*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 1–2
DOI: 10.1214/08-AOAS162


# EDITORIAL: STATISTICS AND "THE LOST TOMB OF JESUS"

BY STEPHEN E. FIENBERG

*Carnegie Mellon University*

What makes a problem suitable for statistical analysis? Are historical and religious questions addressable using statistical calculations? Such issues have long been debated in the statistical community and statisticians and others have used historical information and texts to analyze such questions as the economics of slavery, the authorship of the Federalist Papers and the question of the existence of God. But what about historical and religious attributions associated with information gathered from archeological finds?

In 1980, a construction crew working in the Jerusalem neighborhood of East Talpiot stumbled upon a crypt. Archaeologists from the Israel Antiquities Authority came to the scene and found 10 limestone burial boxes, known as ossuaries, in the crypt. Six of these had inscriptions. The remains found in the ossuaries were reburied, as required by Jewish religious tradition, and the ossuaries were catalogued and stored in a warehouse. The inscriptions on the ossuaries were catalogued and published by Rahmani (1994) and by Kloner (1996) but there reports did not receive widespread public attention.

Fast forward to March 2007, when a television "docudrama" aired on *The Discovery Channel* entitled "The Lost Tomb of Jesus"[1] touched off a public and religious controversy—one only need think about the title to see why there might be a controversy! The program, and a simultaneously published book [Jacobovici and Pellegrino (2007)], described the "rediscovery" of the East Talpiot archeological find and they presented interpretations of the ossuary inscriptions from a number of perspectives. Among these was a statistical calculation *attributed to* the statistician Andrey Feuerverger: "that the odds that all six names would appear together in one tomb are 1 in 600, calculated conservatively—or possibly even as much as one in one million."

At about this time, Feuerverger submitted a paper to *The Annals of Applied Statistics* (AOAS) for review, but its contents remained confidential and only a rough outline of the details of his calculations was publicly available [Mims (2007)]. Commentary regarding Feuerverger's statistical calculation quickly appeared on the web. Was it really a Bayesian calculation? On what assumptions were the statistical arguments based? Most criticism focused not directly on the actual statistical arguments but on how they were portrayed by the documentary's producers

Received January 2008; revised January 2008.

[1] "The Lost Tomb of Jesus." *Discovery Channel*, March 4, 2007. http://dsc.discovery.com/convergence/tomb/tomb.html.

and interpreted by others. And the controversy over the broader interpretation and claims regarding the origin of the East Talpiot tomb raged on.

In July 2007 at the Joint Statistical Meetings in Salt Lake City, Feuerverger gave the first public airing of the details of his work and three discussants presented alternative perspectives. The paper itself underwent an extensive review process and a substantially revised version appears in this issue of AOAS [Feuerverger (2008)]. It includes photographs, detailed discussion of possible data on names from ancient sources, the assumptions upon which the analysis was based, and a novel $p$-value calculation. The paper is accompanied by a series of detailed discussions and critiques, several of which reframe the statistical problem from a Bayesian perspective.

The AOAS editors encourage our readers to judge for themselves the persuasiveness of the assumptions, the data, and the calculations performed by Feuerverger, especially in light of the criticisms voiced in the extended discussion that follows his paper, and his response. Interested readers may then wish to explore the extensive nonstatistical discussion of the East Talpiot available in print and on the web.

## REFERENCES

FEUERVERGER, A. (2008). Statistical analysis of an archeological find (with discussion). *Ann. Appl. Statist.* **2** 3–54.

JACOBOVICI, S. and PELLEGRINO, C. (2007). *The Jesus Family Tomb*: *The Discovery*, *the Investigation, and the Evidence That Could Change History.* Harperone, New York.

KLONER, A. (1996). A tomb with inscribed ossuaries in East Talpiyot, Jerusalem. *Atiqot* **29** 15–22.

MIMS, C. (2007). Q&A with the statistician who calculated the odds that this tomb belonged to Jesus. *Scientific American In Focus*, March 2, 2007. Available at http://sciam.com/article.cfm?articleID=13C42878-E7F2-99DF-3B6D16A9656A12FF.

RAHMANI, L. Y. (1994). *A Catalogue of Jewish Ossuaries in the Collections of the State of Israel.* Israel Antiquities Authority, Israel Academy of Sciences and Humanities, Jerusalem.

DEPARTMENT OF STATISTICS AND MACHINE LEARNING
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213
USA
E-MAIL: fienberg@stat.cmu.edu

*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 3–54
DOI: 10.1214/08-AOAS99


# STATISTICAL ANALYSIS OF AN ARCHEOLOGICAL FIND[1]

BY ANDREY FEUERVERGER

*University of Toronto*

In 1980, a burial tomb was unearthed in Jerusalem containing ossuaries (limestone coffins) bearing such inscriptions as Yeshua son of Yehosef, Marya, Yoseh—names which match those of New Testament (NT) figures, but were otherwise in common use. This paper discusses certain statistical aspects of authenticating or repudiating links between this find and the NT family. The available data are laid out, and we examine the distribution of names (onomasticon) of the era. An approach is proposed for measuring the "surprisingness" of the observed outcome relative to a "hypothesis" that the tombsite belonged to the NT family. On the basis of a particular—but far from uncontested—set of assumptions, our measure of "surprisingness" is significantly high.

**1. Introduction and summary.** In March 1980, the Solel Boneh Construction Company interrupted excavation work at an apartment site complex in the East Talpiyot neighbourhood of Jerusalem, and reported to Israel's Department of Antiquities and Museums that it had accidentally unearthed a previously unknown entrance to a burial cave. This tomb is located approximately 2.5 kilometers south of the site of the Second Temple in the Old City of Jerusalem, destroyed by the Romans in 70 CE.[2]

Shortly after its discovery, this burial site was examined and surveyed and salvage excavations were carried out. Within this cave a number of ossuaries[3] were found, some bearing inscriptions, and these were published by Rahmani (1994), pages 222–224, Nos. 701–709 and by Kloner (1996). Among these ossuaries were found such inscriptions as "Marya," "Yoseh," "Yeshua son of Yehosef," and other inscriptions of related interest.

Since the practice of ossuary burial was prevalent among Jews at the time Jesus of Nazareth was crucified in Jerusalem at the behest of the Romans, archeological

Received April 2007; revised December 2007.

[1]Discussed in 10.1214/08-AOAS99A, 10.1214/08-AOAS99B, 10.1214/08-AOAS99C, 10.1214/08-AOAS99D, 10.1214/08-AOAS99E, 10.1214/08-AOAS99F, 10.1214/08-AOAS99G and 10.1214/08-AOAS99H; rejoinder at 10.1214/08-AOAS99REJ.

*Key words and phrases.* Jesus of Nazareth, distribution of names, onomasticon, data, statistical inference, conditioning, coincidence, "relevance," "rareness" and "surprisingness," tail areas, historical assumptions, a priori, hypotheses and post hoc inference.

[2]CE and BCE are abbreviations for "common era" and "before the common era"—secular versions of the abbreviations AD and BC.

[3]Ossuaries are repositories for bones; see Section 3.

questions arise in respect of the identity of the individuals buried in this tomb. Since names such as Yehosef, Marya, Yeshua, etc., were not uncommon during the era in which such burials took place, the task of assessing whether or not these ossuaries might be those of the New Testament (NT) family is not straightforward.

Several disciplines bear on assessing the authenticity of such findings, including chemical spectroscopy for analyzing and dating patinas, epigraphic and paleographic examination by specialists in ancient semitic script, and DNA analysis of any remains, not to mention historical scholarship of early Christianity. Any tampering with the tombsite or other possibilities for fraud must also be weighed and taken into account.

One purpose of this article is to contribute toward such efforts by developing statistical methods for assessing evidence for and against a "hypothesis" that this tomb belonged to the family of the historical Jesus. In doing so we consider such data as are available on the distribution of names during the era in question, and we compute (on the basis of numerous assumptions detailed explicitly) probabilities and estimates related to such questions as the expected proportion of times that a similarly "surprising" sample of names could be expected to arise by pure chance when sampling from a population having similar characteristics to the one which existed at that time. Our computations were carried out under a specific set of assumptions which are by no means universally accepted. Of course, ultimately, the authenticity of any such find cannot be determined through purely statistical reasoning alone, and it can certainly turn out that this tombsite is not that of the NT family; in that eventuality the validity of our *methods* should remain unaffected. A further purpose of this paper is to lay out this highly interesting data set—together with the novel inferential challenges it poses—for the benefit of the statistical community.

In Section 2 below we describe the unearthed tomb and the ossuaries discovered inside. Background on the practice of ossuary interment is given in Section 3. The genealogy of the NT family—central to our analysis—is discussed briefly in Section 4. Section 5 discusses available data sources and provides some statistical summaries of the Jewish onomasticon, that is, of the distribution of names of the men and women who lived during that era, and Section 6 follows up in further detail for the particular names found in the East Talpiyot tomb. Some statistical "judgement calls" are discussed in Section 7. Because the Talpiyot tomb must be regarded as having been "best" out of many possible observations, in Section 8 we review what is known about the size of the relevant population within which these burials took place. Section 9 addresses some inferential issues which arise in data of this type.

