Title

Designing databases and using generalized estimating equations for entheseal datasets: An example from the Tiwanaku culture (AD 500–1100)

Permalink

https://escholarship.org/uc/item/56s3s89t

Journal

International Journal of Osteoarchaeology, 33(3)

ISSN

1047-482X

Authors

Becker, Sara K
Stull, Kyra E
Chu, Elaine

Publication Date

2023-05-01

DOI

10.1002/oa.3210

Received: 12 September 2022 | Revised: 20 January 2023 | Accepted: 2 February 2023
DOI: 10.1002/oa.3210

SPECIAL ISSUE PAPER

Designing databases and using generalized estimating equations for entheseal datasets: An example from the Tiwanaku culture (AD 500–1100)

Sara K. Becker[1] | Kyra E. Stull[2] | Elaine Chu[2]

[1]Department of Anthropology, University of California, Riverside, Riverside, California, USA

[2]Department of Anthropology, University of Nevada, Reno, Reno, Nevada, USA

Correspondence
Sara K. Becker, Department of Anthropology, University of California, Riverside, 1334 Watkins Hall, 900 University Ave., Riverside, CA 92521, USA.
Email: sara.becker@ucr.edu

Funding information
Hellman Family Foundation; National Science Foundation; Patterson Ashmore UCR Grant

Abstract

Studying entheseal changes (EC) in human skeletal remains involves questions surrounding how researchers should collect, process, and evaluate these data as there are no set standards. Osteoarcheological research should also be able to answer population-level queries using entheses, such as if group A was moving their body in different ways from group B? Or are there age-related tasks or gendered labors? However, not all entheseal data can be easily evaluated in this fashion. If researchers select one area of the body, they may not be able to discuss population-level differences. In addition, if data are evaluated by each muscle attachment area throughout a body, the results can be overwhelming. Further, grouping EC may produce problems with statistical assumptions of independence. To address these design and scalar issues, we discuss proper database construction, including the importance for consistent data collection strategies and in anchoring individuals under a specimen number. We also show how generalized estimating equations (GEE) can address how individual-level scores can be collected and population-level research questions can be answered with entheseal marker data processed in SAS 9.4 or the free alternative, R. We utilize a sample of over 1200 adults from the Tiwanaku culture (AD 500–1100) of Bolivia and Peru. We demonstrate evaluating significant overall differences while also pinpointing specific EC by sex or by age at death to discuss various ways in which bodies moved in the past.

KEYWORDS
Andes, bioarcheology, database design, generalized linear model (GLM) statistics, labor and activity reconstruction, musculoskeletal stress markers, osteoarthritis, R software

1 | INTRODUCTION

Critiques (Jurmain, 1999; Jurmain et al., 2012) aimed at activity pattern studies have helped scholars move beyond biased activity reconstructions (e.g., fitting spear throwing to observed skeletal changes) to instead focus on labor changes at population levels among ancient peoples. What is still unresolved though is how researchers should collect, process, and evaluate activity markers as there are no set standards. This is especially true for one of the most data-generating components of this group, entheses. Entheses are where muscles, ligaments, and tendons anchor to bone, and can alter that underlying connection through movement and stress dissipation at a microstructural or osteon level (Benjamin et al., 2006; Shaw & Benjamin, 2007), thus potentially providing information about daily activities over the life of an individual. Entheses are divided into fibrous and fibrocartilaginous types. Fibrous entheses are where the larger muscle fibers of the skeleton attach directly to bone or the periosteum (Apostolakos et al., 2014; Benjamin et al., 2002, 2006). Fibrocartilaginous entheses attach as gradient areas of fibrous connective tissue, uncalcified or calcified fibrocartilage, and balance bone and tendon elasticity (Benjamin et al., 2002). When healthy, both types should be smooth and devoid of vascular changes or foramina at skeletal attachment points (Benjamin et al., 2002; Schrader, 2019; Villotte & Knüsel, 2013), but enthesopathies may cause new bone to form. Further, fibrocartilaginous entheses may be more susceptible to bony skeletal ridge development because stress dissipates over a larger region because of comparatively thin attachment areas at the subchondral bone, whereas fibrous entheses anchoring at larger cortical bone areas may better mitigate stress response (Benjamin et al., 2002, 2006). In addition, changes found in fibrous entheses may correlate to age and advancing age instead of muscle stress and developmental changes (Cardoso & Henderson, 2013; Michopoulou et al., 2017). However, modern biomechanical literature does not limit skeletal stress changes to only fibrocartilaginous entheses (Benjamin et al., 2002, 2006; Shaw & Benjamin, 2007). Hence, in the last decade, data collection has not been limited to either fibrous or fibrocartilaginous entheses. Additionally, there is also not one methodological standard, but the most utilized approaches to collect entheseal changes (EC) are the following: Hawkey and Merbs (1995), Villotte et al. (2010), Mariotti et al. (2007), or Coimbra (Henderson et al., 2015) (for a concise summary of each method and the information collected, see Schrader, 2019, p. 87–93).

Evaluating or advocating for one EC collection method is not the focus of this paper as all current methods are valid to explore activity and labor. Instead, this paper focuses on collecting EC across an individual's body and from multiple individuals within a sample to answer population-level questions about past groups. This first means constructing a flexible database design that links all EC data to each individual using a consistent specimen number. It also involves utilizing strong statistical analyses that maximize EC information to answer questions about ancient peoples while also not violating assumptions of statistical independence for multiple data points by individual. We describe these as design and scalar issues, which can be found in other osteoarcheological studies not just activity patterns (e.g., Becker, 2013, 2019; Gagnon, 2004; Gagnon & Wiesen, 2013; Greenwald et al., 2022; Nikita, 2014). Though, specifically in activity pattern studies, using one or two EC produces a scalar and design issue because that is not enough information to answer overall questions about past populations. Instead, most EC studies increase the scale and use multiple muscle attachment points; however, the amount of design data produced may overwhelm and make population-level interpretations hard. For example, the Villotte et al. (2010) approach has at least 47 total data points (35 fibrocartilaginous and 12 fibrous) upper and lower limb entheses. Hence, trying to interpret 47 different EC for each person among multiple individuals across a group or subgroup could provide convoluted results about past labor, especially if not everyone was performing the same jobs or there were age-related or gendered labors. Further, decreasing the number of EC evaluated may reduce information about ancient labor, especially if a scholar is not sure which EC to eliminate. Alternatively, subcategorizing EC (e.g., upper limb) may produce problems around assumptions of independence if a scholar is not using appropriate statistical analyses.

To address these design and scalar issues, we specifically focus on two main issues: general overall database design and strong modeling statistics. Our first focus, akin to EC method choice, is not to limit scholars to one database program. Instead, we prefer a design choice that anchors all data under a specific individual or burial number, is relational, is visually easy to use, and works for the simple naming of categories within that database. The reasoning is that this provides both the largest possible sample size and an ease of use when cleaning, programming, or adding to the project with additional individuals or EC. Our second focus is on the use of generalized estimating equations (GEE), population-averaged method accounting for correlation among measures within subjects (Agresti, 2007; Liang & Zeger, 1986). This type of modeling statistic is appropriate for use with skeletal cemetery samples where connections are focused on EC population-level data by different categories (i.e., spatially, temporally, by sex, and by age) or multiple categories (by age and sex). GEE also maximizes data in a database by the linked individual or burial number, but can group EC without violating statistical assumptions of independence. It also accounts for uneven distributions, which are often found in burial samples, small sample sizes, and for randomly missing or unobservable variables.

As a case study of this database and modeling statistics approach, we utilize a sample of 1,203 adults from the Tiwanaku polity (AD 500–1100) of ancient Bolivia and Peru. People in this study lived in the heartland core of the Tiwanaku region in Bolivia, and a lower elevation colony in Moquegua, Peru (Figure 1). Individuals in this sample all represent a shared cultural and genetic history, but the elevation differences of approximately 2,300 m mean there were similar labor specializations in ceramic and textile production but major differences in food and farming between the two disparate climatological regions, and an expectation of activity differences (Becker, 2017, 2020; Somerville et al., 2015; Vallières, 2012). Methodologically, EC were collected utilizing Hawkey and Merbs (1995).[1] In this approach, scores were collected for 37 different muscle attachment points within each individual (Table 1). From this sample of adults, assuming all scores could be collected for all individuals, this could mean that there are 44,511 present/absent EC, and we discuss grouping them by muscular "use-area." Overall, we advocate for forethought on database design, utilizing GEE modeling, and analyzing these data in a free statistical modeling program (although over the years we have used SAS 9.2/9.3/9.4, we currently prefer R). Utilizing our approach, activity pattern scholars can obtain a maximum amount of information to answer population-level questions.

---

[1] This represents what was commonly used circa 2007 when the first author began collecting Tiwanaku data, not an endorsement of any one method.

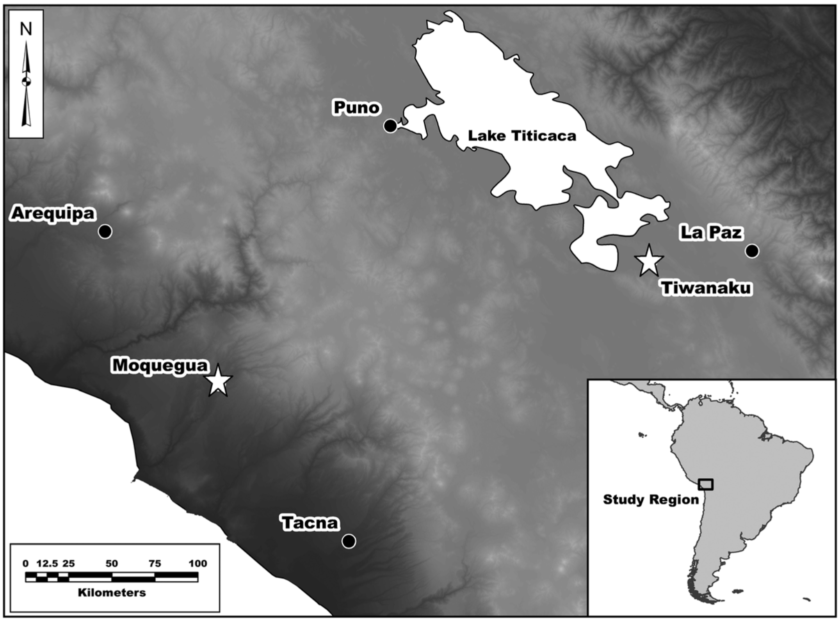


FIGURE 1 Map of the area with core and colony noted by stars.

