Judd, M. A. & Roberts, C. A. Fracture trauma in a medieval British farming village. *Am. J. Phys. Anthropol.* **109**, 229–243 (1999).
26. Cohen, H. et al. Trauma to the skull: A historical perspective from the southern Levant (4300BCE–1917CE). *Int. J. Osteoarchaeol.* **24**, 722–736 (2014).
27. Delgado-Darias, T., Alberto-Barroso, V. & Velasco-Vázquez, J. Violence in paradise: Cranial trauma in the prehispanic population of Gran Canaria (Canary Islands). *Am. J. Phys. Anthropol.* **166**, 70–83 (2018).
28. Gheggi, M. S. Conflict in pre-hispanic Northwest Argentina: Implications arising from human bone trauma patterns. *Int. J. Osteoarchaeol.* **26**, 17–27 (2016).
29. Schulting, R. J. & Fibiger, L. Violence in Neolithic North-west Europe: A population perspective. In *Early Farmers: The View from Archaeology and Science*, vol. 198 (eds. Whittle, A. W. R. & Bickel, P.) (Oxford Univ. Press, 2014).
30. Torres-Rouff, C. Hiding inequality beneath prosperity: Patterns of cranial injury in Middle Period San Pedro de Atacama, Northern Chile. *Am. J. Phys. Anthropol.* **146**, 28–37 (2011).
31. Torres-Rouff, C., Hubbe, M. & Pestle, W. J. Wearing the marks of violence: Unusual trauma patterning at Coyo Oriental, Northern Chile. *Am. J. Phys. Anthropol.* **167**, 32–45 <https://doi.org/10.1002/ajpa.23600> (2018).
32. Montgomery, D. C., Peck, E. A. & Vining, G. G. *Introduction to Linear Regression Analysis* (Wiley, 2021).
33. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, 2006). <https://doi.org/10.1017/CBO9780511790942>.
34. Nikita, E. The use of generalized linear models and generalized estimating equations in bioarchaeological studies. *Am. J. Phys. Anthropol.* **153**, 473–483 (2014).
35. Santon, M., Korner-Nievergelt, F., Michiels, N. K. & Anthes, N. A versatile workflow for linear modelling in R. *Front. Ecol. Evol.* **11**, 745 (2023).
36. Beier, J., Anthes, N., Wahl, J. & Harvati, K. Similar cranial trauma prevalence among Neanderthals and Upper Palaeolithic modern humans. *Nature* **563**, 686–690 (2018).
37. Beier, J., Anthes, N., Wahl, J. & Harvati, K. Prevalence of cranial trauma in Eurasian Upper Paleolithic humans. *Am. J. Phys. Anthropol.* **174**, 268–284 (2021).
38. Knüsel, C. J. & Outram, A. K. Fragmentation: The zonation method applied to fragmented human remains from archaeological and forensic contexts. *Environ. Archaeol.* **9**, 85–97 (2004).
39. Bello, S. M. & Andrews, P. The intrinsic pattern of preservation of human skeletons and its influence on the interpretation of funerary behaviours. In *Social Archaeology of Funerary Remains* (eds. Gowland, R. & Knüsel, C. J.) 1–13 (Oxbow Books, 2006).
40. Stodder, A. L. W. Taphonomy and the nature of archaeological assemblages. In *Biological Anthropology of the Human Skeleton* (eds. Katzenberg, M. A. & Grauer, A. L.) 73–115 (Wiley, 2019). <https://doi.org/10.1002/9781119151647.ch3>.