For statistical inference to be valid, one may not tailor an alternative hypothesis to data that has already been seen. In Section 10 we address such matters and on a best efforts basis we carefully formulate a priori hypotheses for this problem. A paradigm for the inference problem at hand is then developed in Sections 11 and 12. Our method is based on defining an a priori measure of the "surprisingness" of an observation using the "relevance and rareness" of certain name rendi-

tions, and an assumed complex of NT familial relations among them. "Relevance" will refer essentially to membership in an a priori list of *candidates* for inclusion in a NT tombsite, and "rareness" will be defined relative to an a priori list of nested possible name renditions for each such candidate; features of familial interrelations figure prominently in the formulation. Our analysis, implemented for a variety of parameter choices, is reported in Section 13 which first provides a detailed summary of the assumptions underlying our analysis. In Section 14 we provide a detailed discussion of our results, and some concluding remarks. The R computing code on which our results are based may be downloaded from the "statlib" website [Feuerverger (2008)].

We remark that, in assessing the evidence in any way, it is essential to adopt a strictly *historical* viewpoint, and thus to set aside considerations that a NT tombsite cannot exist. In fact, Jewish ritual observances prevalent at the time are entirely consistent with the possible existence of such a tomb. We caution the reader to note, however, that the analysis we present is based on one specific "tradition" of history. These assumptions represent the author's best understanding as at the time the analysis was carried out but they are far from universally agreed upon and they enter into the analysis in a cumulative way. It is anticipated that such points will be revisited in the discussion to this paper.

**2. Description of the find.** The vestibule of the tomb was damaged by the blasting operations that led to its discovery. The tomb had otherwise been covered by earth, apparently undisturbed since antiquity. On the exterior facade above the tomb's entranceway there was carved in relief a circle beneath an upward pointing gable—a rare feature. Within the $2.3 \times 2.3$ m tomb were six *kokhim*[4]—two on each of the other three walls—each just over 1.6 m in length, and under 0.5 m in width, deep enough to store two or three ossuaries in each. Within these *kokhim* a total of ten ossuaries were found,[5] some of them broken. Two ossuary lids, discarded in antiquity, were found beneath the soil fill in the room. Early Roman (Herodian) sherds (i.e., broken pieces of pottery) were also found on the floor which date the site to the late Second Temple period, that is, from the end of the first century BCE or the beginning of the first century CE to approximately 70 CE. Such bones as were within the ossuaries were in an advanced state of disintegration. Two arcosolia (shallow shelves intended for laying out bodies) had been carved in the tomb walls and contained broken and powdered bone remains. Disturbed bones, presumably swept off the arcosolia, were also found on the floor. The *golal* (blocking stone) to the tomb's entrance was not found at the site indicating that the tomb had been accessed by robbers in antiquity.

[4] *Kokhim* (singular *kokh*) are small horizontal tunnels chiseled into the walls of a tomb within which ossuaries could be placed. The Latin terms are *loculus* and *loculi*.

[5] No information is available regarding the placement of the various ossuaries among the *kokhim*.

The ossuaries found within this burial cave are typical of Jewish ossuaries of the first century CE. Six of the ten ossuaries bore inscriptions, five in Jewish script (i.e., Hebrew or Aramaic) and one in Greek. This proportion of inscribed ossuaries (i.e., 6 out of 10) and this proportion of Hebrew to Greek (5 out of 6) are both higher than typical of other tombs previously found in this area. The six inscribed ossuaries and the four uninscribed ones are described below in the order they appear in Kloner (1996); their Israel Antiquities Authority (IAA) identification numbers and dimensions are indicated as well.

OSSUARY #1. IAA 80–500. 68.5 × 26 × 32.5 cm. Inscribed in Greek:

$$Μαριαμηνου\ [η]\ Μαρα$$

This elegantly rendered ossuary (see Figure 1) has multiple possible readings. Mara, an (absolute) contracted form of (the emphatic) Martha, is a rare name, these being feminine versions derived from the Aramaic dominant masculine form mar meaning "lord," "master," or "honorable person." The question of whether Mara was intended here as a title, such as "honorable lady," or whether it was intended only as an alternate (i.e., second) name is disputed. If this inscription were understood as in Hebrew, then Mariamenou would be a diminutive (i.e., endearing) form of Mariamne or Mariamene and the inscription would read "Mariamene [diminutive] the lord/master" provided we also assume also that *Μαρα* (or מרא)

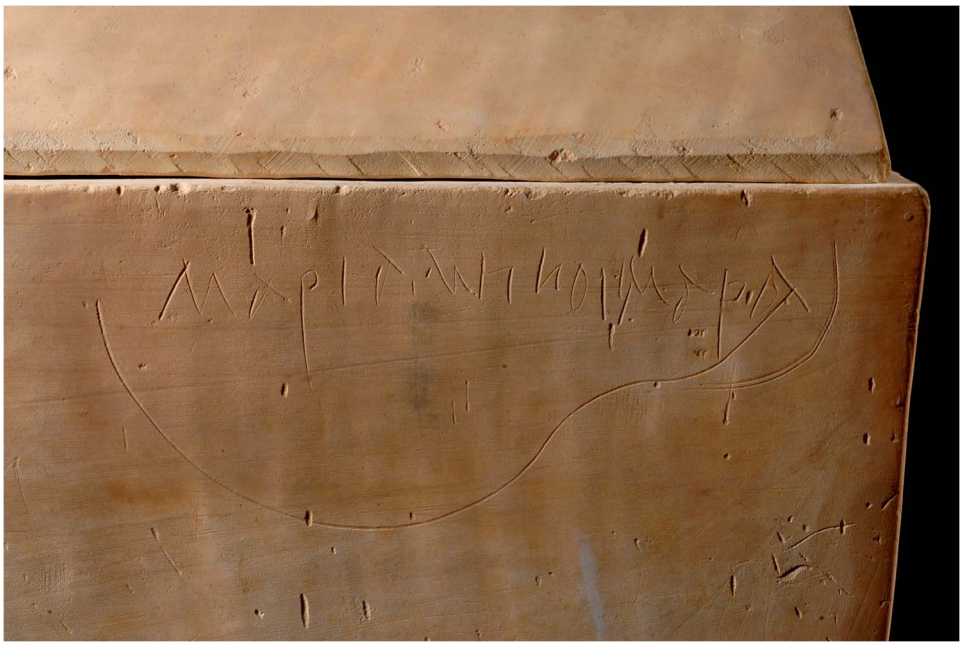

FIG. 1.

is intended as "lord" or "master" and that "$\eta$" is meant as the feminine article "the." An alternate reading requires that one interpret the stroke between "Mariamenou" and "Mara" as representing not an $\eta$, but only a scratch mark; in that case one interpretation is that this ossuary contains the remains of two persons—one called Mariame, and the other called Mara. However, the manner in which these two words run closely together, and on the same line, seems more suggestive of their referring to a single person. Rahmani (1994), pages 14 and 222, reads the inscription as follows: "The stroke between the $\upsilon$ of the first and the $\mu$ of the second name probably represents an $\eta$, standing here for the usual $\eta$ $\kappa\alpha\iota\ldots$ used in the case of double names…" and he posits that the second name is a contracted form [not a contraction] of "Martha" leading to the reading "Mariamene [diminutive] who is also called Mara." According to Greek usage of the time, the first word of the inscription is a genitive/possessive form for Mariamene, rendered in a particular diminutive form understood to be an endearment, so that the inscription then translates as "[the ossuary] of Mariamene [diminutive] also known as Mara." Rahmani's reading, which is the one we adopt, was accepted by Kloner (1996) and has been corroborated by others in the field.

OSSUARY #2. IAA 80–501. 55 × 23 × 27 cm. Inscribed in Hebrew lettering:

יהודה בר ישוע

The lettering is executed clearly—see Figure 2. It translates as "Yehuda son of Yeshua," Yehuda being Hebrew for Judah. Note that "bar" (i.e., בר "son of") is Aramaic, not Hebrew.

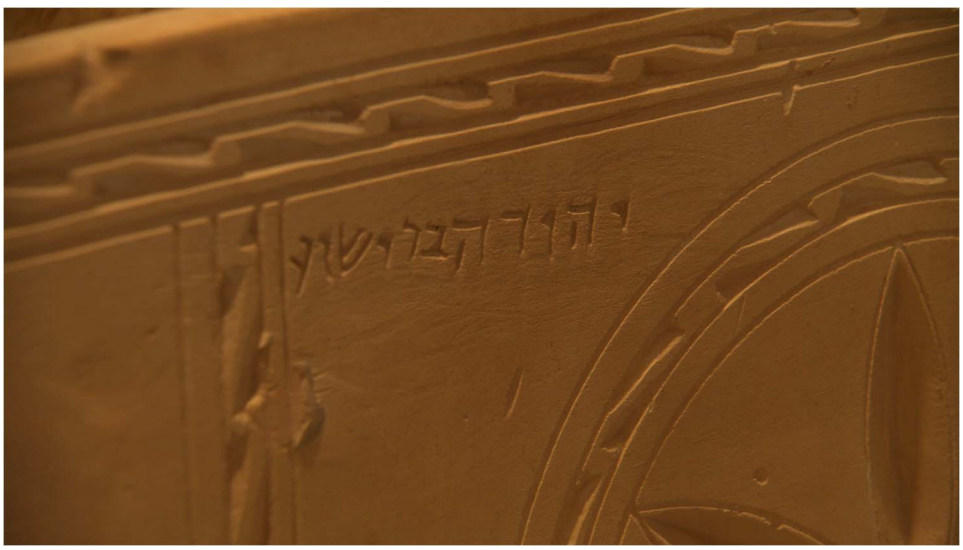

FIG. 2.