2 | BRIEF BACKGROUND ON THE TIWANAKU CULTURE

Tiwanaku is one of the earliest South American Andean complex societies and focused on an agropastoral way of life (Janusek, 2004, 2008; Kolata, 1993, 2003). The core city of Tiwanaku (3,800 m) is located in the high, flat plains or *altiplano* in Bolivia, and Tiwanaku people also colonized other regions, including the Moquegua Valley of Peru (1,500 m) (Blom et al., 1998; Goldstein, 1989, 2000, 2005; Knudson et al., 2014). There were major differences in agriculture between the core and the colony. Raised-field farming was used in the *altiplano*, where dirt embankments were fed by repeatedly digging up canal mud to place on top of the mound as fertilizer in order to feed the local populace (Erickson, 1985; Janusek & Kolata, 2004). In the Moquegua region, Tiwanaku peoples used riverine canal agriculture, where fields were located close to the Moquegua River and alluvial branches were used to supply water to crops, taking advantage of the seasonal flows from rain or meltwater draining from higher elevations. As noted elsewhere in the world but with a similar arid climate, riverine agriculture provides predictable, risk-free cultivation (Nials et al., 2011). Thus, farming in the Moquegua River in Tiwanaku colonial times was likely less intensive than raised-field farming. Tiwanaku people also produced all kinds of crafts, including woven items, detailed ceramics, and carved stone and wood objects. In the highland city of Tiwanaku, there were even neighborhood enclaves dedicated to specific occupations like ceramics (i.e., Ch'iji Jawira) or llama herders (i.e., Mollo Kontu) (Figure 2) (Becker, 2017; Janusek, 2005; Vallières, 2012). Within Moquegua, settlements were divided across the valley by people who may have focused more on an agricultural way of life in the Chen Chen-style settlements or those who were Omo-style pastoralists (Becker & Goldstein, 2017; Goldstein, 2015). Overall, there should be significant EC between the Tiwanaku core and colony, as well as subsets of the population.

3 | CHOOSING AN EC DATABASE SOFTWARE

One software package, Excel, is commonly used by osteological scholars for data entry. As part of the Microsoft Office package, it has an easy learning curve. However, Excel is not a database software; its intended purpose is for accounting, with data entered in rows and columns on each “sheet,” of which each sheet is independent and is not combined. It has mathematical functionality that includes statistics, including basic items such as *t*-test, *z*-test, or a two-by-two contingency test (i.e., chi-square test). Excel includes visually user-friendly graphs, which can help with exploratory data analyses and conditional formatting/data cleaning. Limitations with Excel are that there is a maximum to how much information can be entered on each sheet by row and column, and how many additional saved sheets in each “workbook” document. In general, that total amount of data, with millions of lines of input and hundreds of variables, is more space than most individual scholars working with EC will need. However, a larger project with multiple scholars over a series of years may reach that limit. The larger concern of working in Excel would be in ascertaining interactions, filtering, and querying data. Excel is not designed as a database software with an automatic primary key. For example, burial *B12* has multiple columns of information on various sheets in Excel, but you cannot ask questions (i.e., query) within Excel using that burial *B12* to see all the relationships across all the sheets like you can in software that is designed as a database, potentially noting relationships that may be otherwise missed. Instead, data would need to be exported to a different program. Although Excel data can usually be easily imported into other statistical software packages to run more complex analyses, it does not provide the links that an actual relational database software can. In addition to Microsoft Excel, Google does offer a free Excel-style spreadsheet software called Sheets, albeit with fewer mathematical options, that may be a database

TABLE 1 List of entheses, separated by use area, and list of coded names for statistical analysis.

Movement area	Entheses/location
Upper arm	• Conoid ligament/clavicle = enth1 • Costoclavicular ligament/clavicle = enth2 • Subclavius/clavicle = enth3 • Trapezoid ligament/clavicle = enth4 • Trapezius/scapula = enth5 • Pectoralis minor/scapula = enth6 • Coracobrachialis/humerus = enth7 • Deltoideus/humerus = enth8 • Infraspinatus/humerus = enth9 • Latissimus dorsi/humerus = enth10 • Pectoralis major/humerus = enth11 • Subscapularis/humerus = enth12 • Supraspinatus/humerus = enth13 • Teres major/humerus = enth14 • Teres minor/humerus = enth15
Forearm	• Extensors, common origin/humerus = enth16 • Flexors, common origin/humerus = enth17 • Anconeus/ulna = enth18 • Brachialis/ulna = enth19 • Triceps brachii/ulna = enth20 • Biceps brachii/radius = enth21 • Brachioradialis/radius = enth22 • Pronator quadratus/radius = enth23 • Pronator teres/radius = enth24 • Supinator/radius = enth25
Mid-body	• Gluteus maximus/femur = enth26 • Gluteus medius/femur = enth27 • Gluteus minimus/femur = enth28 • Piriformes/femur = enth29 • Psoas major/iliacus/femur = enth30
Lower body	• Linea aspera (adductor longus and adductor magnus)/femur = enth31 • Quadriceps tendon (superior–anterior, rectus femoris, and vastus intermedius) /patella = enth32 • Patellar ligament/tibia = enth33
Foot	• Soleus/tibia = enth34 • Abductor hallucis/calcaneus = enth35 • Achilles tendon/calcaneus = enth36 • Flexor digitorum brevis/calcaneus = enth37

option if a researcher has budgetary restrictions. Sheets usually needs to be used online, but it has some offline functionality. For both Excel and Sheets, the ultimate concern is with data formatting and data integrity. Relational databases, in contrast, have a data model with typed fields, resulting in each field of a table having a specific data type. Consequently, the data format is homogeneous, which leads to fewer data formatting issues and potentially fewer issues in the data cleaning stages compared with working with nondatabase software (i.e., Excel and Sheets).

Access is the Microsoft answer to database management. As opposed to the ease of learning Excel, Access is more complex and usually requires many training hours. On the positive side, as shown in Figure 3, once mastered, it provides a visually easier form interface for data entry. Data are housed in a way like Excel (called a “table” instead of a sheet), which can be viewed as either in the “form” or by rows and columns in the table for easy querying.[2] Access and other databases autogenerate or permit a user to set a primary key to link all data collected across all tables. For example, Figure 4 shows master context for the first author's EC database specifically for burial AK-00725. If we were to explore the database further, we could see the links between AK-00725 and other data collected in the forms (or tables) on the left, such as how age and sex were estimated and relationships between other activity data collected (i.e., OA or osteoarthritis data). Alternatively, we could see that AK-00725 is a male from the Akapana site and we could do a basic query about how many other males there are from that site or the wider “Region” category of Highland. Advantages of Access are that it was commonly distributed as part of Office and has templates and wizards, and data can be shared or exported to a variety of formats, including to statistical software. Disadvantages of Access, beyond its steep learning curve, are that it is designed as a business software, so free training is not aimed at academics, unless those academics are keeping customer databases. Further, it is finite, not really aimed at working in the cloud, but more for a small group of users or an individual user, and primarily PC users. It is cost prohibitive, especially as more recent versions of Office consider it an add-on at additional cost. Akin to Excel and Sheets, Google has a beta version of a database software called Tables (tables.area120.google.com/u/0/home) modeled after Access. This may be a free and practical consideration for future EC database uses as functionality improves.

If a researcher wants to move beyond Excel into a relational database and feels Access takes too long to master or is not yet ready to work with the beta version of Tables, there are many more recent options for EC collection, some that are free or have free components. AirTable (airtable.com) is a customizable relational database that supports up to 1200 records as a free database and more records as a paid platform. Even though AirTable is a Mac product, it works with various platforms (Windows, macOS, iOS, and Android), which may be helpful to those collecting data on handheld tablets in the field. Some critiques have noted that it also has a steep learning curve, and there is no offline functionality, and therefore, it may not work in every field situation. Another option is MySQL (mysql.com), an Oracle-based open-source database platform that works on Mac, PC, and Linux with a free community edition (Workbench) that has multiuser functionality and password protections with read-only access, allowing an administrator to control access. Another option, although not free but less expensive than some of the other paid platforms, is WorkMap.ai, (hyperoffice.com/access-database-online), a cloud-based database software similar to Access, it works with Access files, and the company has modified the software to fit academic needs. Finally, archeologists have also designed databases specific for our field, such as Ishtar (ishtar-archeo.net/en) and E4/E5 (oldstoneage.com/osa/tech/index). Akin to the entheses methods, the choice of which database to use is up to the researcher.

---

[2] For use by interested scholars, we have included a blank EC Access database and master context following protocols outlined in Data S1 of this article.

