Journal of Archaeological Science: Reports 11 (2017) 750–761

Contents lists available at ScienceDirect

Journal of Archaeological Science: Reports

journal homepage: www.elsevier.com/locate/jasrep

Assessing reproducibility in faunal analysis using blind tests: A case study from northwestern North America

Reno Nims *, Virginia L. Butler

Portland State University, Department of Anthropology, 1721 SW Broadway, Portland, OR 97201, United States

ARTICLE INFO

Article history:
Received 16 August 2016
Received in revised form 25 December 2016
Accepted 7 January 2017
Available online xxxx

Keywords:
Blind reanalysis
Data quality control
Reproducibility
Zooarchaeology
Agreement index
Protocol drift

ABSTRACT

Zooarchaeologists have long recognized that assigning taxonomic identifications to animal remains is a subjective process, and recent studies have highlighted the need for data quality assurance standards in archaeofaunal research. Our study contributes to this growing interest in quality assurance by presenting simple quantitative methods for assessing the reproducibility of analytic results through blind reanalysis of animal remains that we developed during analysis of fishbone from Čḯxʷicən, a large Native American village on the coast of Washington State, U.S.A. Given the large scale of the Čḯxʷicən project – over 112,000 fish remains were documented by five different analysts over three years – there was a real possibility that inconsistencies in laboratory practices affected analytic results (e.g., number of identified specimens, taxa present and relative abundance, elements identified, burning frequencies). To evaluate the reproducibility of the Čḯxʷicən fishbone data, and the possibility of "protocol drift" – changes in how specimens and bone modifications were identified over the course of analysis – we reanalyzed samples of fish remains that were previously documented during three discrete stages (beginning, middle, and end) of the Čḯxʷicən project. The original data and the reanalysis results show close agreement in each stage, with only minor differences in the numbers of recorded specimens, taxonomic representation at family- and finer taxonomic levels, and body part representation assessed for a single taxonomic order. Identifying burning on bone was not very reproducible. Reproducibility studies such as this are useful for highlighting ambiguous identification criteria (e.g., in taxonomic assignment, bone modification), and could stimulate dialog among researchers about ways to address such issues in future studies. Increasing the implementation of these, and other, widely applicable methods should improve zooarchaeological data quality and stimulate further research on quality assurance in archaeology overall.

1. Introduction

Zooarchaeologists have long recognized that assigning taxonomic identifications to animal remains is a subjective process (Driver, 2011; Gobalet, 2001, 2017; Lawrence, 1973; Wolverton, 2013). In one particularly striking demonstration of this fact, Gobalet (2001) and three other ichthyoarchaeologists separately analyzed the same, coastal California fishbone assemblage and found they had major disagreements regarding the number of species, number of identifiable specimens (NISP), and taxonomic identity of specific specimens represented in the collection. For example, one analyst identified 18 species in the assemblage, two researchers identified half that number, and one researcher only identified four species (Gobalet, 2001, p. 378). Such disagreements might be related to differences in the educational background and experience of individual analysts, analysts' familiarity with the biological community linked to the site's context (Gobalet, 2001, p. 384), or yet other factors. Overall the study highlights the pressing need for quality assurance practices in zooarchaeology.

Building on examples from environmental chemistry, Wolverton (2013) has provided a framework for supporting *quality assurance* in zooarchaeology, which includes "activities intended to ensure that the analytical information produced meets the quality requisites…[of the field] in terms of accuracy and representativeness" (Pérez-Bendito and Rubio, 1999:39, *in* Wolverton, 2013, p. 383–384). Quality assurance can be broken into two components: *quality control* (QC) and *quality assessment* (QA). QC is "defined as the specific set of activities intended to examine both the analytical process and its results in terms of quality" (Pérez-Bendito and Rubio, 1999, p. 40), and includes continuous checking of laboratory procedures and organization, equipment, and chain of custody. QA is the set of operations researchers follow to evaluate QC and whether actions were performed correctly through practices such as audits or analyzing blind samples (Pérez-Bendito and Rubio,