FIG. 3.

OSSUARY #3. IAA 80–502. 55 × 28 × 34 cm. Inscribed in Hebrew:

מתיה

This translates as Matya, a shortened form of Mattityahu, that is, Matthew; see Figure 3.

OSSUARY #4. IAA 80–503. 65 × 26 × 30 cm. Inscribed in Hebrew lettering:

ישוע בר יהוסף

This translates as Yeshua son of Yehosef, that is, Jesus son of Joseph. Unlike the other inscribed ossuaries found in this tomb, the incisions here are "messy," "informal," and superficially carved, and each of the four letters of ישוע is faint; see Figure 4. However this reading of the inscription was authenticated (by Rahmani and also Kloner) by comparison with the inscription on Ossuary #2 and is corroborated by others. Also relevant is that no other Hebrew name ends in the letters *vov* and *ayin*. A large, crudely carved rightward-leaning cross, whose purpose or symbolic meaning (if any) is unknown, appears at the head of the inscription. Cross-marks on ossuaries were sometimes carved by masons, most likely to indicate alignment of lid-tops; in this instance the marking does not have the appearance of being an *obvious* scratch mark of this nature. It has been suggested that the "cross" on this ossuary may have been purposeful.

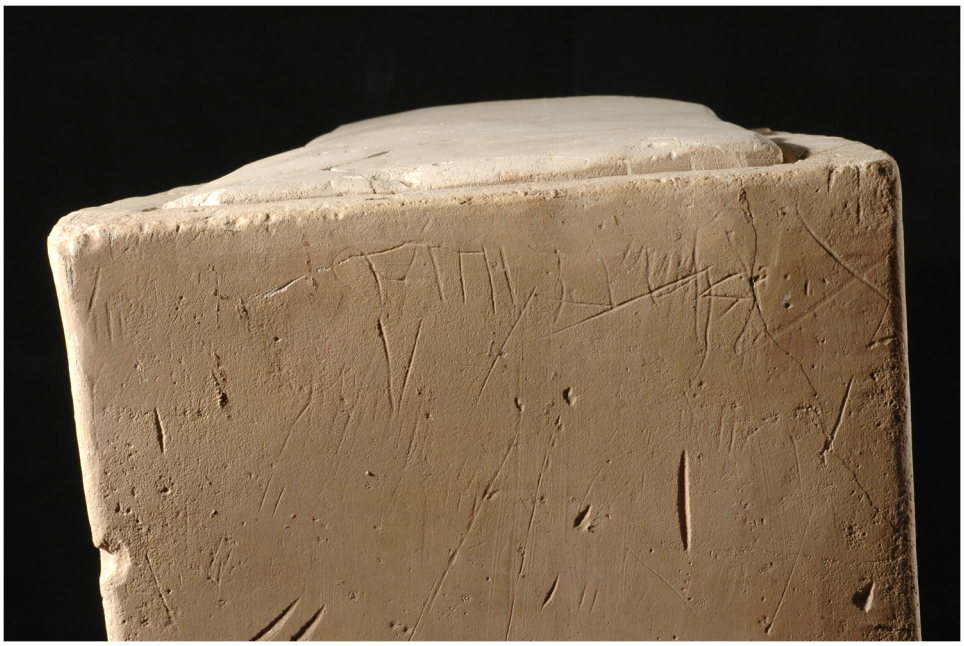

FIG. 4.

OSSUARY #5. IAA 80–504. 54.5 × 26 × 34.5 cm. Inscribed in Hebrew:

יוסה

This translates as Yoseh or Yosa, a relatively rare variant of Yosef or Yehosef (i.e., Joseph). In Hellenized form, this inscription would be read as Yose, Yoses, or Joses. See Figure 5.

OSSUARY #6. IAA 80–505. 52 × 27 × 33 cm. Inscribed in Hebrew:

מריה

This translates as Marya, that is, Maria, a Hellenized form of Miriam or Mariam. See Figure 6.

OSSUARY #7–10. These four ossuaries, the first three of which correspond to IAA numbers 80–506 to 80–508, bear no inscriptions and have dimensions 67 × 31.5 × 38.5 cm, 51 × 27 × 31.5 cm, 61 × 26.5 × 31.5 cm, and (the reported dimensions) 60 × 26 × 30 cm, respectively.

In general appearance, the six inscriptions correspond to four distinct styles. That of Yeshua is unprofessional. The ossuaries of Marya, Yoseh, and Matya are

FIG. 5.

executed in similar plain but neat hands. That of Mariamenou is executed in an "elegant" Greek hand. And finally, the ossuary of Yehuda appears rendered "pro-

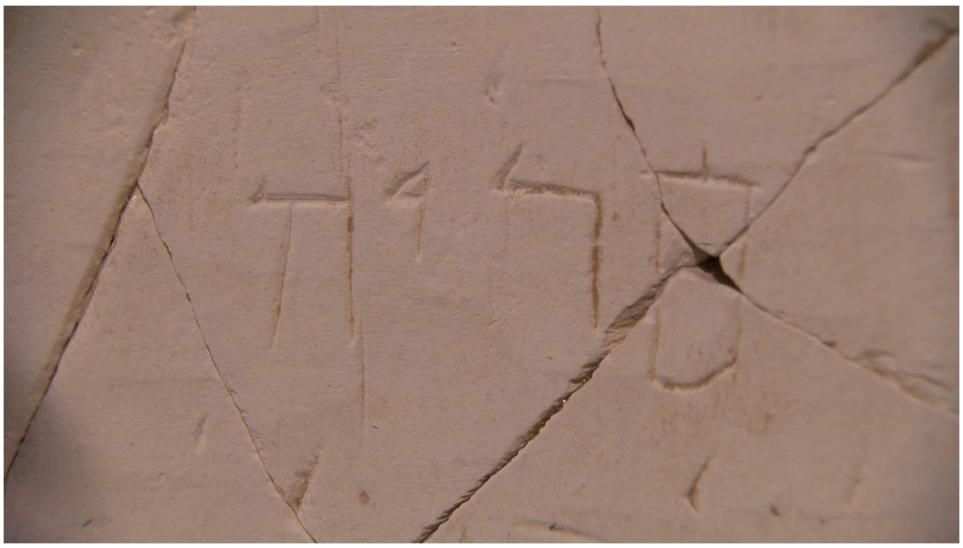

FIG. 6.

fessionally." Rahmani surmised that the similarities between Ossuaries #5 and #6 and their inscriptions, both coming from the same tomb, may indicate that Yoseh and Marya were the parents of Yeshua and the grandparents of Yehuda.[6]

Although the dimensions of the ossuaries differ, each is consistent with the measurements of an adult. Among the inscribed ossuaries, numbers 1 and 4 (Mariamenou and Yeshua) are the longest, possibly corresponding to taller than average persons, and numbers 1 and 2 bore ornamental carvings (rosettes, etc.) as did also the first three of the four uninscribed ossuaries listed; all of the other ossuaries were ornament-free, except for such inscriptions as have been noted.

Finally, we note that the tenth ossuary—that is, the uninscribed, unornamented one with dimensions $60 \times 26 \times 30$ cm—is "missing." The original archeological drawings made at the time of the excavation indicate that ten ossuaries were found at the site, but IAA records show that only nine were retained in its permanent collections. Now, it is not entirely unusual that an ossuary—particularly an uninteresting one—would get "lost" in the comings and goings of such archeological work. However, suggestions have been raised [e.g., Tabor (2006), among others] that the dimensions of the missing ossuary seemingly match those of the disputed ossuary of James.[7] Were this so, statistical dimension matching[8] could easily be used to *prove* that the James ossuary must surely be the one missing from our tomb, with attending consequences that would be startling, particularly if the *full* inscription on the James ossuary were authenticated. Our investigations along these lines, however, did not prove fruitful.

In Sections 5 and 6 we shall discuss the distribution of Jewish names in late antiquity and provide some further details concerning the names found on the Ossuaries #1 through #6. The next two sections provide some background on the practice of ossuary burial, and on the genealogy of the NT family.

**3. Ossuaries and re-interment.** An ossuary is an approximately rectangular chest, typically quarried in the soft limestones common near Jerusalem, containing the bones of one (and sometimes more) deceased persons. The custom of repositing bones of the dead in such bone boxes is not mandated by *halacha*, that is, Jewish ritual law; it was practiced by Jews in and around Jerusalem only from the

[6]If this interpretation is correct, the tombsite cannot be that of the NT family. However Rahmani does not follow up with any explanation for the messy nature of the inscription on Ossuary #4.

[7]An ossuary inscribed "James son of Joseph brother of Jesus" is in the possession of Oded Golan, a private Israeli antiquities collector, under prosecution for alleged forgery at the time this article was written; see, for example, Shanks and Witherington (2003). Israeli prosecutors apparently accept the authenticity of the first component of this inscription but allege that the second component had been forged, although (as of the time of writing) no evidence to that effect has been produced. The statistical aspects of the "James son of Joseph brother of Jesus" inscription were studied by Fuchs (2004).

[8]Rahmani (1994) gives the dimensions of a sample of 897 ossuaries from which the multivariate distribution of dimensions can, for this purpose, be quite reliably inferred.

end of the first century BCE, or from the start of the first century CE, until the year 70 CE. Instead of burial in coffins as had been an earlier custom, bodies were apparently first placed in a pit or a cave and left to decompose for about a year until only bones remained. These bones were then gathered by the deceased's family, deposited into an ossuary, and interred in a tomb. Ossuaries (and tombs in particular) were a more costly form of burial that not all persons could afford. Further information and speculation regarding the religious and politico-historical aspects of this practice, may be found in Hachlili (1994), Kloner (1996) and Rahmani (1994).