41. Fernández-Jalvo, Y. & Andrews, P. Disarticulation and completeness. In *Atlas of Taphonomic Identifications* 311–324 (Springer Netherlands, 2016). [https://doi.org/10.1007/978-94-017-7432-1\\_10](https://doi.org/10.1007/978-94-017-7432-1_10).
42. Fibiger, L., Ahlström, T., Meyer, C. & Smith, M. Conflict, violence, and warfare among early farmers in Northwestern Europe. *Proc. Natl. Acad. Sci. U. S. A.* **120**, 2209481119 (2023).
43. Meyer, C., Lohr, C., Gronenborn, D. & Alt, K. W. The massacre mass grave of Schöneck-Kilianstädten reveals new insights into collective violence in Early Neolithic Central Europe. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11217–11222 (2015).
44. Berryman, H. E. & Symes, S. A. Recognizing gunshot and blunt cranial trauma through fracture interpretation. In *Forensic Osteology: Advances in the Identification of Human Remains* (ed. Reichs, K.) 333–352 (Charles C. Thomas, 1998).
45. Galloway, A. & Wedel, V. L. Bones of the skull, the dentition, and osseous structures of the throat. In *Broken Bones: Anthropological Analysis of Blunt Force Trauma* (eds. Wedel, V. L. & Galloway, A.) 133–160 (Charles C. Thomas, Publisher, Ltd, 2014).
46. Isa, M. I. et al. Experimental investigation of cranial fracture initiation in blunt human head impacts. *Forensic Sci. Int.* **300**, 51–62 (2019).
47. Kroman, A., Kress, T. & Porta, D. Fracture propagation in the human cranium: A re-testing of popular theories. *Clin. Anat. N. Y.* **24**, 309–318 (2011).
48. Symes, S. A., L'Abbé, E. N., Chapman, E. N., Wolff, I. & Dirkmaat, D. C. Interpreting Traumatic Injury to Bone in Medicolegal Investigations. In *A Companion to Forensic Anthropology* (ed. Dirkmaat, D. C.) 340–389 (Wiley, 2012). <https://doi.org/10.1002/9781118255377.ch17>.
49. Haouimi, A. Depressed frontal fracture: Case study (rID: 97910) 1053347–97910 <https://doi.org/10.53347/rID-97910> (2022).
50. Vajuhudeen, Z. Mandibular fractures: Case study (rID: 89829) 1053347–89829 <https://doi.org/10.53347/rID-89829> (2021).
51. Kranioti, E. F. Forensic investigation of cranial injuries due to blunt force trauma: Current best practice. *Res. Rep. Forensic Med. Sci.* **5**, 25–37. <https://doi.org/10.2147/RRFMS.S70423> (2015).
52. Emami, P. et al. Case report: Surviving a tiger attack. *Neurosurg. Rev.* **35**, 621–624 (2012).
53. Bhandari, K., Haque, I. B., Khanal, B., Joshi, S. & Khatri, K. S. Maxillofacial injuries in bear, tiger, and jackal attacks. *Craniomaxillofac. Trauma Reconstr. Open* **3**, 0039–1678671 (2019).
54. Hatwal, A. Complex craniofacial fractures: Case study (rID: 44849). <https://doi.org/10.53347/rID-44849> (2016).
55. Hacking, C. Occipital condyle fracture (type 1) and atlas transverse process fracture (type 5): Case study (rID: 87674). <https://doi.org/10.53347/rID-87674> (2021).

