

RESEARCH PAPER



# What use are ontogenetic data anyway? Challenges in multivariate modelling of primate tooth formation

Christopher A. Wolfe<sup>a</sup>, Julia C. Boughner<sup>b</sup> and Kyra E. Stull<sup>c,d</sup>

<sup>a</sup>Department of Anthropology, East Carolina University, Greenville, NC, USA; <sup>b</sup>Department of Anatomy, Physiology, and Pharmacology, University of Saskatchewan, Saskatoon, Canada; <sup>c</sup>Department of Anthropology, University of Nevada, Reno, Reno, NV, USA; <sup>d</sup>Faculty of Health Sciences, Department of Anatomy, University of Pretoria, Pretoria, South Africa

## ABSTRACT

**Background:** The development of the permanent dentition provides a reliable substrate to infer ontogenetic patterns within and among species. Multivariate methods offer a promising approach to compare taxon-specific patterns.

**Aim:** This study used multivariate statistical approaches to compare ontogenetic patterns by more comprehensively quantifying variation in crypt and tooth formation scores for the permanent dentition in five catarrhine primate taxa, *Homo sapiens*, *Pan paniscus*, *Pan troglodytes*, *Hylobates lar*, and *Papio anubis*.

**Subjects and Methods:** Tooth formation was scored according to published standards for each specimen. Multivariate relationships between teeth were modelled according to a Bayesian multivariate cumulative probit model. Relationships among and between teeth were summarised with correlation matrices, variable loadings plots, and the Frobenius norm. Univariate boxplots were used to contextualise and check the biological salience of the multivariate results.

**Results:** *H. sapiens* results corroborate previous research and show a degree of modularity that separates early forming and later-forming teeth. All four other species may show broad correlative patterns, but clear biological patterns are masked due to small sample sizes and/or sample composition.

**Conclusion:** Even with careful application of statistical procedures, ontogenetic inferences are only as good as the data are comprehensive.

## ARTICLE HISTORY

Received 17 July 2024

Revised 30 September 2024

Accepted 7 October 2024

## KEYWORDS

Tooth mineralisation; crypt formation; multivariate statistics; methodological issues; primate

## Introduction

To begin, we must make an admission. Within these pages, you will not find a neat and tidy explication of primate ontogeny. Instead, what follows is an honest account of the methodological challenges faced as we attempted to quantify multivariate dental development relationships, towards unravelling the complexities of hominin ontogeny. While our objective in this work may seem rather straightforward – to characterise variation in tooth mineralisation among five species of catarrhine primates – the results presented show the difficulty of going from data to inference in comparative studies of primate and hominin ontogeny.

The study of hominin ontogeny has a long if not also fraught history across the biological sciences (e.g. Schultz 1926; Smith BH and Tompkins 1995; Gould 1996; Zollikofer and de León 2010; Bolter and Zihlman 2022). The crux of this body of work is the use of ontogenetic variables to address questions about intra-species variability and inter-species differences (e.g. Alemseged et al. 2006; Cofran and Walker 2017; Rosas et al. 2017; Bolter and Cameron 2020). Recent inertia in this field brought together multidisciplinary stakeholders to

address the evolution of hominin ontogeny using novel methodological and analytical solutions (Bolter et al. 2023). A primary theme extending from Bolter and colleagues is the need to more comprehensively define and quantify ontogenetic variation in extant hominin and adjacent species. In this way, it may be possible to more robustly contextualise and model the ontogenies of fossil specimens and extinct species. Specifically, the publication (Bolter et al. 2023) and related Wenner-Gren workshop entitled, “New approaches to studying sub-adult hominins in the fossil record” (Gr. CONF-900), recommends that researchers combine data sources from living populations of human and non-human primates to experiment with and design new analytical techniques that may more accurately define the growth of fossil hominin subadults. From this recommendation stem the objectives of the present study. We combined data sources from living primate populations in a statistical approach designed specifically to weigh the (co)variation of dental development and capture the multivariate nature of varying growth processes. While broad, this objective leads to three research questions concerning the study of hominin growth and development. First, does examining the covariation

**CONTACT** Christopher A. Wolfe [wolfec23@ecu.edu](mailto:wolfec23@ecu.edu) Department of Anthropology, East Carolina University, Greenville, NC, USA

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

between developing teeth permit a better understanding between species-specific patterns of tooth mineralisation? Second, are there data quality or quantity concerns that may impact the outcome of the statistical approach? Lastly, what parameters are most important to consider during research and study design (i.e. assuring no missing data, gathering an ontogenetically diverse assemblage, sample size, etc.). To those ends, this study grapples with issues of statistical interpretation, notably the relationship of statistical modelling with biological reality, and how researchers may reasonably manage such issues in future studies.

We examine the dental development of five extant catarrhine species: *Homo sapiens*, *Pan paniscus*, *Pan troglodytes*, *Hylobates lar*, and *Papio anubis*. This primate sample of five species provides a taxonomically diverse assemblage to quantify primate ontogenetic variation (Schultz 1926; Smith BH and Tompkins 1995). A myriad of valuable studies examines dental development of one or a few of these groups (e.g. Moorrees et al. 1963a; Demirjian et al. 1973; Dirks 1998; Dean et al. 2001; Liversidge 2003; Dirks and Bowman 2007; Boughner et al. 2012; Boughner et al. 2015). As an extension of this body of work we believe that a more expansive and diverse primate sample representing multiple ages or age groups is foundational to fulfil the primary theme described above.

The development of the dentition provides a rather reliable biological process and subsequent dataset with which to infer ontogenetic patterns within and among species. This reliability is due to several factors, including preservation bias towards teeth as the most heavily mineralised tissue of the primate body (Hillson 1996), the connection of teeth to traditional life history markers across different species (Smith BH and Tompkins 1995; Dirks and Bowman 2007), and the overall stability of teeth as compared to skeletal tissues (Lewis and Garn 1960; Cardoso 2007; Šešelj 2013). Further, analyses of dental development are versatile and span years if not decades of an individual primate's life. Datasets can encompass various micro- and macroscopic aspects including early tooth morphogenesis and crypt formation, crown and root mineralisation, as well as tooth eruption and emergence into occlusion. As much as the timing and/or sequence of these aspects may vary, the methods needed to study said aspects may also vary. Much historical research tends to focus on the variation of alveolar eruption patterns among diverse species, most likely because this process was readily visible in existing skeletal collections and living animals alike (Clements and Zuckerman 1953; Holly Smith et al. 1994; Liversidge and Molleson 2004). In recent decades and particularly with the advent and routine use of 2D and 3D radiographic imaging, more studies have been able to analyse the developmentally earlier and biologically less variable mineralisation of the crown and root (Liversidge 2003; Liversidge and Molleson 2004). These earlier phases of tooth initiation and development are also the focus of our study. Our datasets are founded on conventional methods designed to quantify development where a researcher assigns each tooth one discrete, ordinal score that corresponds to a stage of development (Schour and Massler 1940; Moorrees et al. 1963a; Moorrees et al. 1963b; Demirjian et al. 1973; Dean and Wood

1981). Scores per tooth are then compared among groups or species using a conventional series of descriptive or bivariate statistics (Smith BH and Tompkins 1995; Dean et al. 2022). This valuable approach does present some limitations worth noting in the context of our study aims. First, there is no single standard tooth mineralisation scoring scheme used across all primate species. In *H. sapiens* alone there are a coterie of tooth staging systems that are used based on the research question, available teeth, and worker preference (Schour and Massler 1940; Nolla 1960; Moorrees et al. 1963a; Demirjian et al. 1973; Anderson et al. 1976). The lack of standardisation makes it difficult to complete inter-species comparisons. A second limitation for inter-species studies is the inherent difficulty that arises when making comparisons using ordinal data, such as tooth mineralisation scores (Hannigan and Lynch 2013; Konigsberg 2015; Corron et al. 2021). For example, earlier approaches to manage a complex dataset of ordinal dental scores collapsed or obscured relationships among developing teeth by summing all tooth development scores into a single univariate dental maturity score (DMS) for each individual (Demirjian and Goldstein 1976). The rationale is that DMS data can be compared between individuals (or averaged between species) to make a statement regarding the stage of maturation. Towards capturing and characterising complex ontogenetic patterns, this approach is statistically fraught (Liversidge 2010; Liversidge 2012; Boughner et al. 2015). The summing of discrete, ordinal values ignores the scaling of each dental score relative to another, and assumes that all traits (i.e. all 8 permanent teeth) contribute equally to the variation present. This assumption is problematic because it does not reflect the biological reality of a continuous growth process. We need to know specifically about individual tooth (univariate) variation and how that variation is related (or not) to other components or teeth in the dental arcade. Research suggests that conditional dependence is non-zero between dental traits in modern humans, as well as for correlation between developing teeth after accounting for the fact that development is a time-dependent process (Stull et al. 2021; Sgheiza 2022). In other words, relationships among developing teeth are typical and thus important to capture in ontogenetic studies across primate taxa. In fact, correlation among teeth differs according to variables including age category, tooth class, and biological sex (Stull et al. 2021; Sgheiza 2022). As a result, methods that do not at least consider these differing relationships among teeth across ontogeny also do not fully capture the variation inherent in dental development.

There is strong rationale to invoke more complex multivariate approaches to manage datasets that capture more complex variation in dental development within and across species. However, as noted earlier, comparing single teeth scored using different scoring systems presents a technical obstacle. Another obstacle, rooted in biology, is the integrated multivariate structure of dental development that leads to the presumption that such variation is most accurately examined as a whole dentition versus tooth-by-tooth. A further challenge is that the dentition is developing alongside other body systems within a greater organism (Lewis and Garn 1960; Demirjian et al. 1985; González-José et al.

2004; Cardoso 2007; Gómez-Robles and Polly 2012; Šešelj 2013; Armbruster et al. 2014). Ideally, analyses of primate ontogeny will encompass not only dental scores for the full dentition but also scores for skeletal development and growth, particularly as bones are the other major line of fossil evidence for ancient hominins. As such, looking to the present as well as the future of this field, it is not illogical to conclude that multivariate statistical methods are most appropriate to define and contrast patterns of tooth formation and dental development among species. Boughner et al. (2015) provide one approach to assess multivariate variation, using principal component analyses (PCA). While their study incorporates dependence information in the form of a covariance matrix, PCA is inherently an approach to summarise continuous-valued data. While we can numerically label dental scores as “1,” “2,” “3,” ..., the numbers only represent an order. There is no “magnitude” attached to these scores, as they do not contain real or specific biological information about the size of the interval between scores, such as months between scores that define crown  $\frac{1}{4}$  and crown  $\frac{1}{2}$ . Therefore, the ordinal nature of the tooth scoring system makes interpretation using a method such as PCA difficult due to the lack of measurement or magnitude attached to the scoring system. This difficulty leads to results that do not have a straightforward interpretation related to the variance or covariance among datapoints (Liddell and Kruschke 2018). In other words, are there more robust methods to better address the covariation and provide for straightforward biological and statistical interpretation? Therefore, we modify and test a statistical method that is better suited for ordinal data in the form of multivariate cumulative probit model.

There is a precedent for our approach: the cumulative probit has been used extensively across biological anthropology in applications of ordinal data analysis including dental development research (Liversidge and Molleson 2004; Konigsberg 2015; Stull and Corron 2022; Stull et al. 2023). A probit model is a form of a latent variable model whereby each ordinal score is linked to an underlying latent continuous variable distributed according to a standard normal distribution. As a result, the relationship of each score to another (e.g. the difference between a “1” and a “2”) is defined based on score position in the latent continuous space. In other words, if we assume dental development is a time-continuous process that has historically been measured using discrete staging systems, the probit allows us to connect the discrete stages to an underlying continuous process. Here we focus on the multivariate extension of the cumulative probit model, where we analyse correlated multivariate ordinal dental development data (Chib and Greenberg 1998; Gueorguieva and Agresti 2001; Konigsberg 2015; Stull and Corron 2022; Stull et al. 2023). While traditional probit analyses in biological anthropology tend to be concerned with post-hoc prediction based on parameters associated with age estimation, an often-overlooked component of the model fit is the correlation structure underlying the relationships between each multivariate series of traits (Sgheiza 2022). This correlation value represents the parameter of interest when comparing dental development among extant primate groups. Further, this model offers two guarantees. First, statistical validity, as

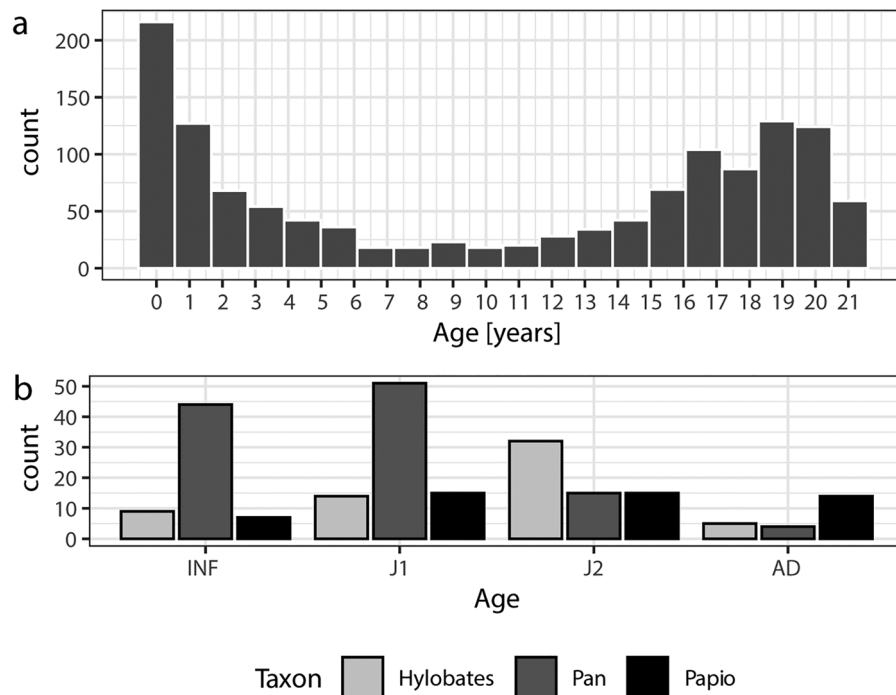
this approach is ubiquitous across the literature. Second, a means to make comparisons not at the score level where perhaps disparate scoring systems may be used, but instead with the correlation matrices learned from the models. This point is important, because the correlation matrices represent the presumed ontogenetic relationships between developing teeth much like the covariance matrix in the PCA from Boughner et al. (2015). While the use of correlations may be trivial to most who study phenotypic integration more broadly (e.g. Schlichting 1989), their use in delineating patterns underlying mineralisation of the dentition is less apparent but presents no less of an exciting opportunity to advance ontogenetic research. This correlation-focused approach is not unlike using the coefficient of variation (standard deviation/mean) to compare continuously measured variables of differing magnitudes. We put forth that the use of a more standardised metric such as a correlation allows for comparisons between disparate datasets that may have differing scoring schemes.

Considering the multivariate ordinal data structure and the desire to include and parse any correlations among developing teeth, we fit a multivariate cumulative probit model to dental mineralisation data from *Homo sapiens*, *Pan paniscus*, *Pan troglodytes*, *Hylobates lar*, and *Papio anubis*. We performed a series of visual comparisons between correlation matrices, completed eigen decomposition of the correlation matrices to summarise variable relationships in multivariate space, and quantified the difference between species-specific relationships using the Frobenius norm. In line with Bolter et al.'s (2023) stated need to better quantify ontogenetic variation, here we strived to provide a methodological tool that could be extended by us and others to future work comparing extinct and extant species. However, in the process of our study, we realised valuable yet frustrating insights about the rigour, complexity, and limits of analysing dental development score datasets. This work shares our insights and offers new ways forward. To help demonstrate such insights into the myriad data limitations, we also present univariate summaries to make sense of the multivariate output. We then discuss the challenges faced in order to refine the methodology needed to effectively parse biological and ontogenetic data from the expanding fossil subadult record.

