



# A Critical Review of Four Efforts to Resurrect MNI in Zooarchaeology

R. Lee Lyman<sup>1</sup> 

Published online: 31 January 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

**Abstract** Evaluation of zooarchaeology’s quantitative units known as NISP (number of identified specimens) and MNI (minimum number of individuals) during the last three decades of the twentieth century suggested neither provided ratio scale measures of taxonomic abundances. Many researchers at that time began to use NISP as often as MNI to measure taxonomic abundances. In part because of a desire to determine the composition of human diet more precisely, and in light of the fact that different taxa have different numbers of identifiable bones, four zooarchaeologists have, since 1990, used assemblages with known ANI (actual number of individuals) to evaluate whether NISP or MNI provides the most accurate measure of ANI. ANI data from ethnoarchaeological and historical contexts suggest taxonomic abundance data quantified as NISP or MNI are ordinal scale at best, something previously shown to be highly probable. Experimental data used to evaluate the accuracy of NISP and MNI as measures of ANI are either inappropriate or not designed to assess which quantitative unit produces the most accurate measure. A new quantitative unit proposed as an alternative to MNE (minimum number of [skeletal] elements)—the NDE (number of distinct elements)—is said to provide proportional abundances of taxa but demonstrably undercounts skeletal parts and fails to provide ratio scale abundance data.

**Keywords** Experimental zooarchaeology · Minimum number of individuals · Number of identified specimens · Quantitative units · Taxonomic abundances · Zooarchaeology

## Introduction

Zooarchaeology—the study of faunal remains recovered from archaeological sites—became an important part of archaeological research during the second half of the

---

✉ R. Lee Lyman  
lymanr@missouri.edu

<sup>1</sup> Department of Anthropology, University of Missouri, 112 Swallow Hall, Columbia, MO 65211, USA

twentieth century (Allen and Nagaoka 2004; Carlson 1999; Horwitz 2002; Lyman 2016; Reitz 1993; Robison 1987; Stewart 2002). Along with important substantive contributions, analytical methods were developed and textbooks were published (Chaplin 1971; Hesse and Wapnish 1985; Klein and Cruz-Uribe 1984; Reitz and Wing 2008). Perhaps not surprisingly because it coincided with growing concern about how the archaeological record was formed and how it may be biased (e.g., Schiffer 1987), taphonomic awareness emerged in zooarchaeology during the last three decades of the twentieth century (e.g., Gifford 1981; Lyman 1994c). At the same time, the quantitative or counting units of zooarchaeology were evaluated (e.g., Casteel 1977a, b; Grayson 1973, 1978, 1979, 1984). The minimum number of individuals (MNI) quantitative unit had been used by zooarchaeologists fairly regularly, though not exclusively, throughout the twentieth century, but its frequency of use increased from the first half of the twentieth century through the 1970s (Lyman 2018a, b). Some of the critical literature suggested the number of identified specimens (NISP) was in some ways a better quantitative unit for tallying taxonomic abundances (e.g., Grayson 1979, 1984), and it gained popularity in the 1980s and 1990s, coming to be used as often as MNI by about 1980 (Lyman 2015a, 2018a, b).

MNI continued to be used frequently during the last several decades of the twentieth century, and this was at least in part the result of perceptions by researchers that NISP was “less meaningful” than MNI (Jochim 1976:167), that MNI was “a much more reliable indicator of the relative importance of various species in the diet than the [NISP]” (Smith 1975: p. 282), and that MNI “is preferable to NISP for comparing the relative representation of the different species present” (Payne 1975: p. 14). But the simultaneous lengthy discussion by multiple commentators on the strengths and weaknesses of MNI as a quantitative unit for measuring taxonomic abundances (e.g., Allen and Guy 1984; Casteel 1977a, b; Clason 1972; Fieller and Turner 1982; Gautier 1984; Grayson 1973, 1978, 1979; Medlock 1976; Nichol and Creak 1979; Perkins 1973; Plug and Plug 1990; Ringrose 1993; Wild and Nichol 1983) contributed to MNI becoming less favored than NISP late in the twentieth century. Grayson and Frey (2004:28), for instance, stated that taxonomic abundances were, early in the twenty-first century, “most often quantified in terms of numbers of identified specimens (NISP),” though they provided no data to substantiate their assertion. Study of the North American zooarchaeology literature for the entire twentieth century indicates both units were used but NISP was employed more frequently than MNI at the end of that century. Both units were often used in the same analyses but each unit answered distinct research questions (Lyman 2018b). Nevertheless, during the 1990s and first several years of the twenty-first century, several authors argued that MNI is a more accurate measure of taxonomic abundances than NISP and sought to revive the former unit’s popularity.

In this paper, I critically review four attempts to resurrect MNI, and seek to understand why these efforts were put forward. I initially suspected it was because these researchers are anthropological archaeologists with deep interests in the composition of hominin diet who hoped to measure each consumed species’ contribution to diet at a ratio scale, rather than merely whether one taxon was more frequent than another. As it turns out, my suspicion is only partially correct. At least two of the researchers seek absolute, ratio scale taxonomic abundance measures.

Because the quantification of zooarchaeological faunal remains provides data upon which numerous and diverse analyses and interpretations rest, it is critically important

to be aware of the strengths and weaknesses of the units used to quantify those remains. By the same token, when new arguments are presented suggesting one unit is somehow “better” than another, those arguments should be subjected to critical evaluation. In this paper, I provide just such an evaluation of the four attempts made since 1990 to demonstrate that MNI is a better measure of taxonomic abundances than NISP. I begin with a review of statistical concepts and aspects of NISP and MNI pertinent to their use as quantitative units aimed at measuring taxonomic abundances. This is followed by a description of the methods used in the critical review. I then turn to the review itself, considering each of the four studies in chronological order of their appearance in the literature.

## **Critical Concepts and Zooarchaeological Quantitative Units**

Because explicit meanings of important terms used in some of the literature discussed below are necessary to understanding, this section presents a discussion of fundamental mathematical and statistical concepts relevant to evaluation of the efforts to resurrect MNI. These concepts are equally critical to full understanding of the NISP and MNI quantitative units as measures of (taxonomic) abundance. In this section, I first introduce and define key mathematical and statistical concepts, and then turn to a summary of earlier observations about NISP and MNI to provide a context for evaluation of the attempts to resurrect MNI.

### **Critical Math and Statistical Concepts**

Statisticians recognize four scales of measurement. Nominal scales register differences in kind. Ordinal scales register differences in rank but not magnitude, such as long, longer, longest, or greater and lesser. Interval scales measure both differences in rank and in magnitude, such as 6 cm long and 3 cm long, or 16 cm long and 13 cm long. The ranks (from least to greatest) of the four length measurements in the order listed are 2, 1, 4, and 3, with 3 cm ranked 1 and 16 cm ranked 4. Importantly, the difference between 6 and 3 cm is the same as the difference between 16 and 13 cm, and the difference between 6 and 16 cm is the same as the difference between 3 and 13 cm. This is a characteristic of interval (and ratio) scales. Ratio scales are the same as interval scales but include a natural zero (Stevens 1946; Zar 1996).

An absolute frequency is a numerical count or tally of the specimens of (usually a kind of) a thing. Ten turkey bones, 15 rabbit bones, and 25 deer bones are absolute frequencies of specimens of three specific kinds of things (species), all of which are the same general kind of thing (bones). Were we to say with respect to this collection that 20% of the identified remains represent turkey, 30% represent rabbit, and 50% represent deer, we would be speaking of the relative abundances of the species at a ratio scale. If the collection comprised only the 25 deer bones, we could say that 100% of the collection was deer. An absolute frequency of something depends solely on the abundance of that something; a relative abundance of something depends both on the abundance of that something and the abundance(s) of all other things with which the abundance of that something is compared. The relative frequency of deer in a collection of 25 deer bones (50%), 15 rabbit bones, and 10 turkey bones would be the same in a

collection of 250 deer bones, 150 rabbit bones, and 100 turkey bones despite the order of magnitude difference in absolute abundances of bones in the two collections. This should make it clear that it can be difficult to compare and detect differences between collections of different absolute sizes; relative abundances eliminate that comparative difficulty by scaling all collections to, say, 100%. Relative abundances recorded as percentages are ratio scale; ordinal scale relative abundances are measured by ranks. In the two collections just described, from least to most abundant, ranks would be turkey rank 1, rabbit rank 2, and deer rank 3. Note that we do not know based on ranks whether rabbits are twice as abundant as turkeys, nor do we know if the differences in absolute abundances between turkey and rabbit, and between rabbit and deer are the same or different. This is a characteristic of ordinal scales of measurement.

Accuracy refers to “the nearness of a measurement to the actual value of the variable being measured” (Zar 1996:5). An inaccurate measure is one that is far from the true value. Accuracy depends in part on the measuring device or tool and its resolution. Exactly accurate measurements may not be necessary to be satisfactory; the elevation of a mountain to the nearest centimeter is unnecessary when to the nearest meter will suffice. “In general, as resolution increases [say, from a length measurement to the nearest 1.0 mm to the nearest 0.1 mm to the nearest 0.01 mm], precision decreases while accuracy increases, up to the limits of the measuring tool” (Lyman and VanPool 2009: p. 487). Accurate measurements should be replicable, but that is not all they need to be. Precision concerns the closeness of repeated measurements of the same phenomenon; precise measurements do not differ greatly from one another and will be consistent or reliable (replicable)—produce the same or nearly the same value—among measurement sessions (Zar 1996). Precise measurements need not be accurate; replicability is a necessary condition of accurate measurements but not a sufficient condition thereof.

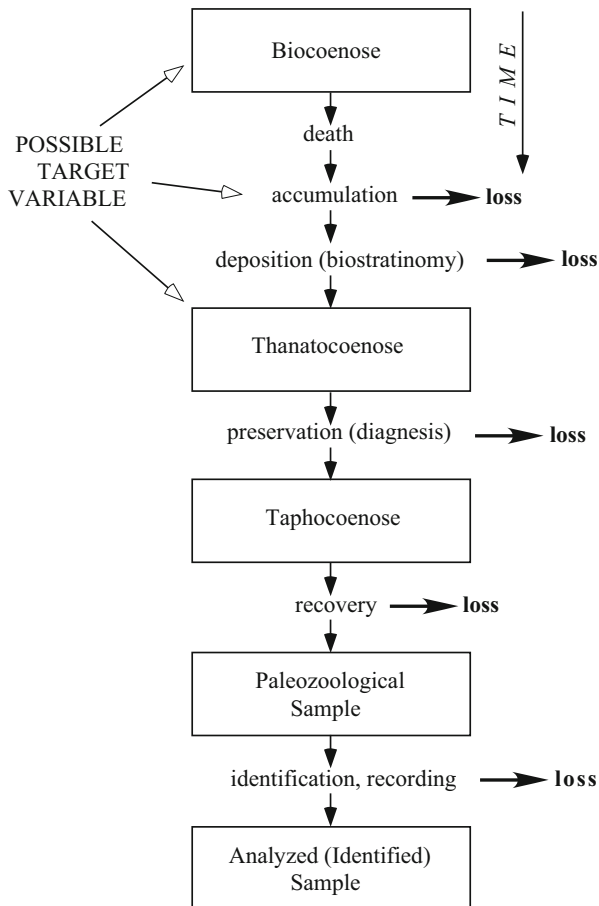
Simplistically, validity concerns the relationship of the variable we are measuring and its empirical manifestations we wish to measure. The maximum dimension of some three-dimensional object is the variable we might measure; this variable is typically labeled length, a concept concerning size, specifically, the maximum linear dimension of a thing. Measuring the length of an object produces a valid measure of size; another valid measure of size, given how we might conceive size, is weight. The color of the object we are focused on is not a valid measure of size insofar as variability in color does not correlate significantly with variability in size. If we measure the length of the object multiple times, and consistently come up with the same values (or nearly so), then the measurement technique is reliable (produces replicate measurements), and it is also likely to be accurate, though the length values would not be a valid measure of the color of the object. If an object’s size is of analytical interest, one might choose length, or weight, or width, or some other variable for measurement that registers the magnitude of the object. The target variable of size would be recorded using one or more measured variables of length, weight, *etc.* Were one to measure the variable color in an effort to record the target variable size, even a very accurate measure would be invalid.

The target variable chosen should depend in large part on the research question asked or hypothesis being tested. The target variable might be the abundances of taxa in the identified assemblage of fossils, the abundances of taxa in the deposit that produced the fossils (the taphocoenose), the abundances of taxa that died in the area (the thanatocoenose), or the abundances of taxa on the landscape at the time when the

fossil assemblage was accumulated and deposited (the biocoenose). The nature of the relationship between the identified sample and any of the possible target variables can be estimated by detailed taphonomic analysis (Fig. 1). In order to keep things relatively simple for purposes of discussion, I assume the relationship is 1:1, regardless of specific target variable.

### Zooarchaeological Quantitative Units

When zooarchaeologists seek to determine the abundances of taxa represented by a collection of animal remains, those abundances are the *target variable*. Researchers hope to estimate as closely as possible the *actual number of individuals* (ANI) represented by a collection of fossils. NISP and MNI per taxon are typical *measured variables* that provide more or less accurate estimates of the ANI. Which measured variable is more accurate when ANI is the target variable is the key concern of the resurrectionists. The measured variables are those the values of which a



**Fig. 1** Schematic outline of the taphonomic history from the time of a living fauna until remains are identified to taxon. Notice that the target variable can vary, but regardless of the target variable chosen, it will be several taphonomic stages removed from the measured variable within the identified assemblage of remains

zooarchaeologist actually determines for any given assemblage of faunal remains. In Table 1, the ANI, MNI, and NISP are listed for a fictional yet realistic paleozoological sample. This fictional data set is constructed to highlight a couple well-known things about NISP and MNI. First, NISP values provide estimates of (almost always) larger magnitude than MNI and ANI; MNI values provide estimates (almost always) smaller than NISP and ANI. Second, inter-taxonomic ratios of MNI values tend to be different than ratios of corresponding NISP values. And third, NISP and MNI tend to be but are not necessarily ordinal scale measures of taxonomic abundances (Grayson 1984; Lyman 2008).