56. Knipe, H. Mandibular fractures with dental trauma: Case study (rID: 56134). <https://doi.org/10.53347/rID-56134> (2017).
57. Rowbotham, S. K. Blunt-Force Trauma: Fatal falls (FT). In *An Atlas of Skeletal Trauma in Medico-Legal Contexts* (eds. Blau, S. et al.) 189–274 (Elsevier, 2018). <https://doi.org/10.1016/B978-0-12-803759-1.00003-5>.
58. Cuete, D. Extradural haematoma—venous: Case study (rID: 23979). <https://doi.org/10.53347/rID-23979> (2013).
59. Foladi, N. Temporomandibular joint ankylosis: Case study (rID: 88120). <https://doi.org/10.53347/rID-88120> (2021).
60. Rasuli, B. Nasal bone fracture: Case study (rID: 91256). <https://doi.org/10.53347/rID-91256> (2021).
61. Haouimi, A. Depressed skull fracture—frontal sinuses: Case study (rID: 90655). <https://doi.org/10.53347/rID-90655> (2021).
62. Rasuli, B. Nasal bone fracture: Case study (rID: 68197). <https://doi.org/10.53347/rID-68197> (2019).
63. Al Salam, H. M. Depressed skull fracture: Case study (rID: 13084). <https://doi.org/10.53347/rID-13084> (2011).
64. Hartung, M. P. Arrow injury to the face: Case study (rID: 73267). <https://doi.org/10.53347/rID-73267> (2020).
65. Hartung, M. P. Arrow injury to the brain: Case study (rID: 72101). <https://doi.org/10.53347/rID-72101> (2019).
66. Blau, S., Ranson, D. & O'Donnell, C. Blunt-Force Trauma (BFT). In *An Atlas of Skeletal Trauma in Medico-Legal Contexts* (eds. Blau, S. et al.) 99–188 (Elsevier, 2018). <https://doi.org/10.1016/B978-0-12-803759-1.00002-3>.
67. Jabaz, D. F. Mandible fracture: Case study (rID: 155152). <https://doi.org/10.53347/rID-155152> (2022).
68. Hacking, C. Mandibular and dental fracture: Case study (rID: 82151). <https://doi.org/10.53347/rID-82151> (2020).
69. Knipe, H. Mandibular fracture with dental trauma: Case study (rID: 57289). <https://doi.org/10.53347/rID-57289> (2018).
70. Murphy, A. Mandibular fracture: Case study (rID: 54433). <https://doi.org/10.53347/rID-54433> (2017).
71. Murphy, A. Zygomatic arch fracture: Case study (rID: 48597). <https://doi.org/10.53347/rID-48597> (2016).
72. Murphy, A. Zygomatic arch fracture: Case study (rID: 47820). <https://doi.org/10.53347/rID-47820> (2016).
73. Cuete, D. Depressed fracture of the anterior wall of the right maxillary sinus: Case study (rID: 34231). <https://doi.org/10.53347/rID-34231> (2015).
74. Cuete, D. Nasal bone fracture: Case study (rID: 23366). <https://doi.org/10.53347/rID-23366> (2013).
75. Errickson, D. et al. A survey of case studies on the use of forensic three-dimensional printing in England and Wales. *Int. J. Legal Med.* **136**, 1605–1619 (2022).
76. Ullrich, H. Patterns of skeletal representation, manipulations on human corpses and bones, mortuary practices and the question of cannibalism in the European Palaeolithic—an anthropological approach. *OPUS: Interdiscip. Investig. Archaeol. Mosc.* **3**, 24–40 (2004).
77. Beier, J. *Data for Publication 'Prevalence of Cranial Trauma in Eurasian Upper Paleolithic Humans': Zenodo Data Repository* (2020). 10.5281/ZENODO.3979485.
78. Walker, P. L. Wife beating, boxing, and broken noses: Skeletal evidence for the cultural patterning of violence. In *Troubled Times: Violence and Warfare in the Past*, vol. 3 (eds. Martin, D. L. & Frayer, D. W.) 145–180 (Gordon & Breach, 1997).
79. Brooks, M. E. et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*. **9**, 378–400 (2017).
80. Hartig, F. DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models: R package version 0.4.6. (2022).
81. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing (2022).
82. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016). <https://doi.org/10.1007/978-3-319-24277-4>.
83. Alonso-Llamazares, C., Blanco-Márquez, B., Lopez, B. & Pardiñas, A. F. Assessing individual and population variability in degenerative joint disease prevalence using generalized linear mixed models. *Am. J. Phys. Anthropol.* **175**, 611–625 (2021).
84. Alonso-Llamazares, C., Lopez, B. & Pardiñas, A. Sex differences in the distribution of enthesal changes: Meta-analysis of published evidence and its use in Bayesian paleopathological modeling. *Am. J. Biol. Anthropol.* **177**, 249–265 (2022).