## Materials and methods

### Extant samples

We evaluated tooth crown and root mineralisation of 1317 modern human subadults (*H. sapiens*), 34 free-lived bonobos (*P. paniscus*), 80 free-lived chimpanzees (*P. troglodytes*), 51 free-lived olive baboons (*P. anubis*), and 44 free-lived lar gibbons (*H. lar*). In modern humans, bonobos, chimpanzees, and baboons, the left-sided mandibular dentition was sampled including the central incisor (I1), lateral incisor (I2), canine (C), third [first] premolar (P3), fourth [second] premolar (P4), first molar (M1), second molar (M2), and third molar (M3). Due to sample size constraints, we sampled only the left-sided maxillary dentition for the gibbons. Previous research shows the relative similarity in sequence of tooth



**Figure 1.** Age profile for all extant specimens in current study. (a) *H. sapiens* age distribution with continuous, chronological age on x-axis and count on the y-axis. (b) Age distribution of non-human primate samples (*Pan*, *Papio*, and *Hylobates*) filled by each species. Ordinal age category is on the x-axis and count in each age category is on the y-axis.

formation between the upper and lower dental arcade (Moorrees et al. 1963a; Liversidge 2003). Further to scoring tooth crown and root mineralisation, the presence and/or absence of tooth crypts was noted in the scoring system for all non-human primates.

The modern human sample was collected from computed tomography (CT) images of subadults (birth to 21 years old) derived from the Subadult Virtual Anthropology Database (SVAD) (Figure 1) (Stull and Corron 2022). The images were generated at the University of New Mexico Health Sciences Centre, Office of the Medical Investigator (Berry and Edgar 2017; Edgar et al. 2020). Mineralisation of tooth crowns and roots were scored for all oral quadrants and all 32 permanent teeth following AlQahtani et al. (2010). The left sided mandibular dentition (eight teeth total) was used for the analyses presented here. Data were scored on an ordinal scale (1–13) based on a numeric adaptation of the original stage names (Table 1). To match the non-human primate scoring system defined below, we recode all human dental scores from AlQahtani et al.'s (2010) 13-stage system to that of Dean and Wood's (1981) 12-stage system (Table 1). While crypt stages (score 1 and 2) were not originally collected in the SVAD sample, the remainder of the scores (3–12) have analogues between systems and were recoded accordingly (Table 1). Original Scoring protocols for dental development can be found at Stull and Corron (2021a).

The *Pan* dental data were initially collected by Boughner et al. (2012). Chimpanzee dentition was scored based on mandibles housed at the Natural History Museum, London, UK and the Powell-Cotton Museum, Kent, UK. Bonobo dentition was scored based on mandibles housed at the Royal Museum of Central Africa, Tervuren, Belgium. All specimens

**Table 1.** Tooth mineralisation scoring protocols.

Scoring codes and definitions for all specimens <sup>a</sup>		Human recode	Original <i>H. sapiens</i> scores and definitions <sup>b</sup>	
1	Crypt initiation	3	1	Initial cusp formation
2	Crypt empty	3	2	Coalescence of cusps
3	Crown initiation	4	3	Cusp outline complete
4	Crown 1/4	5	4	Crown 1/2
5	Crown 1/2	6	5	Crown 3/4
6	Crown 3/4	7	6	Crown complete
7	Enamel buccal/cervical cleft	7	7	Initial root formation
8	Root 1/4	8	8	Root length less than 1/2
9	Root 1/2	9	9	Root length = crown length
10	Root 3/4	10	10	Root length 3/4
11	Root complete, apex open	11	11	Root complete with diverge ends
12	Root complete, apex closed	12	12	Apex closed, wide PDL
			13	Apex closed, normal PDL

<sup>a</sup>Adapted from Boughner et al. (2012) and Dean and Wood (1981). <sup>b</sup>Adapted from Stull and Corron (2022b) and AlQahtani et al. (2010). All analyses were completed based on scores and definitions adapted from Boughner et al. (2012) and Dean and Wood (1981). *H. sapiens* was initially adapted from Stull and Corron (2022b) and AlQahtani et al. (2010). Here we demonstrate the recode from one system to the next.

were wild collected. Mandibles were radiographed during study at each institution resulting in lateral, occlusal, and lingual intraoral views of the whole developing dentition. Tooth mineralisation was scored on the left side of the mandible using a 12-stage ordinal scale described in Boughner et al. (2012) and modified from Dean and Wood (1981) (Table 1). The mineralisation data collected in Boughner et al. (2012) were used in the present study with author permission. Individuals were assigned an estimated age group based on previously published radiographic and histological studies (Holly Smith et al. 1994; Reid et al. 1998). These age groups are infant (INF), younger juveniles (J1), older juveniles (J2), and adults (AD) (Figure 1). To seriate specimens into individual age groups, Boughner et al. (2012) examined dental

eruption patterns (Holly Smith et al. 1994) and crown and root morphology vis-à-vis incremental analysis of dentine and enamel formation (Reid et al. 1998). Note, while it may be more appropriate to delineate age categories independent of the tissue under study (i.e. skeletal growth instead of dental parameters), previous research across primate samples does show the close relationship between dental eruption, mineralisation, skeletal maturity, and sexual maturity (Holly Smith et al. 1994; Reid et al. 1998). Previous research highlights the similarity in development between *P. paniscus* and *P. troglodytes* (Boughner et al. 2012), leading us to pool the *Pan* data to increase overall sample size and coverage across *Pan* ontogeny.

The *Papio* sample was initially collected by Boughner et al. (2015) and used with permission in the current study. Baboon dentition was scored based on mandibles housed at the Natural History Museum and Royal College of Surgeons, London, UK. Similar to the *Pan* sample, all specimens were wild collected, mandibles were radiographed at each institution resulting in lateral, occlusal, and lingual intraoral views, and tooth mineralisation was scored on the left side of the mandible on a 12-stage ordinal scale described in Boughner et al. (2012) and modified from Dean and Wood (1981) (Table 1). Similar to the *Pan* data, chronological age was unknown for this *Papio* sample, and individuals were assigned an estimated age group based on examination of macroscopic dental eruption, microscopic analysis of dentine and enamel formation, and their relationship to patterns of somatic and sexual maturity in *Papio* (Holly Smith et al. 1994; Reid et al. 1998). These age groups are infant (INF), younger juveniles (J1), older juveniles (J2), and adults (AD) (Figure 1).

The *Hylobates* wild collected sample was initially examined in Cofran and Boughner (2025). Specifically, we incorporate *Hylobates lar carpenteri* from the Asiatic Primate Expedition (APE) from 1937 to 1938 (Carpenter 1941; Coolidge 1937; Coolidge et al. 1940; Schultz 1944). This collection is housed at the Museum of Comparative Zoology, Harvard University. Craniodental material was subject to x-ray microtomography (micro-CT or  $\mu$ CT) resulting in a virtual image capturing crypt, crown and root development. Cofran and Boughner (2025) each scored the dentition independently and then compared sets of scores until consensus was met. The scores were used in this study with author permission. Consistent with the approach for staging *Pan* and *Papio* dental development, tooth mineralisation was scored on the left side of the jaw with a 12-stage ordinal scale described in Boughner et al. (2012) and modified from Dean and Wood (1981) (Table 1). As stated above, only the left maxillary dentition is included because of sample size constraints associated with the mandibular dentition. Cofran and Boughner assigned five (versus four) age groups according to an individual's stage of dental eruption as well as status of basicranial suture closure. These groups are infants (INF), younger juveniles (J1), older juveniles (J2), oldest juveniles/subadults (SA), and adults (AD) in whom at least one tooth had not completed development. To match the age categorisation of the non-human primate samples described above, we combined J2 and SA into a single J2 age category (Figure 1).

### Multivariate model specification and model selection

We specified a multivariate probability distribution over dental development variables for each specimen. In total, we fit four separate multivariate cumulative probit models – one fit per taxon. Specifically, we fit the response vector of dental scores to a multivariate cumulative probit model. This model assumed that each discrete ordinal value (e.g. a dental score of “4”) can be represented as a latent continuous value (i.e. because dental development is a continuous biological process), augmented *via* the cumulative distribution function of the standard Gaussian distribution. In other words, we assume the continuous biological process related to dental formation is best modelled as a normal distribution with some average value and standard deviation. While the fit of this model yields various parameter values, of keenest interest to us is the correlation matrix describing the dependence structure between dental scores. Values closer to 1 or  $-1$  represent stronger correlations between developing teeth; values closer to 0 represent weaker correlations. Stronger correlations imply some level of integration during development, where certain teeth develop along a similar pace, magnitude, and direction. Here, we take integration to mean the strength of correlations between pairs of teeth, which may evince more (or less) concurrent development of both teeth. Of note, traditional correlation parameters in a multivariate cumulative probit can be interpreted as either the Pearson correlation coefficient or as a linear relationship between the data. However, the mathematical similarity between a multivariate probit model and other multivariate dispersion models that examine variability among multiple traits or variables suggests that Pearson's  $r$  and Spearman's  $\rho$  are nearly identical and Kendall's  $\tau$  is positively related to both terms (Song 2000). Therefore, the correlation term presented here provides a robust measure of linear and/or rank association among traits.

Parameter estimation was completed in Stan, a probabilistic programming language written in C++ that computes the joint log probability density of a set of continuous parameters up to a proportional constant (Stan Development Team 2024). Bayesian inference was carried out using Stan's dynamic Hamiltonian Monte Carlo algorithm (HMC) to draw from the posterior distribution (Betancourt 2017). The model was fit in the R programming environment (R Core Team 2023) using Stan's lightweight interface to the command line cmdstanr (Gabry and Cesnovar 2022). All models were run with four chains with 1000 warmup iterations and 1000 sampling iterations resulting in 4000 posterior samples per parameter. The tuning parameter associated with the total number of simulation steps (max\_treedpeth) was adjusted to a value of 12 and the target average acceptance probability (adapt\_delta) was adjusted to 0.99 to capture the complexity of the posterior distribution. Parameters across all models showed a split R-hat statistic less than 1.01, bulk and tail effective samples sizes greater than 400, E-FMI statistics greater than 0.2, and no divergent transitions (Gelman et al. 2014; Betancourt 2016a, 2016b; Betancourt 2017; Vehtari et al. 2021). As such, all diagnostic features indicated a statistically valid and reliable series of models. For a complete

mathematical specification of the model, the parametric forms associated with the mean function, and prior choices, see [Appendix A](#).

In the present context, an important feature of ontogenetic data is the extent to which chronological age or age group influences the variation present in the sample population. We ask: should age data be included as a predictor or as an independent variable related to dental score? Or should age not be included at all because age information is intrinsic to all ontogenetic data? This interesting quandary is foundational to consider because chronological age is unknown in many extant specimens and certainly in most extinct individuals. Biological age can be deduced in many specimens but carries an inescapable margin of error that varies according to the data used to calculate age or assign individuals among age groups. To compare the predictive performance of individual models' fit with and without including age as an independent variable, we completed k-fold cross-validation where  $k=5$  folds. We use k-fold cross validation because it allows us to mitigate the impact of data quality issues (e.g. imbalanced data among the five species) by incorporating larger training datasets through the ability to fit the model  $k$  times (Yadav and Shukla 2016; Vehtari et al. 2021). Further, in Bayesian statistics, multivariate probit models can often prove difficult in likelihood-based model selection approaches and therefore we use the advice put forth by Stan developers to compare models *via* cross-validation. To measure the predictive fit, we used the expected log pointwise predictive density (ELPD) (Gelman et al. 2014; Vehtari et al. 2017). Generally, the larger the ELPD, the better performing the model. Ideally, the ELPD difference is 3 to 4 times greater than the standard error of the difference to indicate a quantifiable difference in predictive performance. The general four-step procedure to compare models is as follows: (1) Fit the model to  $k-1$  training data, (2) Estimate the likelihood of the held-out test data based on the previous fit, (3) Repeat for each held out fold ( $k=5$ ) while storing the likelihood each time, and; (4) Using the stored likelihood values, quantify the ELPD and compare the fit of the age versus that of the no age model. This procedure is repeated for each species to determine the best-fitting model per species. Model selection was performed using the `loo` package (Vehtari et al. 2024).

### **Statistical inference: how we examined the models**

All analyses described in this section were performed on the best fitting multivariate probit model. To compare the similarity in patterns of dental scores among all eight developing permanent teeth of an oral quadrant, between extant specimens, we completed posterior inference on the samples of each correlation matrix. Specifically, we used the Frobenius norm to calculate the similarity between posterior correlation matrices (Balgopal 1996; Ramirez et al. 2013). The Frobenius norm is the square root of the sum of the squared difference between two matrices and is a computationally efficient means to compare matrices. The smaller the norm or closer the norm is to zero, the more similar the matrices; the further

the norm is from zero, the more dissimilar the matrices. We compute the norm across all posterior samples among all species resulting in a distribution of values per pairwise comparison where the user can perform any number of summary statistics (*i.e.* mean, mean, mode, range, etc.).

To visualise within-species patterns we present correlation matrices and a loadings plot resulting from eigen decomposition of the posterior mean correlations. For brevity, in eigen decomposition, only the first two dimensions are presented as these capture the greatest amounts of variation. In the loadings plot, the longer the arrow, the more influential that variable to said dimension and the more acute the angle between variables, the stronger the correlation. A point of clarification: in the present analysis, "influential" relates to how strongly correlated variables may or may not be. A lack of correlation implies that the data-generating process cannot be subsumed into a simple mapping from one variable to the next, but instead other information that is external to this study may have to be accounted for. In other words, if the correlations are weak, then perhaps more data are needed to parse any connection between pairs of individual developing teeth. A strong correlative relationship suggests that knowing about one variable tells you a lot of information about the other. A caveat is that our study here assumes a linear (or at minimum, rank) dependence structure, where smaller values equate to smaller values and larger values equate to larger values. It is important to note that other statistical models and copula-based approaches (Song 2000; Danaher and Smith 2011; Joe 2014) can model dependence in other ways and would be an additional avenue for future research. As a last but vital step, to most accurately interpret and contextualise the multivariate analyses, we provide boxplots for each tooth score per species. Using the information stemming from the initial Bayesian model fitting, model selection, and downstream multivariate and univariate interpretations, we aim to describe as comprehensively and as accurately as possible the dental development variation present in a diverse sample of catarrhine primates.

### **Reproducibility**

All code is published at [https://github.com/ChristopherAWolfe/WolfeBoughnerStull\\_AHB](https://github.com/ChristopherAWolfe/WolfeBoughnerStull_AHB) and is available in the open access repository on Zenodo (Wolfe et al. 2024). The human dataset is available at Stull and Corron (2021b). The non-human primate data are available in the original publications reporting these studies (Boughner et al. 2012; Boughner et al. 2015; Cofran and Boughner 2025). All analyses were completed in the R programming environment (v. 4.3.1) (R Core Team 2023). The following packages are necessary to complete the analyses exactly as described in the paper: `tidyverse` (Wickham et al. 2019), `magrittr` (Bache and Wickham 2022), `ggpubr` (Kassambara 2023), `reshape2` (Wickham 2007), `posterior` (Bürkner 2022), `loo` (Vehtari et al. 2024), and `cmdstanr` (Gabry and Cesnovar 2022). Users will also need a working version of `CmdStan` – the command line interface to Stan that can be installed through `cmdstanr`.

## Results

In all five species, the best fitting Bayesian multivariate cumulative probit model incorporated age as a predictor (Table 2). All results stem from this series of best performing models. Pairwise comparisons among species show a nonzero difference between posterior correlation matrices, suggesting species-specific multivariate relationships among the dentition (Figure 2). In the remainder of this results section, we analyse the dependence structure separately by species and provide boxplots to contextualise the output in the multivariate results.

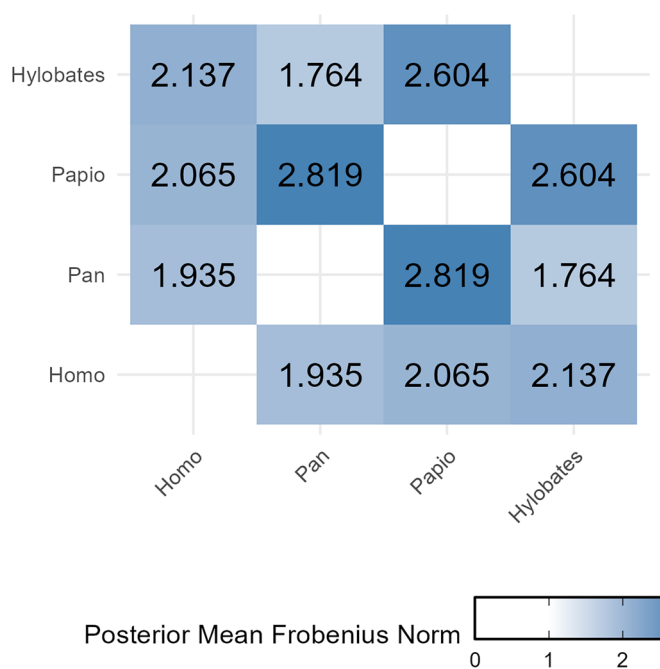
### *Homo sapiens*

Figure 3 visualises the posterior mean correlations among *H. sapiens* mandibular tooth mineralisation and presents the loadings plot to further illustrate relationships among and between teeth. All correlations are positive. The strongest

**Table 2.** Model comparison between multivariate probit models run with and without including an “age” variable as a predictor in the model.