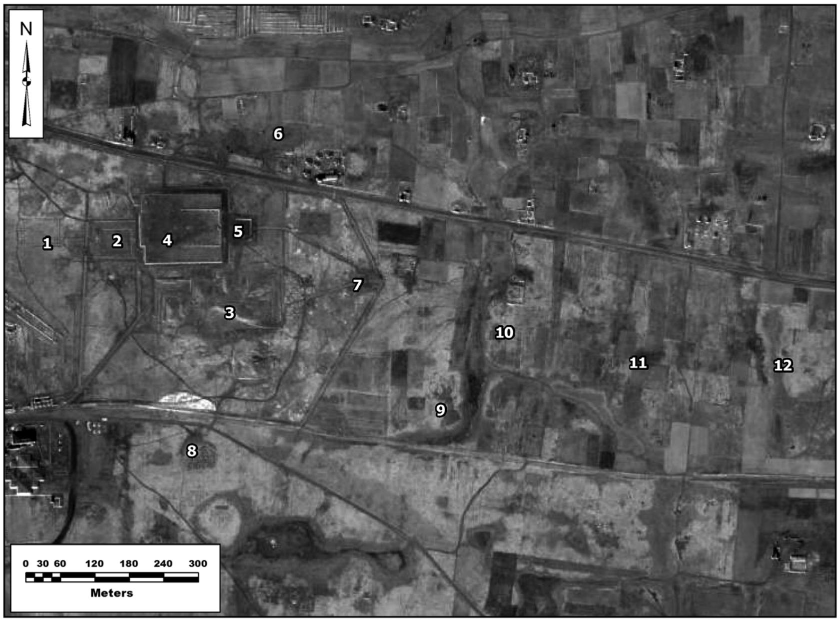
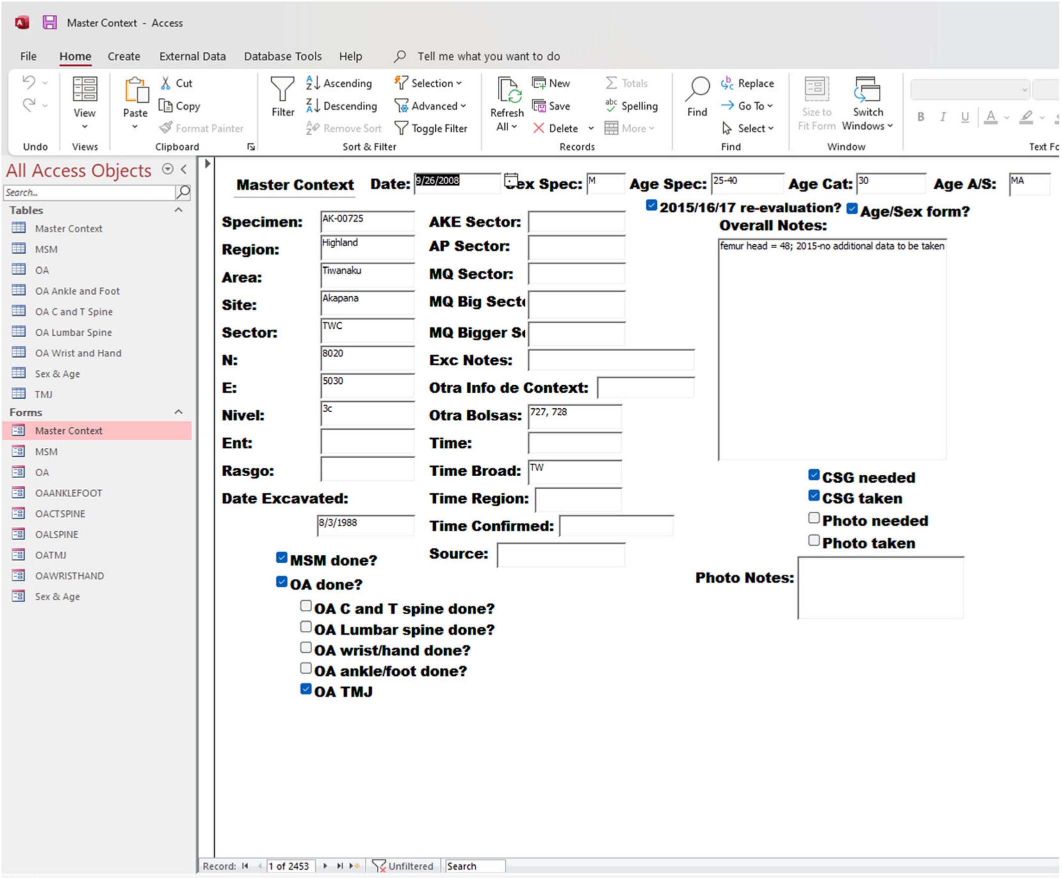


FIGURE 2 Neighborhoods in the city of Tiwanaku: (1) Kerikala, **(2) Putuni, (3) Akapana**, (4) Kalasasaya, (5) Subterranean Temple, **(6) La Karaña,** (7) Kantatallita, **(8) Mollo Kontu**, **(9) Akapana East 1**, **(10) Akapana East 2**, **(11) Marka Pata**, and **(12) Ch'iji Jawira** (sites in bold were included in this study).



Master Context

Date: 9/26/2008
ex Spec: M
Age Spec: 25-40
Age Cat: 30
Age A/S: MA

☑ 2015/16/17 re-evaluation? ☑ Age/Sex form?

Overall Notes: femur head = 48; 2015-no additional data to be taken

Specimen: AK-00725
Region: Highland
Area: Tiwanaku
Site: Akapana
Sector: TWC
N: 8020
E: 5030
Nivel: 3c
Ent:
Rasgo:
Date Excavated: 8/3/1988

AKE Sector:
AP Sector:
MQ Sector:
MQ Big Sect
MQ Bigger S
Exc Notes:
Otra Info de Context:
Otra Bolsas: 727, 728
Time:
Time Broad: TW
Time Region:
Time Confirmed:
Source:

☑ CSG needed
☑ CSG taken
☐ Photo needed
☐ Photo taken

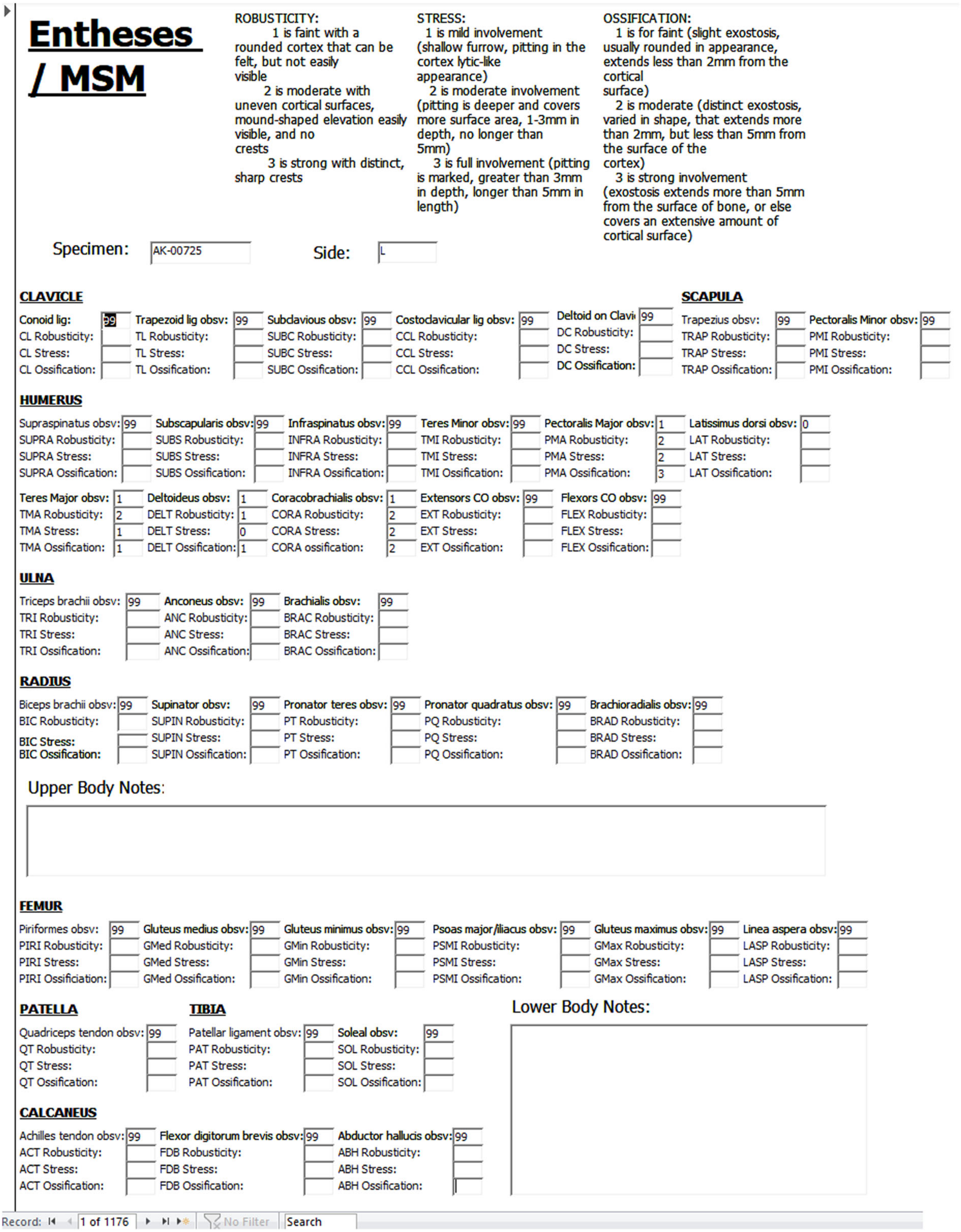
Photo Notes:

☑ MSM done?
☑ OA done?
  ☐ OA C and T spine done?
  ☐ OA Lumbar spine done?
  ☐ OA wrist/hand done?
  ☐ OA ankle/foot done?
  ☑ OA TMJ

FIGURE 3 Example of Access database data entry form. [Colour figure can be viewed at wileyonlinelibrary.com]

4 | BASICS OF EC DATABASE DESIGN

The abundance of EC data produced in any activity study requires a database that can handle all kinds of interactions and queries. In general, before collecting any EC data, an osteoarcheological researcher should consider some best practices: the database's purpose and kinds of information it will contain, database organization from general to specific including lists of fields of data, if there a