* Corresponding author.
E-mail address: rnims1@gmail.com (R. Nims).

http://dx.doi.org/10.1016/j.jasrep.2017.01.012

1999, p. 41).[1] Wolverton (2013) provides numerous zooarchaeological examples of both QC and QA practices, which, he emphasizes, are necessary to the long-term value of our data in archaeology and other research domains. As we seek to apply zooarchaeological research to conservation biology and policy debates, the need for rigorous analytic methods and documentation aligned with quality assurance increases in kind.

Our study contributes to the growing interest in quality assurance in zooarchaeology (Fisher, 2015, p. e.g.; Morin et al., 2016; Twiss et al., 2016) by introducing simple quantitative QA methods for assessing reproducibility in faunal analysis through reanalysis that we developed during study of fish remains from Čḯxʷicən,[2] a large Native American village from coastal Washington State, U.S.A. (Fig. 1) that was excavated extensively in 2004. The fishbone study is part of a larger analysis of mammal, bird, fish and shellfish remains that aims to understand how animal resources and in turn people were affected by great earthquakes and other environmental forces, such as climate change and local bay development, over the past ~2200 years. Given the large scale of the Čḯxʷicən project – over 112,000 fish remains were documented by five different analysts over three years – possible inconsistencies in laboratory practices may have affected analytic results (e.g., number of identified specimens, taxa present and relative abundance, elements identified, burning frequencies). We were concerned that temporal and spatial patterns documented in the fishbone data could reflect analytic error, rather than past human activity.

Here we describe several simple quantitative methods for assessing reproducibility in faunal analysis that we applied in a ‘blind’ reanalysis of fishbone samples which were previously documented during three discrete stages of the Čḯxʷicən project. Though we focus on fish remains in particular and zooarchaeology generally, the methods we outline could also be applied to archaeological analysis broadly. In addition to assessing the reproducibility of our data, we evaluate whether increasing knowledge of, and experience with, the ichthyofaunal materials created ‘protocol drift’ – changes in how specimens and human modifications to bone were identified over the course of analysis. Reproducibility studies such as ours also highlight ambiguous criteria (e.g., in taxonomic assignment, bone modification), and help stimulate dialog among researchers about ways to address such issues in future studies. Finally, it is our hope that increased focus on widely applicable QA methods such as these will help improve zooarchaeological data quality and stimulate further research on quality assurance in archaeology overall (e.g., Banning et al., 2016; Beck and Jones, 1989; Evans et al., 2014; Fish, 1978).

2. Materials and methods

The Čḯxʷicən fishbone assemblage comes from a Lower Elwha Klallam Tribe (LEKT) village in Port Angeles, WA at the base of Ediz Hook on the south shore of the Strait of Juan de Fuca (Fig. 1) that was occupied for the past 2800 years (Larson, 2006). In 2004, Larson Anthropological Archaeological Services (LAAS) and LEKT members excavated 261.4 m^3 of the site's volume (estimated at 6900 m^3) in large open blocks of 1 × 1 m units by natural stratigraphic layers that were subdivided into 10 cm arbitrary levels (Reetz et al., 2006). All excavated matrix was collected in 10 L buckets and wet-screened with nested 1″ (25.6 mm), 1/2″ (12.8 mm), and 1/4″ (6.4 mm) mesh. Every twentieth bucket from a given stratum was also screened to 1/8″ (3.2 mm) mesh and designated a “complete” (“C”) bucket (Reetz et al., 2006). All faunal remains from “C” buckets were sorted into four main animal groups (fish, bird, mammal, shellfish), but for various reasons they were not studied immediately. In 2011, Kristine Bovy (University of Rhode Island), Virginia Butler (Portland State University [PSU]), Sarah Campbell (Western Washington University), Michael Etnier (Western Washington University), and Sarah Sterling (PSU) initiated a large-scale zooarchaeological and geoarchaeological analysis of three structures at Čḯxʷicən.