Grayson (1979: pp. 200, 204; 1984: p. 30) observed paleozoologists rarely treat measured abundances of taxa as *absolute* values but rather consider them to reflect the *ordinal scale relative* abundances of taxa (see also Gifford-Gonzalez and Hildebrandt 2012). Paleozoologists generally seek to know if, say, deer were more abundant than turkeys at one time (or at one site) but turkeys were more abundant than deer at another time (or another site). They might hope to know that in the first instance deer numbered (ANI =) 40 and turkeys 20, whereas in the second instance deer numbered 15 and turkeys 30. Such knowledge is, however, generally perceived to be analytically unattainable; ratios of taxonomic abundances might be calculated but are likely inaccurate to some unknown degree and are best interpreted in ordinal scale terms.

Using NISP values as ratio scale measures of taxonomic abundances requires the assumption that each specimen of a species is independent of every other specimen of that species; each specimen is assumed to be from a different individual animal. It is likely, however, NISP tallies count the same animal(s) multiple times (specimens in an assemblage are to some unknown degree interdependent). There are many reasons to be leery of NISP as a measure of taxonomic abundances. Different taxa have different numbers of *identifiable* bones (e.g., Lyman 2015b). Further, the remains of different taxa may be differentially fragmented by any number of taphonomic processes and agents. Mid-levels of intensity of fragmentation—the number of pieces into which a particular skeletal element is broken—will produce greater NISP per individual animal; high levels of fragmentation will reduce the NISP because diagnostic anatomical landmarks will be missing or destroyed as ever smaller pieces progressively become less identifiable (Cannon 2013). This means the relationship between NISP and the actual number of individuals (ANI) represented in a collection is murky at best and unknown at worst (Fig. 1).

MNI, defined fundamentally as the most abundant skeletal part (e.g., left distal humerus) of a species in an assemblage, overcomes the interdependence problem of NISP because MNI is defined to produce values that are independent of one another. And although it can provide just such values, MNI requires the assumption that each aggregate or assemblage of faunal remains is independent of every other aggregate. If

**Table 1** A set of fictional but realistic zooarchaeological data

Taxon	ANI	MNI	NISP
Deer	10	8	57
Jackrabbit	5	5	43
Turkey	5	3	32

specimens of one or more individual animals are distributed across two or more aggregates, those aggregates are not independent of one another (Grayson 1979, 1984).

MNI per taxon within an aggregate can differ considerably from the frequency of the most abundant skeletal part if the analyst takes into account ontogenetic age differences, individual size differences, sex differences, matching lefts with rights, and other variables pertaining to determination of whether multiple specimens represent one individual animal (are interdependent) or several individuals (are independent). There are no standard rules for how MNI is determined (*e.g.*, are ontogenetic differences accounted for?), and few analysts describe their protocol in detail. NISP per taxon within an aggregate can differ from a simple tally if the researcher decides to not count each discrete specimen in, say, an articulated limb, or to count each individual tooth in a mandible as well as the mandible. If such analytical decisions are made consistently across all aggregates, this problem becomes less significant than some others. However, there are no universally agreed upon standards for tallying specimens making up an articulated limb or teeth in a mandible, and few analysts describe the protocol they follow.

Site total NISP values per taxon are not influenced by consideration of the ontogenetic age or sex of the animals that contributed the bones, or the size of the bones, or anything else (including aggregation), other than the decision about how to count articulated specimens (*e.g.*, teeth in mandibles, articulated humerus and radius). MNI is influenced by all of these variables. Further, although matching left and right specimens of the same skeletal element (*e.g.*, humeri) can sometimes be done with confidence (*e.g.*, Todd 1987; Todd and Frison 1992), it cannot always be done confidently (*e.g.*, Lyman 2006). Additionally, we lack morphometric techniques for determining if, say, a left tibia and a left (or right) femur represent the same individual, a femur and a humerus represent the same individual, *etc.* This, plus the analytical decisions regarding whether to take into account age, sex, and size differences, underscores that MNI is a *derived* measure. This in turn does not mean MNI is inherently a less accurate measure of ANI and NISP is a more accurate measure (regardless of scale). Rather, it means there are more reasons for site total MNI values to fluctuate as a result of variability in zooarchaeological analytical protocols than there are reasons for site total NISP values to fluctuate (ignoring for the moment inter-analyst variability in identification skills). A quantitative unit that measures analytical protocol rather than the variable of interest is one we should hesitate to use (*e.g.*, Marean and Frey 1997).

Grayson (1979, 1984) observed that when multiple aggregates are under analysis, the sum of the NISP per aggregate per taxon does not change if aggregate horizontal and vertical spatial boundaries are changed. This is because NISP is a tally of specimens, regardless of a specimen's assignment to one or the other aggregate. Not so with MNI. Recall that MNI is fundamentally defined as the most abundant skeletal part of a species in an aggregate (*e.g.*, left distal humerus). Grayson made a strong case that as aggregate definitions (horizontal and/or vertical boundaries) change, it is likely that the identity of the most abundant skeletal parts in a series of aggregates will change from aggregate to aggregate. When comparing aggregates we might find the most abundant element for species A is the left distal humerus in stratum I and the right distal tibia in stratum II. If we subdivide stratum I into strata IA and IB, we might find the most abundant element for species A is the left distal humerus in IA and left lower third molar in IB. Grayson reasoned further that MNI values for one species will likely

fluctuate from aggregate to aggregate differently than for other species. As a result, differently defined aggregates will result in varied site total MNI values for many if not all species. As a result, MNI values provide measures of both taxonomic abundances and how faunal remains have been aggregated, and Grayson (1979, 1984) argued such a quantitative unit is undesirable. I return briefly to defining (zoo)archaeological aggregates below.

In real-world paleozoological cases, the relationship of MNI to ANI per taxon per aggregate is unknown. Because MNI is a minimum (it is probable that it is less than ANI), the difference between 20 MNI for species A and 30 MNI for species B might be 10, but it might be some other value, say 7 or 15. And it is unlikely that the observed difference of 10 MNI between species A and B is the same as that between an MNI of 10 for species C and 20 for species D. Thus, MNI can hardly be more than ordinal scale. Similarly, if each identified specimen is independent of every other specimen, then NISP would be the maximum number of individuals. In reality, it is probable that ANI is less than NISP as a result of the interdependence of some specimens. How much less is, however, archaeologically unknowable for any given species. Thus, for reasons similar to those indicating MNI (the minimum-value site total measure of taxonomic abundances) will likely at best be ordinal scale, NISP (the maximum-value site total measure of taxonomic abundances) will likely at best be ordinal scale (Grayson 1979, 1984).

Finally, MNI values per species can often be closely predicted by the NISP values for species in an assemblage (Bobrowsky 1982; Casteel 1977b; Grayson 1979, 1984; Hesse 1982). This, together with the aggregation problem (see above), suggested there was little reason to determine MNI values if one seeks a measure of taxonomic abundances. Grayson (1984: p. 24) hastened to add “no one would argue that measures of taxonomic frequency based on [NISP] can be used as measures of all other variables in which a faunal analyst might be interested.” Explicit definition of a specific target variable is mandatory to making a well-informed choice of a quantitative unit. Equally mandatory is recognizing whether one seeks absolute or relative abundances of species. Given what has been presented in this section (and in more detail in the references cited), it seems overly optimistic to think absolute (ratio scale) abundances of species can be obtained from zooarchaeological remains, regardless of the quantitative unit used. And they may not provide even ordinal scale abundance data.

In the fictional instance in Table 1, neither NISP or MNI presents ratio scale resolution of ANI. And while both NISP and MNI come close to an ordinal scale measure (deer are rank 1, jackrabbits rank 2, and turkey ranks 3 for both quantitative units), neither unit measures the true ordinal scale abundances given the tied ranks of the ANIs of jackrabbits and turkeys. One might say, then, that NISP and MNI provide *estimates* of taxonomic abundances, where an estimate is a measurement based on incomplete data, such as judging how tall a person is without using a tape measure. A *measurement*, in contrast, involves assigning a numerical value to an observation based on rules governing the assignment and using pertinent data. The rules may include the statement that length is measured in linear units of uniform size (e.g., centimeters), and using a tape measure marked off in centimeters to determine (measure) a person's height.

For various reasons, most of which relate to taphonomy, in the late 1970s and early 1980s another quantitative unit emerged in zooarchaeology (see Lyman (2008: p. 215–



218) for historical details). It is known today as the minimum number of (skeletal) elements, or MNE. This label is something of a misnomer because this quantitative unit has been used by both paleontologists (*e.g.*, Brain 1969, 1981; Voorhies 1969) and zooarchaeologists (*e.g.*, Binford 1978, 1981, 1984; Bunn 1982; Lyman 1984, 1985, 1994a) to denote the minimum number of (anatomically incomplete) skeletal parts such as proximal femora, (anatomically complete) skeletal elements such as astragali, or skeletal portions (anatomically comprising multiple discrete skeletal elements) such as the thoracic section of the vertebral column. MNE is less inclusive skeletally than MNI, but it is determined in a manner similar to MNI: the maximum number of any given part of, for instance, the proximal femur—greater trochanter, articular head, trochanteric fossa, lesser trochanter—is the MNE.

Analysts generally do not consider age, sex, and size variability when determining MNE. Similarly, analysts seldom worry about how many left or right MNEs there are for, say, distal humeri. Analysts ignore these variables because MNE is designed to measure a different target variable than MNI or NISP. MNE's target variable is skeletal completeness or representation—for instance, are many upper forelimbs missing whereas distal hindlimbs are relatively abundant?—whereas MNI's target variable is usually taxonomic abundances—deer outnumber rabbits (Lyman 1994b, c, 2008). MNE values for a taxon are often quite similar to the MNI values per skeletal part for a taxon (*e.g.*, Lyman 1994a; Grayson and Frey 2004), for reasons that should be obvious. The two variables are not always identical in value per skeletal part because one might ignore left and right distinctions and simply divide the MNE of distal humeri by two to approximate the MNI of skeletal parts; MNI incorporates left and right side into operationalizing MNI as the most abundant skeletal element.

### Aggregate Definition

Some have argued aggregate definition is a non-issue (*e.g.*, Domínguez-Rodrigo 2012), that stratigraphic boundaries are obvious and clearly demarcate the limits of an aggregate of faunal remains and as such are the clear choice for aggregate definition. This ignores the issue of how faunal remains from storage pits, horizontally discrete occupation surfaces, and any number of other sorts of potentially interdependent (*e.g.*, behaviorally related) archaeological features are to be treated if they all originate in a culturally homogeneous stratum. Whether within-stratum feature-associated faunal remains are aggregated or not will likely influence the site total MNI per taxon (Grayson 1979, 1984). This is so because different aggregates of remains typically have different most abundant skeletal elements.

There are two additional issues concerning the specification of aggregates that are seldom acknowledged by zooarchaeologists. The aggregates specified should depend on the research questions asked, and there has been minimal discussion of this in the archaeological literature in general, let alone the zooarchaeological literature. The few instances of this of which I am aware concern identifying site horizontal boundaries (*e.g.*, Dunnell and Dancy 1983; Foley 1981; Thomas 1975). The second issue is that the analyst must assume the contents of each of the aggregates are independent of one another. That is, one must assume the identified specimens of a species in one aggregate are independent of—come from different individuals than—the identified specimens of that species in all other aggregates. The few efforts made to evaluate this assumption in

the empirical realm indicate it is not necessarily the case (Audouze and Enloe 1997; Enloe 2003; Enloe and David 1992; Rosell et al. 2012; Todd and Frison 1992; Todd and Stanford 1992).

As noted, one might argue using the “natural boundaries” of strata—distinct depositional units—resolves these and related issues, but as archaeologist James Ford (1962: p. 45) noted many years ago, “separation of [archaeological] materials only by the observable breaks in the deposit would be an archaeological variety of cataclysmic geology. By this procedure, we have allowed the history to be separated into periods by chance historical [depositional] events.” This opinion was echoed a decade later (Plog 1973, 1974) and has evolved into concerns over so-called palimpsest assemblages and resulted in what has become known as time perspectivism. The latter involves phrasing our research questions in terms commensurate with the temporal resolution of that portion of the archaeological record we are studying (e.g., Bailey 2007; Holdaway and Wandsnider 2008; see Bennington and Aronson (2012) for a discussion in paleozoology).

On one hand, as Grayson (1984) noted, we likely will not want to give up the advantages that using stratigraphic boundaries brings to archaeological analysis. On the other hand, in retaining those boundaries, we should be cognizant of how they might influence our results. Few studies have explored how shifting aggregate spatial boundaries might influence the total MNI for a site or taxonomic abundances (e.g., Adams 1949; Grayson 1979, 1984; see also Pollock and Ray 1957). Additional research is necessary to thoroughly ascertain not only the influence of aggregation on MNI values but to evaluate Grayson’s (1979, 1984) surmise the most abundant elements of multiple taxa will seldom have similar distributions across aggregates and to examine the degree of interdependence of specimens within assemblages (see Gautier (1984) for an intriguing and typically overlooked consideration of interdependence). I know of no efforts to assess these latter two fundamental variables. Instead, there has been work done that seeks to assess whether NISP or MNI is a more accurate measure of taxonomic abundances.

## Methods

Four efforts have been made over the past three decades to assess whether NISP or MNI provides a more accurate measure of taxonomic abundances (Breitburg 1991; Domínguez-Rodrigo 2012; Hudson 1990, 1993; Morin et al. 2017a, b). One is based on ethnoarchaeological data, one on historical data, and two on experimental data. All are published, including, with few exceptions, all pertinent data. The authors use a variety of statistical techniques to determine how closely a measured variable (NISP or MNI) approximates a target variable (ANI) of known value. I use correlation coefficients (Spearman’s rho; Pearson’s  $r$ ) throughout to evaluate the nature and strength of relationships between relevant variables when analyzing data from all four studies.