Entheses / MSM

ROBUSTICITY:
1 is faint with a rounded cortex that can be felt, but not easily visible
2 is moderate with uneven cortical surfaces, mound-shaped elevation easily visible, and no crests
3 is strong with distinct, sharp crests

STRESS:
1 is mild involvement (shallow furrow, pitting in the cortex lytic-like appearance)
2 is moderate involvement (pitting is deeper and covers more surface area, 1-3mm in depth, no longer than 5mm)
3 is full involvement (pitting is marked, greater than 3mm in depth, longer than 5mm in length)

OSSIFICATION:
1 is for faint (slight exostosis, usually rounded in appearance, extends less than 2mm from the cortical surface)
2 is moderate (distinct exostosis, varied in shape, that extends more than 2mm, but less than 5mm from the surface of the cortex)
3 is strong involvement (exostosis extends more than 5mm from the surface of bone, or else covers an extensive amount of cortical surface)

Specimen: AK-00725
Side: L

CLAVICLE

Conoid lig: 99
CL Robusticity:
CL Stress:
CL Ossification:

Trapezoid lig obsv: 99
TL Robusticity:
TL Stress:
TL Ossification:

Subclavious obsv: 99
SUBC Robusticity:
SUBC Stress:
SUBC Ossification:

Costoclavicular lig obsv: 99
CCL Robusticity:
CCL Stress:
CCL Ossification:

Deltoid on Clavi 99
DC Robusticity:
DC Stress:
DC Ossification:

SCAPULA

Trapezius obsv: 99
TRAP Robusticity:
TRAP Stress:
TRAP Ossification:

Pectoralis Minor obsv: 99
PMI Robusticity:
PMI Stress:
PMI Ossification:

HUMERUS

Supraspinatus obsv: 99
SUPRA Robusticity:
SUPRA Stress:
SUPRA Ossification:

Subscapularis obsv: 99
SUBS Robusticity:
SUBS Stress:
SUBS Ossification:

Infraspinatus obsv: 99
INFRA Robusticity:
INFRA Stress:
INFRA Ossification:

Teres Minor obsv: 99
TMI Robusticity:
TMI Stress:
TMI Ossification:

Pectoralis Major obsv: 1
PMA Robusticity: 2
PMA Stress: 2
PMA Ossification: 3

Latissimus dorsi obsv: 0
LAT Robusticity:
LAT Stress:
LAT Ossification:

Teres Major obsv: 1
TMA Robusticity: 2
TMA Stress: 1
TMA Ossification: 1

Deltoideus obsv: 1
DELT Robusticity: 1
DELT Stress: 0
DELT Ossification: 1

Coracobrachialis obsv: 1
CORA Robusticity: 2
CORA Stress: 2
CORA ossification: 2

Extensors CO obsv: 99
EXT Robusticity:
EXT Stress:
EXT Ossification:

Flexors CO obsv: 99
FLEX Robusticity:
FLEX Stress:
FLEX Ossification:

ULNA

Triceps brachii obsv: 99
TRI Robusticity:
TRI Stress:
TRI Ossification:

Anconeus obsv: 99
ANC Robusticity:
ANC Stress:
ANC Ossification:

Brachialis obsv: 99
BRAC Robusticity:
BRAC Stress:
BRAC Ossification:

RADIUS

Biceps brachii obsv: 99
BIC Robusticity:
BIC Stress:
BIC Ossification:

Supinator obsv: 99
SUPIN Robusticity:
SUPIN Stress:
SUPIN Ossification:

Pronator teres obsv: 99
PT Robusticity:
PT Stress:
PT Ossification:

Pronator quadratus obsv: 99
PQ Robusticity:
PQ Stress:
PQ Ossification:

Brachioradialis obsv: 99
BRAD Robusticity:
BRAD Stress:
BRAD Ossification:

Upper Body Notes:

FEMUR

Piriformes obsv: 99
PIRI Robusticity:
PIRI Stress:
PIRI Ossificiation:

Gluteus medius obsv: 99
GMed Robusticity:
GMed Stress:
GMed Ossification:

Gluteus minimus obsv: 99
GMin Robusticity:
GMin Stress:
GMin Ossification:

Psoas major/iliacus obsv: 99
PSMI Robusticity:
PSMI Stress:
PSMI Ossification:

Gluteus maximus obsv: 99
GMax Robusticity:
GMax Stress:
GMax Ossification:

Linea aspera obsv: 99
LASP Robusticity:
LASP Stress:
LASP Ossification:

PATELLA

Quadriceps tendon obsv: 99
QT Robusticity:
QT Stress:
QT Ossification:

TIBIA

Patellar ligament obsv: 99
PAT Robusticity:
PAT Stress:
PAT Ossification:

Soleal obsv: 99
SOL Robusticity:
SOL Stress:
SOL Ossification:

Lower Body Notes:

CALCANEUS

Achilles tendon obsv: 99
ACT Robusticity:
ACT Stress:
ACT Ossification:

Flexor digitorum brevis obsv: 99
FDB Robusticity:
FDB Stress:
FDB Ossification:

Abductor hallucis obsv: 99
ABH Robusticity:
ABH Stress:
ABH Ossification:

FIGURE 4 Example of Access database master context with specimen (i.e., burial) as the primary key that links all data collected no matter for what type of activity data, part of the body, and what year the first author collected data since research began in 2007. [Colour figure can be viewed at wileyonlinelibrary.com]

primary key (i.e., a unique identifier) and, if yes, what will the primary key be, what will the relationships be between the general and specific information, and after the design is in place, checking functionality and making sure these relationships work as they should, especially before the database is put into use (see Database Best Practices in Data S2). In terms of which database software to use or if

TABLE 2 Demography of the sample population by individual.

Tiwanaku core = 503 adults	By sex
	Total females = 76
	Total males = 102
	By age at death (and by sex)
	Age 15–19 = 21 (4 F, 0 M)
	Age 20–29 = 43 (9 F, 11 M)
	Age 30–39 = 70 (9 F, 17 M)
	Age 40–49 = 56 (7 F, 12 M)
	Age 50+ = 30 (6 F, 5 M)
	Total Tiwanaku city sample = 302
	Tiwanaku city females = 37
	Tiwanaku city males = 54
	Total Lukurmata city sample = 201
	Lukurmata city females = 39
	Lukurmata city males = 48
	Tiwanaku city sample by neighborhood = 162
	• Akapana = 35 adults
	• Females = 7, males = 13
	• Ch'iji Jawira = 6 adults
	• Females = 2, males = 2
	• La K'araña = 2 adults
	• Females = 0, males = 2
	• Marka Pata = 44 adults
	• Females = 11, males = 14
	• Mollo Kontu = 48 adults
	• Females = 9, males = 12
	• Putuni = 27 adults
	• Females = 0, males = 9
Moquegua colony = 700 adults	By sex
	Total females = 248
	Total males = 198
	By age at death (and by sex)
	Age 15–19 = 55 (20 F, 12 M)
	Age 20–29 = 130 (62 F, 42 M)
	Age 30–39 = 211 (92 F, 78 M)
	Age 40–49 = 75 (30 F, 27 M)
	Age 50+ = 50 (27 F, 10 M)
	Total Chen Chen-style sample = 649
	Chen Chen-style females = 231
	Chen Chen-style males = 181
	Total Omo-style sample = 51
	Omo-style females = 17
	Omo-style males = 17

designing an individualized platform, there are both paid and free database options available on various computer platforms. In what follows, we discuss a few options, but as technology is constantly changing, we think once a researcher has considered the research questions to be answered, the design of their database, the project's needs, geographic location, and access to Wi-Fi, flexibility is crucial and simpler is better.

Other basics to impart, first, especially before EC data collection, is to consider preventative data cleaning. No matter which EC, method(s), and database are chosen, before data can be analyzed, there will be a component where EC data will need to be evaluated and tidied. Tidying data is generally the most time-consuming component and is required before cleaning, filtering, subsetting, and analyzing data. Tidying the data means you are organizing your data in such a way that works well in the software programs. Some tidying is inherent to the data analysis pipeline, but some steps can be completed prior to data collection while one is organizing their database. For example, one can remove spaces from column names (i.e., Latissimus_dorsi_observed), be cognizant of capitalizations or varied capitalizations in the column names and within the columns (i.e., "R" and "r" are not the same thing, but you intended both to indicate the right side), and in general be consistent throughout all of data collection. Additional suggestions include shortening the name of variables (e.g., "latdo" for *Latissimus dorsi* observations), coding the entheses/variables with a similar name and numeric sequence (e.g., *Latissimus dorsi* is the 10th observation, code it as "enth10"), and consistently following naming conventions.

Second, when it comes to data collection, data should be anchored to each individual under one burial or specimen number (also shown in Figure 3 as the specimen indicator). Hence, we recommend a master context with unique signifiers for each person observed. This way if additional data are added later, such as another set of entheses or a different method but using the same individual, you have long-term comparable data. In addition, cataloging your data under this one number permits for ease of ways to clean the data, potentially run exploratory data analyses, and explore relationships/correlations. Ultimately, database and data collection strategies are not only designed with the specific researcher and research goal in mind, but also with the intention of future sharing, additions, and/or multiresearcher use. It is important that functionalities of the database, in terms of data entry, management, and maintenance, are intuitive enough for minimal time spent onboarding new researchers or incorporating new data.

5 | SAMPLE, METHODS, AND GEE

The case study for this paper focuses on the 1,203 adults in the **core** in Bolivia and the **colony** in Peru (Table 2). The colony in Moquegua, Peru, has almost 200 more adult individuals, and that is about the near perfect preservation of human remains in the Atacama Desert, whereas the highland Bolivia sample has fair to good preservation. Each individual was evaluated for 37 different entheseal attachment points and given a present/absent score. If present, it was then given an ordinal ranking for robusticity, stress, and ossification following the Hawkey and Merbs (1995) method.