2.1. Original analysis methods

Butler directed analysis of the fishbone assemblage at PSU between 2012 and 2015 following QC/QA protocols that were consistent with Driver's (2011) recommendations (see also Wolverton, 2013): the universe of possible fish taxa was established at the beginning of the project using Strait of Juan de Fuca fisheries survey records (Miller et al., 1980) and north Pacific field guides (Hart, 1973);[3] identification criteria were specified and referred to over the course of the project; difficult to distinguish taxa and elements were specified; and a descriptive summary, a written document outlining the rules, protocols, and criteria used for assigning skeletal elements to taxon, is being prepared (Butler, *in prep.*). PSU Master's students Kathryn Mohlenhoff, Anthony Hofkamp, Shoshana Rosenberg, and Nims identified fish remains using the PSU Anthropology Department comparative collection, which was supplemented with fish skeletons loaned from the personal collections of Bob Kopperl (Willamette Cultural Resources Associates) and Ross Smith (University of Oregon) (Supplementary Table 1). Collections at the University of Victoria, Department of Anthropology, were also consulted with help from Rebecca Wigen (Supplementary Table 1). Butler verified, and often adjusted, all initial identifications under magnification (2×–40× power) with a binocular loupe or microscope.

For each specimen, we documented the finest possible taxon and skeletal element, whether the specimen included a morphological landmark, and whether the specimen was burned. For many specimens, it was difficult to determine whether a dark color reflected exposure to heat, flecks of charcoal attached to the surface of a specimen, or staining from the absorption of minerals found in the surrounding matrix. The decision to call a specimen burned was based on conservative criteria: only those specimens which were at least partially black (charred) or white-to-bluish (calcined) in color were recorded as burned.

Given the large sample size and the small size of many specimens, it was impractical for each specimen to receive a unique catalog number or separate bag. Instead, we tallied and recorded the number of specimens from a given provenience and mesh size that shared the same attributes (e.g., taxon, skeletal element, burned) together. All fish remains from a given context with similar attributes were bagged together, with acid-free labels summarizing analytic information. Specimens that could not be identified any more specifically than as fish were recorded as unidentified, with one exception. The category ‘non-salmonid’ was used to tally unidentifiable vertebra fragments that were definitively not from salmonids in an effort to counter the potential over-documentation of salmon specimens, which can be easily identified to genus from very small vertebra fragments (e.g. Casteel, 1976; Ewonus, 2011; Grier and Lukowski, 2012). The primary counting units used for fish faunal analysis were NISP and number of specimens (NSP), which includes all remains designated as fish, including identified fish, unidentified fish, and non-salmonid specimens. In addition, we counted the minimum number of elements (MNE) using the number of specimens that bear morphological landmarks. Specimens identified as mammal,

[1] The concepts of validity and reproducibility/reliability/replicability are central to research seeking to improve data quality. In simple terms, *validity* is the likelihood that our measure measures what it purports to measure. For example, in zooarchaeology we ask, is Bone Specimen X that we designate as Taxon Y, *truly from* Taxon Y? *Reproducibility* (or replicability/reliability) concerns whether one (or several researchers) would assign Bone Specimen X to Taxon Y in repeated trials. Quality assurance plans that incorporate explicit laboratory procedures (QC) and directly address whether laboratory practices are followed (QA), have the effect of increasing both validity and reproducibility, which are at the core of data quality.

[2] An alternative spelling for the site name, Tse-whit-zen, has been used in previous reports and publications. The Klallam language spelling, Čḯxʷicən, is preferred by the Lower Elwha Klallam Tribe. The site name is pronounced *ch-WHEET-son*.

[3] We have since updated the taxonomic designations in these sources following Pietsch and Orr (2015).