I assume ANI is ratio scale in all four studies. Granting this assumption has significant implications for evaluating the accuracy of NISP and MNI as measures of ANI. In short, if the calculated correlation coefficient between ANI and either NISP or MNI is  $< 1.0$ , then NISP or MNI is less than perfectly accurate at a ratio scale (Pearson’s  $r < 1.0$ ) or less than perfectly accurate at an ordinal scale (Spearman’s rho

< 1.0). And this is despite the fact that the correlation coefficient might be statistically significant (*e.g.*,  $p \leq .05$ ). A Pearson's  $r < 1.0$  means the NISP or MNI values do not consistently differ in magnitude; for instance, the difference between 10 and 15 NISP or MNI would not be the same as the difference between 5 and 10 ANI. A Spearman's  $\rho < 1.0$  means the NISP or MNI values do not consistently differ in rank the same way the ANI values do. The acceptable degree of inaccuracy when using NISP or MNI to estimate ANI (how far either of the latter two is from ANI) at either ratio or ordinal scales is unknown and not agreed upon by zooarchaeologists. Because the resurrectionists want the most accurate, ratio scale estimate, I take any correlation coefficient < 1.0 to indicate the quantitative unit (NISP, MNI) used is undesirable to some unknown degree.

### A Note on Particularistic Results

There is something to keep in mind as we work through the four studies alleging MNI is a more accurate measure of taxonomic abundances than NISP. MNI may in fact be more accurate than NISP as a measure of taxonomic abundances in some instances. However, given the variability in taphonomic histories across assemblages (but see Rogers et al. 2017), finding a collection in which NISP or MNI accurately predicts ANI will demonstrate only that, for the particular collection under study, NISP or MNI is a more accurate measure. It will say little about the universal accuracy of that quantitative unit for measuring taxonomic abundances in any and all assemblages (Grayson 1979, 1984). We will not know in nearly all paleozoological cases whether NISP or MNI provides the more accurate measure of taxonomic abundances.

### Why Resurrect MNI?

Earlier commentators argued MNI provides a more “representative” measure of taxonomic abundances than the NISP (*e.g.*, Payne 1975) and that MNI measures the composition of human (meat) diet much more accurately than NISP (Parmalee 1985). Why have more recent researchers sought to determine whether NISP or MNI is more accurate?

Hudson (1990) is interested in human subsistence, particularly how “procurement strategies” are reflected by taxonomic abundances. Breitburg (1991:153), too, is explicitly concerned with determining “the relative importance of species to human subsistence,” but he also hopes to have a quantitative unit that allows him “to reconstruct the environment that surrounded the site” on the basis of taxonomic abundances. Domínguez-Rodrigo (2012:48) wants a quantitative unit that allows him “to determine the original number of animals and taxonomic representation [proportions] in any given assemblage.” He alleges (Domínguez-Rodrigo 2012: p. 51) the “faunal analyst aspires to have his derived MNI be the most accurate, most closely approximating estimate of the *actual* number of animals represented in the analyzed sample.” His general interest is “paleoecological studies” (Domínguez-Rodrigo 2012: p. 59). Morin et al.'s (2017b) major interest is in a quantitative unit that provides the most accurate measure of skeletal part frequencies, not taxonomic abundances. But

they argue a robust quantitative unit for measuring skeletal part abundances will also provide a robust measure of taxonomic abundances. Their more general interest seems to be human subsistence given the experiments they perform (Morin et al. 2017a, 2017b). In sum, there are at least two driving forces behind efforts to resurrect MNI as a more accurate measure of taxonomic abundances than NISP: a desire to measure ancient hominin diet and a desire to do paleoecological and paleoenvironmental research.

Two of the researchers want to measure “relative abundances” of taxa (Breitburg 1991; Hudson 1990, 1993), and the other two hope to measure “proportional” abundances of taxa (Domínguez-Rodrigo 2012; Morin et al. 2017b). As noted earlier, the former term—relative abundance—could mean either ordinal scale abundances such as deer are more abundant than turkey, or ratio scale percentage abundances such as deer make up 60% of the assemblage whereas turkey make up 40%. Proportional abundances are exclusively the latter. All four researchers present analyses in which the ANI values of one or more assemblages are known, and they compare those values to NISP and MNI values generated in a zooarchaeological-like way to determine which quantitative unit provides the most accurate measure of taxonomic abundances.

## Efforts to Resurrect MNI

As noted earlier, the question the resurrectionists posed is: Does NISP or does MNI provide a more accurate measure of the ANI or true taxonomic abundances represented by an assemblage of faunal remains? A related question I am concerned with here is: does MNI gain any resolution (ratio vs ordinal scale) or accuracy (closer to ANI) relative to NISP? In this study, I provide an answer to both questions by critically evaluating four studies that have stated MNI is more accurate than NISP for measuring ANI. I consider the four studies in the chronological order of their appearance in the literature and note that none of the early studies seems to have significantly influenced any of the later studies.

## Ethnoarchaeological Evaluation

Hudson (1990, 1993) recorded ethno(zoo)archaeological data among the Aka of central Africa. Descriptive zooarchaeological data for the three sites she studied are summarized in Table 2. For all three sites, she kept records of animals procured directly and animal portions traded into and out of the site by human occupants. After a site was abandoned, she excavated 100% of the deposits using standard archaeological techniques. She knew how many animals (and which carcass portions) site occupants procured, traded away, or added *via* trade to their larder. The ANI values per taxon for each site are summations of that information, and the NISP and MNI values are standard zooarchaeological variables.

Hudson (1990) found the summed three-site NISP and ANI values for the 16 total taxa are correlated (Pearson's  $r = 0.921$ ,  $p < .001$ ; Spearman's  $\rho = 0.786$ ,  $p < .01$ ) (Fig. 2). She also found the summed three-site MNI and ANI values for the 16 taxa are correlated (Pearson's  $r = 0.963$ ,  $p < .001$ ; Spearman's  $\rho = 0.918$ ,  $p < .01$ ). She observed that divergences of NISP and MNI from ANI result from differential recovery,

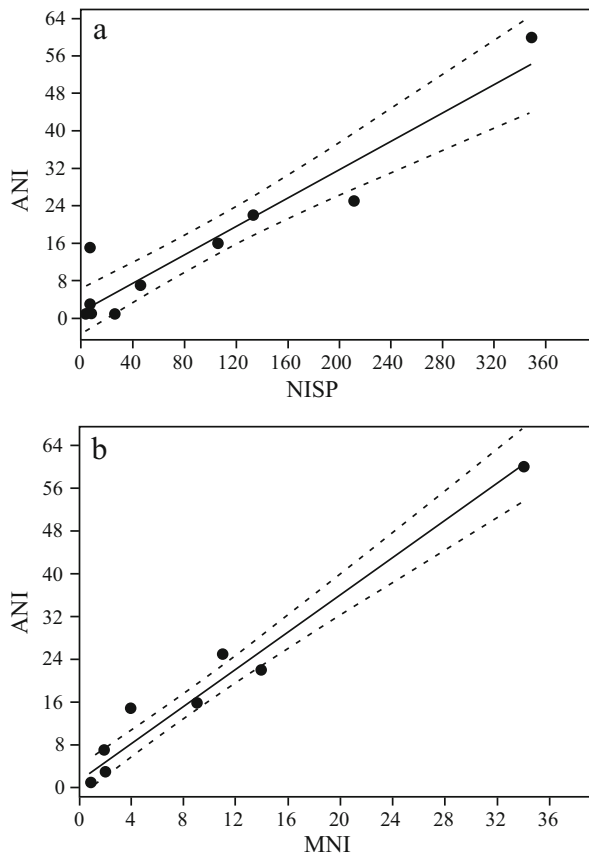
**Table 2** Hudson's (1990, 1993) data. NISP/MNI/ANI

Taxon	Miseteke 6	Miseteke 10	Mokumbokumbo (K1)
Blue/small duiker	40/3/3	258/27/47	51/4/10
Medium duiker	26/1/3	118/7/13	67/3/9
Yellow-backed duiker			26/1/1
Porcupine		56/7/8	77/7/14
Forest rat			106/9/16
Squirrel			7/4/15
Mongoose			3/1/1
Genet			0/0/1
Civet		0/0/1	7/2/2
Monkey			3/1/1
Pangolin		0/0/1	
Bushpig		0/0/1	
Tortoise		0/0/2	46/2/5
Python		0/0/1	
Monitor lizard			7/1/1
Birds			0/0/2
Total	66/4/6	432/41/74	400/35/78

Some data are not presented in Hudson (1990), and some data presented in both Hudson (1990) and Hudson (1993) do not match precisely. When differences occurred, data in Hudson (1993) were given precedence

and sharing or distribution such that some carcass portions move away from the site and other carcass portions move in and are deposited on the site. Importantly, Hudson (1990: p. 197) found that neither NISP nor MNI provide ratio scale taxonomic abundances when compared to ANI. She observed that even the rank order (ordinal scale) abundances of taxa differed across NISP and MNI relative to ANI (Table 2). Hudson (1990: p. 198) concluded that "Interpretations can be made with most confidence for cases where the sample size is large and the differences in proportional representation are dramatic. Small samples with high evenness are the most problematic." This is basically what Grayson (1979, 1984) had concluded a decade earlier; in specific terms, NISP and MNI may not even be ordinal scale reflections of taxonomic abundances because taxa with very similar abundance values of NISP or MNI are likely to become tied in abundance or reverse rank orders when shifting from one quantitative unit to the other, or from one set of aggregates to another. Faunas with taxa of similar abundances are more even than faunas with taxa of disparate abundances, and the former are the ones in which ordinal scale abundances are most likely to shift as the quantitative unit used changes, sample size changes, aggregate definition changes, NISP tallying protocol changes, or how MNI is determined changes.

Hudson (1990: p. 344) argued that MNI provides a better measure of the ANI than NISP because of the higher Pearson's  $r$  correlation coefficient between MNI and ANI than that between NISP and ANI. Two things are worthy of mention in this regard. Most importantly, 95% confidence intervals for the two Pearson's  $r$  correlation coefficients indicate the two relationships are not significantly different; NISP to ANI



**Fig. 2** The relationships between NISP and ANI (a), and MNI and ANI (b) in Hudson's (1990, 1993) data (from Table 2). Solid line is best fit regression; dashed lines represent 95% confidence intervals

Pearson's  $r$  ( $= .921$ ) 95% CI = .783–.973, and MNI to ANI Pearson's  $r$  ( $= .963$ ) 95% CI = .894–.987. MNI is no better than NISP as a predictor of ANI. Of secondary importance, NISP accounts for ( $r^2 =$ ) 84.8% of the variability in ANI whereas MNI accounts for ( $r^2 =$ ) 92.7% of the variability in ANI. These results are not surprising because of the much greater potential range of values that NISP per taxon can take relative to MNI in the summed three-site totals (NISP per taxon can range from 0 to 349 [blue duiker]; MNI per taxon can range from 0 to 34 [blue duiker]). One could use this observation to argue that MNI provides much better, that is, more accurate (closer to ANI) taxonomic abundances than does NISP. That would overlook the statistically insignificant differences between correlation coefficients and also the fact both NISP and MNI at best provide only ordinal scale measures of taxonomic abundances.

Neither NISP nor MNI provides a perfectly ordinal scale measure of taxonomic abundances. The ordinal scale correlations between ANI and NISP ( $\rho = 0.786$ ), and between ANI and MNI ( $\rho = 0.918$ ) are both less than 1.0, indicating the ranks of taxonomic abundances registered by NISP and by MNI do not progress in exactly the same order of taxa as do the ranks of ANI. Not surprisingly given Grayson's (1979, 1984) previous observations, it is the ranks of the dozen or so rarest taxa (whether rarity

is measured by NISP or by MNI) that diverge from their ANI ranking. If NISP and MNI do not provide ordinal scale taxonomic abundance data, they cannot provide ratio scale data.

Although Hudson's (1990, 1993) ethnoarchaeological data are provocative, they indicate MNI provides no greater resolution (*e.g.*, ratio scale values) than does NISP. Both provide ordinal scale data, and even those data are imperfect at ordinal scales. In short, there is little to recommend either value, and the evidence suggests for this assemblage (Table 2), taxa of similar value should be treated as attributes (present only) rather than as variables that can vary in rank or magnitude.

## Historic Documents

Using three historic archaeological sites (one includes two distinct periods of occupation), Breitburg (1991) examined the relationship of NISP and MNI to the documented parent populations (ANI) of three species of domestic animals—cattle (*Bos taurus*), swine (*Sus scrofa*), and sheep (*Ovis aries*). Breitburg (1991: p. 154) assumed the “number of animals reported in letters, journals, census schedules, or other documentary sources is an accurate record of the [animal] populations that existed in reality.” I make the same assumption in what follows and refer to the number of documented animals as the ANI. Breitburg calculated chi-squared statistics between numerous pairs of the percentage abundances of taxa measured as NISP, MNI, and ANI for various site-specific assemblages or aggregates of remains. In virtually all cases, his statistical results indicated that MNI “cannot be said to be sampled from a different statistical universe” than the historically documented ANI (Breitburg 1991: p. 157). He went on to state that “similarity in percentages between MNI and ANI show that MNI is a valid, if not superior, measure of ANI than is NISP” (Breitburg 1991: p. 157). He concluded “MNI is a more valid measure of taxonomic abundance than NISP” because NISP is particularly influenced by fragmentation of bones and the resulting statistical interdependence of fragments. And, he implied MNI data were quite likely ratio scale (Breitburg 1991: pp. 160, 161).

There are a number of things in Breitburg's analysis to be concerned about. Many of the relative (%) ANI values pertain to a site as a whole (its entire spatio-temporal context), yet Breitburg compares them to relative (%) NISP and relative (%) MNI values for particular archaeological features such as an individual structure or garbage dump within a site. Only comparisons between ANI for an occupational period of the site, or for the site as a whole, and corresponding NISP or MNI values for the site as a whole are valid for testing which of the latter two variables might provide a more accurate measure of ANI.