TABLE 3 Generalized estimating equations entheseal changes results by each enthesis and combined use-area entheses between core and colony during the Tiwanaku culture (AD 500–1100).

Enthesis	Insertion (I), origin (O), or ligament (L)/bone	*n*	Core modeled % frequency	Colony modeled % frequency	χ^2 value (*df* = 1)	*p*-value
Conoid ligament	L/clavicle	416	85	82	0.13	0.71
Costoclavicular ligament	L/clavicle	400	96	85	3.26	0.07
Subclavius	**I/clavicle**	**428**	**44**	**19**	**12.23**	**0.0005**
Trapezoid ligament	**L/clavicle**	**402**	**67**	**36**	**15.94**	**<0.0001**
Trapezius	I/scapula	247	44	45	0.02	0.90
Pectoralis minor	I/scapula	259	63	40	3.14	0.08
Coracobrachialis	**I/humerus**	**528**	**42**	**22**	**10.62**	**0.001**
Infraspinatus	I/humerus	251	29	17	2.48	0.12
Latissimus dorsi	**I/humerus**	**247**	**22**	**3**	**14.34**	**0.0002**
Pectoralis major	I/humerus	517	91	90	0.04	0.83
Subscapularis	I/humerus	269	53	58	0.20	0.65
Supraspinatus	I/humerus	269	22	24	0.04	0.84
Teres major	**I/humerus**	**268**	**80**	**56**	**6.04**	**0.01**
Teres minor	**I/humerus**	**257**	**50**	**27**	**5.12**	**0.02**
Extensors CO	**O/humerus**	**458**	**24**	**41**	**4.61**	**0.03**
Flexors CO	O/humerus	454	20	18	0.06	0.80
Anconeus	I/ulna	498	58	40	3.74	0.053
Brachialis	I/ulna	538	87	82	2.04	0.15
Triceps brachii	**I/ulna**	**477**	**68**	**26**	**21.1**	**<0.0001**
Biceps brachii	I/radius	483	92	90	0.08	0.78
Brachioradialis	**I/radius**	**326**	**53**	**28**	**5.34**	**0.02**
Pronator quadratus	I/radius	366	42	40	0.08	0.78
Pronator teres	**I/radius**	**485**	**55**	**34**	**6.94**	**0.008**
Supinator	**I/radius**	**484**	**47**	**18**	**15.22**	**<0.0001**
Gluteus maximus	I/femur	532	94	89	1.83	0.18
Gluteus medius	**I/femur**	**341**	**58**	**29**	**8.54**	**0.004**
Gluteus minimus	I/femur	360	65	54	1.39	0.24
Piriformes	I/femur	341	52	37	1.93	0.16
Psoas major/iliacus	I/femur	370	83	69	3.29	0.07
Linea aspera	**I/femur**	**554**	**71**	**28**	**34.01**	**<0.0001**
Quadriceps tendon	Patella	253	31	34	0.08	0.78
Patellar ligament	L/tibia	369	69	59	1.34	0.25
Soleal (soleus)	**O/tibia**	**479**	**83**	**50**	**12.98**	**0.0003**
Abductor hallucis	O/calcaneus	321	33	32	0.01	0.94
Achilles tendon	Calcaneus	389	65	59	0.30	0.59
Flexor digitorum brevis	O/calcaneus	321	39	46	0.27	0.60
Entheseal use area						
Upper arm		**5801**	**63**	**47**	**15.99**	**<0.0001**
Forearm		**4569**	**56**	**43**	**12.0**	**0.0005**
Midbody		**1941**	**77**	**58**	**8.54**	**0.004**
Lower body		**1449**	**62**	**39**	**18.78**	**<0.0001**
Foot		1510	62	48	3.33	0.07

Note: Results in bold indicate a significant *p*-value at 0.05 or lower.

TABLE 4 Generalized estimating equations entheseal changes results between core and colony by sex during the Tiwanaku culture (AD 500–1100).

Entheseal use-area (females)	*n*	Core modeled % frequency	Colony modeled % frequency	χ^2 value ($df = 1$)	*p*-value
Upper arm	**2933**	**62**	**47**	**6.01**	**0.01**
Forearm	2310	49	43	1.64	0.20
Midbody	1025	75	58	3.48	0.06
Lower body	**557**	**59**	**35**	**8.32**	**0.004**
Foot	739	57	48	0.97	0.33
Entheseal use-area (males)	***n***	**Core modeled % frequency**	**Colony modeled % frequency**	**χ^2 value ($df = 1$)**	***p*-value**
Upper arm	**2223**	**69**	**47**	**11.17**	**0.001**
Forearm	**1695**	**66**	**16**	**17.04**	**<0.0001**
Midbody	**713**	**79**	**57**	**4.04**	**0.04**
Lower body	**467**	**67**	**43**	**8.03**	**0.005**
Foot	589	57	46	0.79	0.37

Note: Results in bold indicate a significant *p*-value at 0.05 or lower.

To evaluate the numerous EC and understand their patterns of expression within individuals at the population level, we advocate using GEE, which is a generalized linear model that accommodates correlated data and is appropriate for nonnormal, clustered data (Becker, 2019; Ghislatta & Spini, 2004; Hardin & Hilbe, 2013; Michopoulou et al., 2017; Nikita, 2014). GEE model estimates of population parameters are calculated using individually recorded data points, allowing for the largest possible sample size. However, each of these data points remains linked to the individual, preserving individual-level information (Ghislatta & Spini, 2004). The procedure retains the categorical dependent variable, while keeping the data points (each EC) of an individual linked. It also solves a common statistical issue in osteoarcheology, scalar issues of choosing between either individuals or each pathology measurement as the unit of comparison (Gagnon, 2004, 2006; Gagnon & Becker, 2019; Gagnon & Wiesen, 2013). If data are reduced to an overall present/absent count per individual, very specific pathology data can be lost and/or be insufficient to address research questions. However, if the condition is calculated on a per data point basis, one individual may skew statistical results when looking for patterns or violate statistical independence.

The GEE used in this study was applied to a logistic regression model (logit-linear) that simultaneously explores relationships between categorical, dependent variables and any number of nominal or quantitative predictor variables that cannot be assessed using bivariate analysis. Interactions among predictor variables can be examined for significance of whether or not the effect of one variable outcome is conditioned on the value of another at a 0.05 level or greater. The algebraic equation for this is $\text{Log}(\pi / 1 - \pi) = x\theta$, where x = an independent variable and the GEE estimates the change in θ (theta). The outcome variables were the binary EC data. We then used GEE regression analysis to look for significant differences ($p < 0.05$) between the independent categories of site location (e.g., core vs. colony) and time period and the dependent categories of each EC per specimen or groups of specimens. Finally, the Wald test was calculated as part of the logistic regression to test the null that groups in this study had the same EC rates.

GEE can also incorporate within-subject and between-subject variations, such as regional location, time period, sex, or age of individuals, or all of these, providing information on gendered labors or age-related activities (Djukic et al., 2015; Emslander et al., 1998; Molnar et al., 2011). For this reason, GEE models are usually ideal for group- or population-level inferences as it averages over all subjects on the within-subject covariance structure modeling response relationships. GEE is also flexible in that it accommodates small sample sizes and missing or unobservable data (Ballinger, 2004; Liang & Zeger, 1986), especially important when working with human skeletal remains where not all bones may be present and not all EC can be collected. Thus, GEE can provide the largest possible sample size for comparisons among the information collected. In the past, GEE was only available in costly statistical platforms (e.g., SAS); a current benefit to scholars, especially students, is that GEE is free in the open-source statistical R program (found under **gee**, Carey et al., 2022; **geepack**, Højsgaard et al., 2005; Yan, 2002; Yan & Fine, 2004; or **multigee**, Touloumis, 2015).

When analyzing EC, researchers usually analyze by each enthesis (i.e., *Latissimus dorsi* observation) as part of the larger methodological approach (e.g., 37 EC in the Hawkey and Merbs method). However, we advocate for use-areas for clearer presentation of data. This also removes some of the potential effects of dependence, because we know certain anatomical elements move as a unit and muscles work in tandem with other muscles. Therefore, we also present the data by five use-areas as per muscle location and movement areas (i.e., as noted in Table 1; upper arm, forearm, mid-body, lower body, and foot). By grouping EC in this fashion, it can be easier to evaluate things like geographic or chronological differences, as we will show in Section 6. The data were analyzed using *the chi-square statistic* (not to be confused with a two-by-two contingency table) and noting that *p*-values of 0.05 or less were considered significant. Overall, we built models using a variety of factors, such as age, sex, or geographical region, age-at-death was held as a constant in each calculation, and

TABLE 5 Generalized estimating equations entheseal changes results between core and colony by age during the Tiwanaku culture (AD 500–1100).

Entheseal use-area	*n*	Core modeled % frequency	Colony modeled % frequency	χ^2 value ($df = 1$)	*p*-value
15–19 years					
Upper arm	560	55	37	2.73	0.10
Forearm	**395**	**40**	**26**	**9.67**	**0.002**
Midbody	**153**	**71**	**39**	**20.20**	**<0.0001**
Lower body	**106**	**63**	**12**	**25.4**	**<0.0001**
Foot	**126**	**80**	**34**	**4.79**	**0.03**
20–29 years					
Upper arm	**2479**	**70**	**51**	**8.04**	**0.005**
Forearm	**1943**	**61**	**48**	**5.44**	**0.02**
Midbody	**840**	**90**	**66**	**9.84**	**0.002**
Lower body	**515**	**64**	**46**	**6.36**	**0.01**
Foot	395	43	33	0.54	0.46
30–39 years					
Upper arm	**2479**	**70**	**51**	**8.04**	**0.005**
Forearm	**1943**	**61**	**48**	**5.44**	**0.02**
Midbody	**840**	**90**	**66**	**9.84**	**0.002**
Lower body	**515**	**64**	**46**	**6.36**	**0.01**
Foot	692	70	54	1.80	0.18
40–49 years					
Upper arm	**648**	**82**	**50**	**12.82**	**0.0003**
Forearm	**559**	**73**	**49**	**7.96**	**0.005**
Midbody	229	—	—	—	—[a]
Lower body	**123**	**91**	**47**	**13.51**	**0.0002**
Foot	149	80	65	0.62	0.43
50+ years					
Upper arm	385	52	58	0.09	0.76
Forearm	294	71	54	2.35	0.13
Midbody	119	—	—	—	—[a]
Lower body	72	—	—	—	—[a]
Foot	97	87	51	0.06	0.8

[a] The generalized Hessian matrix is not positive definite and would not converge. No calculation could be obtained in generalized estimating equations for this sample.