Sample	Age model ELPD	Intercept-only ELPD	ELPD difference	SE of difference
Homo	<b>-1290.468</b>	-3924.612	-2634.1	142.5
Pan	<b>-912.2736</b>	-1753.946	-841.7	93.9
Papio	<b>-429.2444</b>	-511.7192	-82.5	61.8
Hylobates	<b>-473.4557</b>	-957.4855	-484	43.2

The larger the expected log predictive density (ELPD), the better performing the model. The model(s) incorporating an estimate of age performed best across all taxa. The ELPD Difference is the difference between the best performing model and the worst performing model (comparing 2 models), and the SE difference is the standard error of this difference. Ideally, the ELPD difference is 3–4 times greater than the SE difference to indicate a quantifiable difference in predictive performance.



**Figure 2.** Pairwise comparisons among posterior mean correlation matrices between species. The values are representative of the Frobenius norm or the square root of the sum of the squared differences between the matrices. The smaller the norm or closer to zero, the more similar the matrices and the further from zero, the more dissimilar the matrices. The diagonal is left blank because it represents the comparison of each matrix with itself.

correlations occur between teeth that generally mineralise along similar trajectories. In other words, the dental scores of these teeth are more similar when compared to the scores of other teeth. Pairwise comparisons between I1, I2, C, and M1 return the strongest positive correlations. While reduced, a moderately strong positive relationship was found between P3 and P4. The loadings plot corroborates the patterns. The first two dimensions explain approximately 74% of the variation in the matrix. Specifically, the second dimension represents a probable ontogenetic trajectory where the earliest forming dentition has more positive values and the later forming dentition has more negative values.

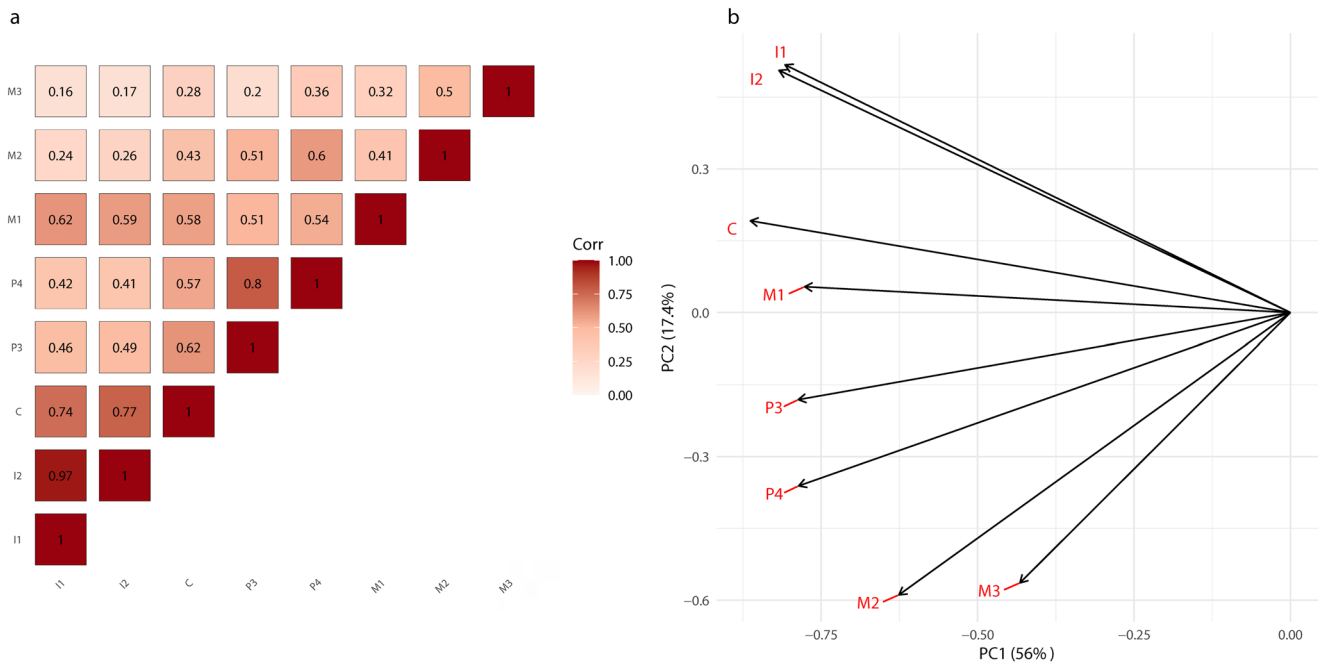
Boxplots of mineralisation scores for each tooth by an individual’s age further clarify the patterns described above (Figure 4). The distribution of scores shows that I1, I2, C, and M1 all begin mineralisation (stage 3, initial cusp formation) early in infancy and finish development (stage 12) at the outset of the adolescence period. In contrast, P3 and P4, which develop about concurrently, do not begin mineralisation until later in infancy to early childhood (3–4 years old). Both M2 and M3 begin development late(r) in childhood and complete development long into adolescence. This later development is particularly acute for M3, which may not complete formation until early adulthood.

### *Pan*

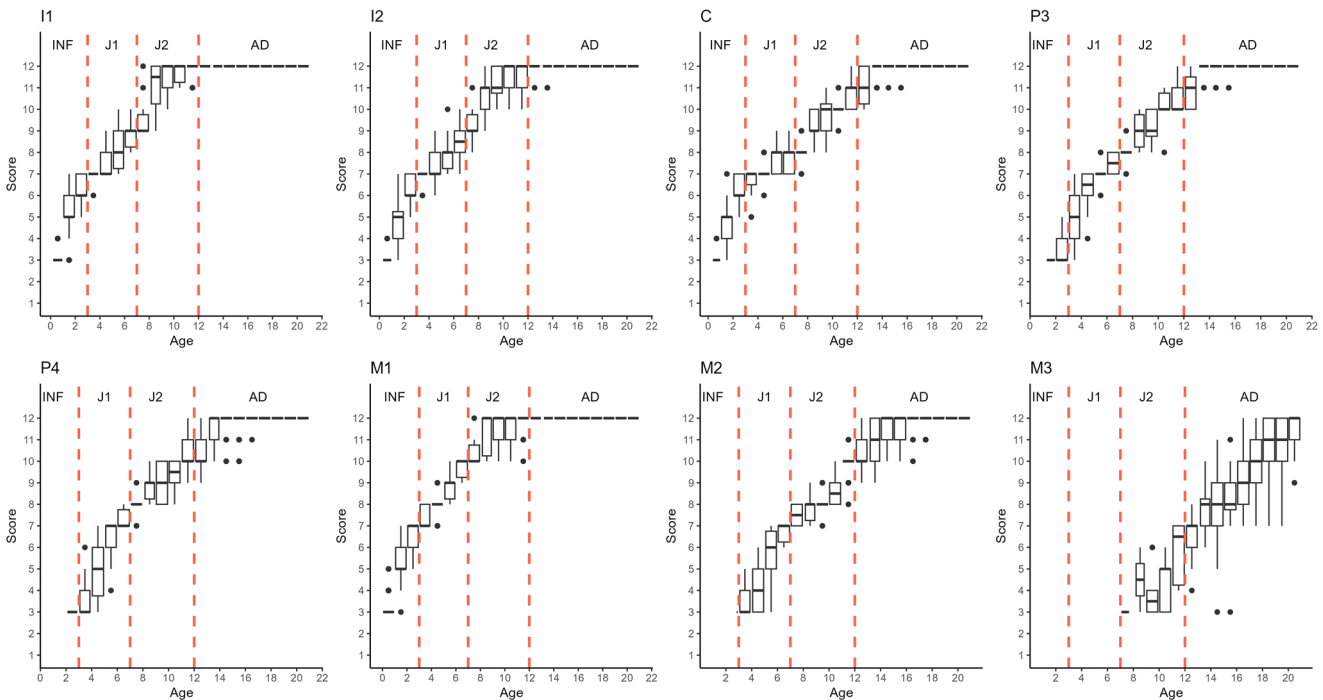
Correlations among all variables (i.e. dental scores for all eight teeth per specimen) are stronger in *Pan* as compared to *Homo* (Figure 5). That said, no clear pattern arises to indicate any module(s) related to ontogeny, function, or tooth class. For instance, M1 has weaker correlations among and between all other teeth. This outcome is interesting because in primates M1 typically begins development first, and often spans pre- and postnatal parts of ontogeny, including dental ontogeny when other teeth such as incisors are forming. Similar to M1 as the first-forming tooth, M3 as the last-forming tooth shows weak correlations with most other teeth. An examination of each tooth by itself further illuminates possible explanations for any perceived discrepancies suggested in Figure 6. Specifically, in the M1 boxplot, given that M1 development starts around birth, we would expect the bulk of scores in the infant age category to be lower. Yet, on average most scores are around a stage 6 (crown  $\frac{3}{4}$ ), with a large degree of variation around the median. Further, M1 scores in the J1 age category are rather high as compared to scores of other developing teeth – even the later forming M3. However, these outcomes reflect the composition of our *Pan* sample, with fewer infants overall. Altogether, the distribution of dental scores for all teeth show the expected increase in scores as an individual ages, but the bias in the sample caused by a lack of very young infants is reflected in relatively few teeth demonstrating the earliest (stages 1–3) developmental stage.

### *Papio anubis*

Compared to *Pan* and *H. sapiens*, the *P. anubis* dentition is overall less strongly correlated (Figure 7). The strongest



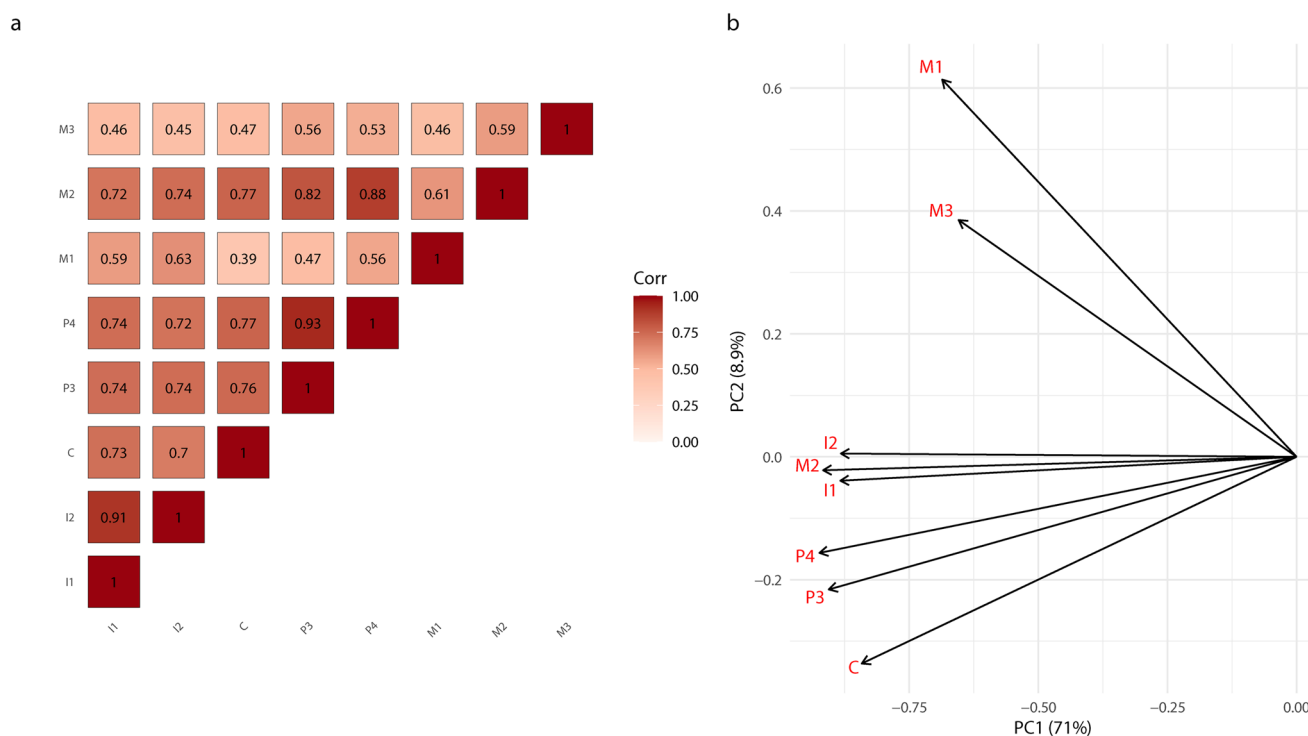
**Figure 3.** Multivariate comparisons of *H. sapiens* tooth formation. (a) Correlation plot of the posterior mean correlation values between the mandibular dentition. Values closer to 1 are strongly positive and darker red in colour. (b) Loadings plot of the eigendecomposition of the posterior mean correlation matrix. Dimension 1 and Dimension 2 explain approximately 73% of the variation in the data. The more acute the angle between variables, the more strongly correlated. The longer the arrows, the more influential in dimensions 1 and 2.



**Figure 4.** Boxplots visualising tooth mineralisation in *H. sapiens*. Scores adapted from protocols in Stull and Corron (2022b) and modified from AlQahtani et al. (2010). Black circles are outliers, the horizontal line is the median, and the whiskers encompass the minimum and maximum values excluding any outliers. For comparison to other species, I have included the age categories from other groups. INF = 0–3 years old, J1 = 3–7 years old, J2 = 7–12 years old, and AD = >12 years old. These stages roughly follow similar stages defined in Bogin (2020).

positive correlations are between I1 and I2, and P3 and P4, suggesting ontogenetic modules related to tooth class. However, the positive correlation strength is weaker among the rest of the pairwise tooth comparisons. Only 61% of the total variation is captured in the first and second dimensions, corroborating reduced correlations between most *Papio* teeth

compared to the other four species. The reduced arrow size in the loadings plots suggests that the C, M1, M2, and M3 all play a smaller role in describing the variation present as compared to the higher strength of positive correlation between pairs of incisors and of premolars. An examination of the boxplots shows that C, M1, M2, and M3 are more



**Figure 5.** Multivariate comparisons of *Pan* tooth formation. (a) Correlation plot of the posterior mean correlation values between the mandibular dentition. Values closer to 1 are strongly positive and darker red in colour. (b) Loadings plot of the eigendecomposition of the posterior mean correlation matrix. Dimension 1 and Dimension 2 explain approximately 80% of the variation in the data. The more acute the angle between variables, the more strongly correlated. The longer the arrows, the more influential in dimensions 1 and 2.

disjointed in their data distributions, with a gap between scores indicating the youngest age group versus the older age groups (Figure 8). In comparison, there appears to be slightly more variability in the I1, I2, P3, and P4 scores, with more coverage across all possible tooth scores from 1 (crypt initiation) to 12 (apical root closure) at a given age. Similar to *Pan*, the *Papio* age distribution (Figure 1) shows far more older individuals, which may help explain the dependency structure, and imbalanced nature, of the data distributions across teeth.