Using the ten species-specific assemblages for which ANI data are available (Table 3), neither NISP ( $r = 0.348$ ,  $p = 0.32$ ) nor MNI ( $r = 0.188$ ,  $p = 0.6$ ) correlates with ANI, suggesting neither quantitative unit provides an accurate measure of taxonomic abundances, not even at an ordinal scale. These results contradict Breitburg's (1991) and, for reasons mentioned in the preceding paragraph, seem more valid. Breitburg's data for distinct features within Fort Loudoun nevertheless demonstrate the most serious problem with MNI—aggregation. It is unclear which aggregates of faunal remains Breitburg identifies should be lumped together and which should be tallied separately. For instance, the degree of independence (remains from different

**Table 3** Summary of Breitburg's (1991) data. NISP/MNI/ANI

Taxon	Fort Loudoun I	Fort Loudoun II	Wynnewood	A. Jackson's Hermitage
Cattle	6499/50/183	387/4/104	283/6/70	249/3/259
Swine	1896/38/123	179/4/102	869/16/225	3475/25/850
Sheep			57/4/100	76/3/396

animals) of faunal remains recovered from structure 7 relative to those from structure 10, or either of these structures with the remains from the dump in occupational phase I (which includes structures 7 and 10), is unknown. Refitting remains will not completely resolve this conundrum because, as noted earlier, not all remains can be empirically shown to be from the same (or a different) individual animal. Data in Table 4 indicate that the maximum distinction method (Grayson 1979, 1984) of determining MNI—determining MNI for each distinct feature, then summing all values for a site—produces different site-wide results than the minimum distinction method—determining MNI for all specimens as one aggregate regardless of context.

Breitburg's (1991) analyses suggest both MNI and NISP provide ordinal scale taxonomic abundance data at best, and on occasion, they do not even do that. That Breitburg's data are imperfectly ordinal scale is apparent in the wide confidence intervals for both the relationship between NISP and ANI, and the relationship between MNI and ANI (Fig. 3). And, the rank-order correlation coefficients between NISP and ANI ( $\rho = 0.321$ ,  $p = 0.36$ ) and between MNI and ANI ( $\rho = -0.012$ ,  $p = 0.97$ ) are both  $< 1.0$ , indicating dissimilar ranks between the variable pairs. Neither quantitative unit is more accurate than the other, and one does not provide greater resolution than the other. In sum, there is little reason to accept Breitburg's (1991: p. 160) conclusion that "MNI [values] are a more valid measure of taxonomic abundance than NISP [values]." There is also little evidence in the analyses I performed to suggest NISP provides an accurate measure of taxonomic abundances. It again seems that monitoring taxonomic presences rather than abundances might be the best option.

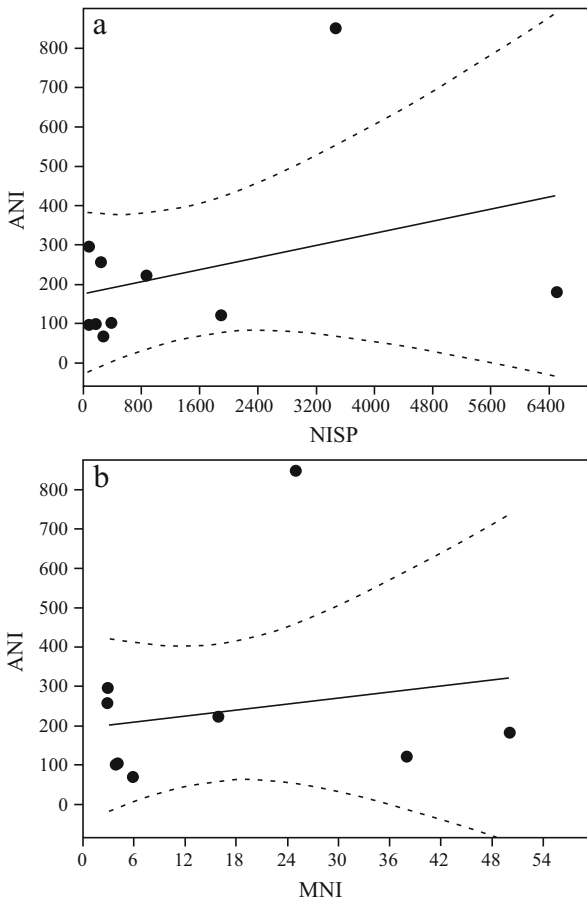
## Experiments: Part 1

Previous work on the nature of NISP and MNI as quantitative units used paleozoological data to model and explore how those quantitative units behaved (took on the values they do) and why they behaved the way they do (Casteel 1977a, 1977b; Grayson 1978, 1979, 1984). This revealed that both units are typically at best ordinal scale and often

**Table 4** Breitburg's (1990) Fort Loudoun MNI data by aggregate. *FLI*, Fort Loudoun occupation I; *FLII*, Fort Loudoun occupation II; *max*, maximum distinction method *sensu* Grayson (1979, 1984); *min*, minimum distinction method *sensu* Grayson (1979, 1984)

Taxon	FLI max	FLI min	FLII max	FLII min	Site max	Site min
Cattle	112	50	7	4	119	94
Swine	88	38	7	4	95	64





**Fig. 3** The relationships between NISP and ANI (a), and MNI and ANI (b) in Breitburg's (1991) data (from Table 3). Solid line is best fit regression; dashed lines represent 95% confidence intervals

likely are nominal scale with respect to measuring taxonomic abundances. What captured Domínguez-Rodrigo's (2012) attention were the unknown values of ANI in that earlier work. In his view, only experiments in which ANI was known would provide a robust evaluation of whether NISP or MNI was a more accurate measure of ANI.

Domínguez-Rodrigo (2012) performed two experiments, each with known NISP, MNE, and ANI values, or what I will hereafter refer to as the experimental parameters. His experimental protocol was to provide student analysts (some working as teams) with a collection of faunal remains; actual NISP (ANISP), actual MNE (AMNE), and ANI values per taxon making up the collection were known to Domínguez-Rodrigo but not the students. Student analysts identified the remains and tallied NISP and determined MNI values, mimicking zooarchaeological protocol. Domínguez-Rodrigo used those values to examine various things, many of which need not concern us here. On the basis of the experimental results, he concluded "MNI still remains one of the most accurate estimators of animals in any given bone assemblage" (Domínguez-Rodrigo 2012: p. 59). This conclusion was based on a simple fact. NISP results the student

**Table 5** Experimental results produced by Domínguez-Rodrigo's (2012) student research teams

Experiment ( <i>N</i> of teams)	ANISP	ANI	NISP (mean ± SD)	NISP range	MNI (mean ± SD)	MNI range
1 ( <i>N</i> = 10)	78	7	63.5 ± 6.4	54–74	6.9 ± 0.32	6–7
2 ( <i>N</i> = 6)	1138	7	945 ± 135	792–1099	6.8 ± 0.41	6–7

research teams produced were, on average, farther from the ANISP than were the MNI results from the ANI values (Table 5). The fact that different groups of analysts identified different total NISPs during experiment 1 is not surprising; that there will be inter-analyst variability in what is identified has long been known based on both practical considerations and empirical work (e.g., Driver 1992; Gobalet 2001; Grayson 1979, 1984). Perhaps not surprisingly, statistical tests indicated with respect to measuring ANI, there was a “wider margin of error for NISP than for MNI” (Domínguez-Rodrigo 2012: p. 50). Although this result cannot be denied, the relevant questions here are: Do Domínguez-Rodrigo's experiments demonstrate that MNI is a more accurate measure of ANI than NISP? Do they indicate which quantitative unit provides greater resolution?

The seven individual animals in Domínguez-Rodrigo's experiment 1 represent six species (Table 6), and one of those species is represented by both an adult and a juvenile, for a total of seven morphotypes (species-specific age class per taxon), if you will. Presuming the analyst takes into account ontogenetic differences in the remains of the species represented by two individuals, seven individual animals must be the MNI for the collection in Domínguez-Rodrigo's experiment 1; MNI per taxon (= 1) must equal ANI per taxon (= 1) as there is no possibility for inter-taxonomic variability in MNI (assuming the student analysts correctly identify the faunal remains as to species and morphotype). NISP per morphotype varies from 2 to 35. MNI per taxon cannot differ from ANI per taxon whereas NISP per taxon can (and quite likely will) vary from ANI per taxon by values ranging from 1 to 34. Experiment 1 cannot fail to suggest MNI is a better measure of taxonomic abundances than NISP simply given the experimental parameters with which Domínguez-Rodrigo begins. It is highly probable the experimental parameters (Table 6) will produce NISP values far from the ANI values and MNI values that are close to the ANI values. This point warrants elaboration.

**Table 6** Domínguez-Rodrigo's (2012) data for experiment 1. All individuals are adults unless otherwise indicated

Taxon	NISP	MNE	MNI	ANI
Deer	9	9	1	1
Goat	35	15	1	1
Sheep	15	15	1	1
Donkey	10	9	1	1
Horse	4	4	1	1
Juvenile horse	2	2	1	1
Cow	3	2	1	1
Total	78	56	7	7

Think about Domínguez-Rodrigo's experiment 1 in terms of the summed NISP, MNI, and ANI values, regardless of taxon. A  $\sum$ MNI of six would result (and did for one of ten sets of analysts, but for unspecified reasons) if ontogenetic differences were not taken into account—a distinct possibility with MNI (Grayson 1984; Lyman 2008). On one hand, the  $\sum$ MNI for experiment 1 simply cannot vary greatly—and one suspects it would not at all—from the  $\sum$ ANI of 7. The total NISP ( $\sum$ NISP), regardless of taxon, on the other hand, can vary from 1 to 78 for any number of reasons (*e.g.*, inter-analyst variation in identification). Again, one could predict that given experiment 1 parameters,  $\sum$ MNI will quite likely appear to provide a better measure of  $\sum$ ANI than  $\sum$ NISP. From either perspective (NISP per taxon, MNI per taxon, ANI per taxon values; or  $\sum$ NISP,  $\sum$ MNI,  $\sum$ ANI values), because of its parameters, experiment 1 provides no robust test of the central question: Does NISP or does MNI provide a more accurate measure of ANI? The experimental parameters disallow such a test because they automatically (predictably will) indicate MNI (or  $\sum$ MNI) more closely measures ANI (or  $\sum$ ANI) than NISP. For the same reasons, the experiment 1 parameters preclude assessment of whether NISP or MNI provides greater resolution when estimating ANI.

Domínguez-Rodrigo reported that analysts with more experience identified more total NISP (regardless of taxon, mean  $\sum$ NISP =  $72.33 \pm 2.08$ ; range = 70–74) during experiment 1 than analysts with less experience (mean  $\sum$ NISP =  $59.71 \pm 2.98$ ; range = 54–62; Student's  $t = 6.56$ ,  $p < .001$ ). Because NISP varied more between analysts than did MNI, Domínguez-Rodrigo concluded the latter quantitative unit is to be preferred for measuring taxonomic abundances. This variability too is not surprising because the possible range of variation in NISP (from 0 to 78) is, as noted earlier, much greater than the possible range of variation in MNI (0 to 7). Further, as also noted above, it has long been known that different analysts are unlikely to produce exactly the same NISP (or MNI) for myriad reasons including experience, quality of comparative collections, intensity or thoroughness of study of individual specimens, and the like (*e.g.*, Driver 1992; Gobalet 2001). The detection of variability in NISP between analysts, given the near impossibility of variability in MNI (or  $\sum$ MNI) between analysts given experiment 1 parameters, means this part of this experiment is no rigorous test of whether NISP or MNI provides a more accurate measure of ANI.

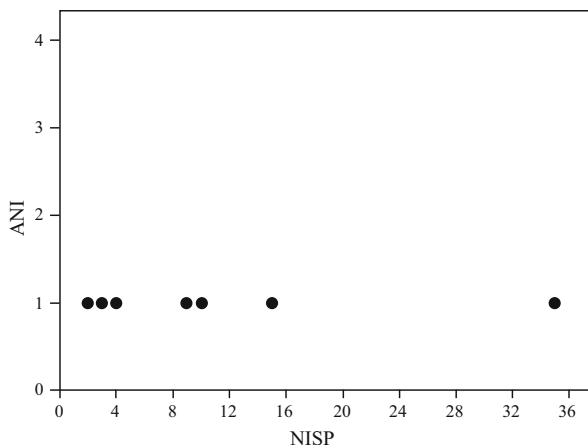
In his experiment 2 (Table 7), Domínguez-Rodrigo found that analysts provide accurate estimates of ANI (= 7) using  $\sum$ MNI values (range = 6–7; five of six analysts came up with an MNI of 7), whereas their total identified NISP (range per analyst = 792–1099) values were far off the mark of the actual  $\sum$ NISP in the collection (= 1138). As with experiment 1, Domínguez-Rodrigo used results of experiment 2 to argue experienced analysts tend to identify more specimens than inexperienced analysts (means = 1063, and 827, respectively, out of 1138 NISP). That “the derivation of NISP estimates varies [between analysts] more [extremely] than the derivation of MNI” (Domínguez-Rodrigo 2012: pp. 51, 52) is (as with experiment 1 results) believed by him to strengthen his conclusion that MNI is a “potentially more accurate estimator of taxonomic variability [than] NISP” (p. 52). But again, the parameters of experiment 2 are inappropriate for testing whether NISP or MNI provides a more accurate measure of ANI.  $\sum$ NISP can vary markedly in value (0 to 1138) but  $\sum$ MNI cannot (0 to 7). The experimental parameters alone lead one to predict  $\sum$ NISP is highly likely to provide a wider range of (and thus less accurate) estimates of  $\sum$ ANI than  $\sum$ MNI.

**Table 7** Domínguez-Rodrigo's (2012) data for experiment 2

Taxon	NISP	MNE	ANI
Adult sheep	627	147	4
Subadult sheep	511	139	3
Total	1138	286	7

Domínguez-Rodrigo also evaluates the argument that because NISP and MNI often provide the same (redundant) information on taxonomic abundances, one need not determine MNI but can instead utilize NISP. To show the two quantitative units provide the same information, researchers have calculated correlation coefficients (typically Spearman's rho) between the two (*e.g.*, Grayson 1979, 1984; Lyman 2008). Domínguez-Rodrigo follows suit and calculates Spearman's rho between NISP and MNI (= ANI) per taxon for his experiment 1 data. He interprets the insignificant correlation between experiment 1's NISP and MNI data ( $\rho = 0.07, p > .8$ ) as indicating the two units do *not* contain redundant information regarding taxonomic abundances. In his view, this reinforces his argument that MNI is a more accurate measure of taxonomic abundances than NISP. But yet again an insignificant correlation coefficient is predictable given the experimental parameters. Figure 4 illustrates the relationship between experiment 1 data with NISP per taxon as the independent variable and MNI per taxon as the dependent variable; the point scatter and the statistics would be no different were the roles of the variables reversed (the point scatter would be perfectly linear but vertically arranged rather than horizontally). The lack of variation in the dependent variable (MNI) and marked variation in the independent variable (NISP) means knowing the latter tells us little about the former. But that is a result of the experimental parameters alone, not some inherent property of the quantitative units.