Note: Results in bold indicate a significant *p*-value at 0.05 or lower.

the independent variables and each person's specimen number was the categorical dependent variable. Both the **PROC GENMOD** package in SAS and **gee** in R were used, and an independent correlation structure was specified.

6 | RESULTS

Previously published data show temporal differences in EC with significantly higher rates before the Tiwanaku culture (Becker, 2017, 2020); therefore, we only compare EC from AD 500–1100. EC data utilizing GEE are used to show spatial differences and specific data by age and sex. Table 3 contains all 37 entheseal scores for a Tiwanaku core and colony comparison where sex and age-at-death were run as part of the covariance matrix and held as constant. The bold denotes a significant relationship. Visually, it may be hard to see a pattern in the entheses that are significant between these two groups for all 37 results. Instead, the bottom of Table 3 shows the results in the combined **use-areas** of the body. Visually, it is much easier to note that four of the five are significantly different and that all the modeled use-area percentages are higher for individuals who lived in the core than those who lived in the colony. Further, although EC results are combined for both sides of the body, we can also look at results by left or right side of the body (Tables S1 and S2). By sex (Table 4), females have significant differences in the upper arm and lower body between core and colony. Males show significant results for the upper body, forearm, mid-body, and lower body comparisons with higher rates in the core than the colony. By age (Table 5), there are many

TABLE 6 Generalized estimating equations entheseal changes results within the neighborhoods in the city of Tiwanaku. Results in bold indicate a significant p-value at 0.05 or lower.

Entheseal use area	Site/modeled %	Akapana east	Ch'iji Jawira	La K'araña	Marka Pata	Mollo Kontu	Putuni
Upper arm (*n* = 468)	Akapana/67%	**$\chi^2 = 3.92$ p** = 0.048	**11.95** **0.0005**	**8.70** **0.003**	0.63 0.43	2.93 0.09	**10.68** **0.001**
	Akapana East/85%	-	**29.9** **<0.0001**	**25.4** **<0.0001**	**7.27** **0.01**	0.06 0.81	**24.3** **<0.0001**
	Ch'iji Jawira/40%	-	-	**11.79** **0.0006**	**5.41** **0.02**	**26.03** **<0.0001**	1.42 0.23
	La K'araña/44%	-	-	-	3.36 0.07	**21.84** **<0.0001**	2.70 0.10
	Marka Pata/59%	-	-	-	-	**5.89** **0.02**	**6.15** **0.01**
	Mollo Kontu/83%	-	-	-	-	-	**21.69** **<0.0001**
	Putuni/31%	-	-	-	-	-	-
Forearm (*n* = 340)	Akapana/50%	**6.63** **0.01**	**55.13** **<0.0001**	1.63 0.20	0.08 0.78	0.65 0.42	0.03 0.87
	Akapana East/83%	-	0.69 0.41	2.48 0.12	**5.99** **0.01**	3.16 0.08	**7.34** **0.01**
	Ch'iji Jawira/89%	-	-	**18.48** **<0.0001**	**59.05** **<0.0001**	**18.13** **<0.0001**	**67.47** **<0.0001**
	La K'araña/64%	-	-	-	1.16 0.28	0.12 0.73	2.11 0.15
	Marka Pata/52%	-	-	-	-	0.37 0.54	0.20 0.65
	Mollo Kontu/59%	-	-	-	-	-	0.91 0.34
	Putuni/48%	-	-	-	-	-	-
Entheseal use area	**Site/modeled %**	**Akapana east**	**Ch'iji Jawira**	**La K'araña**	**Marka Pata**	**Mollo Kontu**	**Putuni**
Midbody (*n* = 389)	Akapana/68%	—[a]	—[a]	—[a]	$\chi^2 = 1.48$ $p = 0.22$	0.71 0.40	**7.82** **0.01**
	Akapana East/%	-	—[a]	—[a]	—[a]	—[a]	—[a]
	Ch'iji Jawira/%	-	-	—[a]	—[a]	—[a]	—[a]
	La K'araña/%	-	-	-	—[a]	—[a]	—[a]
	Marka Pata/84%	-	-	-	-	0.01 0.90	**18.14** **<0.0001**
	Mollo Kontu/82%	-	-	-	-	-	**7.87** **0.005**
	Putuni/33%	-	-	-	-	-	-
Lower body (*n* = 224)	Akapana/52%	2.69 0.10	**5.03** **0.02**	0.32 0.57	0.16 0.69	**4.43** **0.04**	—[a]
	Akapana east/89%	-	**9.08** **0.003**	2.21 0.14	2.16 0.14	0.04 0.85	—[a]
	Ch'iji Jawira/25%	-	-	**40.72** **<0.0001**	**12.97** **0.0003**	**14.79** **0.0001**	—[a]
	La K'araña/60%	-	-	-	0.02 0.90	**4.06** **0.04**	—[a]
	Marka Pata/59%	-	-	-	-	3.82 0.051	—[a]
	Mollo Kontu/90%	-	-	-	-	-	—[a]
	Putuni/%	-	-	-	-	-	-
Foot (*n* = 154)	Akapana/66%	—[a]	—[a]	—[a]	0.03 0.87	0.03 0.86	**18.52** **<0.0001**

(Continues)

TABLE 6 (Continued)

Entheseal use area	Site/modeled %	Akapana east	Ch'iji Jawira	La K'araña	Marka Pata	Mollo Kontu	Putuni
	Akapana east/%	-	—[a]	—[a]	—[a]	—[a]	—[a]
	Ch'iji Jawira/%	-	-	—[a]	—[a]	—[a]	—[a]
	La K'araña/%	-	-	-	—[a]	—[a]	—[a]
	Marka Pata/69%	-	-	-	-	0.11 0.74	**14.67** **0.0001**
	Mollo Kontu/63%	-	-	-	-	-	**26.37** **<0.0001**
	Putuni/13%	-	-	-	-	-	-

[a] The generalized Hessian matrix is not positive definite and would not converge. No calculation could be obtained in generalized estimating equations for this sample.

Note: Results in bold indicate a significant *p*-value at 0.05 or lower.

significant differences for most age group, with rates higher in the core than colony.

Overall, we can delve into smaller sample sizes within these same sets of recorded observations. Looking at results from the core sites within the Tiwanaku city (Table 6), there are a variety of significant differences in the use areas between neighborhood enclaves potentially indicating occupational activity differences. For the Moquegua colony, we have run EC GEE by the 30–39 category, the largest age-at-death category in this area, and combined this with sex comparing Omo- and Chen Chen-style subgroups. Results for females (Figure 5) show one significant difference in the mid-body EC and for males (Figure 6) in the upper body EC. These are just a few of the different comparisons that could be run using these sample populations, but they provide differing magnitudes of comparison as case study examples.

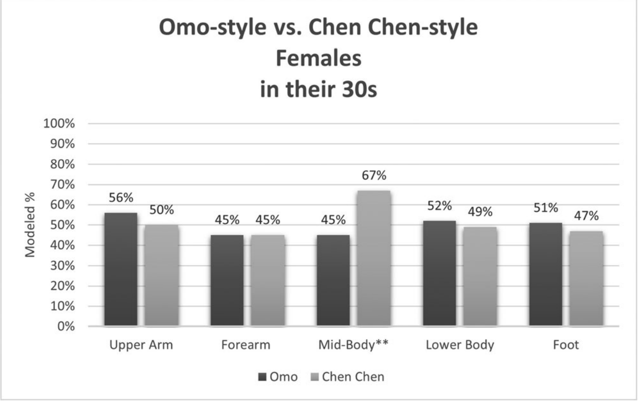


FIGURE 5 Generalized estimating equations entheseal changes results within the Moquegua colony between Omo- and Chen Chen-style females in the 30–39 age-at-death category. Stars indicate significant differences at the midbody entheseal changes, with $\chi^2 = 8.26$ and $p = 0.004$.