The approximate dimensions of ossuaries are usually recorded in centimeters in the order length $\times$ width $\times$ height. Typical ossuary boxes are somewhat tapered so that the length $\times$ width dimensions at the top will be slightly larger than at the bottom. Being quarried and chiseled artifacts, the shape, and hence the dimensions, of ossuary boxes are not entirely precise. The length of an ossuary had to be sufficient to house the femur (thighbone) which is the human body's longest bone, and the two other dimensions had to be sufficient to house the skull, pelvis, and other bones.

Ossuaries were frequently carved with ornamental motifs such as lattices, friezes, triglyphs, or rosettes. Such markings would typically have helped identify the persons lying within, especially for ossuaries that were uninscribed (as might occur, e.g., in families lacking literacy).

Rahmani (1994) notes that 233 of the 897 ossuaries in the State of Israel collections as of 1989 bear inscriptions[9] meant to identify the individuals within, with these inscriptions being in one or more of the languages in common use at the time—primarily Aramaic, Hebrew and Greek. About two thirds of these inscriptions are in Hebrew/Aramaic, while about one third are in Greek, or a combination of Greek and Hebrew. The use of inscriptions evidence some degree of literacy on the part of the family to whom the tomb complex belonged. In virtually all cases, such inscriptions consist only of the individual's first name,[10] or of their first name together with the name of their father. Inscriptions for women occasionally included the name of the husband in lieu of the father. Only a single case among the ossuaries catalogued by Rahmani includes the name of a brother, and only one mentions the name of a son; such rare mentions presumably occurred only when the other mentioned persons were individuals of particular distinction. Contractions of names appear also to have been used, and were likely intended as

[9]Because plain ossuaries are of lesser interest and often become "discarded," these figures significantly overstate the inscription rate.

[10]According to *halachah* the name marked on a grave must correspond to the actual name by which that person was known during their lifetime. In particular, if an individual had been known by a nickname, that form of their name must be used on their coffin. Note that although *halachah* postdates the era of Jesus we are assuming here that this basic tenet was already essentially being observed at that time.

endearments. Note that the use of inscriptions was intended solely to assist members of the immediate family to identify the remains within; they served no public or other function.

Ossuary burial was practiced primarily within the environs of Jerusalem in part, no doubt, because of the availability of suitable stone there. In fact [Ilan (2002), page 52], of the 712 names in Ilan derived from ossuaries sources, only 66 were found outside of the Jerusalem region, with 24 of these having come from a single burial cave in Jericho. Rahmani (1994), page 21, notes that ossuaries quarried at Jerusalem were also used by Jews living as far away as 25 km from Jerusalem (including Jericho).

**4. A brief NT genealogy.** The names in the genealogy of the NT family bear on the statistical analysis; however, our discussion here will be brief. We caution the reader that our analysis relies on a specific "tradition" for this geneology but that such historical details cannot be regarded as being certain.

Jesus was born a few years before the turn of the millennium and was crucified in (most likely) April of 30 CE. The earliest known historical record of the names of Jesus' siblings is provided by Mark 6:3 (written around 70 CE) who lists the names of Jesus' brothers in the order James, Joses, Judah, and Simon. Since it was customary to name the eldest first, it is reasonable to assume that James was Jesus' eldest brother, and Joses was next eldest. Matthew 13:54–56, usually believed to be a historically later source, records the names in the order James, Joseph, Simon, and Judah—using Joseph in place of Joses for the second brother, and reversing the order of the last two names. It seems likely that Judah was actually the youngest, for upon Jesus' death James took over the ministry, and upon James' death Simon (and not Judah) did—Joses thus having likely no longer been alive at the time. These sources also refer to sisters of Jesus in the plural but do not name them.

The earliest extant versions of Mark and Matthew were originally written in Greek, with Mark being considered here to be the earlier and therefore more authoritative source. Hence the earliest known written record refers to the second brother as Joses, and not as Joseph. We shall take Joses as having been the actual name of that brother.

It is commonly believed that Jesus had two sisters and that they were called Mariam and Salome. A single (and later) source whose reliability seems less certain suggests there may have been a third sister named Joanna.

Joseph was the son of Jacob (i.e., Yaakov, or James), and 2nd century sources name the parents of Mary as Joachim and Anna of Sepphoris[11]—the largest city in the vicinity of Nazareth at the time. Concerning further ancestry, at the start of the NT there is a lengthy series of "begats" (i.e., geneological lists) whose purpose is to trace the lineages of Mary and Joseph back to King David; these can arguably

[11] Sepphoris was savagely destroyed by the Romans in 4 BCE and later rebuilt by Herod Antipas.

be used to study their genealogies. In particular, the name Matya appears several times in the lineage of Jesus (as recorded in Luke) and some scholars attribute this name to the lineage of Mary. The tracing of ancestries back to the house of David relates to the theme of the NT since it may have been commonly held that the lineage of the Messiah would trace back along a "Davidic line."

Concerning the ultimate fate of the siblings of Jesus, only a small amount is known. Paul 1, Cor. 9:1 ff refers to the brothers as traveling with their wives which suggests that they were married and likely had children. The names of these women and any children are not known, although a reference is known to grandsons of Yehuda named Zoker and James.

Josephus Flavius (1943) records the execution of James in 62 CE in the vicinity of the Temple, stating that this James was "the brother of the man known as Jesus who is called the Messiah." Consequently, James may be regarded as an a priori candidate whom one might not be surprised to find in a NT family tomb, if one such existed (although early historical records appear to indicate that James was buried at the place of execution). The two youngest brothers Simon and Judah are both surmised to have lived beyond the year 70 CE, into the reign of Trajan (tenth emperor of Rome who ruled between 98–117 CE) and are therefore not a priori candidates for such a tomb. The fate of Joses is unknown; after he is mentioned by name in the gospels he is never heard of again. However, because it was Simon who succeeded as leader when James died, it is generally assumed that Joses was no longer alive at the time. Joses is therefore an a priori candidate for a NT tomb. As for Judah, the manner of his death is not known.

Concerning any possible "wife" of Jesus, nothing is known except that had one existed she would likely have been interred in the family tomb if there were one.[12] Jesus too is, of course, a candidate for a NT tombsite, and we also know—from the NT passages concerning Joseph of Arimathea—that persons who pre-deceased Jesus are not candidates for such a tombsite since the family evidently did not possess one prior to Jesus' death.

**5. Statistics of the Jewish onomasticon.** At least three resources are available for studying the distribution of names during the era relevant to this study. The first is the catalogue of Jewish ossuaries in the collections of the State of Israel compiled by Rahmani (1994) who details all ornamented and inscribed ossuaries held by the Israel Antiquities Authorities (IAA) and by the Israel Museum as of 1989—a total of 897 specimens in all. Of these, 233 bear inscriptions identifying the names of a total of 241 male persons and a substantially lesser (but undetermined) number of female persons. Of the 233 inscribed ossuaries, 143 are in Jewish script (i.e., Hebrew or Aramaic), 73 in Greek script, and the remainder

[12]The only "viable" *candidate* for a "wife," assuming one existed, appears to be Mary Magdalene although we shall make no such assumption. Mary Magdalene does, however, turn out to be an a priori candidate for inclusion in a NT tombsite based on other grounds.

in a mix of both scripts or in other languages (such as Latin). A total of 147 unique *names* (male and female) occur among them. The compilation in Rahmani is not arranged by either tomb groups or by gender, and only limited summary information is provided on the distribution of names. Although it is, in principle, possible to do so by working with an index of names provided, it is not straightforward to abstract statistical information from this source.

The second resource, and by far the most comprehensive one currently available, is the lexicon of Jewish names of late antiquity compiled by Tal Ilan (2002). It covers the period between 330 BCE (marking the Hellenistic conquest of Palestine) and 200 CE (which marks the closing of the Mishnaic period and of the early Roman Empire). Ilan's compilation includes the names of 2509 males and 317 females taken from all available sources, including not only ossuaries from both within as well as outside the State of Israel collections, but also from literary sources, epigraphic and papyrological documents, and many other sources. Detailed source information and some statistical compilations are also provided. Although Ilan includes all recorded names used by Jews of Palestine during the stated period, it also includes a further 86 names of women and 685 names of men regarded as fictitious, that is, not corresponding to persons who had actually lived. Fictitious names will be excluded from our analysis.

A third resource is Hachlili (2005); in particular, Tables V-2, (a) and (b) of Hachlili (page 200) provide frequencies for the most common personal names among Jews, by gender and by source, in the late Second Temple period. These tabulations are based only on the most common names—for a total of 1091 males and 192 females—taken from ossuaries, Masada ostraca, and other sources. The sample sizes of which these common names constitute subsets are not provided. These tables essentially coincide with subsets of names in Ilan (2002) but dating to the late Second Temple period.