In summary, the parameters of Domínguez-Rodrigo's experiments 1 and 2 are inappropriate for testing whether NISP or MNI provides a more accurate measure of ANI because they include no variability in the ANI or MNI per taxon. This use of experimental data to evaluate the focal question provides little clear evidence as to

**Fig. 4** The relationship between NISP and ANI for each of seven morphotypes (species) in Domínguez-Rodrigo's (2012) data for experiment 1. Data from Table 6

whether NISP or MNI as a measure of ANI is to be preferred. Because NISP per taxon (or morphotype) varies, and because different analysts will tend to identify different frequencies of specimens in a collection, the experimental parameters dictate that NISP will provide a less accurate measure of taxonomic abundances than will MNI. But no other outcome is possible. And, unlike Hudson's (1990, 1993) and Breitburg's (1991) data, Domínguez-Rodrigo's experimental data do not allow evaluation of whether either quantitative unit provides nominal, ordinal, or ratio scale results, so the second question of which quantitative unit provides greater resolution cannot be evaluated. Fortunately, there are other experimental data that may bear on both issues.

## Experiments: Part 2

Morin et al. (2017a, 2017b) performed two experiments. One replicated prehistoric breaking of bones to extract marrow (the Marrow Cracking Experiment [MCE]), the other replicated breaking bones (into smaller pieces) to render grease from interstitial spaces in bone tissue (the Bone Grease Extraction Experiment [BGRE]). In both experiments, the original number of (unbroken) skeletal elements (ANE, actual number of elements) was known as was the ANI per skeletal element. Morin et al. were interested in examining how differential fragmentation influenced identification and evaluating how well NISP and MNE measured the ANE of skeletal parts. They found what they took to be serious problems with both NISP and MNE as measures of skeletal part frequencies, and proposed a different counting unit they termed the *number of distinct elements*, or NDE. NDE values are “simply tallies [of] the number of times a diagnostic [anatomical] landmark is represented in a sample of specimens attributed to the same [skeletal] element and taxon” (Morin et al. 2017b: p. 952).

After undertaking a variety of analyses of their experimental data, Morin et al. (2017b: p. 967) conclude that “ $\sum$ NDE counts produce more accurate estimates of species representation than NISP [or MNI].” I presume by “species representation” they mean taxonomic abundances because of how MNI is defined—the most abundant skeletal part. Because their NDE unit is meant to provide accurate measures of the abundances of each of various skeletal parts, the most abundant NDE should, one suspects, be equal to the most abundant skeletal part (assuming lefts and rights are distinguished) and thus represent the MNI. Paraphrasing Morin et al. (2017b) (see quotations that follow), they argue (i) MNI is equivalent to the most abundant skeletal part, (ii) their NDE quantitative unit provides the most accurate measure of each kind of skeletal part, and therefore (iii) NDE provides the most accurate measure of taxonomic abundances. For this reason, considerable space is devoted to evaluating their experimental results.

Morin et al.'s (2017a, 2017b) central research question concerns the influence of fragmentation on the NISP and MNE quantitative units. The “main goal of [their experiments] was to measure how identifiability and estimates of relative abundance vary *between skeletal elements*” (Morin et al. 2017a: p. 890, emphasis added). Their experimental data involve two species, one of which was represented by numerous skeletal elements (red deer [*Cervus elaphus*] ANI = 30; ANE = 501; NSP [number of specimens, both identifiable and unidentifiable] = 15,829) and the other by very few elements (domestic cow [*Bos taurus*] ANI = 2; ANE = 5; NSP = 47). Experimental samples MCE and BGRE were given to three individuals with instructions to identify

the remains and to keep track of NISP, MNE, and MNI per skeletal element. In other words, the specimens making up each sample were treated (experimentally) as if they comprised a zooarchaeological collection.

Based on the experimental results, Morin et al. (2017a: p. 902) found “greater fragmentation increases variation in the proportion of identified remains between analysts.” Skeletal part frequencies measured by either NISP or MNE varied across participants (not unexpectedly, as should by now be clear) but to a statistically insignificant degree. Further, neither NISP nor MNE produced estimates of skeletal part abundances that were statistically significantly different than the ANE values (Morin et al. 2017a: p. 910). Because there are only two taxa of quite disparate abundances in the experimental samples, the experimental data are not very robust for assessing whether NISP or MNI (or MNE) produces the most accurate measure of ANI, even at an ordinal scale. Those data can, however, be used to indirectly evaluate which unit *might* provide the most accurate measure.

Morin et al.’s (2017a) MCE data are summarized in Table 8 and their BGRE data are summarized in Table 9. I (i) use the average value of NISP, MNE, and MNI of all three analysts who identified the bones (inter-analyst variability is not of interest here), (ii)

**Table 8** Morin et al.’s (2017a) marrow cracking experiment (MCE) data

Skeletal part	ANE <sup>a</sup>	ANI <sup>a</sup>	NISP <sup>b</sup>	MNE <sup>b</sup>	MNI <sup>b</sup>	NDE <sup>c</sup>
Cranium	1	1	8	1.7	1.7	–
Mandible	2	1	8	2	1	2
Hyoid <sup>d</sup>	2	1	2	2	1	2
Rib	2	1	2.7	2	1	2
Scapula	7	4	13.3	4.3	3	4
Humerus	23	13	177	23	13	23
Radius	21	12	257.3	21.3	12	21
Carpals <sup>d</sup>	11	2	11	11	2	11
Metacarpal	26	13	200.7	24.7	12.7	26
Innominate	3	2	2.3	2.3	2	2
Femur	24	12	224.7	23	12.3	20
Tibia	21	13	261.3	21	12.3	19
Malleolus	7	4	7	6.7	3.7	6
Calcaneus	3	2	3.3	2.7	2	2
Talus	3	2	2	2	1	2
Other tarsals <sup>d</sup>	5	1	5.3	5	1	5
Metatarsal	26	13	264.7	22.7	13	22
Vestigial Metacarpal <sup>d</sup>	46	12	39.7	20.3	10.7	46
Phalange	7	1	6	4.3	1	5

<sup>a</sup> Data from Morin et al. 2017a: Table 1

<sup>b</sup> Mean value of three analysts calculated based on data in Morin et al. 2017a: Table 19

<sup>c</sup> Data from Morin et al. 2017b: Table 7

<sup>d</sup> All specimens are anatomically complete

**Table 9** Morin et al.'s (2017a) bone grease extraction experiment (BGRE) data

Skeletal part	ANE <sup>a</sup>	ANI <sup>a</sup>	NISP <sup>b</sup>	MNE <sup>b</sup>	MNI <sup>b</sup>	NDE <sup>c</sup>
Cranium	1	1	6.3	1	1	–
Mandible	3	2	7	2.7	1.7	2
Hyoid <sup>d</sup>	2	1	2.3	1.7	1	1
Rib	2	1	1.7	1	1	1
Scapula	6	3	11.7	2.7	1.7	1
Humerus	20	10	529.7	23.3	12.3	17
Radius	20	10	547.7	19	9.7	15
Carpals <sup>d</sup>	7	2	7	7	2	7
Metacarpal	24	12	286.3	23	11.7	19
Innominate	2	2	5	2	2	2
Femur	29	17	730	29.7	16.7	20
Tibia	21	12	529.7	22.3	13	18
Malleolus	7	4	7	5.5	3	4
Calcaneus	3	2	3	3	2	2
Talus	2	1	3.5	2.5	1.5	2
Other Tarsals <sup>d</sup>	3	2	4	4	2	4
Metatarsal	21	11	360.7	21.7	12	18
Vestigial Metacarpal	40	10	34.3	31	11.5	40
Phalange	6	3	12.7	6.3	1.3	5

<sup>a</sup> Data from Morin et al. 2017a: Table 1

<sup>b</sup> Mean value of three analysts calculated based on data in Morin et al. 2017a: Table 20

<sup>c</sup> Data from Morin et al. 2017b: Table 7

<sup>d</sup> All specimens are anatomically complete

include only remains of red deer, and (iii) calculate Pearson's  $r$  between test case data (NISP, MNE, MNI) and experimentally known data (ANE, ANI) using the skeletal element categories in each experiment on the presumption that the frequency of each unique skeletal part can serve as a proxy for the frequency of a unique species. This presumption is warranted on the observation that in zooarchaeological contexts, different species are often represented by different most abundant skeletal parts (Grayson 1979, 1984; Lyman 2008).

The correlation coefficients (Table 10) indicate two things of interest pertinent to the focal question of which quantitative unit provides the most accurate measure of taxonomic abundances (recalling that those abundances are in this case skeletal elements, not taxa). First, all correlations are significant ( $p < .0025$  for all). Second, for both sets of experimental data (MCE and BGRE), NISP provides the weakest correlations with ANI, MNE is of middle strength, and MNI provides the strongest correlation. Granting different skeletal elements are a valid proxy of values that might be taken by different taxa, the correlation coefficients suggest MNI provides a more accurate measure of ANI than MNE or NISP. But three things warrant notice. First, the ordering of the correlation coefficients, from least to greatest, is not unexpected given how NISP,

**Table 10** Correlation coefficients (Pearson's  $r$ ) between select pairs of variables. Data from Tables 8 and 9

Variable pair	MCE		BGRE	
	Pearson's $r$	$p$	Pearson's $r$	$p$
NISP:ANE	0.6539	.0024	0.6688	.0017
NISP:ANI	0.9022	<<.001	0.8921	<<.001
MNE:ANE	0.8907	<<.001	0.9780	<<.001
MNE:ANI	0.9713	<<.001	0.9540	<<.001
MNI:ANI	0.9960	<<.001	0.9870	<<.001
MNE:ANE <sup>a</sup>	0.9933	<<.001	0.9909	<<.001
NDE:ANE <sup>a</sup>	0.9899	<<.001	0.9822	<<.001

<sup>a</sup> Anatomically complete specimens (whole bones) not included

MNE, and MNI are defined. Second, all correlation coefficients are statistically significant, so ANI is predictable using NISP, MNE, or MNI, though with increasing accuracy. However, the 95% confidence intervals for the correlation coefficients indicate that, in fact, MNE is no better than NISP for predicting ANI in the MCE experiment, and MNI is better than either MNE or NISP in the MCE experiment (Table 11). For the BGRE experiment, NISP and MNE are equally good estimators of ANI, MNE is as good as MNI, but NISP is not as good as MNI.

The third thing to notice with respect to the correlation coefficients (Table 10) concerns the values of the quantitative units. As Grayson (1979, 1984) pointed out long ago, where problems arise with any of the quantitative units (NISP, MNE, MNI) as measures of ANI is with values that are close to (or tied with) one another. Such values are likely to swap order or become tied (take the same value) for any of myriad analytical reasons (*e.g.*, variability in identification skill, how remains are aggregated, use of different counting protocols), rendering them not even ordinal scale. This observation is important because Morin et al. (2017b: p. 967) state “experimental results seem to confirm that NDE provides estimates that are proportional to known abundances” with respect to ANE. I presume by “proportional” they mean if, say, tibiae are twice as abundant as humeri in ANE values, the best quantitative unit to use, whether NISP, MNE, or MNI per element, is the one that most closely mirrors those ANE values in ratio scale terms. The fact the correlation coefficients between NISP and ANE, and between MNE and ANE are both  $< 1.0$  indicates proportional abundances of NISP and MNE are imperfect measures of the proportional abundances of ANE.

Morin et al. (2017b: p. 940) state MNE “provides the foundations for the derivation of MNI values.” Indeed, this is the case according to the typical definition of MNI—the most abundant skeletal part. This is the case whether the analyst ignores variability in ontogeny, size, and sex or incorporates it, just like left or right side of bilaterally paired elements, into the definitions of skeletal element categories. To measure taxonomic abundances using their newly proposed NDE quantitative unit, Morin et al. (2017b: p. 954) indicate one must account for the fact that NDE values do not include an indication of whether a particular specimen is from a left or right skeletal element and also account for the fact that species “differ in frequencies of bones.” To overcome



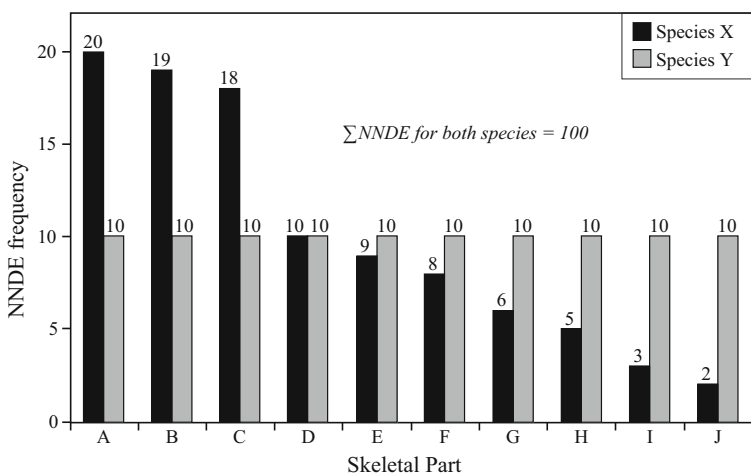
**Table 11** Confidence intervals (95%) around correlation coefficients between one quantitative unit and ANI, and a second quantitative unit and ANI. Correlation coefficients are listed in Table 10

Variable pair	MCE	BGRE
NISP and MNE	0.759–0.962 and 0.925–0.989	0.736–0.958 and 0.882–0.982
MNE and MNI	0.925–0.989 and 0.990–0.998*	0.882–0.982 and 0.966–0.995
NISP and MNI	0.759–0.962 and 0.990–0.998*	0.736–0.958 and 0.966–0.995*

\*Correlation coefficients significantly different

this problem, they recommend calculating a “normed NDE” (NNDE) value for each landmark “by dividing the NDE count for an element by the abundance of the same element in a living animal” (p. 954). For instance, the NDE for all bilaterally paired bones, such as femora, is divided by two to produce an NNDE. Morin et al. (2017b: p. 954) suggest the “NNDE values can be summed to obtain a total standardized element count for each taxon” ( $\sum \text{NNDE}$ ) and imply this sum “can be used to estimate both skeletal and taxonomic abundances.”