7 | DISCUSSION OF CASE STUDY DATA

Results by the 37 EC (Table 3) between core and colony show some significant differences, with modeled frequencies generally higher in the core. Use-area results also summarize these data likely showing raised-field agriculture as a heavier workload overall. Riverine or terrace farming, as practiced in the Tiwanaku colony in Moquegua, Peru, may have required work to establish fields and build channels, but it may not have been as repetitively intensive as raised-field farming. In the male and female comparisons (Table 4), rates are higher for females in the core but only significantly so in the upper arm and lower body EC. Males are significantly different for the upper arm, forearm, midbody, and lower body, again with higher modeled percentages in the core. In thinking about raised-field agriculture, it may be that men are performing all types of tasks from digging out the channels to carrying the dirt to the top of the raised mounds, and this type of agricultural work may require an almost whole-body effort (Erickson, 1985). In comparison, women are participating in this work, but it may be that instead of doing the heavier dirt hauling work, women may have performed work using tools, as has been noted in

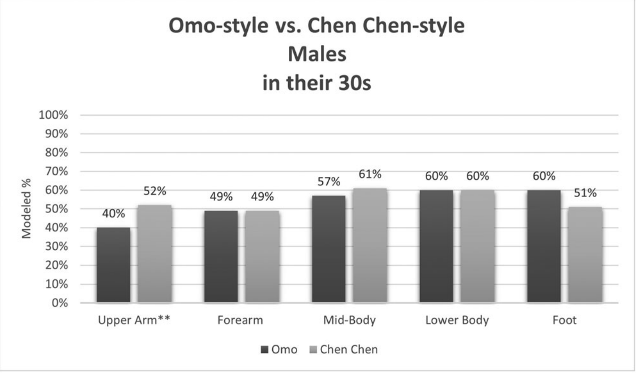


FIGURE 6 Generalized estimating equations entheseal changes results within the Moquegua colony between Omo- and Chen Chen-style males in the 30–39 age-at-death category. Stars indicate significant differences at the upper arm entheseal changes, with $\chi^2 = 5.34$ and $p = 0.02$.

modern observations and recreations of Andean raised-field labor (Erickson, 1988). The caution in these interpretations is that although age was held as a constant in these models, we may still have some age-related effects because a small portion of the sample is older than 50 years and specific age could not be estimated for some individuals. However, when we do look at results divided by age (Table 5), we find a similar pattern to the overall results and results divided by sex where modeled GEE EC rates are higher in the core than the colony. In most cases, they are significantly higher for all age groups except for the oldest individuals.

Results for the smaller sample comparisons among the different Tiwanaku city neighborhoods (Table 6) show some significant differences among the five EC use-areas that may have to do with differing occupations within the city, although sample sizes are rather small. One site, Ch'iji Jawira, has been noted archeologically as a ceramics production center (Rivera, 2003). When evaluating individuals from this site, they show evidence of strong hand musculature and EC, as well as repetitive movement damage (Becker, 2016). When comparing this neighborhood to others within Tiwanaku, these people show significant differences from many others in terms of high rates of upper arm and forearm EC but significantly lower rates of lower body musculature, which could be consistent with an occupation as potters. In addition, within these comparisons, the Mollo Kontu site archeologically has been described as home to *llameros* or llama keepers, herders, and weavers of camelid fiber (Couture, 2003; Vallières, 2016), and the upper and lower body EC results could support this as an occupation (e.g., drop spindle and weaving camelid fiber, as well as walking long distances between the core and colony). Finally, the Omo-style and Chen Chen-style EC comparisons (Tables 5 and 6) among 30- to 39-year-old females and males each have one significant difference, mid-body and upper arm, respectively, both with higher rates among Chen Chen-style peoples. Although less is known about specific tasks performed among Tiwanaku colonists in Moquegua, it may be those practicing a farming way of life were working those areas of musculature more intensely than their Omo-style pastoralists counterparts.

8 | LIMITATIONS IN THIS STUDY

Overall, our study is limited in that we have not backed a specific EC collection method nor a specific database for researchers to use. However, we feel the information provided educates osteoarcheologists on current available options, which is better for each scholar's individual circumstance. We have advocated for using GEE and note that GEE is limited in that it is not subject-specific, but population-averaged. Hence, it does not provide specific data about an individual within the study, but models using the population-averaged data. In addition, very small sample sizes may not be appropriate for GEE, as modeling may not function because of too many covariances or there will be a Type I error (Gunsolley et al., 1995). This may be shown in the results within the city of Tiwanaku (Table 6) when there are small sample sizes, either a Type I error is possible or the modelling was not possible.

9 | CONCLUSIONS

EC research requires complex thinking before any data are ever collected. Although there are multiple methods that can be used to collect data (i.e., Hawkey and Merbs, Villotte et al., Mariotti et al., and/or the Coimbra method), working on the mechanics of how these data will be collected, a database design, and what kinds of statistics will be used are important. After choosing one of the aforementioned EC methods, or some combination thereof, we advocate for researchers thinking strongly about scale and design issues in order to answer population-level questions about past peoples, especially questions that can address larger regional and chronological comparisons, but may also answer specifics such as those by zone, site, age-at-death, and/or sex about occupation or activity. To help with this, we have also provided information on best practices with database design, suggesting that researchers use a relational database program, and consider using one that is queryable and can hold large amounts of connected data, usually via a unique identifier such as specimen or burial number. Further, we have also advocated for utilizing GEE, which can combine any number of scores and factors and is resilient as it models population-level data with estimates of population parameters that are calculated using individually recorded data points, allowing for the largest possible sample size, while still linking these to the individual, preserving individual-level information (Ghislatta & Spini, 2004). GEE helps with scalar issues in osteoarcheology and maximizes data points collected, while also still permitting missing data, nonnormally distributed data, and smaller than normal sample sizes. Finally, GEE is a good choice for EC data because it can also control for things like age-at-death and by sex, to help analyze and interpret the way people moved their bodies in the past.

CONFLICT OF INTEREST

The authors report no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data can be made available upon reasonable request.

ORCID

Sara K. Becker https://orcid.org/0000-0002-3395-309X

REFERENCES

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). John Wiley & Sons, Inc.. https://doi.org/10.1002/0470114754

Apostolakos, J., Durant, T. J. S., Dwyer, C. R., Russell, R. P., Weinreb, J. H., Alaee, F., Beitzel, K., McCarthy, M. B., Cote, M. P., & Mazzocca, A. D. (2014). The enthesis: A review of the tendon-to-bone insertion. *Muscles, Ligaments and Tendons Journal*, *4*, 333–342. https://doi.org/10.11138/mltj/2014.4.3.333

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods*, *7*, 127–150. https://doi.org/10.1177/1094428104263672

Becker, S. K. (2013). Health consequences of contact on two seventeenth century native groups from the mid-Atlantic region of Maryland. *International Journal of Historical Archaeology*, *17*, 713–730. https://doi.org/10.1007/s10761-013-0240-3

Becker, S. K. (2016). Skeletal evidence of craft production from the Ch'iji Jawira site in Tiwanaku, Bolivia. *Journal of Archaeological Science: Reports*, *9*, 405–415.

Becker, S. K. (2017). Community labor and laboring communities within the Tiwanaku state (C.E. 500–1100). *Archaeological Papers of the American Anthropological Association*, *28*, 38–53. https://doi.org/10.1111/apaa.12087

Becker, S. K. (2019). Evaluating elbow osteoarthritis within the prehistoric Tiwanaku state using generalized estimating equations (GEE). *American Journal of Physical Anthroplogy*, *169*, 186–196. https://doi.org/10.1002/ajpa.23806

Becker, S. K. (2020). Why heterarchy? A view from the Tiwanaku state's (AD 500–1100) labor force. *American Anthropologist*, *122*, 934–939. https://doi.org/10.1111/aman.13499

Becker, S. K., & Goldstein, P. S. (2017). Evidence of osteoarthritis in the Tiwanaku colony, Moquegua, Peru (AD 500–1100). *International Journal of Osteoarchaeology*, *28*(1), 54–64.

Benjamin, M., Kumai, T., Milz, S., Boszczyk, B. M., Boszczyk, A. A., & Ralphs, J. R. (2002). The skeletal attachment of tendons—Tendon 'entheses'. *Comparative Biochemistry and Physiology Part a: Molecular & Integrative Physiology*, *133*, 931–945. https://doi.org/10.1016/S1095-6433(02)00138-1

Benjamin, M., Toumi, H., Ralphs, J., Bydder, G., Best, T., & Milz, S. (2006). Where tendons and ligaments meet bone: Attachment sites ("entheses") in relation to exercise and/or mechanical load. *Journal of Anatomy*, *208*(4), 471–490. https://doi.org/10.1111/j.1469-7580.2006.00540.x

Blom, D. E., Hallgrímsson, B., Keng, L., Lozada, M. C., & Buikstra, J. E. (1998). Tiwanaku 'colonization': Bioarchaeological implications for migration in the Moquegua Valley, Peru. *World Archaeology*, *30*(2), 238–261. https://doi.org/10.1080/00438243.1998.9980409

Cardoso, F. A., & Henderson, C. Y. (2013). The categorisation of occupation in identified skeletal collections: A source of bias? *International Journal of Osteoarchaeology*, *23*(2), 186–196. https://doi.org/10.1002/oa.2285

Carey, V.J., Lumley, T.S., Moler, C., and Ripley, B. (2022). Generalized estimation equation solver_. R package version 4.13–23. https://CRAN.R-project.org/package=gee

Couture, N. C. (2003). Ritual, monumentalism, and residence at Mollo Kontu, Tiwanaku. In A. Kolata (Ed.), *Tiwanaku and its hinterland: Archaeology and paleoecology of an Andean civilization* (pp. 202–225). Smithsonian Institution Press.

Djukic, K., Milovanovic, P., Hahn, M., Busse, B., Amling, M., & Djuric, M. (2015). Bone microarchitecture at muscle attachment sites: The relationship between macroscopic scores of entheses and their cortical and trabecular microstructural design. *American Journal of Physical Anthropology*, *157*, 81–93. https://doi.org/10.1002/ajpa.22691

Emslander, H. C., Sinaki, M., Muhs, J. M., Chao, E. Y. S., Wahner, H. W., Bryant, S. C., Riggs, B. L., & Eastell, R. (1998). Bone mass and muscle strength in female college athletes (runners and swimmers). *Mayo Clinic Proceedings of the Staff Meetings of the Mayo Clinic*, *73*(12), 1151–1160. https://doi.org/10.4065/73.12.1151

Erickson, C. L. (1985). Applications of prehistoric Andean technology: Experiments in raised field agriculture, Huatta, Lake Titicaca, 1981–82. In I. Farrington (Ed.), *Prehistoric intensive agriculture in the tropics, volume 232 (i) of BAR international series*. British Archaeological Reports.