Ilan's more extensive compilation allows less variable estimation of the incidence of names, although estimates meant to pertain only to the population of ossuaries, but based on all of Ilan, may be somewhat biased not only because nonossuary sources are thereby included, but also because Ilan's compilation includes periods some 300 years prior to when ossuary burials became prevalent as well as 130 years after that practice had ceased. Estimates based on the samples of Rahmani or Hachlili will be much more variable, but presumably less biased, based as they are, in the first instance, on names appearing on actual ossuaries only, and in the second, on names from the late Second Temple period only. It is possible to extract from Ilan's lexicon names obtained only from ossuaries, and these constitute a superset of the sample in Rahmani. Of course, one could argue that no population assembled from such sources can be regarded as valid for the inference at hand, however, we regard that viewpoint as nihilistic and shall not adopt it.

Although the information in Ilan (2002) is not arranged specifically for our purposes, the compilations there include names taken from ossuaries as well as from many other sources, and further, many more names taken from ossuaries appear in

Ilan than in Rahmani since Rahmani catalogues only ossuaries in the State of Israel collections while Ilan includes names on ossuaries from all available sources. As already mentioned, Ilan contains the names of 2509 male persons and 317 female persons. These comprise 721 unique male names and 110 unique female names. Furthermore, Ilan states that, of these, the names of 519 male and 193 female persons (712 persons in all) had been derived from ossuary inscriptions (numbers substantially higher than Rahmani). From this it *appears* that about 27% of inscribed ossuaries bear female names, while 73% bear the names only of males; however the relative frequency of ossuaries of females is underrepresented in these numbers due to the custom of sometimes naming fathers on both male as well as female ossuaries, and of occasionally naming husbands on female ossuaries. Note also that 61% of the female names in Ilan are derived from ossuary sources while only 21% of the male names are so derived, numbers that reflect the patriarchal nature of society at the time.

Our presentation of these distributions of names is laid out in Tables 1 through 5. Table 1 gives the total number of unique male and female *persons* in each of Ilan and Rahmani, as well as the corresponding number of unique male and female *names*. The fourth column gives Ilan's counts when restricted to names obtained only from ossuary inscriptions. In this table, as in some of the others below, not all tabulations or computations were completed, either for reasons of feasibility or for constraints of time; this will be indicated throughout by dashed lines at the affected table entry positions. It will be important to bear in mind that *dashes in the tables do not represent zeros*.

Table 2 gives the ten most common female names according to Ilan, together with their frequencies in Ilan, Rahmani, and among Ilan's ossuary sources only. There are (very) slight variations between the numbers in our table and a similar one in Ilan, ours having been corrected for a small number of additional entries Ilan had later added to her lexicon. Fictitious name counts are shown separately, with "F" labels attached; for example, Ilan lists 63 Salomes, but two were fictitious.

TABLE 1
*Onomastic gender distribution*

| **Gender** | **Ilan** | **Rahmani** | **Ilan ossuaries** |
|---|---|---|---|
| Male persons | 2509 | 241 | 519 |
| Female persons | 317 | – | 193 |
| Total persons | 2826 | – | 712 |
| Male names | 721 | – | – |
| Female names | 110 | – | – |
| Total names | 831 | 147 | – |

TABLE 2
*Jewish female names of late antiquity*

| **Generic name** | **Ilan** | **Rahmani** | **Ilan ossuaries** |
|---|---|---|---|
| Mariam/Mary | 74 + 6F | 18 | 44 |
| Salome | 61 + 2F | – | 41 |
| Shelamzion | 25 + 0F | – | 19 |
| Martha | 21 + 0F | – | 17 |
| Joanna | 12 + 0F | – | 7 |
| Shiphra | 12 + 0F | – | 9 |
| Berenice | 9 + 1F | – | 2 |
| Sarah | 8 + 1F | – | 5 |
| Imma | 8 + 0F | – | 6 |
| Mara | 7 + 0F | 2 | 5 |
| No. females | 317 + 86F | – | 193 |
| No. female names | 110 | – | – |

Note that names obtained from ossuaries are never fictitious. Here again dashed lines represent undetermined entries (not zeros).

Table 3 gives the 21 most common male names appearing in Ilan, together with their frequencies in Ilan, Rahmani, and among Ilan's ossuary sources only, with slight updates having again been made to a similar table of Ilan. There are also minor differences between the Rahmani column of our table, as determined by us, and a table based on Rahmani given by Fuchs (2004). The fictitious name counts in the "Ilan" column again occur only on nonossuary sources; in one instance (an Eleazar) the fictional status is uncertain.

A number of difficulties occur in producing such tables. In Rahmani (1994), the gender of several of the names is ambiguous. (Presumably one could try to resolve these by cross-referencing to Ilan where most names are categorized by gender.) Furthermore, some inscriptions are uncertain due to problems of legibility. The resulting tables therefore depend somewhat on what conventions one adopts toward the various problems of this nature.

Ilan (2002) and Hachlili (2005) give considerable further information concerning the customs of naming as well as about the distribution of names in that era. By way of general comment, one can say that the pool of names in use was not unlimited. For that reason different renditions of a generic name category often acted as distinct names so as to help distinguish among individuals. Names associated with the Hasmonean dynasty were especially popular. For men, these include the names Mattathias, Yochanan, Simon, Judah, Eleazar, and Yonathan. As for Hasmonean women, only two of their Hebrew names are known—one called Mariam, and the other Shelamzion. It is possible that the name Salome was popular for

TABLE 3
*Jewish male names of late antiquity*

| **Generic name** | **Ilan** | **Rahmani** | **Ilan ossuaries** |
|---|---|---|---|
| Shimon/Simon/Peter | 249 + 8F | 24 | 62 |
| Yehosef/Yosef/Joseph | 221 + 10F | 19 | 45 |
| Yehudah/Judah/Judas | 171 + 8F | 20 | 45 |
| Eleazar/Lazarus | 169 + 7F + 1? | 14 | 30 |
| Yochanan/John | 124 + 5F | 8 | 26 |
| Yehoshua/Yeshua/Jesus | 101 + 2F | 10 or 11 | 23 |
| Hananiah/Ananias | 83 + 3F | 11 | 19 |
| Yonathan/John | 72 + 3F | 6 | 14 |
| Mattathias/Matthew | 62 + 1F | 7 | 17 |
| Menachem | 44 + 2F | 0 | 4 |
| Yaakov/Jacob/James | 43 + 2F | 5 | 6 |
| Hanan | 36 + 3F | 4 | 7 |
| Alexander | 30 + 1F | 4 | – |
| Dositheus | 30 + 1F | 6 | – |
| Zachariah | 25 + 6F | 1 | – |
| Ishmael | 31 + 0F | 2 | – |
| Levi | 25 + 4F | 1 | – |
| Saul | 29 + 0F | 10 | – |
| Choni/Onias | 27 + 0F | 0 | – |
| Shmuel/Samuel | 21 + 5F | 0 | – |
| Hezekiah | 23 + 3F | 0 | – |
| No. of males | 2509 + 685F | 241 | 519 |
| No. of male names | 721 | – | – |

the same reason, but its origin is uncertain. Biblical names, particularly of the secondary characters, were also popular, with the names of primary biblical characters being less prevalent than might have been expected.

The counts shown for each of the *generic* names in Tables 2 and 3 include all *renditions* or variants of that name. However, we shall require more detailed statistical information regarding the *renditions* within the *generic* categories for certain names relevant to this study. Three variants will interest us particularly, namely the variants Mariamenou and Marya for Mariam, and the variant Yoseh for Yosef. Such breakdowns are provided in Tables 4 and 5. We see from Table 4 that there are (in all) 16 variants for Mariam, and from Table 5 that there are 22 variants for Joseph if language differences are also allowed for. In Table 4, horizontal lines demark two groups of Mariam renditions relevant for us, with Mariamenou and Mariamne isolated at the top of the table and versions "equivalent" to Marya isolated at the bottom; close-sounding versions are placed close to, but on the opposite sides, of these lines. Likewise, in Table 5, the renditions considered relevant to the biblical brother Joses appear in the five rows isolated at the bottom.

TABLE 4
*Mariam renditions*

| Rendition of name | Ossuary sources | Nonossuary sources | Combined sources |
|---|---|---|---|
| *Μαριαμηνου* | 1 | 0 | 1 |
| *Μαριαμνη* | 0 | 0 + 1F | 0 + 1F |
| *Μαριαμην* | 0 | 1 | 1 |
| *Μαριαμ* | 2 | 2 | 4 |
| *Μαριαμη* | 10 | 8 | 18 |
| *Μαραμη* | 0 | 1 | 1 |
| *Μαριαμης* | 0 | 1 | 1 |
| *Μαριαδος* | 1 | 0 | 1 |
| *Μαριεαμη* | 1 | 0 | 1 |
| מרים | 12 | 10 + 4F | 22 + 4F |
| מרימ | 3 | 0 | 3 |
| מרימא | 0 | 1 | 1 |
| *Μα*[*ρ*]*ιας* | 1 | 0 | 1 |
| *Μαρια* | 4 | 6 + 1F | 10 + 1F |
| מריה | 8 | 0 | 8 |
| סריה | 1 | 0 | 1 |
| Total of above | 44 | 30 + 6F | 74 + 6F |

We note the following important differences between ossuary and nonossuary sources. For Mariam, the rendition מריה apparently occurs only on ossuaries. For renditions of Joseph, the form יוסי never appears on ossuaries, while the Greek form *Ιωσηπος* and the Hebrew form יוסף are also greatly underrepresented on ossuaries. The rendition יהוסף is the most common one appearing on ossuaries, although it is well represented among nonossuary sources as well. In the five renditions (at the bottom of Table 5) consistent with the biblical brother, their "free use" on ossuaries, and relative rareness on nonossuaries, appears consistent with the notion that they act much like a separate name category. Of these five, the Hebrew rendition יוסה has never been found on any ossuary other than at Talpiyot.