A  $\sum \text{NNDE}$  value entails significant problems as a measure of taxonomic abundances, the main one being inter-taxonomic variability in the evenness of NDE (or NNDE) across skeletal parts. In a set of multiple assemblages with equal  $\sum \text{NNDE}$  values, uneven assemblages of skeletal parts will entail higher numbers of individuals (e.g., MNI) whereas even assemblages will represent lower numbers of individuals. Consider Fig. 5, which illustrates a fictitious data set made up of two species with equal  $\sum \text{NNDE}$  values. Notice that species *X* includes remains of at least 20 individuals (= MNI) based on the maximum  $\sum \text{NNDE}$  value for skeletal part A, whereas species *Y* includes remains of 10 individuals (= MNI). And it should be obvious that the two frequency distributions of skeletal parts would likely be interpreted as indicating important interspecific variability in taphonomic histories.



**Fig. 5** A fictitious example of how variability in the evenness of frequencies of skeletal parts of two species measured as NNDE can influence measures of taxonomic abundances tallied as  $\sum \text{NNDE}$ . Note that for both species,  $\sum \text{NNDE} = 100$ , but for species *X*, MNI = 20 whereas for species *Y*, MNI = 10

Another problem plagues NNDE values. Dividing the NDE of anatomically paired bones (for instance) by two ignores the fact that left and right elements may be of unequal abundances and differ to a statistically significant degree (Lyman 2008). The net result is that dividing NDE by two will produce a measure of taxonomic abundance that is  $< \text{MNI}$  when lefts and rights of a skeletal element are of unequal abundances. Morin *et al.* do not acknowledge this problem, but one can anticipate their response to it by identifying a problem they do acknowledge and summarily discount. They note a skeletal element “known to be present in a sample may become analytically absent if its specimens show only NDE landmarks [represented by]  $< 50\%$  [of the landmark]” (Morin *et al.* 2017b: p. 967). I will return to this “ $< 50\%$ ” notion shortly. Morin *et al.* (2017b: p. 967) discount this potential of under counting by arguing “what is critical is not the magnitude of the values but *their proportionality to the actual number of elements, individuals and species in a sample*” (emphasis in original). Morin *et al.* expect that if the ratio of ANE is, say, 1 distal humeri to 2 distal tibiae, then that same ratio of 1 to 2 should be evident using NDE as the quantitative unit.

Do NDE values more accurately reflect the proportionality of ANE values than, say, MNE values? They do not. Initially, I deleted the anatomically complete (unbroken) bones such as carpals and tarsals unlikely to be fractured for marrow extraction from Morin *et al.*'s two sets of experimental data. For the MCE sample, MNE is more strongly correlated with ANE (Pearson's  $r = 0.9933$ ) than NDE is with ANE ( $r = 0.9899$ ), but the correlation coefficients are statistically insignificantly different. The 95% CI for the MCE sample's MNE:ANE coefficient ( $r = 0.9933$ ) is 0.979–0.998, and the 95% CI for that samples' NDE:ANE coefficient ( $r = 0.9899$ ) is 0.969–0.997. There is no statistically significant difference between the two coefficients, indicating MNE is just as accurate an estimate of ANE as is NDE for the marrow extraction experiment. The 95% CI for the BGRE sample's MNE to ANE coefficient ( $r = 0.991$ ) is 0.973–0.997, and the 95% CI for that sample's NDE:ANE coefficient ( $r = 0.982$ ) is 0.945–0.994. Again, there is no statistically significant difference between the two coefficients, indicating MNE is just as accurate an estimate of ANE as is NDE for the grease extraction experiment.

Including the unbroken (anatomically complete) hyoids, carpals, other tarsals, and vestigial metapodials results in different correlation coefficients. For the BGRE sample, MNE:ANE  $r = 0.98$  and NDE:ANE = 0.98; inclusion of the anatomically complete bones in this case does not influence the result (respective 95% CIs are identical at 0.942–0.992). MNE and NDE provide equally accurate estimates of ANE. For the MCE sample, MNE:ANE  $r = 0.89$  and NDE:ANE  $r = 0.99$ ; inclusion of the anatomically complete bones strongly influences the result. The 95% CIs for the correlation coefficients indicate they are significantly different (MNE:ANE = 0.734–0.958; NDE:ANE = 0.983–0.997). NDE seems to provide a more accurate measure of ANE than does MNE. Exactly why this should be the case in the bone marrow extraction experiment (MCE) but not the bone grease extraction experiment (BGRE) is unclear. This result does highlight the potential influence of including anatomically complete skeletal elements on measures of fragmentation (see also Lyman 1994b).

Recall the “ $< 50\%$ ” rule mentioned above regarding whether or not to tally a particular NDE. Morin *et al.* (2017b) adopt this rule for their diagnostic anatomical landmarks to ensure that the same bone is not counted twice. A tally of 1 NDE is made for a specific anatomical landmark “*if and only if the fragment shows at least 50% of*

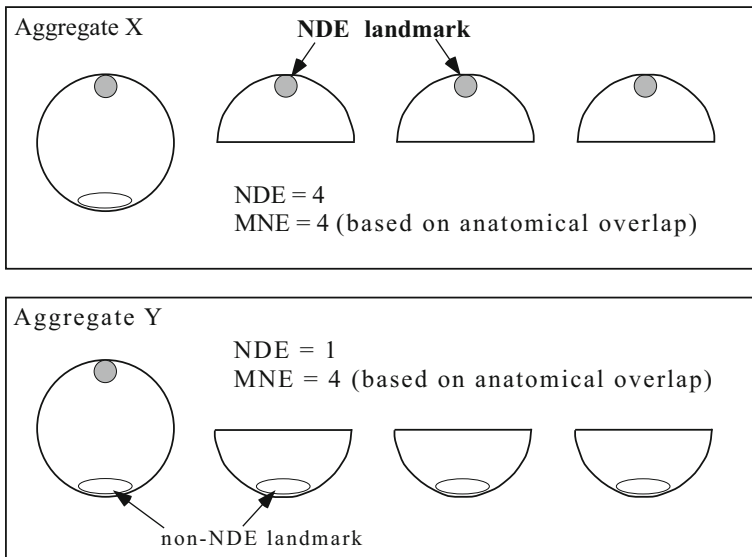
the cortical surface of that landmark” (Morin et al. 2017b: p. 952). Morin et al. (2017b: pp. 952, 953) recommend the use of a “small square cutout [to provide] a control for assessing whether at least 50% of the cortical surface of the landmark is preserved.” It is unknown exactly how well this protocol facilitates determination of whether or not a landmark is > 50% present given that  $\pm 1\%$  makes all the difference, and many complete landmarks Morin et al. specify are millimeters in maximum dimension. Morin et al. (2017b: p. 954) would likely counter that the landmarks are relatively small and “were chosen on the basis of their susceptibility of being identified in highly fragmented assemblages.” An experiment in which Morin et al.’s protocol for assessing whether > 50% of a landmark or < 50% is present in a series of specimens could be designed using measures of surface area (e.g., Cannon 2013), but the utility of doing so is in question because the flaws with the NDE quantitative unit mentioned thus far cannot be overcome. And there are others as well.

Morin et al. (2017b) list what they take to be five “advantages” of NDE over other zooarchaeological quantitative units. First, they suggest NDE is easier to calculate than MNE, but present no data comparing, say the time taken to determine both for a particular assemblage. Although calculating NDE may be easy, one wonders why this is important and if the loss of accuracy (undercounting, and see below) negates any gain in ease or efficiency. The second advantage is that MNE values are calculated various ways, but not so with NDE, making data for the latter more comparable than data for the former across analysts. But Morin et al. (2017b: p. 959) also indicate “criteria of sex, age, and size are not pertinent to the calculation of NDE values,” meaning the NDE will likely for yet another reason produce undercounts of the ANI (and also the MNI) in a collection.

As a third advantage, Morin et al. (2017b: p. 961) indicate NDE “does not inflate the representation of rare elements [because] NDE values are expected to increase linearly with NISP because the probability of identifying a new element is *independent of previous identifications*.” MNE values, on the other hand, are interdependent (identifying an additional MNE depends on previous identifications) and thus, MNE “tallies increase at a decelerating rate relative to NISP” (Morin et al. 2017b: p. 961). NISP values are potentially interdependent, meaning each identified specimen could come from or represent the same skeletal element or animal. MNI values are designed to be independent of one another at the scale of individual (Grayson 1979, 1984); MNE values are designed to be independent of one another at the scale of skeletal element (Lyman 2008); NISP values are not designed to be independent of one another at any scale. The > 50% rule of NDE values means those values must, by definition, be independent of one another (no two specimens can both represent more than 50% of an anatomical landmark and simultaneously represent the same animal). This was part of the reason Watson (1979) developed his similar *diagnostic zone* protocol and why others have used it (Dobney and Rielly 1988; Knüsel and Outram 2004; Münzel 1986, 1988; Rackham 1986; see Morlan (1994) for a similar quantitative technique). Although a reasonable advantage to NDE, one wonders if the accompanying undercounting that results is worth the gain of independence, particularly realizing that NDE values are not always ratio scale (are not proportionate relative to ANE and ANI; see above).

The fourth advantage is said to be that NDE values will not vary when aggregates of remains change, unlike with MNI or MNE. This allows NDE values from different

aggregates to be summed without fear of summing interdependent aggregates, unlike with MNE and MNI. Morin et al. do not provide a reason for tallying skeletal parts across multiple aggregates or why this potential might be important. Figure 6 is a simplified version of Morin et al.'s (2017b) Fig. 15 illustrating their suggested counting protocol. Each rectangle represents the boundaries of an aggregate, each circle represents an ungulate proximal metatarsal, a complete circle is 100% anatomically complete, each incomplete circle represents 49% of an ungulate proximal metatarsal, and all specimens are from the left rear leg. In both aggregates, MNE = 4, but in aggregate X, NDE = 4 whereas in aggregate Y, NDE = 1. Independence of tallies of NDE across the two aggregates is maintained. Why independence of tallies across aggregates X and Y is necessary is unclear, though in tallying skeletal parts this way the analyst is assuming skeletal parts across the aggregates are not independent of one another. This in turn implies there is some mixing of materials comprising those aggregates, in which case one might question how and why those aggregates are defined the way they are. The implicit warrant for counting this way (tallying the same NDE across multiple aggregates) seems to be that it avoids the problem of aggregation that afflicts MNE and MNI (tallies are likely to shift in value), and it circumvents the interdependence problem, but only with respect to multiple aggregates. The four specimens in aggregate X of Fig. 6 are clearly independent of one another, as are the four specimens in the aggregate Y. If there is no clear evidence of inter-aggregate mixing or sharing of skeletal parts of the



**Fig. 6** An example of how NDE maintains inter-aggregate independence of tallies of skeletal parts; only those specimens with > 50% of an NDE landmark are tallied. Assume each rectangle represents the boundaries of an aggregate, each circle represents an ungulate proximal metatarsal, a complete circle is 100% anatomically complete, each incomplete circle represents 49% of an ungulate proximal metatarsal, and all specimens are from the left rear leg. This example is a precise, if schematic, replica of one presented by Morin et al. (2017b) in their Fig. 15

same individuals, it would make more sense to tally an MNE of four for each aggregate, and this would avoid the undercounting of NDE.

The final advantage is said to be that NDE mimics the “non-repetitive element [NRE]” tallies of archaeomalacologists (e.g., Giovas 2009; Harris et al. 2015; Thomas and Mannino 2017). According to Morin et al. (2017b: p. 963), the procedural similarities of the two “should enable sounder comparisons of vertebrate and invertebrate tallies than previous counting methods.” Well, yes, but one wonders why this might be necessary. Molluscs and mammals and birds, for instance, have not altered their skeletal anatomies so much over, say, the past five million years that we must worry that there were systematic gains or losses in the ANE per individual organism of one taxon or the other that would skew zooarchaeological tallies. Thus, the precise nature of what will be gained by using NDE for mammals and NRE for molluscs is unclear. The major analytical weakness of molluscan NRE values is that they undercount (Giovas 2009), and as we have seen, Morin et al.’s NDE values suffer from precisely the same problem. Recognizing this, some archaeomalacologists have developed a modified NRE technique that is the equivalent of a (vertebrate) zooarchaeologist’s traditional MNI—the most abundant skeletal part (Harris et al. 2015).

A paleozoologist might have as a target variable not taxonomic abundances *per se*, but rather biomass per taxon. Returning to Morin et al.’s desire for sound comparisons, a typical way to measure biomass in paleontology and also in zooarchaeology is to determine the MNI per taxon, and then multiply that value by the average biomass per individual (see Lyman (2008) for discussion and references). Given that NDE for mammals and NRE for molluscs both undercount, the analyst should be concerned that undercounting may differ inter-taxonomically. Further, NRE may undercount by 7% and NDE undercount by 10% in one assemblage but undercount by different amounts in different assemblages given variability in fragmentation and other taphonomic influences. Undercounting is, for many other zooarchaeologists, an extremely problematic issue. Given basic probability theory, zooarchaeologists tend to assume large samples are more representative than small samples of the target variable, whether taxonomic richness, abundances or heterogeneity, butchering patterns, seasonality, hunting selectivity, skeletal part transport, and a plethora of other possible target variables. Why else would they be concerned about sample size effects, discuss analytical techniques such as rarefaction and sampling to redundancy, and be worried about the influence of screen mesh size on recovery (e.g., Crabtree 2018; Eck 2007; Gifford-Gonzalez and Hildebrandt 2012; Giovas 2018; Jammiczky et al. 2008; Joslin 2012; Lyman and Ames 2007; Travouillon et al. 2007; Trusler 2014)? Morin et al.’s (2017a, 2017b) thinking on the matter is thus particularly noteworthy.