### *Hylobates lar*

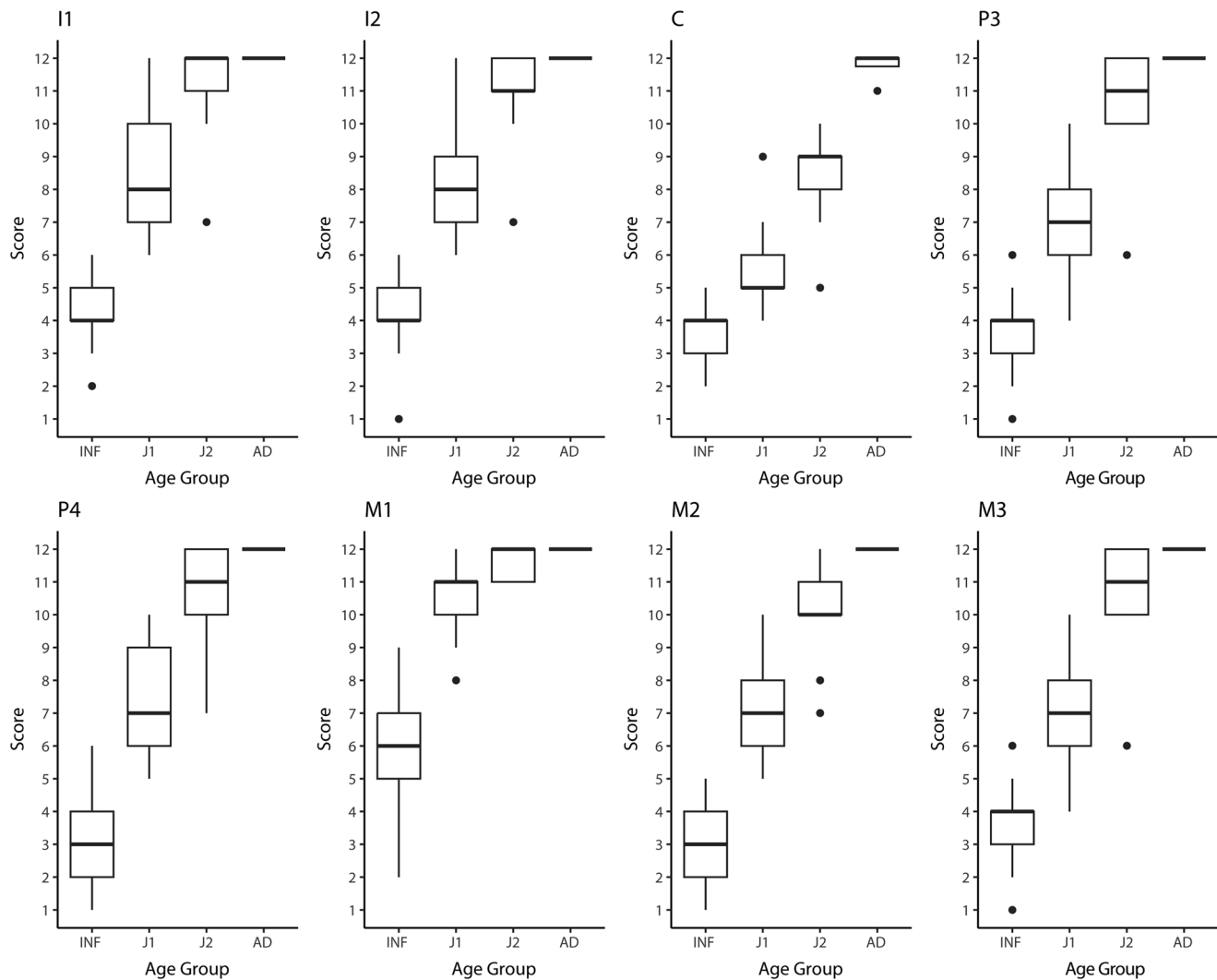
The overall strength of correlations among variables in *H. lar* is intermediate to *Pan* and *H. sapiens* (Figure 9). That is, development among all 8 permanent teeth is on average less strongly correlated as compared to *Pan*, but more strongly correlated as compared to *H. sapiens*. The strongest developmental relationship appears to be between P3 and P4. Their similar ontogenetic trajectory (i.e. pattern of dental scores) indicates a module within this tooth class. With relative homogeneity across mean correlation values, no additional evidence of patterning according to tooth class ontogeny is present. Relatedly, while the loadings plot does suggest distinct groupings according to [I1, I2, M1], [P3, P4, M3], and [C, M2], this result may reflect imbalanced ages that bias the resulting distribution of dental development scores (Figure 10). Specifically, the gibbon sample has very few individuals in the youngest age category and even these infants are towards the older end of the age group. This sample composition issue is

reflected in M1 and I1 data, where we would expect the youngest individuals to have the smallest scores, but instead these older infants have M1 and I1 scores that are more on par with later forming teeth such as P3, P4, and M2. Therefore, the sample is missing infants of younger/est ages. The absence of the earliest scores for I1 and M1 and the concomitant absent ontogenetic variance introduces sampling bias that makes it difficult to parse fully the strengths and directions of relationships among all teeth as they develop.

In summary, the four datasets (*Homo*, *Pan*, *Papio*, *Hylobates*) present differences in the underlying dependence structure among the five species. In *H. sapiens*, the largest sample, a clear ontogenetic trajectory is shown across the univariate and the multivariate analyses. This clear trajectory is also evidenced by stronger correlations among groups of early forming versus later forming teeth. However, the remainder of the specimens/species in this study do not share a similar “*Homo*” pattern.

### Discussion

The results present a broad view of multivariate tooth mineralisation in five catarrhine species. In alignment with themes in Bolter et al. (2023), this study put forth a multivariate statistical method that expands upon previous multivariate work (e.g. Boughner et al. 2015). The intention of this work was to test the applicability of a multivariate method more appropriate for the ordinal data presented in dental development and to address the ability of this method to quantify covariation between the dentition in each taxon. However,



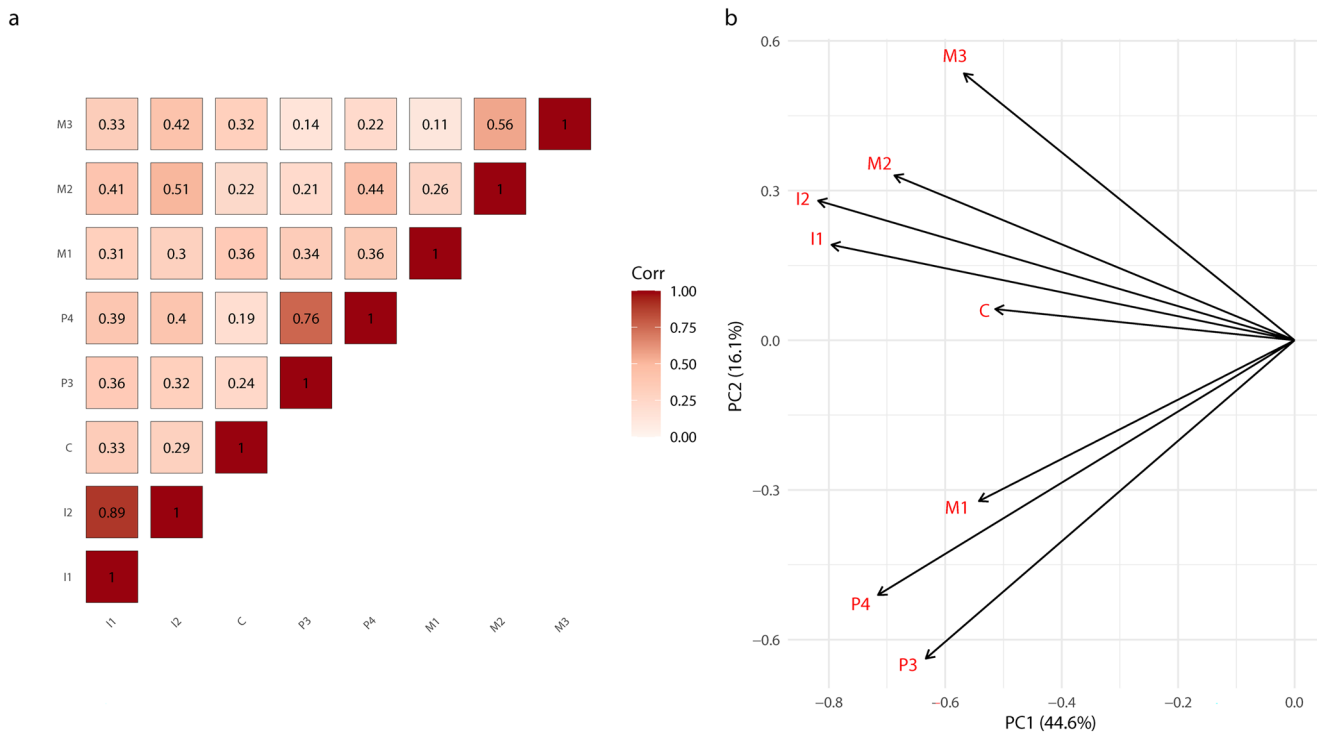
**Figure 6.** Boxplots visualising tooth mineralisation in *Pan*. Scores adapted from protocols in Boughner et al. (2012) and modified from Dean and Wood (1981). Black circles are outliers, the horizontal line is the median, and the whiskers encompass the minimum and maximum values excluding any outliers.

the approach explored here could provide only broad correlative patterns that may or may not relate clearly to published information about tooth development pattern of each species. We explore this discrepancy in this section, including sample composition artefacts and variation in developmental processes.

Regardless of the primate taxon, all permanent teeth run in the same sequence along the dental arcade (i.e. I1, I2, C, P3, P4, M1, M2, M3). Yet primate permanent dentition typically mineralises in the following sequence: M1, I1, I2, C, P3, P4, M2, M3 (Swindler and Meekins 1991; Smith BH and Tompkins 1995; Boughner and Dean 2004). In general, this sequence is demonstrative of the relative timing of tooth formation with M1 tending to initiate earliest and M3 latest. As such, and interestingly, the order of tooth initiation is decoupled from the order of tooth position, because the permanent dentition tends to initiate in the “middle” of the presumptive permanent dental row, starting with M1. Importantly, there is species-specific variability to this pattern, such as early I1 initiation in *H. lar* at about the same time as M1 initiation (Cofran and Boughner 2025). Yet, the general expectation is that the multivariate patterns described

here correspond to an ontogenetic pattern where teeth that develop at the same time (e.g. early forming or late forming) will correlate most strongly. In this study, only *H. sapiens* shows clear evidence of this pattern that is expected based on published literature. Specifically, the results presented above (Figure 3) corroborate work by Liversidge (2003) and Moorrees et al. (1963a) where the pattern of multivariate development in human teeth reflects two modules associated with early-forming teeth and later-forming teeth: [M1, I1, I2, C] and [P3, P4, M2, M3], respectively.

The correlative patterns described in *Pan* (Figure 5) do not fit the expectation described in previously published research. Boughner et al. (2012) show that between-tooth relationships are related to both tooth class and when during ontogeny individual teeth may develop. Further, previous multivariate analyses suggest that variation reflects ontogenetic sequence, supported by strong correlations with known ages of individual specimens (Boughner et al. 2015). However, in our study, the *Pan* results visualised in Figure 5 do not corroborate this expectation. At best, there is a small degree of similarity in the multivariate analyses between Boughner et al.’s (2015) and those presented here. Specifically, both studies highlight



**Figure 7.** Multivariate comparisons of *P. anubis* tooth formation. (a) Correlation plot of the posterior mean correlation values between the mandibular dentition. Values closer to 1 are strongly positive and darker red in colour. (b) Loadings plot of the eigendecomposition of the posterior mean correlation matrix. Dimension 1 and Dimension 2 explain approximately 60% of the variation in the data. The more acute the angle between variables, the more strongly correlated. The longer the arrows, the more influential in dimensions 1 and 2.

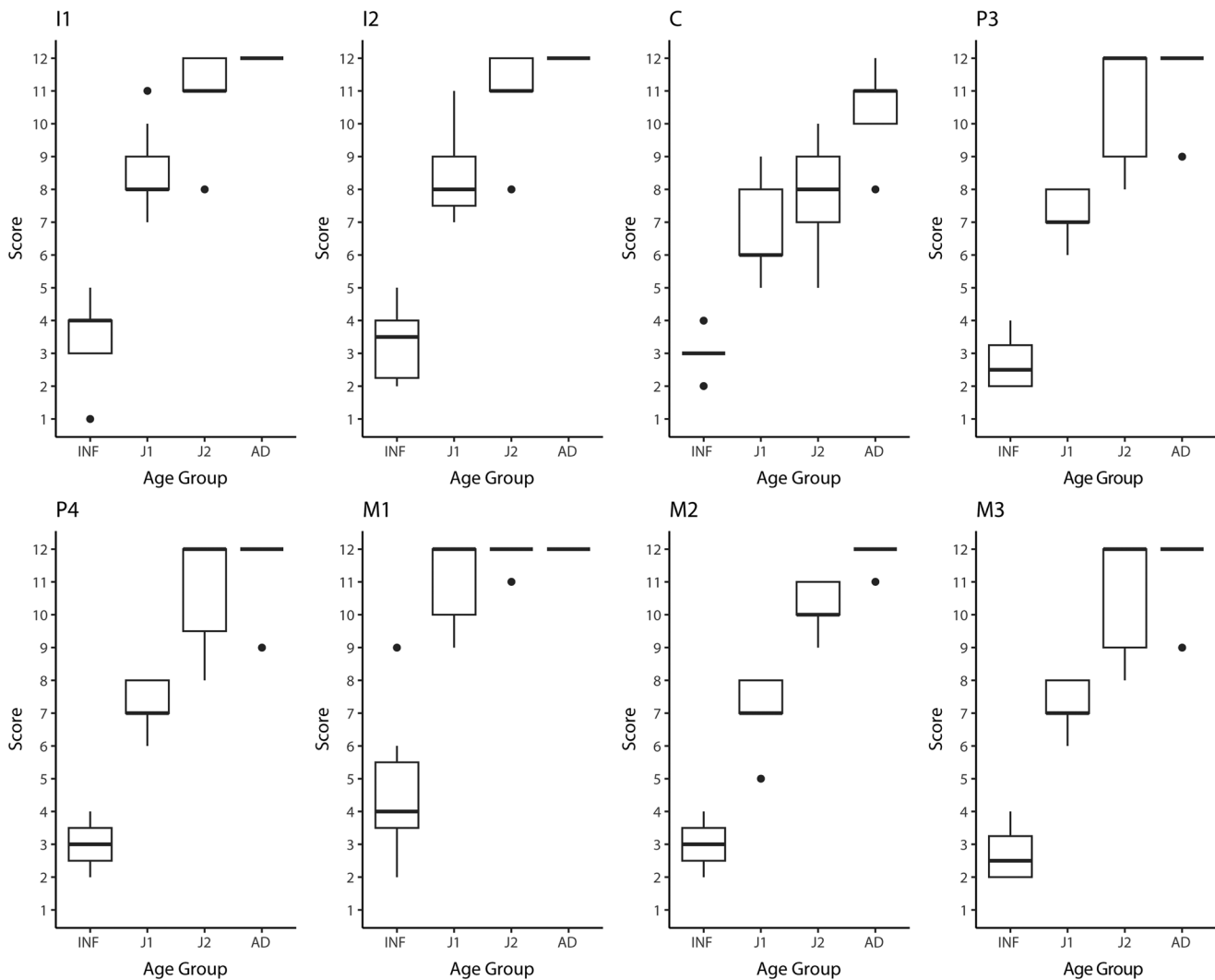
the outsized influence of C, M1, and M3 development in the second dimension from the principal component analyses/eigen decomposition. Furthermore, in general, *Pan* results in this study do not fit the same pattern seen in *H. sapiens* where earlier-forming dentition and later-forming dentition are most strongly correlated. While there is some agreement between our results and the literature, the match is not exact. Our results here indicate that testing the validity of a multivariate approach against published studies as well as against knowledge of the dental score data/biology of the system is a vital part of this (and any) exploratory multivariate statistical process.

The picture is even more muddled in both *P. anubis* and *H. lar*. Previous research shows a relative similarity between the sequences of *Papio* and *Pan* tooth mineralisation, with most differences tied to variation in later-forming teeth (Boughner et al. 2015). Further, work by Cofran and Boughner (2025) suggests a similar mineralisation pattern that matches the expectation above with the strongest relationships among developing teeth patterned according to an ontogenetic sequence defined as [M1, I1], [I2, P3, P4, M2], [C, M3]. The results in our study corroborate neither information in Boughner et al. (2015) related to *Papio* nor information in Cofran and Boughner (2025) related to *Hylobates*. Instead, *Papio* dentition in the current study shows overall weaker correlations among developing teeth with no notable patterns apparent in early forming or later forming teeth (Figure 7). Further, *Hylobates* tooth formation as described in this study corroborates in a limited capacity the results in Cofran and Boughner (2025). The main difference appears to be one of kind if not also one of degree,

where patterns arising from our analysis are less well defined or simply undetected compared to patterns reported in published work. This outcome suggests that either past studies were less sensitive to variation in sample composition and size, or that the current multivariate approach is more sensitive to such sample biases.

Importantly, while there does appear to be a difference between previously published works and the results presented in this study, there are consistent correlative relationships among the dentitions that emerge in each taxon. Perhaps most consistent is the moderate to strong positive correlations between incisors (I1 and I2) and between premolars (P3 and P4) among all species studied here. Thus, maybe the most interesting question is not “why do certain teeth develop at different times?,” but “why do certain teeth develop concurrently?” This topic is particularly interesting from a systemic perspective of ontogeny, such as hormones and concomitant growth spurts, and what variable(s) catalyse the initiation of a given tooth (Boughner and Hallgrímsson 2008; Gómez-Robles and Polly 2012). Not all permanent teeth initiate at once and when each tooth initiates can differ among species. For instance, in *H. lar*, I1 development is concurrent with M1 and thus much earlier than other primate taxa including *Papio* and *Pan* (Cofran and Boughner 2025). Also, in *H. sapiens*, M3 tends to initiate much later as compared to other taxa (Dean and Lucas 2009). Meanwhile, across all species, development of both C and M2 can be highly variable (Schwartz and Dean 2001).