Erickson, C. L. (1988). Raised field agriculture in the Titicaca Basin: Putting ancient agriculture back to work. *Expedition*, *30*(12), 8–16.

Gagnon, C. M. (2004). Food and the state: Bioarchaeological investigations of diet in the Moche Valley of Peru. *Dental Anthropology Journal*, *17*(2), 45–54. https://doi.org/10.26575/daj.v17i2.145

Gagnon, C. M. (2006). *Daily life and the development of the state in the Moche Valley of north coastal Peru: A bioarchaeological analysis* (Unpublished Ph.D. dissertation). University of North Carolina.

Gagnon, C. M., & Becker, S. K. (2019). Susquehannock lives 1575–1675: Insights from skeletal remains. *Historical Archaeology*, in press, *53*(2).

Gagnon, C. M., & Wiesen, C. (2013). Using general estimating equations to analyze oral health in the Moche Valley of Perú. *International Journal of Osteoarchaeology*, *23*, 557–572.

Ghislatta, P., & Spini, D. (2004). An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational Behavior Statistics*, *29*(4), 421–437. https://doi.org/10.3102/10769986029004421

Goldstein, P. S. (1989). The Tiwanaku occupation of Moquegua. In D. Rice, C. Stanish, & P. Scarr (Eds.), *Ecology, settlement, and history in the Osmore drainage, Peru* (pp. 219–255). British Archaeological Reports.

Goldstein, P. S. (2000). Communities without borders: The vertical archipelago and diaspora communities in the Southern Andes. In M. A. Canuto & J. Yaeger (Eds.), *The archaeology of communities: A new world perspective* (pp. 182–209). Routledge.

Goldstein, P. S. (2005). *Andean diaspora: The Tiwanaku colonies and the origins of the South American empire*. University of Florida Press.

Goldstein, P. S. (2015). Multiethnicity, pluralism, and migration in the South Central Andes: An alternate path to state expansion. *Proceedings of the National Academy of Sciences*, *112*(30), 9202–9209. https://doi.org/10.1073/pnas.1500487112

Greenwald, A. M., Burns, G. R., Eerkens, J. W., & Bartelink, E. J. (2022). A multi-scalar reassessment of dietary isotopic variability in bioarchaeology: Differentiating secular and inter-individual variation. *Journal of Archaeological Science*, *141*, 105592. https://doi.org/10.1016/j.jas.2022.105592

Gunsolley, J. C., Getchell, C., & Chinchilli, V. M. (1995). Small sample characteristics of generalized estimating equations. *Communications in Statistics - Simulation and Computation*, *24*, 869–878. https://doi.org/10.1080/03610919508813280

Hardin, J. W., & Hilbe, J. (2013). *Generalized estimating equations* (2nd ed.). CRC Press.

Hawkey, D. E., & Merbs, C. F. (1995). Activity-induced musculoskeletal stress markers (MSM) and subsistence strategy changes among ancient Hudson Bay Eskimos. *International Journal of Osteoarchaeology*, *14*(4), 7–17. https://doi.org/10.1002/oa.1390050403

Henderson, C. Y., Mariotti, V., Pany-Kucera, D., Villotte, S., & Wilczak, C. A. (2015). The new 'Coimbra method': A biologically appropriate method for recording specific features of fibrocartilaginous entheseal changes. *International Journal of Osteoarchaeology*., (4).

Højsgaard, S., Halekoh, U., & Yan, J. (2005). The R package geepack for generalized estimating equations. *Journal of Statistical Software*, *15*, 1–11.

Janusek, J. W. (2004). *Identity and power in the ancient Andes: Tiwanaku cities through time*. Routledge. https://doi.org/10.4324/9780203324615

Janusek, J. W. (2005). Residential diversity and the rise of complexity in Tiwanaku. In C. Stanish, A. Cohen, & M. Aldenderfer (Eds.), *Advances in Titicaca Basin archaeology 1* (pp. 143–171). Cotsen Institute of Archaeology at UCLA. https://doi.org/10.2307/j.ctvhhhfn9.16

Janusek, J. W. (2008). *Ancient Tiwanaku*. Cambridge University Press.

Janusek, J. W., & Kolata, A. L. (2004). Top-down or bottom-up: Rural settlement and raised field agriculture in the Lake Titicaca Basin, Bolivia. *Journal of Anthropological Archaeology*, *23*(4), 404–430. https://doi.org/10.1016/j.jaa.2004.08.001

Jurmain, R. (1999). *Stories from the skeleton: Behavioral reconstruction in human osteology*. Gordon and Breach.

Jurmain, R., Alves Cardoso, F., Henderson, C., & Villotte, S. (2012). Bioarchaeology's holy grail: The reconstruction of activity. In A. L. Grauer (Ed.), *A companion to paleopathology* (pp. 531–552). Wiley-Blackwell. https://doi.org/10.1002/9781444345940.ch29

Knudson, K. J., Goldstein, P. S., Dahlstedt, A. C., Somerville, A. D., & Schoeninger, M. J. (2014). Paleomobility in the Tiwanaku diaspora: Biogeochemical analyses at Rio Muerto, Moquegua, Peru. *American*

Journal of Physical Anthroplogy, *155*, 405–421. https://doi.org/10.1002/ajpa.22584

Kolata, A. L. (1993). *The Tiwanaku: Portrait of an Andean civilization*. Blackwell.

Kolata, A. L. (2003). Tiwanaku ceremonial architecture and urban organization. In A. Kolata (Ed.), *Tiwanaku and its hinterland: Archaeology and paleoecology of an Andean civilization* (pp. 175–201). Smithsonian Institution Press.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22. https://doi.org/10.1093/biomet/73.1.13

Mariotti, V., Facchini, F., & Belcastro, M. G. (2007). The study of entheses: Proposal of a standardized scoring method for twenty-three entheses of the postcranial skeleton. *Collegium Antropologicum*, *31*, 291–313.

Michopoulou, E., Nikita, E., & Henderson, C. Y. (2017). A test of the effectiveness of the Coimbra method in capturing activity-induced entheseal changes. *International Journal of Osteoarchaeology*, *27*, 409–417. https://doi.org/10.1002/oa.2564

Molnar, P., Ahlstrom, T. P., & Leden, I. (2011). Osteoarthritis and activity—An analysis of the relationship between eburnation, musculoskeletal stress markers (MSM) and age in two Neolithic hunter–gatherer populations from Gotland, Sweden. *International Journal of Osteoarchaeology*, *21*, 283–291. https://doi.org/10.1002/oa.1131

Nials, F. L., Gregory, D. A., & Hill, J. B. (2011). The stream reach concept and the macro-scale study of riverine agriculture in arid and semiarid environments. *Geoarchaeology*, *26*, 724–761. https://doi.org/10.1002/gea.20371

Nikita, E. (2014). The use of generalized linear models and generalized estimating equations in bioarchaeological studies. *American Journal of Physical Anthropology*, *153*, 473–483. https://doi.org/10.1002/ajpa.22448

Rivera, C. (2003). Ch'iji Jawira: A case of ceramic specialization in the Tiwanaku urban periphery. In A. Kolata (Ed.), *Tiwanaku and its hinterland: Archaeology and paleoecology of an Andean civilization, vol 2: Urban and rural archaeology* (pp. 296–315). Smithsonian Institution Press.

Schrader, S. A. (2019). *Activity, diet and social practice: Addressing everyday life in human skeletal remains*. Springer. https://doi.org/10.1007/978-3-030-02544-1

Shaw, H. M., & Benjamin, M. (2007). Structure–function relationships of entheses in relation to mechanical load and exercise. *Scandinavian Journal of Medicine & Science in Sports*, *17*, 303–315. https://doi.org/10.1111/j.1600-0838.2007.00689.x

Somerville, A. D., Goldstein, P. S., Baitzel, S. I., Bruwelheide, K. L., Dahlstedt, A. C., Yzurdiaga, L., Raubenheimer, S., Knudson, K. J., & Schoeninger, M. J. (2015). Diet and gender in the Tiwanaku colonies: Stable isotope analysis of human bone collagen and apatite from Moquegua, Peru. *American Journal of Physical Anthropology*, *158*, 408–422. https://doi.org/10.1002/ajpa.22795

Touloumis, A. (2015). R package multgee: A generalized estimating equations solver for multinomial responses. *Journal of Statistical Software*, *64*, 1–14. https://doi.org/10.18637/jss.v064.i08

Vallières, C. (2012). *A taste of Tiwanaku: Daily life in an ancient Andean urban center as seen through cuisine (unpublished PhD dissertation)*. McGill University.

Vallières, C. (2016). Camelid pastoralism at ancient Tiwanaku: Urban provisioning in the highlands of Bolivia. In J. Capriles & N. Tripcevich (Eds.), *The archaeology of Andean pastoralism*. University of New Mexico Press.

Villotte, S., Castex, D., Couallier, V., Dutour, O., Knüsel, C. J., & Henry-Gambier, D. (2010). Enthesopathies as occupational stress markers: Evidence from the upper limb. *American Journal of Physical Anthropology*, *142*, 224–234. https://doi.org/10.1002/ajpa.21217

Villotte, S., & Knüsel, C. J. (2013). Understanding entheseal changes: Definition and life course changes. *International Journal of Osteoarchaeology*, *23*, 135–146. https://doi.org/10.1002/oa.2289

Yan, J. (2002). Geepack: Yet another package for generalized estimating equations R-news. *The R Journal*, *2*(3), 12–14.

Yan, J., & Fine, J. (2004). Estimating equations for association structures. *Statistics in Medicine*, *23*, 859–874. https://doi.org/10.1002/sim.1650

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Becker, S. K., Stull, K. E., & Chu, E. (2023). Designing databases and using generalized estimating equations for entheseal datasets: An example from the Tiwanaku culture (AD 500–1100). *International Journal of Osteoarchaeology*, 1–15. https://doi.org/10.1002/oa.3210