In sum, Morin et al. (2017a, 2017b) perform no robust tests of whether NISP, MNE, MNI, or NDE provides the most accurate measure of ANI. That, however, was not their primary research goal and thus their experimental data are less than ideal for testing which unit is most accurate. The experimental data they generated seem, on further study, to contradict their conclusion that NDE provides the most accurate measure of ANI. Nor does the NDE quantitative unit provide measures of skeletal part abundances

of greater resolution (ratio scale) than MNE, and it does not measure proportionate taxonomic abundances more accurately than MNE at ratio scales. That NDE will sometimes not provide even ordinal scale taxonomic abundances should be obvious.

## Discussion and Conclusion

The history of quantification of zooarchaeological remains is an intriguing one. Striking parallels are apparent in paleontology (e.g., Badgley 1986; Gilinsky and Bennington 1994; Moore et al. 2007) and in forensic anthropology (e.g., Adams and Konigsberg 2008; Lambacher et al. 2016; Robb 2016), though the volume of literature on grappling with which quantitative unit might provide the most accurate measure of taxonomic abundances is greatest in zooarchaeology. In forensic anthropology, there is typically only one taxon of interest—*Homo sapiens*—and in paleontology, MNI or at least not NISP has been the preferred quantitative unit for the last 100 years or more (Lyman 2016, 2018a, 2018b). How taxonomic abundances are quantified seems less important to paleontologists than to zooarchaeologists (see Holtzman (1979) for a rare exception, and see Vermeij and Herbert (2004) for a discussion of variables that influence NISP and MNI not yet considered by zooarchaeologists). Contention over which unit is the best for measuring taxonomic abundances is not much of an issue in forensics and paleontology.

Hudson (1990, 1993) and Breitburg (1991) are interested in assessing human diet. Their efforts to validate the utility of MNI rested in part on that interest. As Theodore White (1953) demonstrated more than six decades ago, MNI was a much more logical quantitative unit than NISP (or MNE) to use when the target variable was human diet. Although biomass estimates might be made for purposes of paleoecological inference (e.g., Guthrie 1984; Staff et al. 1985), fluctuation in taxonomic abundances of species sensitive to, say, climatic conditions is often interpreted as an indication of paleoenvironmental conditions without conversion to biomass (Lyman 2017). Domínguez-Rodrigo (2012) seems to argue that MNI is best for any and every analytical question concerning taxonomic abundances, whether paleodiet, paleoenvironments, or something else. Morin et al. (2017a, 2017b) are interested in the influence of the degree of fragmentation on the identifiability of specimens and abundances of skeletal parts. Their definition of a new quantitative unit (NDE) was a response to their experimental results, but that unit demonstrably undercounts and fails to provide ratio scale measures of the original frequencies of skeletal elements (ANE). Although a robust test of NDE's potential for measuring ANI accurately and at a ratio scale cannot be performed, there is little in Morin et al.'s experimental data to recommend NDE over MNE or MNI.

The review presented here of efforts to resurrect MNI reveals both good and bad things about various quantitative units and also about the nature of a robust experimental sample, should someone care to build one. No fewer than ten taxa should be represented (20 would be better), and perhaps a third to half of them should be within one or two actual numbers of identifiable specimens and numbers of individuals of one another to simulate rank order abundances that could become tied, or swap relative abundances as analysis shifts from NISP to MNI, or from not considering age, sex, and size differences to considering them. The number of identifiable specimens per taxon

should range from 1 to, say, 200, so that the total sample NISP is  $\sim 1000$ . More species, more specimens, and more individuals would make a more robust (and statistically powerful) experiment, but sacrificing skeletons and the hours necessary for analysts to simulate a real zooarchaeological identification project make for the prohibitive nature of such an experiment. And, we already know a great deal about the nature of the NISP and MNI counting units and how they relate to each other. We know enough to be able to predict many of the outcomes of experiments that have been performed. And, even with the results a more robust experiment might produce, we would be left wondering about the universality of those results given the contingent nature of taphonomic histories and variability in counting protocols.

The logic of various researchers who explored the nature, validity, and statistical properties of NISP and MNI in the 1970s and 1980s cannot be faulted. Many of those researchers worked to validate MNI and invalidate NISP, or to design a different quantitative unit that was something like but more accurate than the traditional MNI. Those efforts faded away for lack of success, but never fully disappeared. Efforts to resurrect MNI (or a similar quantitative unit) over the past three decades have failed to show that MNI is a more accurate ratio scale measure of taxonomic abundances than NISP. Even if it is eventually demonstrated that MNI is a better measure than NISP, perhaps by an experiment like the one sketched in the preceding paragraph, that demonstration will be for one particularistic, contingent experiment. It will not demonstrate that MNI is always and everywhere a more accurate measure of taxonomic abundances than NISP, something the theoretical/mathematical modeling of quantification and knowledge of taphonomic histories in decades past showed.

Lest my conclusions regarding the attempted resurrection of MNI as a measure of taxonomic abundances (ANI) be misinterpreted, my final comment is that I do think NISP is to be preferred for measuring ANI for a number of reasons. This decidedly does not mean I think MNE or MNI (or even NDE) have no analytical value. These two quantitative units are necessary to estimation of such things as skeletal part frequencies and biomass, variables that are necessary to analyses requiring estimates of target variables *other than taxonomic abundances*. Having an explicit conception of the target variable one hopes to measure is, as I argued early in this paper, critical to a valid choice of the variable measured and the choice of zooarchaeological quantitative unit used to measure that variable.

**Acknowledgements** Helpful comments by five anonymous reviewers, particularly the two with very sharp eyes and pencils, helped me polish the text.

## References

- Adams, W. R. (1949). Faunal remains from the Angel Site. Master of Arts thesis. Bloomington: Indiana University.
- Adams, B. J., & Konigsberg, L. W. (2008). How many people? Determining the number of individuals represented by commingled human remains. In B. J. Adams & J. E. Byrd (Eds.), *Recovery, analysis, and identification of commingled human remains* (pp. 241–255). Totowa: Humana Press.
- Allen, J., & Guy, J. B. M. (1984). Optimal estimations of individuals in archaeological faunal assemblages: how minimal is the MNI? *Archaeology in Oceania*, 19, 41–47.

- Allen, M. S., & Nagaoka, L. A. (2004). "In the footsteps of von Haast. . . the discoveries something grand": the emergence of zooarchaeology in New Zealand. In L. Furey & S. Holdaway (Eds.), *Change through time: 50 years of New Zealand archaeology* (pp. 193–214). Auckland: New Zealand Archaeological Association Monograph 26.
- Audouze, F., & Enloe, J. G. (1997). High resolution archaeology at Verberic: limits and interpretations. *World Archaeology*, 29, 195–207.
- Badgley, C. (1986). Counting individuals in mammalian fossil assemblages from fluvial environments. *PALAIOS*, 1, 328–338.
- Bailey, G. (2007). Time perspectives, palimpsests and the archaeology of time. *Journal of Anthropological Archaeology*, 26, 198–223.
- Bennington, J. B., & Aronson, M. F. J. (2012). Reconciling scale in paleontological and neontological data: dimensions of time, space, and taxonomy. In J. Louys (Ed.), *Paleontology in ecology and conservation* (pp. 39–67). Berlin: Springer-Verlag.
- Binford, L. R. (1978). *Nunamiut ethnoarchaeology*. New York: Academic Press.
- Binford, L. R. (1981). *Bones: ancient men and modern myths*. Orlando: Academic Press.
- Binford, L. R. (1984). *Faunal remains from Klasies River mouth*. Orlando: Academic Press.
- Bobrowsky, P. T. (1982). An examination of Casteel's MNI behavior analysis: a reductionist approach. *Midcontinental Journal of Archaeology*, 7, 173–184.
- Brain, C. K. (1969). The contribution of the Namib Desert Hottentots to an understanding of australopithecine bone accumulations. *Scientific Papers of the Namib Desert Research Station*, 39, 13–22.
- Brain, C. K. (1981). *The hunters or the hunted? An introduction to African cave taphonomy*. Chicago: University of Chicago Press.
- Breitburg, E. (1991). Verification and reliability of NISP and MNI methods of quantifying taxonomic abundance: a view from historic site zooarchaeology. In J. R. Purdue, W. E. Klippel, & B. W. Styles (Eds.), *Beamers, bobwhites, and blue-points: tributes to the career of Paul W. Parmalee* (pp. 153–162). Springfield: Illinois State Museum Scientific Papers 23.
- Bunn, H. T. (1982). *Meat-eating and human evolution: studies on the diet and subsistence patterns of Plio-Pleistocene hominids in east Africa*. Doctoral dissertation. Berkeley: University of California.
- Cannon, M. D. (2013). NISP, bone fragmentation, and the measurement of taxonomic abundance. *Journal of Archaeological Method and Theory*, 20, 397–419.
- Carlson, C. C. (1999). History of zooarchaeology in New England. In M. A. Levine, K. E. Sassaman, & M. S. Nassaney (Eds.), *The archaeological Northeast* (pp. 171–181). Westport: Bergin & Garvey.
- Casteel, R. W. (1977a). Characterization of faunal assemblages and the minimum number of individuals determined from paired elements: continuing problems in archaeology. *Journal of Archaeological Science*, 4, 125–134.
- Casteel, R. W. (1977b). A consideration of the behaviour of the minimum number of individuals index: a problem in faunal characterization. *OSSA*, 3(4), 141–151.
- Chaplin, R. E. (1971). *The study of animal bones from archaeological sites*. London: Seminar Press.
- Clason, A. T. (1972). Some remarks on the use and presentation of archaeozoological data. *Helenium*, 12, 139–153.
- Crabtree, P. J. (2018). The value in studying large faunal collections using traditional zooarchaeological methods: a case study from Anglo-Saxon England. In C. M. Giovas & M. J. LeFebvre (Eds.), *Zooarchaeology in practice: case studies in methodology and interpretation in archaeofaunal analysis* (pp. 173–188). Cham: Springer.
- Dobney, K., & Rielly, K. (1988). A method for recording archaeological animal bones: the use of diagnostic zones. *Circa*, 5, 79–96.
- Dominguez-Rodrigo, M. (2012). Critical review of the MNI (minimum number of individuals) as a zooarchaeological unit of quantification. *Archaeological and Anthropological Sciences*, 4, 47–59.
- Driver, J. C. (1992). Identification, classification and zooarchaeology. *Circa*, 9(1), 35–47.
- Dunnell, R. C., & Dancey, W. S. (1983). The siteless survey: a regional scale data collection strategy. In M. B. Schiffer (Ed.), *Advances in archaeological method and theory* (Vol. 6, pp. 267–287). New York: Academic Press.
- Eck, G. G. (2007). The effects of collection strategy and effort on faunal recovery: a case study of the American and French collections from the Shungura Formation, Ethiopia. In R. Bobé, Z. Alemseged, & A. K. Behrensmeyer (Eds.), *Hominin environments in the East African Pliocene: an assessment of the faunal evidence* (pp. 183–215). Dordrecht: Springer.
- Enloe, J. G. (2003). Acquisition and processing of reindeer in the Paris basin. In S. Costamagno & V. Laroulandie (Eds.), *Zooarchaeological insights into Magdalenian lifeways* (pp. 23–31). Oxford: BAR International Series 1144.



- Enloe, J. G., & David, F. (1992). Food sharing in the Paleolithic: carcass refitting at Pincevent. In J. L. Hofman & J. G. Enloe (Eds.), *Piecing together the past: applications of refitting studies in archaeology* (pp. 296–315). Oxford: BAR International Series 578.
- Fieller, N. R. J., & Turner, A. (1982). Number estimation in vertebrate samples. *Journal of Archaeological Science*, 9, 49–62.
- Foley, R. (1981). Off-site archaeology: an alternative approach for the short-sited. In I. Hodder, G. Isaac, & N. Hammond (Eds.), *Pattern of the past: studies in honour of David Clarke* (pp. 157–183). Cambridge: Cambridge University Press.
- Ford, J. A. (1962). *A quantitative method for deriving cultural chronology. Technical Manual no. 1*. Washington, DC: Pan American Union.
- Gautier, A. (1984). How do I count you, let me count the ways? Problems in archaeozoological quantification. In C. Grigson & J. Clutton-Brock (Eds.), *Animals and archaeology 4: husbandry in Europe* (pp. 237–251). Oxford: BAR International Series 227.
- Gifford, D. P. (1981). In M. B. Schiffer (Ed.), *Taphonomy and paleoecology: a critical review of archaeology's sister disciplines* (Vol. 4, pp. 365–438). New York: Academic Press.
- Gifford-Gonzalez, D., & Hildebrandt, W. R. (2012). If mussels weighed a ton: problems with quantifying Santa Barbara Channel archaeofaunas. In M. A. Glassow & T. L. Joslin (Eds.), *Exploring methods of faunal analysis: insights from California archaeology* (pp. 97–107). Perspectives in California Archaeology Vol. 9. Los Angeles: Cotsen Institute of Archaeology, University of California.
- Gilinsky, N. L., & Bennington, J. B. (1994). Estimating numbers of whole individuals from collections of body parts: a taphonomic limitation of the paleontological record. *Paleobiology*, 20, 245–258.
- Giovas, C. M. (2009). The shell game: analytic problems in archaeological mollusc quantification. *Journal of Archaeological Science*, 36, 1557–1564.
- Giovas, C. M. (2018). Impact of analytic protocols on archaeofish abundance, richness, and similarity: a Caribbean-Pacific crossover study. In C. M. Giovas & M. J. LeFebvre (Eds.), *Zooarchaeology in practice: case studies in methodology and interpretation in archaeofaunal analysis* (pp. 59–89). Cham, Switzerland: Springer.
- Gobalet, K. W. (2001). A critique of faunal analysis: inconsistency among experts in blind tests. *Journal of Archaeological Science*, 28, 377–386.
- Grayson, D. K. (1973). On the methodology of faunal analysis. *American Antiquity*, 38, 432–439.
- Grayson, D. K. (1978). Minimum numbers and sample size in vertebrate faunal analysis. *American Antiquity*, 43, 53–65.
- Grayson, D. K. (1979). On the quantification of vertebrate archaeofaunas. In M. B. Schiffer (Ed.), *Advances in archaeological method and theory* (Vol. vol. 2, pp. 199–237). New York: Academic Press.
- Grayson, D. K. (1984). *Quantitative zooarchaeology: topics in the analysis of archaeological faunas*. Orlando, FL: Academic Press.
- Grayson, D. K., & Frey, C. J. (2004). Measuring skeletal part representation in archaeological faunas. *Journal of Taphonomy*, 2, 27–42.
- Guthrie, R. D. (1984). Alaskan megabucks, megabulls, and megagrams: the issue of Pleistocene gigantism. In H. H. Genoways & M. R. Dawson (Eds.), *Contributions in Quaternary vertebrate paleontology: a volume in memorial to John E. Guilday* (pp. 482–510). Special Publication no. 8, Pittsburgh, PA: Carnegie Museum of Natural History.
- Harris, M., Weisler, M., & Faulkner, P. (2015). A refined protocol for calculating MNI in archaeological molluscan shell assemblages: a Marshall Island case study. *Journal of Archaeological Science*, 57, 168–179.
- Hesse, B. (1982). Bias in the zooarchaeological record: suggestions for interpretation of bone counts in faunal samples from the Plains. In D. H. Ubelaker & H. J. Viola (Eds.), *Plains Indian studies: a collection of essays in honor of John C. Ewers and Waldo R. Wedel* (pp. 157–172). Washington, DC: Smithsonian Contributions to Anthropology no. 30.
- Hesse, B., & Wapnish, P. (1985). *Animal bone archeology: from objectives to analysis. Manuals in archeology 5*. Washington, DC: Taraxacum.
- Holdaway, S., & Wandsnider, L. (Eds.). (2008). *Time in archaeology: time perspectivism revisited*. Salt Lake City: University of Utah Press.
- Holtzman, R. C. (1979). Maximum likelihood estimation of fossil assemblage composition. *Paleobiology*, 5, 77–90.
- Horwitz, L. K. (2002). The development of archaeozoological research in Israel and the West Bank. *Archaeofauna*, 11, 131–145.
- Hudson, J. L. (1990). Advancing methods in zooarchaeology: an ethnoarchaeological study among the Aka. Doctoral dissertation. Santa Barbara: Department of Anthropology, University of California.