A possible scenario to explain why we would expect a particular developmental timing of anterior versus posterior teeth is the distinction between the successional dental



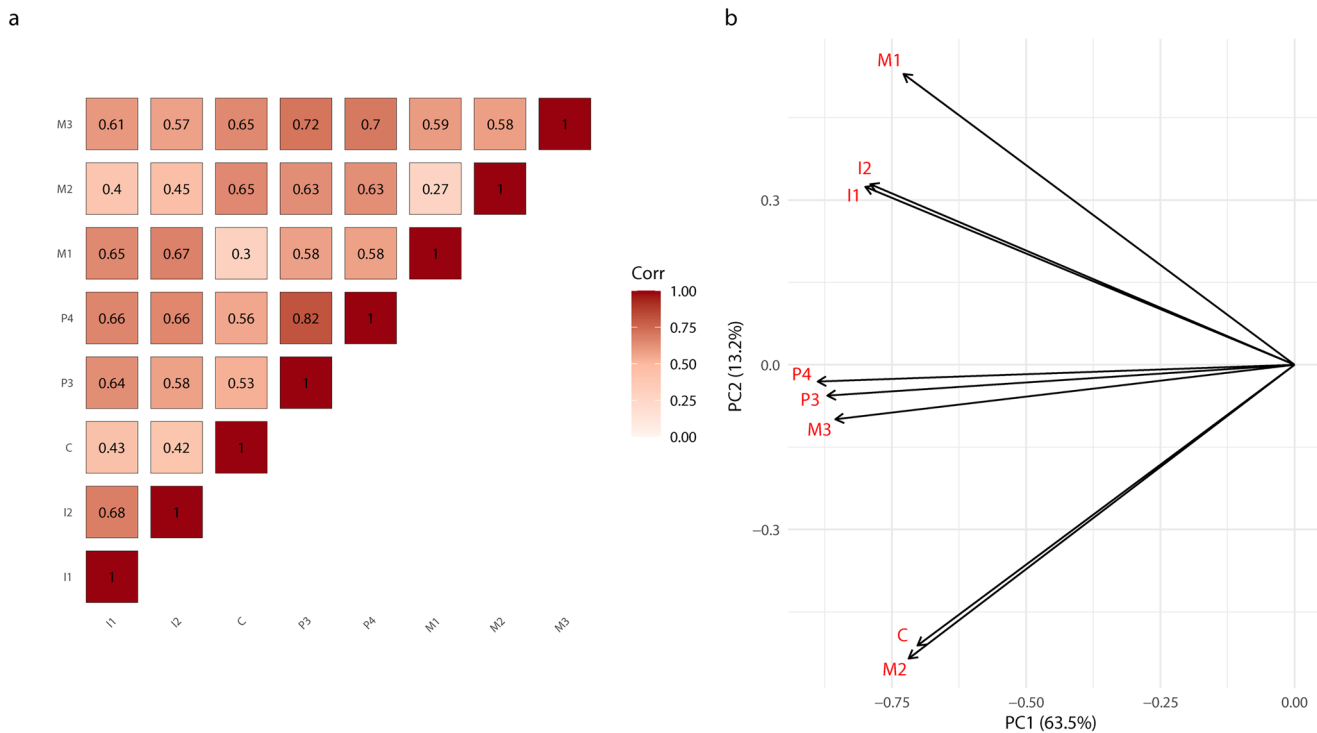
**Figure 8.** Boxplots visualising tooth mineralisation in *P. anubis*. Scores adapted from protocols in Boughner et al. (2012) and modified from Dean and Wood (1981). Black circles are outliers, the horizontal line is the median, and the whiskers encompass the minimum and maximum values excluding any outliers.

lamina (DL), which gives rise to the permanent teeth that “succeed” their deciduous precursors, and the additional DL, which gives rise to all three molars (Berkovitz et al. 2009; Jernvall and Thesleff 2012; Square et al. 2021; and reviewed in Appelt et al. 2021). Posteriorly, M1 is the first tooth to initiate within the additional DL, where there are no deciduous teeth to replace. Anteriorly, I1 is the first replacement tooth to initiate within the successional DL. By the time that both permanent incisors and the permanent canine initiate, the successional DL has extended sufficiently for both premolars to start forming – which they tend to do concurrently in all five species. Also, around this time, the additional DL has extended posteriorly and catalysed somehow, the M2 starts forming. The last stage is the continued backward extension of the additional DL and, catalysed by something at some point, the initiation of M3. Thus, while the beginning of this discussion highlights the intrinsic differences between previously published work and that presented here, perhaps the presence of even the smallest correlative statistical relationship (e.g. I1/I2 or P3/P4) foreshadows the underlying developmental mechanisms at play and their relation to the sequence and timing of tooth initiation, development, and completion.

Even with the optimism brought forth by similar correlations between the incisors or the premolars, and the possible scenario inspired by growth of the dental lamina, the fact remains that the results presented here do not completely match the expectation set forth in previous literature. This mismatch is especially true in *Pan*, *Papio*, and *Hylobates*. Thus, the discussion above precipitates a series of questions, with perhaps the most obvious being: are the results presented here biologically informative and statistically valid? Related, why do the results not necessarily corroborate what is previously published? And, perhaps most unsettling: are the analyses appropriate to the questions at hand?

### **Methodological challenges in primate dental development**

Grounded in the recommendations of Bolter et al. (2023), the present study applied a Bayesian multivariate probit model to tooth mineralisation scores of five extant primate species. This specific statistical approach was chosen to account for the complexity of ordinal data in a multivariate setting. While the initial intent following Bolter et al. (2023) is to apply the



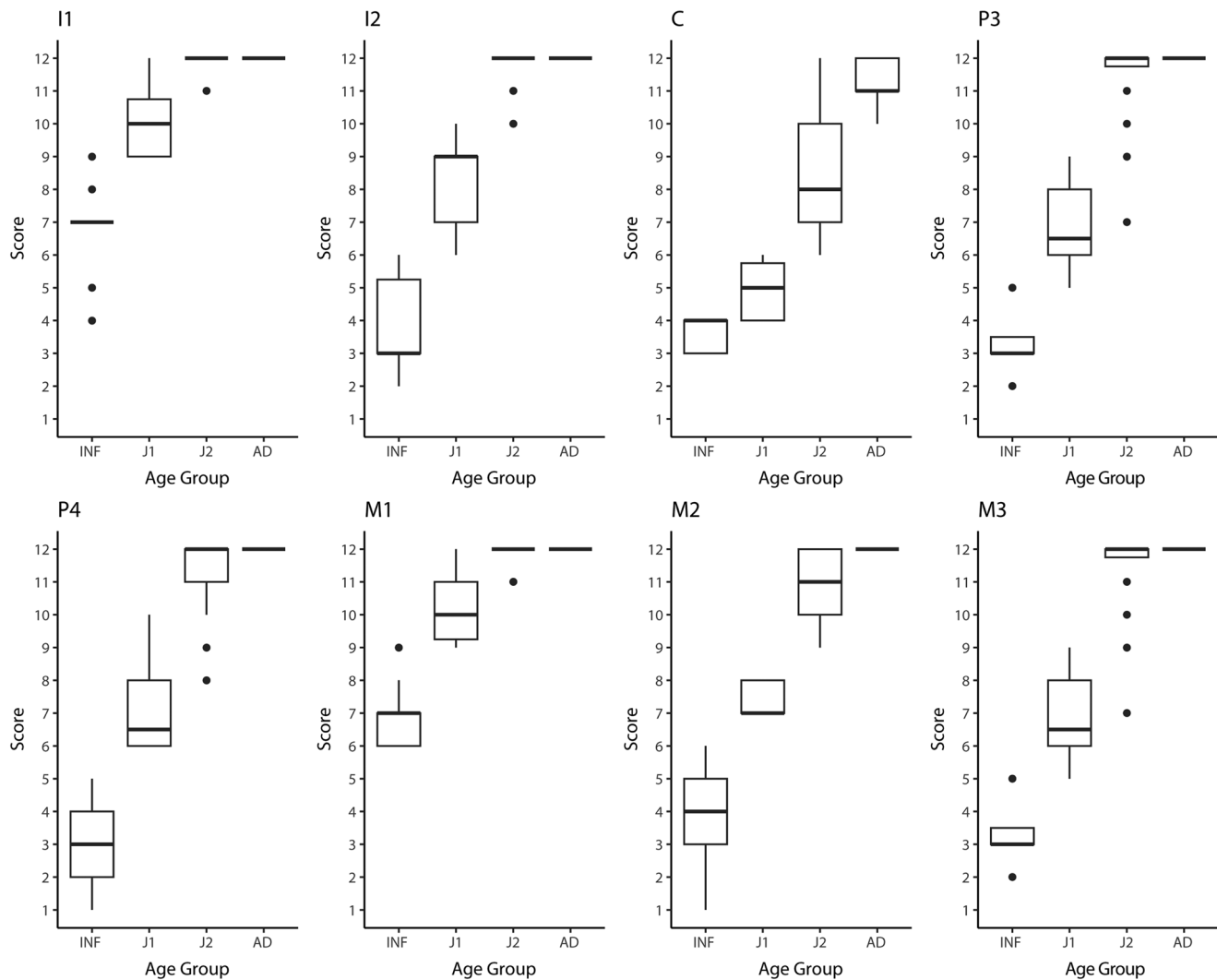
**Figure 9.** Multivariate comparisons of *H. lar* tooth formation. (a) Correlation plot of the posterior mean correlation values between the mandibular dentition. Values closer to 1 are strongly positive and darker red in colour. (b) Loadings plot of the eigendecomposition of the posterior mean correlation matrix. Dimension 1 and Dimension 2 explain approximately 77% of the variation in the data. The more acute the angle between variables, the more strongly correlated. The longer the arrows, the more influential in dimensions 1 and 2.

method to extinct hominin data, we must first ensure the results are valid. Validity in this work can be defined in two ways. First, statistical validity relates to the ability of the statistical model to be appropriate to the research question and the observed data, and accurately recover the structure of the observed data. Second, biological validity relates to the ability of the results to reflect accurately what previous research suggests the biology should show. For example, if the results were to predict that M3 (the latest forming tooth) is beginning development first, then one could question the biological validity of the results. During data analysis, we were faced with several challenges that may bring into question the statistical and/or biological validity of the results. In this section, we discuss specifically the limitations and challenges related to checking the validity of our study, and in this light, we offer our interpretation of the way forward in future studies.

We first contextualise the results in relation to validity of the multivariate probit model. In 1972, CJ Kowalski lamented the improper usage of multivariate statistics in biological anthropology. This work is one of many over the last few decades that has cautioned against perceived overuse of “complex,” “uninterpretable” methods in the field (Kowalski 1972; Corruccini 1975; Vark and Howells 1984). Yet recent technological advances matched with a long history of robust research demonstrates the necessity of using multivariate statistics to screen for and identify patterns of human variation important to the evolution, development, and behaviour of extant modern primates (Lande 1979; Boughner et al. 2015; Konigsberg 2015; Adams and Collyer 2019; Stull et al. 2023). However, it remains valid to ask if multivariate analyses could

illuminate the types of variation we might be interested in, relative to dental development. Previous work suggests that, yes, multivariate analyses are important to studying dental development (Braga and Heuze 2007; Boughner et al. 2015). In fact, the impetus behind this study was to improve upon such previous work in order to more stringently account for the complex, ordinal nature of dental development scores. The multivariate probit model is a valid statistical representation of the data analysed here. Figure 11 shows the posterior predictive fit of P3 scores in *Pan*. We used the parameters from the best fitting *Pan* model to predict the data to be expected. If the statistical model is a good representation of the data-generating process, then the predicted and observed data will be similar. Indeed, there is clear agreement between the observed data and the model predictions. In other words, the model is recovering what we initially inputted. This statement agrees with a myriad of research studies demonstrating the validity of probit models in biological anthropology (Liversidge and Molleson 2004; Konigsberg 2015; Stull et al. 2022, 2023).

In this discussion of statistical challenges, another important component is the degree to which our univariate analyses would necessarily predict any multivariate relationships present in the data. We began this discussion by highlighting the general expectation of relative tooth formation timing across primate species. This expectation led to another presumption that any multivariate relationships we see (i.e. stronger correlations between teeth), will in some way reflect this pattern. Certainly in *H. sapiens* this is the case, with earlier-forming teeth and later-forming teeth displaying a degree of modularity. But this expectation does not hold in

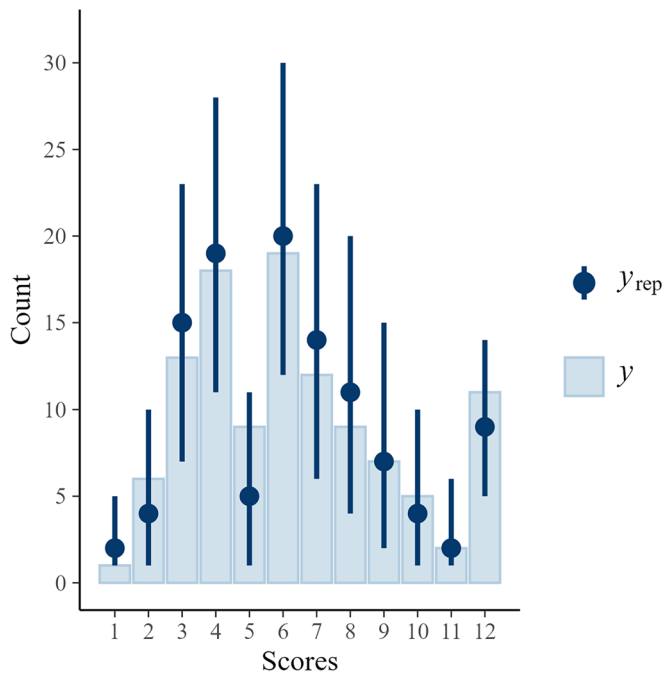


**Figure 10.** Boxplots visualising tooth mineralisation in *H. lar*. Scores adapted from protocols in Boughner et al. (2012) and modified from Dean and Wood (1981). Black circles are outliers, the horizontal line is the median, and the whiskers encompass the minimum and maximum values excluding any outliers.

the remaining specimens in this study. Thus, it may be inappropriate to assume that the univariate dental scores would predict the exact characterisation of the multivariate structure. For example, it may be possible that multivariate methods highlight the differential effect of age (or age group) on characterising the variation present in dental scores. This variation may be missed in traditional single-tooth approaches (see discussion below). Altogether, we suggest that further multivariate analyses appropriate to ordinal dental scores be done, in order to more comprehensively define the variation that may be present beyond that highlighted in traditional single-tooth and/or univariate approaches. In such studies we also suggest continued exploration of both uni- and multivariate analyses, towards building a body of evidence about whether and under what conditions these approaches may and may not return comparable results. These insights will be valuable in terms of directing the field towards the most powerful statistical approaches to characterising and comparing primate dental development data, particularly if sample sizes are small and/or homogenous.

Figure 11 demonstrates the validity of the statistical model to the observed data. However, statistical validity does not

guarantee biological validity. We argue that the results presented in this study may differ from other published works because of issues associated with the underlying data that are used to fit the statistical models. Put simply, the model is simply outputting that which it is input. While we can recover the structure of said input (Figure 11), there is no guarantee that this structure is generalisable to the biology as a whole – especially when dealing with convenience samples, which are effectively all that we have access to as biological anthropologists, primatologists, and paleoanthropologists. The present study suffers from data quality issues that plague much cross-sectional growth and development research – particularly in studies of (extinct and extant) non-human primates. These data quality issues relate to sample size, sample composition, response data structure, and predictor data structure. It is no surprise that the most interpretable sample in our study is *H. sapiens*. This group contains the largest sample size, is a fairly robust representation across all of ontogeny, and uses a continuous predictor in chronological age. In comparison, none of the four non-human primate samples has relatively equal representation across ontogeny; thus, individual age categories may be biased. For example, M1



**Figure 11.** Posterior predictive bars demonstrating the fit of P3 development according to the *Pan* model. The bars  $y$  are the counts of observed data by score and the error bars  $y_{rep}$  are the 95% predicted values from the model. Note the overlap between  $y$  and  $y_{rep}$  indicating the model does recover the true data and hence, is a good fit to the observed data.

development in gibbons (Figure 10) has the lowest score of 6 for an infant specimen. With the expectation that M1 begins development earliest (sometimes beginning prenatally), then in this sample we are missing an entire portion of ontogeny captured by scores less than 6 for the M1. In effect, our sample has “older” infants, but not “younger” infants. This type of bias is present across each non-human primate sample in our study and is likely to be found in samples more broadly due to underrepresentation of very young animals in institutional skeletal collections of wild-shot primates collected decades if not centuries ago. Conversely, similar bias arises when a specimen has scores that represent only the completion of development (stage 12, root apex closed). Effectively, a dataset that is only made up of older individuals will also generate a skewed picture of biological relationships. Aside from the problem of absent data, this skew occurs because we often characterise the relationship between teeth based on a shared trajectory (i.e. I1 and I2 both develop at the same time). The problem with this approach is two-fold. First, if development is complete, we can no longer track how dental scores may change over time. This lack of tracking includes not knowing whether a tooth completed mineralisation/root closure recently or a while ago. Second, in the absence of data (e.g. a missing tooth), we may be tempted to assume a relationship between, say, I1 and I2, or P3 and P4, that may be common among species but not hold true for the species under investigation (e.g. the early-forming I1 in *H. lar*). Regardless of whether the sample is skewed older or younger, the multivariate model is objectively picking up on this lack of variability in sample composition and as a result, is displaying a biased view of ontogenetic and biological relationships.