**6. More about the Talpiyot inscriptions.** In this section we provide some further details for the particular names occurring on the Ossuaries #1–6 described in Section 2. Our primary resource here, again, is Ilan (2002).

**Mariam & Marya:** The name Mariam or Miriam, and its variants, was the most common female name of the Second Temple era.[13] We note also that starting

[13]We are following the statistics of Ilan's onomasticon here; some sources put Salome as the most common female name, with Mariam as the second most common.

TABLE 5
*Joseph renditions*

| Rendition of name | Ossuary sources | Nonossuary sources | Combined sources |
|---|---|---|---|
| *Ιωσηφος* | 0 | 0 + 5F | 0 + 5F |
| *Ιωσηπος* | 4 | 38 | 42 |
| *Ιωσιπος* | 0 | 1 | 1 |
| *Ιωσηπου* | 1 | 6 | 7 |
| *Ιωσηφ* | 2 | 6 | 8 |
| *Ιωσηπ* | 0 | 3 | 3 |
| *Ιωσιας* | 0 | 1 | 1 |
| *Ιωσιου* | 0 | 1 | 1 |
| Ioseph | 0 | 0 + 1F | 0 + 1F |
| Iosepu | 0 | 1 | 1 |
| יוסף | 2 | 17 + 2F | 19 + 2F |
| יהוסף | 27 | 61 | 88 |
| יהוסה | 1 | 0 | 1 |
| יהסף | 2 | 1 | 3 |
| יוסי | 0 | 29 + 2F | 29 + 2F |
| איסי | 0 | 6 | 6 |
| אסי | 0 | 1 | 1 |
| *Ιωση* | 1 | 1 | 2 |
| *Ιωσε* | 2 | 0 | 2 |
| *Ιωσης* | 2 | 0 | 2 |
| יסה | 1 | 0 | 1 |
| יוסה | 1 | 2 | 3 |
| Total of above | 46 | 175 + 10F | 221 + 10F |

with the earliest gospels of Mark, Marya is the principal form by which the name of the historical Mary has been handed down; it is therefore likely that this is the form of the name by which she was known. (We remark that this contention is not universally accepted.)

**Mariamenou** [*η*] **Mara:** Of the occurrences of the generic Mariam in Ilan (2002) only one instance consists of the "full" and highly unusual form *Μαριαμηνου*; it corresponds to our Ossuary #1 on which the additional detail "[*η*] *Μαρα*" is inscribed. The form *Μαριαμνη* also occurred only once but does not correspond to a person who actually lived, while *Μαριαμην* also occurred once, although not on an ossuary. We remark that Mariamenou and Mara are each individually quite rare names so that either of these should have sufficed for purposes of identification by family members if referring to a single individual.

An argument can be put forth that the actual name of Mary Magdalene was Mariamne. For some background, we refer to Bovon (2002) and references therein. In a 4th century version of the Acts of Philip, a woman who is thought to be Mary

Magdalene is referred to throughout as Mariamne, and Bovon surmises that Philip was her brother.[14] This version of these Acts is the earliest and most complete one known and is also one of the earliest known historical sources explicitly citing Mary Magdalene's name. These Acts also indicate that she died in Palestine, thus potentially allowing that an ossuary of hers might be found in Jerusalem. James Tabor [private communication] has recently found a still earlier reference. Hippolytus, a second century Christian writer, wrote in *Refutations* 5.2: "These are the heads of very numerous discourses which the Nassenes assert that James the brother of the Lord handed down to Mariamne." This reference dates to approximately 175 CE, some 100 years after the destruction of Jerusalem, and furthermore suggests that "Mariamne" was, at one time, the head of a ministry thereby entitling her to be addressed as "lord" or "honorable lady." The family buried at Talpiyot appears to have understood Aramaic over a period of some two generations (in view of their use of בר) and is therefore likely to have known the Aramaic meaning of "mara."

As her name indicates, Mary Magdalene came from Magdala (or Migdal); she herself likely spoke Greek and is believed by some to have preached extensively among Greek-speaking Jews. It has been speculated that she was also an apostle and a key contributor to the early Christian movement, and explanations have been advanced (revolving around patriarchal intrigues) as to why she may have later been portrayed as a "sinner." Ossuary #1 is the only one in the Talpiyot tomb in Greek script. Since Mary Magdalene was not a descendent of the same bloodlines as the family of Jesus, it is at least plausible—if this really were her ossuary—that it might have been rendered in Greek script even while the others may not have been. The inscription on Ossuary #1 will be regarded in our analysis as an appropriate rendering of her name. As an inscription, Mariamenou [$\eta$] Mara is extraordinary, and—all previous considerations aside—among the 74 Mariams whose names are currently known to us, it provides arguably the "closest fit" to Mary Magdalene.

Our analysis will be based on the following specific assumptions concerning the inscription on Ossuary #1: First, we will assume that it refers to only one person and that it represents an appropriate appellation for Mary Magdalene. Second, we will assume that this rare rendition is not applicable to many other "Mariams." Further—inferring from the remarkable detail of this inscription—we will assume that even if a larger sample of Mariams could somehow be obtained, it is unlikely that so *specifically* appropriate a name (for Mary Magdalene) would arise with frequency greater than occurs in Ilan's sample. The reader should note that these assumptions are far from universally accepted. We shall revisit this matter in Section 14.

[14]The mentioned "argument" then only requires us to assume that a brother would know his own sister's name.

**Yeshua:** The name Yeshua is a derivative of Yehoshua and is the sixth most common Jewish male name of the Hellenistic and Roman periods. Its popularity derives from the fact that Yehoshua was the successor to Moses. Note that the shortened form Yeshua is the one by which the name of Jesus is known, and all literary records—whether based on the NT or on its Hebrew versions—use that form for the name. Jesus quite likely preached in Aramaic and is, in any case, known to have been able to speak it; in this respect, the use of Aramaic on Ossuary #4 is therefore not implausible.

**Yehosef & Yoseh:** The name Yehosef was the second most common male name in the Second temple period. The form Yoseh which appears on Ossuary #5, however, is an uncommon version for this name. Among the 46 ossuaries bearing some version of the name Yehosef, only one (corresponding to our Ossuary #5) bears the Hebrew form יוסה; furthermore, this version of the name is one that corresponds to that used in the gospel of Mark.[15] In our analysis, we will assume that the (father) Yehosef named on Ossuary #4 is not the same individual as the Yoseh named on Ossuary #5, and that the two name versions were intended for deliberate distinction. The rationale behind this lies, first, in the seemingly special characteristics of the name יוסה, and second, in the fact that *halacha* (although a later tradition) mandates that the name by which a person was actually known in life is the form that must appear on their gravesite. Third, the use of the somewhat informal Yeshua (instead of the more formal Yehoshua) in the patronym of the Yehuda ossuary suggests that the Talpiyot tomb family may have respected "nicknames." We note again, however, that these assumptions are not universally accepted.

**Matya:** This is a shortened form of Matityahu (Matthew), a common name having Maccabean and Hasmonean origins. According to Luke and Matthew, this name occurs in the genealogy of Jesus several times, through Mary's lineage in particular.

**Yehuda:** This translates as Judah, a strong Maccabean name, and the third most common Jewish name in the Hellenistic and Roman periods. It is also the name of a younger brother (or half-brother) of Jesus.

**7. Some statistical "judgement calls."** In this section we indicate some statistical "judgement calls" and approximations which we propose to apply. The first is a specialized assumption concerning the independence of assignment of names. In particular, we shall assume that fathers called Yehosef would name a son Yeshua with frequency comparable to that in the general population (although subject to the proviso that the names of fathers and sons ought normally to differ); likewise, we shall assume that men called Yehosef would marry women called Mariam in the same frequency as that name occurs generally; and so on. Assignment of names

[15] In the earliest extant (Greek) version of Mark, the name of the brother Joses is written only as $I\omega\sigma\epsilon$ or as $I\omega\sigma\eta\sigma$. It translates into Hebrew pronounced as Yoseh (rather than Yosa).

within families is well known to be dependent *time-longitudinally,* with children frequently named after earlier "nodes" on their family tree. However in the present context this assumption is applied primarily on a *time-cross-sectional* basis. Although this assumption is unlikely to be accurate with respect to very rare and/or very unusual names, for the types of names which concern us the dimension of the underlying distribution here seems small enough that modest time-cross-sectional dependencies should not have excessive impact. Much as we would prefer to avoid such an assumption, an incisive analysis without it does not seem feasible. We shall, however, revisit this in Section 14.

We shall also occasionally ignore certain small (and generally negligible) corrections to joint frequencies for such facts as that brothers ought normally to bear different names, and so on. In contexts where these could matter more substantively (as in our computing code [Feuerverger (2008)], for example) appropriate corrections will be taken into account.

We next address the question of biases in the samples available for assessing the name frequencies. We first consider the situation for the generic name categories and afterward for the renditions occurring within them. There are several potential sources of bias if Ilan's complete lexicon is used. One is the usual selection bias relating to representativeness of the sources. Difficulties of this type affect many surveys and here little can be done to correct them.