- Hudson, J. L. (1993). The impacts of domestic dogs on bone in forager camps; or, the dog-gone bones. In J. [L.] Hudson (Ed.), *From bones to behavior: ethnoarchaeological and experimental contributions to the interpretation of faunal remains* (pp. 301–323). Occasional Paper no. 21. Carbondale: Center for Archaeological Investigations, Southern Illinois University.
- Jamniczky, H. A., Brinkman, D. B., & Russell, A. P. (2008). How much is enough? A repeatable, efficient, and controlled sampling protocol for assessing taxonomic diversity and abundance in vertebrate microfossil assemblages. In J. T. Sankey & S. Baszio (Eds.), *Vertebrate microfossil assemblages: their role in paleoecology and paleobiogeography* (pp. 9–16). Bloomington: Indiana University Press.
- Jochim, M. A. (1976). *Hunter-gatherer subsistence and settlement: a predictive model*. New York: Academic Press.
- Joslin, T. L. (2012). Analytical sampling strategies for marine fish remains: measuring taxonomic diversity and abundance in central California middens. In M. A. Glassow & T. L. Joslin (Eds.), *Exploring methods of faunal analysis: insights from California archaeology* (pp. 135–147). Perspectives in California Archaeology Vol. 9. Los Angeles: Cotsen Institute of Archaeology, University of California.
- Klein, R. G., & Cruz-Uribe, K. (1984). *The analysis of animal bones from archeological sites*. Chicago: University of Chicago Press.
- Knüsel, C. J., & Outram, A. K. (2004). Fragmentation: the zonation method applied to fragmented human remains from archaeological and forensic contexts. *Environmental Archaeology*, 9(1), 85–97.
- Lambacher, N., Gerdau-Rodnic, K., Bonthorne, E., & de Tarazaga Montero, F. J. V. (2016). Evaluating three methods to estimate the number of individuals from commingled context. *Journal of Archaeological Science: Reports*, 10, 674–683.
- Lyman, R. L. (1984). Bone density and differential survivorship of fossil classes. *Journal of Anthropological Archaeology*, 3, 259–299.
- Lyman, R. L. (1985). Bone frequencies: differential transport, *in situ* destruction, and the MGUI. *Journal of Archaeological Science*, 12, 221–236.
- Lyman, R. L. (1994a). Quantitative units and terminology in zooarchaeology. *American Antiquity*, 59, 36–71.
- Lyman, R. L. (1994b). Relative abundances of skeletal specimens and taphonomic analysis of vertebrate remains. *PALAIOS*, 9, 288–298.
- Lyman, R. L. (1994c). *Vertebrate taphonomy*. Cambridge: Cambridge University Press.
- Lyman, R. L. (2006). Identifying bilateral pairs of deer (*Odocoileus* sp.) bones: how symmetrical is symmetrical enough? *Journal of Archaeological Science*, 33, 1256–1265.
- Lyman, R. L. (2008). *Quantitative paleozoology*. Cambridge: Cambridge University Press.
- Lyman, R. L. (2015a). The history of “laundry lists” in North American zooarchaeology. *Journal of Anthropological Archaeology*, 39, 42–50.
- Lyman, R. L. (2015b). On the variable relationship between NISP and NTA in bird remains and in mammal remains. *Journal of Archaeological Science*, 53, 291–296.
- Lyman, R. L. (2016). *Theodore E. White and the development of zooarchaeology in North America*. Lincoln: University of Nebraska Press.
- Lyman, R. L. (2017). Paleoenvironmental reconstruction from faunal remains: ecological basics and analytical assumptions. *Journal of Archaeological Research*, 25, 315–371.
- Lyman, R. L. (2018a). The history of MNI in North American zooarchaeology. In C. M. Giovas & M. LeFebvre (Eds.), *Zooarchaeology in practice: case studies in methodology and interpretation in archaeofaunal analysis* (pp. 13–33). New York: Springer.
- Lyman, R. L. (2018b). Observations on the history of zooarchaeological quantitative units: why NISP, then MNI, then NISP again? *Journal of Archaeological Science: Reports*, 18, 43–50.
- Lyman, R. L., & Ames, K. M. (2007). On the use of species-area curves to detect the effects of sample size. *Journal of Archaeological Science*, 34, 1985–1990.
- Lyman, R. L., & VanPool, T. (2009). Metric data in archaeology: a study of intra-analyst and inter-analyst variation. *American Antiquity*, 74, 485–504.
- Marean, C. W., & Frey, C. J. (1997). Animal bones from caves to cities: reverse utility curves as methodological artifacts. *American Antiquity*, 62, 698–711.
- Medlock, R. C. (1976). Determining the minimum number of individuals in archeological faunal analysis. Master of Arts thesis. Fayetteville: Department of Anthropology, University of Arkansas.
- Moore, J. R., Norman, D. B., & Upchurch, P. (2007). Assessing relative abundances in fossil assemblages. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 253, 317–322.
- Morin, E., Ready, E., Boileau, A., Beauval, C., & Coumont, M.-P. (2017a). Problems of identification and quantification in archaeozoological analysis, part I: insights from a blind test. *Journal of Archaeological Method and Theory*, 24, 886–937.

- Morin, E., Ready, E., Boileau, A., Beauval, C., & Coumont, M.-P. (2017b). Problems of identification and quantification in archaeozoological analysis, part II: presentation of an alternative counting method. *Journal of Archaeological Method and Theory*, 24, 938–973.
- Morlan, R. E. (1994). Bison bone fragmentation and survivorship: a comparative method. *Journal of Archaeological Science*, 21, 797–807.
- Münzel, S. C. (1986). Coding system for bone fragments. In L. H. Van Wijngaarden-Bakker (Ed.), *Database management and zooarchaeology* (pp. 193–195). PACT 14. Strasbourg.
- Münzel, S. C. (1988). Quantitative analysis and archaeological site interpretation. *ArchaeoZoologia*, 2(1, 2), 93–110.
- Nichol, R. K., & Creak, G. A. (1979). Matching paired elements among archaeological bone remains: a computer procedure and some practical limitations. *Newsletter of Computer Archaeology*, 14, 6–17.
- Parmalee, P. W. (1985). Identification and interpretation of archaeologically derived animal remains. In R. I. Gilbert & J. H. Mielke (Eds.), *The analysis of prehistoric diets* (pp. 61–95). New York: Academic Press.
- Payne, S. (1975). Partial recovery and sample bias. In A. T. Clason (Ed.), *Archaeozoological studies* (pp. 7–17). Amsterdam: North-Holland.
- Perkins Jr., D. (1973). A critique on the methods of quantifying faunal remains from archaeological sites. In J. Matolcsi (Ed.), *Domestikationsforschung und geschichte der haustiere* (pp. 367–370). Budapest: Academiai Kiado.
- Plog, F. T. (1973). Diachronic anthropology. In C. L. Redman (Ed.), *Research and theory in current archeology* (pp. 181–198). New York: John Wiley and Sons.
- Plog, F. T. (1974). *The study of prehistoric change*. New York: Academic Press.
- Plug, C., & Plug, I. (1990). MNI counts as estimates of species abundance. *South African Archaeological Bulletin*, 45, 53–57.
- Pollock, H. E. D., & Ray, C. E. (1957). Notes on vertebrate animal remains from Mayapan. Department of Archaeology, Carnegie Institution of Washington. Current Reports, 41, 633–656.
- Rackham, D. J. (1986). Assessing the relative frequency of species by the application of a stochastic model to a zooarchaeological database. In L. H. van Wijngaarden-Bakker (Ed.), *Database management and zooarchaeology* (pp. 185–192). Strasbourg: PACT 14.
- Reitz, E. J. (1993). Zooarchaeology. In J. K. Johnson (Ed.), *The development of Southeastern archaeology* (pp. 109–131). Tuscaloosa: University of Alabama Press.
- Reitz, E. J., & Wing, E. S. (2008). *Zooarchaeology* (second ed.). Cambridge: Cambridge University Press.
- Ringrose, T. J. (1993). Bone counts and statistics. *Journal of Archaeological Science*, 20, 121–157.
- Robb, J. (2016). What can we really say about skeletal part representation, MNI and funerary ritual? A simulation approach. *Journal of Archaeological Science: Reports*, 10, 684–692.
- Robison, N. D. (1987). Zooarchaeology: its history and development. In A. E. Bogan & N. D. Robison (Eds.), *The zooarchaeology of eastern North America: history, method and theory, and bibliography* (pp. 1–26). Miscellaneous Paper no. 12. Knoxville: Tennessee Anthropological Association.
- Rogers, R. R., Carrano, M. T., Rogers, K. A. C., Perez, M., & Reagan, A. K. (2017). Isotaphonomy in concept and practice: an exploration of vertebrate microfossil bonebeds in the upper cretaceous (Campanian) Judith River Formation, north-central Montana. *Paleobiology*, 43, 248–273.
- Rosell, J., Blasco, R., Fernández-Laso, M. C., Vaquero, M., & Carbonell, E. (2012). Connecting areas: faunal refits as a diagnostic element to identify synchronicity in the Abric Romani archaeological assemblages. *Quaternary International*, 252, 56–67.
- Schiffër, M. B. (1987). *Formation processes of the archaeological record*. Albuquerque: University of New Mexico Press.
- Smith, B. D. (1975). *Middle Mississippi exploitation of animal populations*. Anthropological Papers no. 57. Ann Arbor: Museum of Anthropology, University of Michigan.
- Staff, G., Powell, E. N., Stanton, R. J., & Cummins, H. (1985). Biomass: is it a useful tool in paleocommunity reconstruction? *Lethaia*, 18, 209–232.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677–680.
- Stewart, K. M. (2002). Past and present zooarchaeology in Canada. *Archaeofauna*, 11, 147–157.
- Thomas, D. H. (1975). Nonsite sampling in archaeology: up the creek without a site? In J. W. Mueller (Ed.), *Sampling in archaeology* (pp. 61–81). Tucson: University of Arizona Press.
- Thomas, K. D., & Mannino, M. A. (2017). Making numbers count: beyond minimum numbers of individuals (MNI) for the quantification of mollusc assemblages from shell matrix sites. *Quaternary International*, 427, 47–58.
- Todd, L. C. (1987). Taphonomy of the Horner II bone bed. In G. C. Frison & L. C. Todd (Eds.), *The Horner Site: the type site of the Cody Cultural Complex* (pp. 107–198). Orlando: Academic Press.

- Todd, L. C., & Frison, G. C. (1992). Reassembly of bison skeletons from the Horner Site: a study in anatomical refitting. In J. L. Hofman & J. G. Enloe (Eds.), *Piecing together the past: applications of refitting studies in archaeology* (pp. 63–82). Oxford: BAR International Series 578.
- Todd, L. C., & Stanford, D. J. (1992). Application of conjoined bone data to site structural studies. In J. L. Hofman & J. G. Enloe (Eds.), *Piecing together the past: applications of refitting studies in archaeology* (pp. 21–35). Oxford: BAR International Series 578.
- Travouillon, K. J., Archer, M., Legendre, S., & Hand, S. J. (2007). Finding the minimum sample richness (MSR) for multivariate analyses: implications for palaeoecology. *Historical Biology*, *19*, 315–320.
- Trusler, A. K. (2014). The impact of recovery methods on taxonomic richness in Roman faunal assemblages. *Archaeometry*, *56*, 1075–1084.
- Vermeij, G. J., & Herbert, G. S. (2004). Measuring relative abundance in fossil and living assemblages. *Paleobiology*, *30*, 1–4.
- Voorhies, M. R. (1969). *Taphonomy and population dynamics of an early Pliocene vertebrate fauna, Knox County, Nebraska*. Contributions to Geology, Special Paper no. 1. Laramie: University of Wyoming.
- Watson, J. P. N. (1979). The estimation of the relative frequencies of mammalian species: Khirokitia 1972. *Journal of Archaeological Science*, *6*, 127–137.
- White, T. E. (1953). A method of calculating the dietary percentage of various food animals utilized by aboriginal peoples. *American Antiquity*, *19*, 396–398.
- Wild, C. J., & Nichol, R. K. (1983). Estimation of the original number of individuals from paired bone counts using estimators of the Krantz type. *Journal of Field Archaeology*, *10*, 1337–1344.
- Zar, J. H. (1996). *Biostatistical analysis* (third ed.). Upper Saddle River, New Jersey: Prentice Hall.