This error is compounded by the fact that our age predictor is ordinal in the non-human primate groups. Previous work suggests that ordinal categories are far less informative as compared to continuous-valued data (Wolfe et al. 2022). It is possible that if more nuanced information were present about age, such as knowing continuous, true chronological age (e.g. 10 years and 8 months old), then perhaps a clearer picture of the multivariate relationships is possible to achieve. The obvious challenge here is that for many collections or specimens, chronological age is not known and cannot be reconstructed precisely and/or reliably. For this reason, we tested the applicability of the model by not including age *a priori*. There is compelling rationale for methods that do not require age data exactly because the true or precise chronological age may not be known or knowable for historical human populations, non-human primate populations, and, particularly, extinct hominin or even extinct non-hominin primate individuals. However, in every case that we ran in our study, we found that model performance was worse in models that excluded age as a predictor (Table 2). In these cases, model performance was quantified based on k-fold cross validation and the expected log posterior density (ELPD). The larger the ELPD, the better fitting the statistical model to the observed data. The simple fact is that dental development, like many traits associated with ontogeny, is function valued. This means that individual tooth scores depend upon some continuous index (Gomulkiewicz et al. 2018). Normally, we would incorporate a measure of time (continuous or ordinal) to account for said change in scores across an individual’s childhood; or perhaps more to the point, when studying cross-sectional samples, to account for differences in scores across individuals of various ages. As previous work has adumbrated, perhaps it is better to set the timing of stages in relation to other biological and/or skeletal variables, but not focus on an estimated age variable that may be fraught with error (Roksandic and Armstrong 2011; Liversidge 2015). Plainly, more information is necessary to highlight the complex ontogenetic relationships between disparate developmental processes (i.e. skeletal maturation and dental development) and the degree to which these processes could be used to delineate the timing or determine the age categories and model the time-dependence of growth processes. This requires not only increasing data coverage across the full period of dental development, but also a large, quality sample of other matched variables across each species related to skeletal growth, maturation, and development (Stull and Corron 2021a). *Vis-à-vis* rare and fragmented fossil primate materials, this recommendation becomes a large ask. We appreciate that this methodological ideal may not be realistic at this time. However, we raise this point such that workers may remain well informed about all the caveats and conditions of studying ontogeny of extinct species, towards tempering conclusions appropriate to the level of uncertainty involved in the analysis.

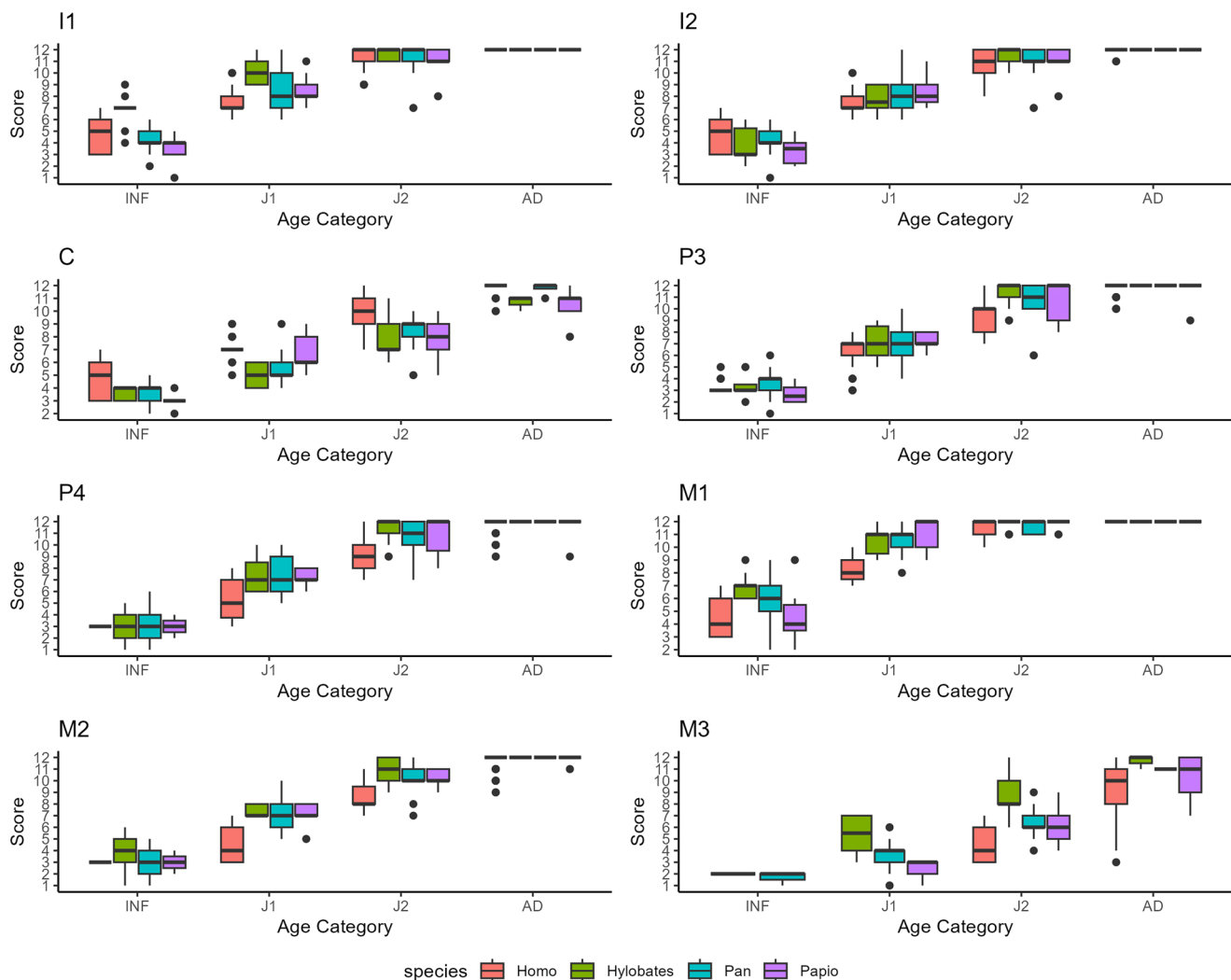
Importantly, we stress that notwithstanding the inclusion of some sort of “age” variable, a large and ontogenetically diverse sample size is paramount to accurate comparative studies. As a secondary set of analyses to further explore and

highlight the importance of an “age” variable, we collapsed the continuous human age values to match the age categories in *Pan*, *Papio*, and *Hylobates*. Even after collapsing the continuous-valued ages, human dental scores were still far more all-encompassing of the range of values (scores 1–12) as compared to species with smaller sample sizes (Figure 12). This outcome highlights even further that it is important to consider sample size and composition when drawing valid biological inferences from our statistical models. To show the importance of sample composition, in another set of secondary analyses we randomly sampled the *Homo* sample at different sizes ( $n=50, 200, 500, 1317$ ), and on each random sample re-ran the analyses. The loadings plots show the relationships with varying amounts of data and one can appreciate the varying relationships based on sample size (Figure 13). The point of this exercise is to explore why the human pattern may have matched our expectations, and why the other samples did not match our expectations. As we increased total sample size, the variable relationships more similarly matched what we would expect biologically. Therefore, regardless of the validity of the underlying statistical approach, biologically meaningful results are only

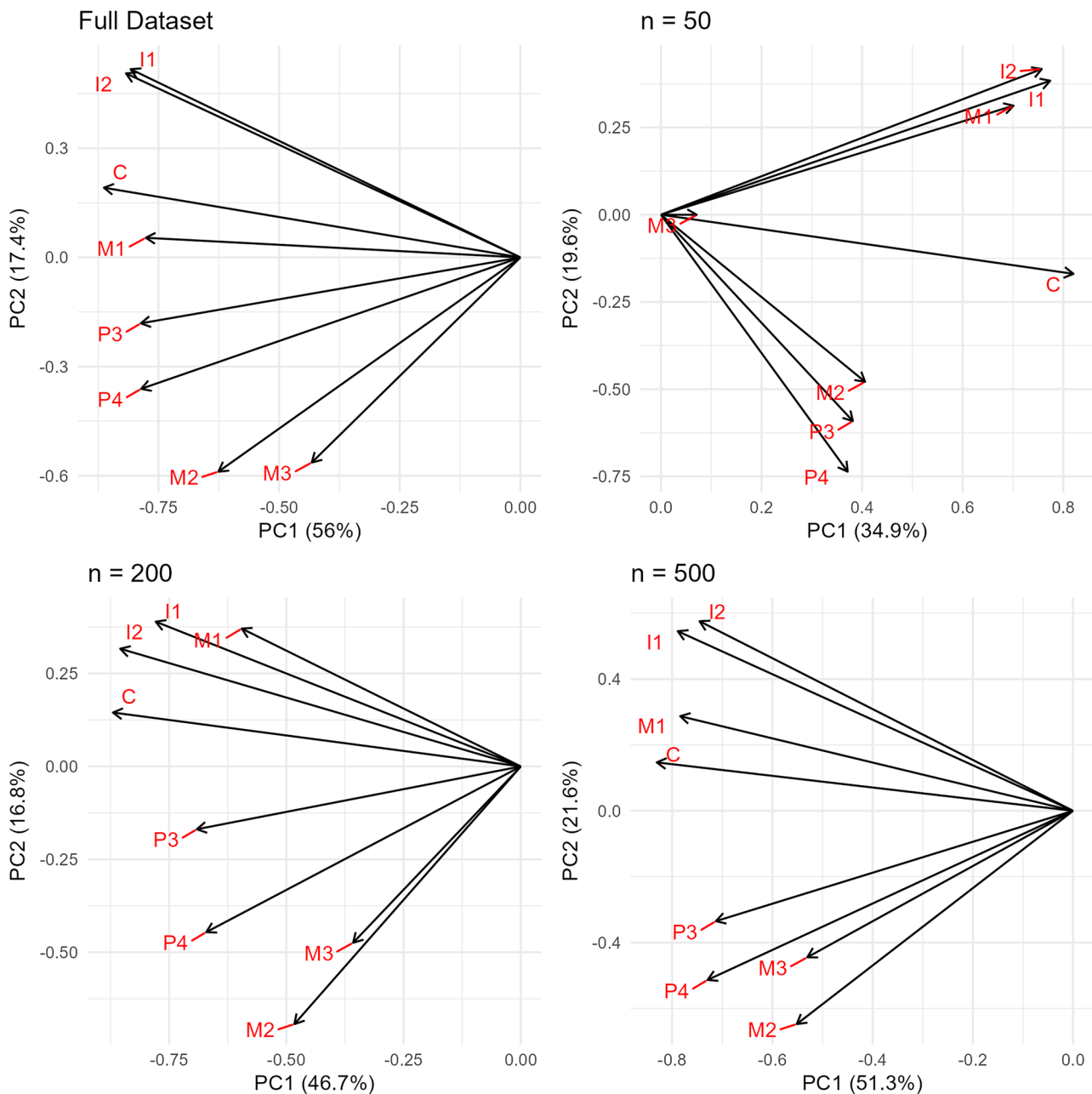
obtainable with biologically meaningful samples that are large and/or diverse enough to capture the full range of variation.

Of note, these secondary analyses are exploratory in nature and meant to further demonstrate the importance of thinking about one’s data when designing statistical models and interpreting biological relationships. There is not a single sample size that we could offer as “best” across all studies. In fact, given a series of constraints, it may be impossible to imagine a scenario where we could even get 500+ *Pan* specimens from the same group, during the same time period, covering all of ontogeny. Instead, we suggest effort be made to consider data quality as much as data quantity. For the current research question, an ontogenetically diverse assemblage of each species is likely superior. For example, 10 individuals from 5 different age groups (e.g. 2 infants, 2 children, 2 juveniles, 2 adolescences, and 2 adults) as compared to 100 individuals from 2 age groups (50 infants and 50 adolescences). The ideal situation is samples that are large and diverse.

The multivariate probit is an appropriate statistical approach to model primate dental development. We



**Figure 12.** Boxplots visualising tooth mineralisation across all species. The continuous age values in the *H. sapiens* data were collapsed to match the non-human primate data: INF = 0–3 years old, J1 = 3–7 years old, J2 = 7–12 years old, and AD = >12 years old.



**Figure 13.** Loadings plot adapted from matrices output from individual multivariate probit models on random subsets of  $n=50$ ,  $n=200$ ,  $n=500$ , and  $n=1316$  (full dataset) individuals. Note the overall difference in location of variables in multivariate space according to sample size. The more acute the angle between variables, the more strongly correlated. The longer the arrows, the more influential in dimensions 1 and 2.

suggest the probit be used in other research projects to characterise variation in extant and extinct populations. Given this statistical validity, it is likely that with a diverse enough sample, biological validity is also possible. The *Homo* results (Figure 3 and 4) are an example of this as they agree with previously published works (e.g. Liversidge 2003 and Moorrees et al. 1963a). Yet, even with the data biases described above, the results do pick out patterns with relevance to all species, such as the strong correlations between paired incisors (I1, I2) and premolars (P3, P4). Given the embryological and anatomical relationship within each tooth pair, this result is unsurprising. That said, unexpected outcomes such as the strong correlation between I1 and M1 in *H. lar* are also revealed using this approach. Altogether, biases aside, valid biological information can still be

obtained. Therefore, the present study is statistically valid with valid biological information that can be further expounded upon with an even more ontogenetically diverse assemblage in future studies.

### Conclusion and future directions

There has been a flourish of renewed interest in subadult individuals in both extant and extinct primates. This interest culminated in the Wenner-Gren workshop entitled, "New approaches to studying sub-adult hominins in the fossil record," which assembled a multi-disciplinary group of scholars to discuss and define a new paradigm for studying immature fossil specimens in human evolution (Bolter et al. 2023). An omnipresent theme of this workshop and the resulting

outcome of this paper is that neither biology, the data describing the biology, nor the statistical procedures summarising the data, produce a neat and tidy result. Instead, using an analytical method specifically tailored to data types used in dental development, we demonstrate that even the most statistically valid approach may not always guarantee the biological reality we see in our specimens and species for which we have relatively sound datasets and robust literature. The initial intent of this study was to provide a “novel” analytical approach to hominin and/or primate dental development (Bolter et al. 2023). What followed was an introspective analysis of the results, how valid they may or may not be in relation to previous work, and how the diversity of approaching subadult research in humans and non-human primate combines the limitations of all approaches. The multivariate probit is a statistically valid model and given the previously cited work and the results herein, should be an important component of future studies in dental development. The *Homo* results corroborate previous research and can be taken as biologically valid. The *Pan*, *Papio*, and *Hylobates* results are invalid inasmuch as the underlying data must be expanded to represent the entire ontogenetic period. Even still, there are biological relationships between certain teeth that do fit expected patterns that can be further explored. Therefore, we suggest with an appropriate accounting of error and limitations, the results here can be interpreted and expounded upon in later studies of primate dental development.

More data for *Pan*, *Papio* and *Hylobates* (and other primate taxa) will help illuminate the patterns described here to a similar effect of the *Homo* data. However, unless the data are available to quantify in sum, then our analyses truly are only as good as the data we are able to access. Thus, more than anything else, the workshop and this present work signifies the absolute importance of diverse, multi-disciplinary research teams and an open-science initiative where multivariate-literate teams of people can not only put forth their individual expertise concerning primate biology, but also bring together diverse multivariate data sources and critical minds to interpret these data. Importantly, such an interpretation requires that we be transparent about the challenges we face in data collection, data cleaning, data analysis, and presentation of the results. Only with this type of approach can we truly make headway in quantifying a complex, multivariate, function-valued process that not only describes how or why extant species may have evolved to grow up in particular ways, but perhaps more importantly to our own species, how and why a uniquely human life history evolved in the first place.