Another source of bias arises if nonossuary listings are included in the frequencies. One may attempt to address this (for the generic names) by comparing their frequencies by ossuary and nonossuary sources; these may be determined from the second and fourth columns of Tables 2 and 3. Such comparisons do not suggest biases of great consequence; tests for the equality of proportions between ossuary and nonossuary sources proved to be nonsignificant, although among generic names not relevant to this study there are one or two instances among the more unusual names where the relative frequencies between ossuary and nonossuary sources appear to differ more substantively. As it seems preferable to allow some element of bias in return for reduced variability (in hope of obtaining estimates with smaller overall error) we shall use Ilan's full lexicon to estimate the relative frequencies for the *generic* name categories relevant to our work.

As will be evident later, *smaller* frequencies for *relevant* names *in-sample* are "advantageous" for driving tests toward "significance," while *smaller* frequencies for *relevant* names *out-of-sample* will drive tests away from significance. In these respects, the frequencies for such names as Simon, Yehudah, and Matthew will ultimately not matter for us, and those for the names Joanna and Martha will matter rather little. For Mariam, Salome, and Joseph, the combined versus the ossuaries-only relative frequencies are essentially identical. For Yeshua and Yaakov the frequency differences each fall in their nonconservative directions although not significantly so, and the effects of this can be studied in experimentation.

A third source of bias stems from the fact that Ilan's lexicon covers a broader range of dates than relevant for us, this being the case (although very much less

so) even if Ilan's data were restricted to ossuary sources alone. One could, in principle, study this effect by laboriously categorizing the individual entries in Ilan, however the ossuary versus nonossuary comparisons do already largely address this concern.

For the *renditions* of names within the *generic* categories the situation is, however, altogether different as Tables 4 and 5 have shown, presumably reflecting variations in the popularity of specific renditions over time. Allowances for this are necessary. To obtain estimates for name *rendition* frequencies we propose to use the overall proportion (i.e., including nonossuaries) for the generic categories—these being judged the most stable in terms of bias-variance tradeoff—but to correct "internally" for differences in the ossuary versus nonossuary rendition frequencies. Thus for the rendition Yoseh, we estimate its frequency as

$$\frac{(7/46) \times 221}{2509} = \frac{33.63}{2509},$$

since there are 221 (nonfictitious) Josephs among Ilan's 2509 males, while among the 46 Josephs whose names are derived from ossuaries, 7 were versions deemed consistent with Yoseh. Note that the frequency derived above is considerably higher (hence more conservative) than the value $10/2509$ obtained "directly." Likewise the frequency for Marya will be estimated as

$$\frac{(13/44) \times 74}{317} = \frac{21.86}{317},$$

and not as $19/317$, and so on. Needless to say, it is the fraction from within the generic categories that will primarily drive the variability of such estimates.

**8. Size of the relevant population.** We require estimates of the size of the relevant population of Jerusalem and of the number of ossuary burials that took place overall. The estimates in this section draw on various sources. In particular, in a paper on the James ossuary, Camil Fuchs (2004) carefully estimated the population of Jerusalem in a sequence of steps which we summarize here.

First, citing studies by Hachlili (1994) and Kloner (1980), Fuchs notes that the maximum range of dates during which Jews practiced ossuary burial was between 20 BCE and 70 CE, an interval of approximately 90 years. These, however, are outside limits, and since the practice of ossuary burial was undoubtedly introduced gradually, a reasonable, but still conservative estimate, is to assume that the custom was prevalent between 6 CE and 70 CE, an interval of some 65 years.

Second, citing studies by Broshi (1977, 1978) and Levine (2002) who estimate the habitable areas of Jerusalem and their population densities, and the study by Wilkinson (1974) on the capacity of water supply systems, Fuchs argued (following Broshi) that around 20 BCE, the population of Jerusalem was about 38,500, while around 70 CE the population was about 82,500 (corresponding to a growth rate of about 1% per annum). Levine's estimate for around 70 CE was between

60,000 to 70,000, while Wilkinson's estimates for around 70 CE was about 75,000 persons. These are all in reasonably good agreement; to be conservative, Fuchs adopted Broshi's estimates.

Third, citing various sources, Fuchs estimated the birth rate to have been between 4% and 4.5% per year—corresponding to an average fertility rate of about 6 to 7 children per woman—and he estimated juvenile mortality to have been between 35% and 50%. Fuchs used the midranges in his computations, and a truncated Poisson distribution to model the number of children per woman estimating that approximately 132,200 Jerusalemites died in the period between 6 CE and 70 CE.

Of these, approximately 66,100 were male and 66,100 were female, counts which include infants, juveniles, adults, and non-Jews. Conservative estimates are that 5% of the population were non-Jews and that 42% of the deceased were juveniles, leaving 36,420 male and an approximately equal number of female deceased Jewish adults during this period.

Next, to afford a tomb-site and other costs associated with ossuary burial required some degree of affluence. As well, ossuaries bearing inscriptions evidence some degree of literacy on the part of the family involved. Literacy and affluence were no doubt correlated attributes, and Fuchs concluded, using a sequence of relatively conservative estimates, that at most 12% of the population satisfied these dual criteria. This led him to a "relevant population size" of around 4,370 males at most buried in inscribed ossuaries in the Jerusalem area during the relevant era. To place Fuchs' estimate in context, recall that the State of Israel collections (as itemized by Rahmani) contained only 233 ossuaries bearing inscriptions (with some being of women) and that in Ilan the names of 519 male persons were derived from ossuaries (with some only being fathers on mens' as well as on womens' ossuaries). Fuchs' estimates thus appear to be both reasonable and conservative.

Fuchs did not require nor did he estimate the number of ossuaries of females bearing inscriptions. Since among Ilan's ossuary sources 519 male and 193 female names were found, it appears that 27% of inscribed ossuaries bear female names—a male to female ratio of about 2.7 to 1. Of course, this underestimates the proportion of inscribed female ossuaries. While one could more accurately estimate this proportion by pursuing fine detail in Ilan we propose instead to use a crude estimate based on a ratio of 2 to 1, namely that 2,185 females were buried in inscribed ossuaries. This estimate appears adequate for our purposes and (conveniently) corresponds with the ratio found in the Talpiyot tomb.

Relative to questions of whether or not the Talpiyot tombsite could be that of the NT family, the data from that site must be viewed as the "best" of many trials. So far, about 100 tombsites have already been explored, but the mere existence of others that have not been must somehow also be accounted for. The Talpiyot site consists of 4 male and 2 female inscriptions. When divided into Fuchs' estimates for the total number of inscribed adult ossuaries, we obtain approximately 1,100; this appears to be an appropriate number of trials out of which the Talpiyot observation could be considered as being the "best."

**9. Inferential issues.** This section concerns whether or not statistical reasoning applies to this problem, and whether the available data permit meaningful analysis of an archeological find such as this. Remarks regarding the interpretation of "tail areas" are postponed to Section 14.

Several issues need to be addressed. First is the "fear-factor" connected with proposing an analysis on a controversial topic; it seems fair to say (and certainly in hindsight) that the intensity with which any analysis of this data set will be scrutinized constitutes an arguably unprecedented feature of this problem. Faced with this one may be tempted to adopt so highly conservative a stance that all evidence becomes masked. We side-step this and try to analyze the data as in an ordinary statistical problem; the resulting computations must then to be interpreted by each "consumer" for themselves. Second are "theological" considerations which if rigorously adhered to void any possibilities for analysis. The approach we adopt is to analyze the data from a purely "*historical*" *viewpoint*, by which we also mean that all persons referred to are assumed to have been real and subject to all considerations real persons are subject to. Third, there is the question of whether the available data bear adequately on the problem at hand. One could argue that the available onomastica cannot be authenticated (i.e., matched to the actual populations) and so on. We bypass such viewpoints and adopt the position that considerable and relevant data are available for the problem at hand.

Harder to dismiss is the role of "coincidence" [see Diaconis and Mosteller (1989)], the issue being that this data did not originate in a planned experiment; coincidences occur all the time, and their a priori probabilities can be extremely small, even though the probability is not small that *some* coincidence will happen to *someone*, *somewhere*, *sometime*. It could be argued that such data cannot be analyzed, or that extremely minute levels of "significance" are required to carry evidentiary value. A kind of "relativity" operates here toward which the analyst must adopt a stance. For our problem, to an "observer on the ground" in Jerusalem interested only in results from digs, these data originate in a standard way. It is tempting to argue that because this find concerns the most well-known family that ever lived it actually might exempt us—*purely on technical grounds*—from the limitations of coincidence. In any case, our analysis will be carried out from the vantage of the aforementioned "observer on the ground" in Jerusalem.

There are also certain subconscious and/or widely held misperceptions that "interfere" in our attempts to assess the evidence in these data. In particular, one needs to face the fact that it does seem extraordinary, at first, to contemplate that an ossuary that may have been intended for Jesus of Nazareth could ever possibly be found. The following *historical* point therefore needs to be made: Jesus was a Jew—a devout man who followed the letter and the spirit of the Jewish laws, as did other members of the NT family. Unless prevented by *force majeure*, the family (and followers) of Jesus would have certainly seen to a quick and proper burial in accordance with the Jewish ritual customs prevalent at the time. Roman authorities saw to Yeshua's crucifixion because they deemed it against their interests to allow

a man proclaimed as being "King of the Jews" to live, and for the same reason would have certainly executed any son(s) of such a "King." But there is little reason for Roman authorities to have stood in the way of families of crucified persons from subsequently conducting proper burials, and there are in any case accounts of how release of the body was secured through the influence of Joseph of Arimathea. In fact, Joseph of Arimathea offered a burial site, in Jerusalem, for that purpose (as evidently the NT family did not yet have one of its own) and the single most likely eventuality, from a purely historical stance, is that the remains of Jesus were intended for interment in an ossuary—although possibly as much out of the sight and knowledge of Roman authorities as possible. Moving the remains to (say) Nazareth—a trek of some three or four days—may hardly have been feasible considering logistics at the time; indeed the Talpiyot location is among the many where one might reasonably expect such a tomb—if one existed—to be found.