85. Laffranchi, Z., Charisi, D., Jiménez-Brobeil, S. A. & Milella, M. Gendered division of labor in a Celtic community? A comparison of sex differences in enthesal changes and long bone shape and robusticity in the pre-Roman population of Verona (Italy, third-first century BC). *Am. J. Phys. Anthropol.* **173**, 568–588 (2020).
86. Michopoulou, E., Nikita, E. & Valakos, E. D. Evaluating the efficiency of different recording protocols for enthesal changes in regards to expressing activity patterns using archival data and cross-sectional geometric properties. *Am. J. Phys. Anthropol.* **158**, 557–568 (2015).
87. Nikita, E., Mattingly, D. & Lahr, M. M. Methodological considerations in the statistical analysis of degenerative joint and disc disease. *Int. J. Paleopathol.* **3**, 105–112 (2013).
88. Winburn, A. P. & Stock, M. K. Reconsidering osteoarthritis as a skeletal indicator of age at death. *Am. J. Phys. Anthropol.* **170**, 459–473 (2019).
89. McCool, W. C. et al. Climate change-induced population pressure drives high rates of lethal violence in the Prehispanic central Andes. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 2117556119 (2022).
90. McCool, W. C., Vernon, K. B., Yaworsky, P. M. & Codding, B. F. Subsistence strategy mediates ecological drivers of human violence. *PLOS ONE* **17**, 0268257 (2022).
91. McCool, W. C., Wilson, K. M. & Vernon, K. B. Ecological Constraints on Violence Avoidance Tactics in the Prehispanic Central Andes. *Environ. Archaeol.* **2022**, 1–14 (2022) <https://doi.org/10.1080/14614103.2022.2137652>.
92. Snyder, T. J. & Haas, R. Climate change intensified violence in the south-central Andean highlands from 1.5 to 0.5 ka. *Quat. Res.* **115**, 109–119 (2023).
93. Bolker, B. M. et al. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends Ecol. Evol.* **24**, 127–135 (2009).
94. Bolker, B. M. Linear and generalized linear mixed models. In *Ecological Statistics* (eds. Fox, G. A. et al.) 309–333 (Oxford University Press, 2015). <https://doi.org/10.1093/acprof:oso/9780199672547.003.0014>.
95. Gelman, A., Hill, J. & Vehtari, A. *Regression and Other Stories* (Cambridge University Press, 2021). <https://doi.org/10.1017/9781139161879>.
96. Harrison, X. A. et al. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* **6**, 4794 (2018).
97. Korner-Nievergelt, F. et al. *Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and Stan* (Academic Press, 2015).
98. Silk, M. J., Harrison, X. A. & Hodgson, D. J. Perils and pitfalls of mixed-effects regression models in biology. *PeerJ* **8**, 9522 (2020).
99. Zuur, A. F., Hilbe, J. M. & Ieno, E. N. *A Beginners Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists* (Highland Statistics Ltd, 2013).
100. Zuur, A. F. & Ieno, E. N. A protocol for conducting and presenting results of regression-type analyses. *Methods Ecol. Evol.* **7**, 636–645 (2016).
101. Pearl, J. & Mackenzie, D. *The Book of Why: The New Science of Cause and Effect* (Penguin Books, 2019).

## Acknowledgements

This research was funded by the German Research Foundation (DFG FOR 2237, awarded to Katerina Harvati, University of Tübingen) and the Senckenberg Society for Nature Research (Senckenberg Gesellschaft für Naturforschung), Frankfurt/Main. MS was supported by a MSCA 2021 postdoctoral fellowship (101066328) funded via the Engineering and Physical Sciences Research Council (grant number EP/X020819/1). We acknowledge funding from the Open Access Publication Fund of the University of Tübingen as well as the Federal Ministry

of Education and Research (BMBF) and the Baden-Württemberg Ministry of Science as part of the Excellence Strategy of the German Federal and State Governments (awarded to JB). We are grateful to Marina Tselioudis (Binghamton University) for support in literature research and data collection on the forensic and clinical reference trauma cases.

### Author contributions

JB: Conceptualization; methodology; investigation; data curation; visualization; writing—original draft; writing—review & editing MS: Methodology; software; writing—review & editing HR: Conceptualization; methodology; software; formal analysis; data curation; visualization; writing—original draft; writing—review & editing.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-76231-1>.

**Correspondence** and requests for materials should be addressed to J.B. or H.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024