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### Funding

We thank the Wenner-Gren Foundation for funding [Gr. CONF-900]. J.C.B. would also like to acknowledge the Natural Sciences and Engineering

Council of Canada (NSERC) Discovery Grant [#2022-03716] and Canada Foundation for Innovation Grant [#29037].

### References

- Adams DC, Collyer ML. 2019. Phylogenetic comparative methods and the evolution of multivariate phenotypes. *Annu Rev Ecol Evol Syst.* 50(1):405–425. doi: [10.1146/annurev-ecolsys-110218-024555](https://doi.org/10.1146/annurev-ecolsys-110218-024555).
- Alemseged Z, Spoor F, Kimbel WH, Bobe R, Geraads D, Reed D, Wynn JG. 2006. A juvenile early hominin skeleton from Dikika, Ethiopia. *Nature.* 443(7109):296–301. doi: [10.1038/nature05047](https://doi.org/10.1038/nature05047).
- AlQahtani SJ, Hector MP, Liversidge HM. 2010. Brief communication: the London Atlas of Human Tooth Development and Eruption. *Am J Phys Anthropol.* 142(3):481–490. doi: [10.1002/ajpa.21258](https://doi.org/10.1002/ajpa.21258).
- Anderson DL, Thompson GW, Popovich F. 1976. Age of attainment of mineralization stages of the permanent dentition. *J Forensic Sci.* 21(1):191–200. doi: [10.1520/JFS10353J](https://doi.org/10.1520/JFS10353J).
- Appelt C, Van Ankum E, Marchiori D, Boughner J. 2021. Cell processes underpinning the evolution of primate dental form and formula. In: Pitirri MK, Richtsmeier JT, editors. *Genes, skin, energetics, breathing, and feeding, volume II, evolutionary cell processes in primates*. Boca Raton (FL): CRC Press; p. 55–81. doi: [10.1201/9781003206231-3](https://doi.org/10.1201/9781003206231-3).
- Armbruster WS, Pélabon C, Bolstad GH, Hansen TF. 2014. Integrated phenotypes: understanding trait covariation in plants and animals. *Philos Trans R Soc Lond B Biol Sci.* 369(1649):20130245. doi: [10.1098/rstb.2013.0245](https://doi.org/10.1098/rstb.2013.0245).
- Bache SM, Wickham H. 2022. magrittr: a Forward-Pipe Operator for R [Internet]. Available from: <https://CRAN.R-project.org/package=magrittr>.
- Balgopal R. 1996. Applications of the Frobenius norm criterion in multivariate analysis: the University of Alabama. Available from: <https://www.mathgenealogy.org/id.php?id=40517>
- Berkovitz BKB, Holland GR, Moxham BJ. 2009. *Oral anatomy, histology and embryology*. Amsterdam: Mosby/Elsevier.
- Berry SD, Edgar HJH. 2017. Development of a large-scale, whole body CT image database. Paper presented at: AMIA Annual Symposium Proceedings, Washington, DC, USA, 6–8.
- Betancourt M. 2016a. Identifying the optimal integration time in Hamiltonian Monte Carlo [Internet]. Available from: <https://arxiv.org/abs/1601.00225>.
- Betancourt M. 2016b. Diagnosing suboptimal cotangent disintegrations in Hamiltonian Monte Carlo [Internet]. Available from: <https://arxiv.org/abs/1604.00695>.
- Betancourt M. 2017. A conceptual introduction to Hamiltonian Monte Carlo [Internet]. Available from: doi: [10.48550/ARXIV.1701.02434](https://doi.org/10.48550/ARXIV.1701.02434).
- Bogin B. 2020. *Patterns of Human Growth*. 3rd ed. Cambridge University Press.
- Bolter DR, Cameron N. 2020. Utilizing auxology to understand ontogeny of extinct hominins: a case study on *Homo naledi*. *Am J Phys Anthropol.* 173(2):368–380. doi: [10.1002/ajpa.24088](https://doi.org/10.1002/ajpa.24088).
- Bolter DR, Cameron N, Hawks J, Churchill SE, Berger L, Bernstein R, Boughner JC, Elton S, Leece AB, Mahoney P, et al. 2023. Addressing the growing fossil record of subadult hominins by reaching across disciplines. *Evol Anthropol.* 32(4):180–184. doi: [10.1002/evan.21995](https://doi.org/10.1002/evan.21995).
- Bolter DR, Zihlman AL. 2022. Chapter 15. Evolution of human growth. In: Cameron N, Schell LM, editors. *Human growth and development*. 3rd ed. Boston: Academic Press; p. 425–448. doi: [10.1016/B978-0-12-822652-0.00016-X](https://doi.org/10.1016/B978-0-12-822652-0.00016-X).
- Boughner JC, Dean MC. 2004. Does space in the jaw influence the timing of molar crown initiation? A model using baboons (*Papio anubis*) and great apes (*Pan troglodytes*, *Pan paniscus*). *J Hum Evol.* 46(3):255–277. doi: [10.1016/j.jhevol.2003.11.007](https://doi.org/10.1016/j.jhevol.2003.11.007).
- Boughner JC, Dean MC, Wilgenbusch CS. 2012. Permanent tooth mineralization in bonobos (*Pan paniscus*) and chimpanzees (*P. troglodytes*). *Am J Phys Anthropol.* 149(4):560–571. doi: [10.1002/ajpa.22166](https://doi.org/10.1002/ajpa.22166).
- Boughner JC, Der J, Kuykendall KL. 2015. A multivariate approach to assess variation in tooth mineralization using free-lived and captive-raised

- chimpanzees (*P. troglodytes*). *Am J Phys Anthropol.* 158(3):452–462. doi: [10.1002/ajpa.22800](https://doi.org/10.1002/ajpa.22800).
- Boughner JC, Hallgrímsson B. 2008. Biological spacetime and the temporal integration of functional modules: a case study of dento–gnathic developmental timing. *Dev Dyn Off Publ Am Assoc Anat.* 237(1):1–17.
- Braga J, Heuze Y. 2007. Quantifying variation in human dental development sequences: an EVO-DEVO perspective. In: Bailey SE, Hublin J-J, editors. *Dent Perspect Hum Evol.* Dordrecht: Springer; p. 247–261.
- Bürkner P-C. 2022. Posterior: tools for working with posterior distributions. Available from: <https://mc-stan.org/posterior/>.
- Bürkner P-C, Charpentier E. 2020. Modelling monotonic effects of ordinal predictors in Bayesian regression models. *Br J Math Stat Psychol.* 73(3):420–451. doi: [10.1111/bmsp.12195](https://doi.org/10.1111/bmsp.12195).
- Cardoso HFV. 2007. Environmental effects on skeletal versus dental development: using a documented subadult skeletal sample to test a basic assumption in human osteological research. *Am J Phys Anthropol.* 132(2):223–233. doi: [10.1002/ajpa.20482](https://doi.org/10.1002/ajpa.20482).
- Carpenter CR. 1941. A field study in Siam of the behavior and social relations of the gibbon (*Hylobates lar*). Baltimore (MD): The Johns Hopkins Press.
- Chib S, Greenberg E. 1998. Analysis of multivariate probit models. *Biometrika.* 85(2):347–361. doi: [10.1093/biomet/85.2.347](https://doi.org/10.1093/biomet/85.2.347).
- Clements EMB, Zuckerman S. 1953. The order of eruption of the permanent teeth in the Hominoidea. *Am J Phys Anthropol.* 11(3):313–337. doi: [10.1002/ajpa.1330110309](https://doi.org/10.1002/ajpa.1330110309).
- Cofran Z, Boughner JC. 2025. “Smile-Obates”: Permanent dental development in the white-handed gibbon (*Hylobates lar carpenteri*). *Am J Biol Anthropol.* 186:e70019. doi: [10.1002/ajpa.70019](https://doi.org/10.1002/ajpa.70019).
- Cofran Z, Walker CS. 2017. Dental development in *Homo naledi*. *Biol Lett.* 13(8):20170339. doi: [10.1098/rsbl.2017.0339](https://doi.org/10.1098/rsbl.2017.0339).
- Coolidge HJ. 1937. The Asiatic primate expedition. *Science.* 85:11–12. (2192.), doi: [10.1126/science.85.2192.11.b](https://doi.org/10.1126/science.85.2192.11.b).
- Coolidge HJ, Allen J, Peters JL. 1940. Mammal and bird collections of the Asiatic primate expedition. *Mammals. Bulletin of The Museum of Comparative Zoology.* 87:121–211.
- Corron LK, Stock MK, Cole SJ, Hulse CN, Garvin HM, Klaes AR, Stull KE. 2021. Standardizing ordinal subadult age indicators: testing for observer agreement and consistency across modalities. *Forensic Sci Int.* 320:110687. doi: [10.1016/j.forsciint.2021.110687](https://doi.org/10.1016/j.forsciint.2021.110687).
- Corruccini RS. 1975. Multivariate analysis in biological anthropology: some considerations. *J Hum Evol.* 4(1):1–19. doi: [10.1016/0047-2484\(75\)90086-X](https://doi.org/10.1016/0047-2484(75)90086-X).
- Danaher PJ, Smith MS. 2011. Modeling multivariate distributions using copulas: applications in marketing. *Mark Sci.* 30(1):4–21. doi: [10.1287/mksc.1090.0491](https://doi.org/10.1287/mksc.1090.0491).
- Dean C, Leakey MG, Reid D, Schrenk F, Schwartz GT, Stringer C, Walker A. 2001. Growth processes in teeth distinguish modern humans from *Homo erectus* and earlier hominins. *Nature.* 414(6864):628–631. doi: [10.1038/414628a](https://doi.org/10.1038/414628a).
- Dean MC, Lim S-Y, Liversidge HM. 2022. Patterns of permanent incisor, canine and molar development in modern humans, great apes, and early fossil hominins. *Arch Oral Biol.* 143:105549. doi: [10.1016/j.archoralbio.2022.105549](https://doi.org/10.1016/j.archoralbio.2022.105549).
- Dean MC, Lucas VS. 2009. Dental and skeletal growth in early fossil hominins. *Ann Hum Biol.* 36(5):545–561. doi: [10.1080/03014460902956725](https://doi.org/10.1080/03014460902956725).
- Dean MC, Wood BA. 1981. Developing pongid dentition and its use for ageing individual crania in comparative cross-sectional growth studies. *Folia Primatol.* 36(1–2):111–127. doi: [10.1159/000156011](https://doi.org/10.1159/000156011).
- Demirjian A, Buschang PH, Tanguay R, Patterson DK. 1985. Interrelationships among measures of somatic, skeletal, dental, and sexual maturity. *Am J Orthod.* 88(5):433–438. doi: [10.1016/0002-9416\(85\)90070-3](https://doi.org/10.1016/0002-9416(85)90070-3).
- Demirjian A, Goldstein H. 1976. New systems for dental maturity based on seven and four teeth. *Ann Hum Biol.* 3(5):411–421. doi: [10.1080/03014467600001671](https://doi.org/10.1080/03014467600001671).
- Demirjian A, Goldstein H, Tanner JM. 1973. A new system of dental age assessment. *Hum Biol.* 45(2):211–227.
- Dirks W. 1998. Histological reconstruction of dental development and age at death in a juvenile gibbon (*Hylobates lar*). *J Hum Evol.* 35(4–5):411–425. doi: [10.1006/jhev.1997.0185](https://doi.org/10.1006/jhev.1997.0185).
- Dirks W, Bowman JE. 2007. Life history theory and dental development in four species of catarrhine primates. *J Hum Evol.* 53(3):309–320. doi: [10.1016/j.jhev.2007.04.007](https://doi.org/10.1016/j.jhev.2007.04.007).
- Edgar HJH, Daneshvari Berry S, Moes E, Adolphi NL, Bridges P, Nolte KB. 2020. New Mexico decedent image database. Available from: doi: [10.25827/5s8c-n515](https://doi.org/10.25827/5s8c-n515).
- Gabry J, Cesnovar R. 2022. cmdstanr: R Interface to “CmdStan” [Internet]. Available from: <https://mc-stan.org/cmdstanr>.
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2014. Bayesian data analysis. 3rd ed. Boca Raton (FL): Chapman & Hall/CRC.
- Gelman A, Hwang J, Vehtari A. 2014. Understanding predictive information criteria for Bayesian models. *Stat Comput.* 24(6):997–1016. doi: [10.1007/s11222-013-9416-2](https://doi.org/10.1007/s11222-013-9416-2).
- Gómez-Robles A, Polly PD. 2012. Morphological integration in the hominin dentition: evolutionary, developmental, and functional factors. *Evolution.* 66(4):1024–1043. doi: [10.1111/j.1558-5646.2011.01508.x](https://doi.org/10.1111/j.1558-5646.2011.01508.x).
- Gomulkiewicz R, Kingsolver JG, Carter PA, Heckman N. 2018. Variation and evolution of function-valued traits. *Annu Rev Ecol Evol Syst.* 49(1):139–164. doi: [10.1146/annurev-ecolsys-110316-022830](https://doi.org/10.1146/annurev-ecolsys-110316-022830).
- González-José R, Van Der Molen S, González-Pérez E, Hernández M. 2004. Patterns of phenotypic covariation and correlation in modern human as viewed from morphological integration. *Am J Phys Anthropol.* 123(1):69–77. doi: [10.1002/ajpa.10302](https://doi.org/10.1002/ajpa.10302).
- Goodrich B. 2017. Truncated multivariate normal variate in Stan. Available from: <https://discourse.mc-stan.org/uploads/short-url/i9PwTnrcYK3J5BgY1rksltgKNw8.pdf>
- Gould SJ. 1996. Mismeasure of man. New York: WW Norton & Company.
- Gueorguieva RV, Agresti A. 2001. A correlated probit model for joint modeling of clustered binary and continuous responses. *J Am Stat Assoc.* 96(455):1102–1112. doi: [10.1198/016214501753208762](https://doi.org/10.1198/016214501753208762).
- Hajivassiliou VA, Ruud PA. 1994. Chapter 40. Classical estimation methods for LDV models using simulation. In: *Handbook of econometrics.* Vol. 4. Amsterdam: Elsevier; p. 2383–2441. doi: [10.1016/S1573-4412\(05\)80009-1](https://doi.org/10.1016/S1573-4412(05)80009-1).
- Hannigan A, Lynch CD. 2013. Statistical methodology in oral and dental research: pitfalls and recommendations. *J Dent.* 41(5):385–392. doi: [10.1016/j.jdent.2013.02.013](https://doi.org/10.1016/j.jdent.2013.02.013).
- Hillson S. 1996. Dental Anthropology. 1st ed. Cambridge: Cambridge University Press.
- Holly Smith B, Crummett TL, Brandt KL. 1994. Ages of eruption of primate teeth: a compendium for aging individuals and comparing life histories. *Am J Phys Anthropol.* 37(S19):177–231. doi: [10.1002/ajpa.1330370608](https://doi.org/10.1002/ajpa.1330370608).
- Jernvall J, Thesleff I. 2012. Tooth shape formation and tooth renewal: evolving with the same signals. *Development.* 139(19):3487–3497. doi: [10.1242/dev.085084](https://doi.org/10.1242/dev.085084).
- Joe H. 2014. Dependence Modeling with Copulas. Boca Raton (FL): Chapman & Hall.
- Kassambara A. 2023. ggpubr: “ggplot2” based publication ready plots [Internet]. Available from: <https://CRAN.R-project.org/package=ggpubr>.
- Konigsberg LW. 2015. Multivariate cumulative probit for age estimation using ordinal categorical data. *Ann Hum Biol.* 42(4):368–378. doi: [10.3109/03014460.2015.1045430](https://doi.org/10.3109/03014460.2015.1045430).
- Kowalski CJ. 1972. A commentary on the use of multivariate statistical methods in anthropometric research. *Am J Phys Anthropol.* 36(1):119–132. doi: [10.1002/ajpa.1330360114](https://doi.org/10.1002/ajpa.1330360114).
- Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution.* 33(1Part2):402–416. doi: [10.2307/2407630](https://doi.org/10.2307/2407630).
- Lewandowski D, Kurowicka D, Joe H. 2009. Generating random correlation matrices based on vines and extended onion method. *J Multivar Anal.* 100(9):1989–2001. doi: [10.1016/j.jmva.2009.04.008](https://doi.org/10.1016/j.jmva.2009.04.008).
- Lewis AB, Garn SM. 1960. The relationship between tooth formation and other maturational factors. *Angle Orthod.* 30:70–77.
- Liddell TM, Kruschke JK. 2018. Analyzing ordinal data with metric models: what could possibly go wrong? *J Exp Soc Psychol.* 79:328–348. doi: [10.1016/j.jesp.2018.08.009](https://doi.org/10.1016/j.jesp.2018.08.009).
- Liversidge HM. 2010. Interpreting group differences using Demirjian’s dental maturity method. *Forensic Sci Int.* 201(1–3):95–101. doi: [10.1016/j.forsciint.2010.02.032](https://doi.org/10.1016/j.forsciint.2010.02.032).

- Liversidge HM. 2003. Variation in modern human dental development. In: *Patterns of growth and development in the genus Homo*. Cambridge: Cambridge University Press; p. 73–113.
- Liversidge HM. 2012. The assessment and interpretation of Demirjian, Goldstein, and Tanner's dental maturity. *Ann Hum Biol.* 39(5):412–431. doi: [10.3109/03014460.2012.716080](https://doi.org/10.3109/03014460.2012.716080).
- Liversidge HM. 2015. Controversies in age estimation from developing teeth. *Ann Hum Biol.* 42(4):397–406. doi: [10.3109/03014460.2015.1044468](https://doi.org/10.3109/03014460.2015.1044468).
- Liversidge HM, Molleson T. 2004. Variation in crown and root formation and eruption of human deciduous teeth. *Am J Phys Anthropol.* 123(2):172–180. doi: [10.1002/ajpa.10318](https://doi.org/10.1002/ajpa.10318).
- Moorrees CFA, Fanning EA, Hunt EE Jr. 1963a. Age variation of formation stages for ten permanent teeth. *J Dent Res.* 42(6):1490–1502. doi: [10.1177/00220345630420062701](https://doi.org/10.1177/00220345630420062701).
- Moorrees CFA, Fanning EA, Hunt EE Jr. 1963b. Formation and resorption of three deciduous teeth in children. *Am J Phys Anthropol.* 21(2):205–213. doi: [10.1002/ajpa.1330210212](https://doi.org/10.1002/ajpa.1330210212).
- Nolla CM. 1960. The development of the permanent teeth. *J Dent Child.* 27:254–266.
- R Core Team. 2023. R: a language and environment for statistical computing [Internet]. Available from: <https://www.R-project.org/>.
- Ramirez D, Via J, Santamaria I, Scharf LL. 2013. Locally most powerful invariant tests for correlation and sphericity of Gaussian vectors. *IEEE Trans Inform Theory.* 59(4):2128–2141. doi: [10.1109/TIT.2012.2232705](https://doi.org/10.1109/TIT.2012.2232705).
- Reid DJ, Schwartz GT, Dean C, Chandrasekera MS. 1998. A histological reconstruction of dental development in the common chimpanzee, *Pan troglodytes*. *J Hum Evol.* 35(4-5):427–448. doi: [10.1006/jhev.1998.0248](https://doi.org/10.1006/jhev.1998.0248).
- Roksandic M, Armstrong SD. 2011. Using the life history model to set the stage(s) of growth and senescence in bioarchaeology and paleodemography. *Am J Phys Anthropol.* 145(3):337–347. doi: [10.1002/ajpa.21508](https://doi.org/10.1002/ajpa.21508).
- Rosas A, Ríos L, Estalrich A, Liversidge H, García-Taberner A, Huguet R, Cardoso H, Bastir M, Lalueza-Fox C, Rasilla M, et al. 2017. The growth pattern of Neandertals, reconstructed from a juvenile skeleton from El Sidrón (Spain). *Science.* 357(6357):1282–1287. doi: [10.1126/science.aan6463](https://doi.org/10.1126/science.aan6463).
- Schlichting CD. 1989. Phenotypic integration and environmental change: what are the consequences of differential phenotypic plasticity of traits? *BioScience.* 39(7):460–464. doi: [10.2307/1311138](https://doi.org/10.2307/1311138).
- Schour I, Massler M. 1940. The development of the human dentition. *J Am Dent Assoc.* 28:1953–1160.
- Schultz AH. 1926. Fetal growth of man and other primates. *Q Rev Biol.* 1(4):465–521. doi: [10.1086/394257](https://doi.org/10.1086/394257).
- Schultz AH. 1944. Age changes and variability in gibbons. A Morphological study on a population sample of a man-like ape. *American J Phys Anthropol.* 2(1):1–129. doi: [10.1002/ajpa.1330020102](https://doi.org/10.1002/ajpa.1330020102).
- Schwartz GT, Dean C. 2001. Ontogeny of canine dimorphism in extant hominoids. *Am J Phys Anthropol.* 115(3):269–283. doi: [10.1002/ajpa.1081](https://doi.org/10.1002/ajpa.1081).
- Šešelj M. 2013. Relationship between dental development and skeletal growth in modern humans and its implications for interpreting ontogeny in fossil hominins. *Am J Phys Anthropol.* 150(1):38–47. doi: [10.1002/ajpa.22209](https://doi.org/10.1002/ajpa.22209).
- Sgheiza V. 2022. Conditional independence assumption and appropriate number of stages in dental developmental age estimation. *Forensic Sci Int.* 330:111135. doi: [10.1016/j.forsciint.2021.111135](https://doi.org/10.1016/j.forsciint.2021.111135).
- Smith BH, Tompkins RL. 1995. Toward a life history of the Hominidae. *Annu Rev Anthropol.* 24(1):257–279. 1995): doi: [10.1146/annurev.an.24.100195.001353](https://doi.org/10.1146/annurev.an.24.100195.001353).
- Song PX-K. 2000. Multivariate dispersion models generated from Gaussian copula. *Scandinavian J Stat.* 27:305–320.
- Square TA, Sundaram S, Mackey EJ, Miller CT. 2021. Distinct tooth regeneration systems deploy a conserved battery of genes. *Evodevo.* 12(1):4. doi: [10.1186/s13227-021-00172-3](https://doi.org/10.1186/s13227-021-00172-3).
- Stan Development Team. 2024. Stan modeling language users guide and reference manual [Internet]. Available from: <https://mc-stan.org>.
- Stull KE, Chu EY, Corron LK, Price MH. 2022. Subadult age estimation using the mixed cumulative probit and a contemporary United States population. *Forensic Sci.* 2(4):741–779. doi: [10.3390/forensicsci2040055](https://doi.org/10.3390/forensicsci2040055).
- Stull KE, Chu EY, Corron LK, Price MH. 2023. Mixed cumulative probit: a multivariate generalization of transition analysis that accommodates variation in the shape, spread and structure of data. *R Soc Open Sci.* 10(3):220963. doi: [10.1098/rsos.220963](https://doi.org/10.1098/rsos.220963).
- Stull KE, Corron LK. 2021a. Subadult Virtual Anthropology Database (SVAD) data collection protocol: epiphyseal fusion, diaphyseal dimensions, dental development stages, vertebral neural canal dimensions. Zenodo. doi: [10.5281/zenodo.5348392](https://doi.org/10.5281/zenodo.5348392).
- Stull KE, Corron LK. 2021b. SVAD\_US (1.0.0) [Data Set]. Zenodo. doi: [10.5281/zenodo.5193208](https://doi.org/10.5281/zenodo.5193208).
- Stull KE, Corron LK. 2022. The Subadult Virtual Anthropology Database (SVAD): an accessible repository of contemporary subadult reference data. *Forensic Sci.* 2(1):20–36. doi: [10.3390/forensicsci2010003](https://doi.org/10.3390/forensicsci2010003).
- Stull KE, Corron LK, Price MH. 2021. Subadult age estimation variables: exploring their varying roles across ontogeny. In: Algee-Hewitt BFB, Kim J, editors. *Remodel Forensic Skelet Age Mod Appl New Res Dir*. London: Academic Press; p. 50–74.
- Swindler DR, Meekins D. 1991. Dental development of the permanent mandibular teeth in the baboon, *Papio cynocephalus*. *Am J Hum Biol.* 3(6):571–580. doi: [10.1002/ajhb.1310030606](https://doi.org/10.1002/ajhb.1310030606).
- Vark GN, Howells WW. 1984. Multivariate statistical methods in physical anthropology: a review of recent advances and current developments. Cham: Springer.
- Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner P-C, Paananen T, Gelman A. 2024. loo: efficient leave-one-out cross-validation and WAIC for Bayesian models [Internet]. Available from: <https://mc-stan.org/loo/>.
- Vehtari A, Gelman A, Gabry J. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput.* 27(5):1413–1432. doi: [10.1007/s11222-016-9696-4](https://doi.org/10.1007/s11222-016-9696-4).
- Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. 2021. Rank-normalization, folding, and localization: an improved `\widehat{R}` for assessing convergence of MCMC (with discussion). *Bayesian Anal.* 16(2):667–718. doi: [10.1214/20-BA1221](https://doi.org/10.1214/20-BA1221).
- Wickham H. 2007. Reshaping data with the reshape package. *J Stat Soft.* 21(12):1–20. doi: [10.18637/jss.v021.i12](https://doi.org/10.18637/jss.v021.i12).
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, et al. 2019. Welcome to the tidyverse. *JOSS.* 4(43):1686. doi: [10.21105/joss.01686](https://doi.org/10.21105/joss.01686).
- Wolfe CA, Chu EY, Corron LK, Price MH, Stull KE. 2022. Advances in subadult age estimation: using information theory to explore the relationship between growth indicators and age. Paper presented at: 91st Business Meeting of the American Association of Biological Anthropologists, Denver, CO.
- Wolfe CA, Boughner JC, Stull KE. 2024. WolfeBoughnerStull\_AHB [Internet]. Available from: doi: [zenodo.12738701](https://doi.org/zenodo.12738701).
- Yadav S, Shukla S. 2016. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. Paper presented at: 2016 IEEE 6th International Conference on Advanced Computing IACC. New York: IEEE; p. 78–83.
- Zollhofer CPE, de León, MSP. 2010. The evolution of hominin ontogenies. *Semin Cell Dev Biol.* 21(4):441–452. doi: [10.1016/j.semcdb.2009.10.012](https://doi.org/10.1016/j.semcdb.2009.10.012).

## Appendix A: Statistical model specification

This section describes in detail the multivariate cumulative probit model implemented in the current study. We start by defining the exact statistical model. Then, we describe the Bayesian inference used to estimate the posterior density of desired parameters. This includes the specification of prior probabilities and how we handled missing response data in the analyses. All code, datasets, and specific procedures are cited in the main text.

### Multivariate cumulative probit model

Let  $y_{ij}$  be the observed polychotomous ordinal outcome of dental development scores for individual  $i$  where  $(i=1, \dots, N)$  and outcome  $j$  where  $(j=1, \dots, J)$ . The observed  $y_{ij}$  is associated with the latent variable  $y_{ij}^*$  where the relationship can be defined as:

$$y_{ij} = k \text{ if } \gamma_{j,k-1} < y_{ij}^* \leq \gamma_{j,k}, \quad (1)$$

for  $k=1, \dots, K_j$ , where  $K_j$  is the number of categories [mineralisation scores] for the  $j$ -th outcome, and  $\gamma_{j,k}$  are the threshold parameters with  $\gamma_{j,0} = -\infty$  and  $\gamma_{j,K_j} = \infty$ .

The probability that the observed outcome  $y_{ij}$  equals  $k$  is:

$$P(y_{ij} = k | \theta_j, \gamma_j, X_i) = \Phi(\gamma_{j,k} - \mu_{ij}) - \Phi(\gamma_{j,k-1} - \mu_{ij}), \quad (2)$$

where  $\Phi(\cdot)$  is the cumulative distribution function of the univariate standard normal distribution,  $\theta_j$  is some series of parameters of the  $j$ -th outcome that parametrises the latent variable  $y_{ij}^*$  and  $\mu_{ij}$  is the mean of the latent variable distribution parameterised by  $\theta_j$  and, depending on the context, some predictor  $X_i$ . In this work we parameterise the latent variable in one of three ways:

$$\text{if } X_i \text{ is areal, continuous value } \mu_{ij} = X_i \beta_b + \varepsilon_{ij}, \quad (3)$$

$$\text{if } X_i \text{ is a monotonically ordered ordinal value } \mu_{ij} = \alpha_j + mo(X_i, \zeta) \beta_j + \varepsilon_{ij} \quad (4)$$

$$\text{if } X_i \text{ is not incorporated in a threshold-only model } \mu_{ij} = 0 + \varepsilon_{ij} \quad (5)$$

Above,  $\beta_j$ ,  $\alpha_j$ , and  $\zeta$  are parameters learned in the statistical model. In *H. sapiens*, we specify the continuous mean functions. In all other specimens we specify the monotonic mean function. During model selection we specify the threshold-only model. In (4),  $X_i$  can take on any value  $X_i \in \{0, \dots, D\}$  where  $D$  is the number of age categories. Further,  $X_i \rightarrow mo(X_i, \zeta) = D \sum_{i=1}^X \zeta_i$  where  $mo(\cdot)$  is a monotonic transform and  $\zeta$  is a vector defined as a simplex. Specification of this transform is adapted from (Bürkner and Charpentier 2020). The error terms  $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iJ})$  follow a multivariate normal distribution with a mean vector zero and  $J \times J$  positive definite correlation matrix  $\Sigma$  where the diagonal is set to 1

for identifiability and the off diagonals are the correlation coefficient  $\rho$  between pairs of response variables.

The joint distribution of the multiple outcomes requires that we integrate over the multivariate normal distribution such that:

$$P(y_i | \mu_i, \gamma, \Sigma) = \int_{A_j} \phi_j(y_i^* | \mu_i, \Sigma) dy_i^*, \quad (6)$$

where  $\phi_j$  is the  $J$ -dimensional multivariate normal density function and  $A_j$  is defined by the thresholds  $\gamma$ . There is no closed form solution to the integral defined above. Instead, we adapt a Markov Chain Monte Carlo approach to the Geweke, Hajivassiliou, and Keane (GHK) algorithm to simulate probabilities in the multivariate probit model (Hajivassiliou and Ruud 1994). The GHK algorithm takes advantage of the fact that the latent data model  $y_{ij}^*$  can be simplified using Cholesky factorisation where the correlation matrix  $\Sigma = LL'$  or the product of the lower triangular matrix and its transpose. We can rewrite the latent variable as  $y_{ij}^* = \mu_{ij} + LZ_i$  where  $Z_i$  is distributed  $N(0,1)$ . Effectively, this approach simplifies the process and allows us to draw from a truncated multivariate normal distribution using draws from a univariate normal distribution. Truncation results from the threshold parameters described above representing the transition from one ordinal stage to the next.

### Parameter estimation

Parameter estimation is completed in Stan, a probabilistic programming language written in C++ that computes the joint log probability density of a set of continuous parameters up to a proportional constant (Stan Development Team 2024). Bayesian inference is carried out using Stan's dynamic Hamiltonian Monte Carlo algorithm (HMC) to draw from the posterior distribution (Betancourt 2017). Specification of the multivariate probit model in Stan using an adaptation of the GHK algorithm is described at (Goodrich 2017). The general form for any set of Bayesian analyses is posterior  $\propto$  likelihood  $\times$  prior. The likelihood is described in Equation 6 and is parameterised according to the Cholesky factorisation.

All random variables were given weakly informative priors related to the scale of the data. In Equation 3,  $\beta \sim normal(0,5)$ . In Equation 4,  $\beta \sim normal(0,5)$  and  $\alpha \sim normal(0,5)$ . Further, in Equation 4,  $\zeta \sim dirichlet([1,1,1])$ . In all models, the Cholesky factor is distributed according to  $L \sim LKJ(4)$  (Lewandowski et al. 2009). Lastly, in all models, the threshold parameters are distributed  $\gamma \sim normal(k+1,1)$  where  $k$  is the  $k$ -th category – this construction just ensures that the thresholds are an ordered vector such that the value of  $k=12$  is larger than that at  $k=1$ . The GHK approach in Goodrich (2017) requires we introduce a parameter  $u_j$  to absorb inequality constraints. This is implied  $u \sim uniform(0,1)$ .

The model does allow for missing data in the response variable  $y_{ij}$ . We take the approach most commonly employed across Bayesian analyses and treat a missing value as a parameter to be learned as part of the joint distribution in the model (Gelman et al. 2014). While there is slight complexity due to the ordinal nature of the observed data, we assume that missing values  $y_{miss_{ij}}$  has no constraint and thus has a density of 1 or log density of 0 (Goodrich 2017).