Next, the ossuary inscribed "Yehuda son of Yeshua" plays an unsolicited role in the inference because at least this much is true: If this tombsite really were that of the NT family, then there did live a person named Yehuda whose father happened also to bear the name Yeshua. In that eventuality, the possibility arises that the two Yeshuas may have been the same person. It would not have been considered unusual for a Jewish man to have a child, and if that child was believed to be a target of the Romans, it would not have been unusual to try to protect it. However, other possibilities exist as well, with the time elapsed between the crucifixion and the destruction of Jerusalem allowing other scenarios to have played out. If, on the other hand, an ossuary inscribed "Yehuda son of Yeshua" may (for whatever reason) not be located in a NT family tombsite, then the Talpiyot site cannot be that of the NT family and the names found there must be purely coincidental. In our analysis, this ossuary will initially be "set aside," but we revisit this in Section 14.

Experimental design issues (as well as their absence) also play a role as there are several hypothetical scenarios under which our data could, *in principle*, have been collected. Furthermore, we do not know a priori whether or not a NT tomb site actually exists, the individuals who might have been within it, or the renditions of their names—considerations which each subtly affects the character of our inferences.

Conditioning and/or ancillarity, which are standard statistical practice, play an especially important role in our analysis. It seems reasonable, and perhaps even a practical necessity, in analyzing these data, to condition on the number of inscribed ossuaries found in this tomb, and also to condition on the fact that two were female and four were male. In some respects, these values carry little "information" relevant to the questions of interest here. We also condition on the fact that two of the inscribed male ossuaries are aligned in the generational sequence "C son of B" and "B son of A." The fact that there were a total of ten ossuaries in the tomb may or may not be viewed as ancillary, but not the ratio $6/10$ of inscriptions, for that ratio carries information concerning the "literacy" of the family that owned the tomb. Likewise, the specific languages used on the inscriptions cannot be regarded

as entirely ancillary because some information is available about the languages used by NT family members. Conditioning will thus play a significant role in our analysis, with even our "test statistics" permitted to depend on certain observed configurational aspects of the find.

A further inferential issue is that more than one reasonable analysis may be proposed (even by the same statistician) leading to somewhat differing "$p$-values." C. R. Rao recently referred (2007, at Cochin) to a 1992 Leiden Ph.D. thesis by Van den Berg which consisted of sending the identical data set to ten renowned statisticians, resulting in ten different analyses. Andrews and Feuerverger (2005) have argued that examining a collection of models allows the variations among their results to speak for the true inherent uncertainties without trivializing a problem.

As a final point, we mention that NT genealogical data is subject to considerable ambiguity, with names having frequently changed in form across sources, across time, and across translations. Care must therefore be exercised to assure that any proposed analysis is not influenced unduly by prior examination of the data, a principle well enough understood, but difficult to incorporate in practice.

**10. Our "a priori" hypotheses.** In Sections 11 and 12 we develop a statistical approach based on "relevance" and "rareness," or "surprisingness," for addressing questions such as those raised by the Talpiyot site. Here—*on a best efforts basis*—we attempt to formulate a reasonable set of a priori alternative hypotheses. Our approach is strictly "*historical*" and with no claim made, of course, that the data has not been seen. We propose eight a priori hypotheses (APH) in all.

- **APH 1:** An ossuary intended for Jesus was likely to have been produced in the Jerusalem area. He was first laid to rest near the site of the crucifixion under the initiative of Joseph of Arimathea,[16] and it is unlikely that followers would have dishonoured the body in any way.
- **APH 2:** It is likely that one or more among the more affluent followers of this Messianic movement would have seen to a tombsite for the NT family and/or for some of its key leaders.
- **APH 3:** Inferring from biblical accounts, if there were a NT tombsite, no one who predeceased Jesus may be in it. One such *person* is Joseph, the father. (This does not preclude the *name* Joseph from occurring in the tomb.) Another such person is John the Baptist.[17]
- **APH 4:** No one who died after 70 CE may be found in such a tomb. Hence Simon and Yehuda will be excluded (although their names are not). This also excludes most—although not all—of the apostles, many of whom lived beyond 70 CE.

[16] This NT account suggests, incidentally—and it is an important point for us—that the NT family did not yet have a tombsite of its own.

[17] According to Josephus, John the Baptist died [was beheaded] at Machaerus before Jesus. He was thus most likely buried at Qumran, or in the vicinity the Dead Sea.

- **APH 5:** Closest relations, particularly closest blood relatives, are among those who might be expected to be in the tomb. Among those whose names are essentially known, are the mother Mary, brothers James and Joses, sisters[18] Mariam and Salome, and as a more remote possibility, a third sister Joanna. Potential blood relations or others *very* close to the family can also be identified from among those present at the burial ritual. This includes a Marya (referred to as the mother of James and Joses); it includes Mary Magdalene[19] whose presence at the burial ritual is consistent across all gospel accounts; and it includes a Salome who might be a sister of Jesus.[20,21] The list of family intimates might also include a sister of Mary and/or possibly her spouse Cleopas (generally assumed to be the brother of Joseph).
- **APH 6:** The tomb might include close associates and/or others mentioned prominently or strategically in the NT (e.g., some apostles, especially if related through blood and/or marriage), close friends, and/or slightly more distant relations of the family. It would exclude anyone whose tomb has already been found elsewhere, or who is known to have lived and/or died elsewhere. A brief discussion of potential such persons is given below. The a priori probability of inclusion for individuals in this group is less than for those in APH 5, and their number would be related to the size of the tomb complex. Because the genealogy of the NT family is not known fully, such a tomb might also contain individuals whose names are unknown (or would not have occurred) to us.
- **APH 7:** It would be expected that if a NT tomb existed it might be unusual or distinctive in some way, reflecting the prominence or other characteristics of the family via some feature(s) of the site; exactly how, one cannot say. As the NT family does not appear to have been large it is plausible that their tombsite might also not be large.
- **APH 8:** There is no a priori hypothesis as to the number of ossuaries that might be found in such a tomb, as to their configuration, or as to the renditions of names that might appear on them,[22] but it might be expected that these ossuaries would in some respects be unusual, with some bearing distinctive or unusual

[18]The likelihood of a sister being in a NT tomb depends in part upon whether or not she was married.

[19]Although Mary Magdalene is sometimes cited as a possible *candidate* for a "spouse" on the basis of her presence at the burial (confirmed in Mark 15 and Luke 8), and on the basis of later gnostic sources which refer to her as a companion of Jesus, our analysis does not assume this; it only assumes that she is on a "short list" of persons close enough to the family to be a candidate for inclusion in a NT tomb, an assumption which is by no means universally accepted.

[20]The brothers are not named as having been at the burial and most likely fled (as did the other apostles) for fear of their lives; none was present at the crucifixion.

[21]A woman called Martha (whose brother was Lazarus) may also have been present at the burial, however her ossuary is believed to have been found at Dominus Flevit.

[22]On the other hand, an ossuary inscribed "Shimon bar Yonah" found at Dominus Flevit and believed to correspond to one of the apostles helps us to infer what a NT inscription should look like.

inscriptions and/or ornamentation, and perhaps more detail in the rendering of names than typical.

Let us next consider, in a little further detail, the persons (or names) that might be viewed as candidates for inclusion under APH 6. Those present at the funeral have already been discussed. Among others mentioned prominently in the NT are individuals named Joanna and Suzana mentioned in Luke 24:10 as providers of financial support. The name Martha also appears in the NT as a close friend but she came from Bethany and would likely have been interred in her own family's tombsite there. As concerns the apostles—many of whom are believed to have survived beyond 70 CE—there are no substantive a priori reasons for any of them to be found in a NT family tombsite—especially if it were a small one—unless related by blood to the family; this would be the case if the apostle happened also to be a brother.

As evident from the discussion, the a priori candidates for a NT tombsite are not unlimited. Of course, from this information, more than one plausible a priori list can be constructed. However, we will work with different possible lists as well as with different numbers of (and frequencies for) candidates.

We can now write down our a priori list of candidates for a NT tombsite. In alphabetical order, for the women, this list includes, initially, the *persons*[23]

Mariam, Mary, Mary Magdalene and Salome.

For the men, it includes

James, Jesus and Joses.

In expanded versions, the lists may include

Cleopas, Joanna and Martha,

although these persons are considered to be more remote possibilities. The list of *persons* (but not necessarily *names*) that would *disqualify* the tombsite as belonging to the NT family includes

Joseph, Simon and Yehuda,

as well as many rather specific and/or unusual *names*[24] thought not to be associated with the NT family in any way. The consequences of not specifying a disqualifier list more fully will be statistically conservative. Finally, the list of names that do not disqualify the find, but that otherwise offer little or no "evidentiary value" is

[23]The four lists given here, are not lists of *names*, but of NT *persons;* here Mariam, Salome and Joanna refer to (possible) sisters, James, Joses, Simon and Yehudah to brothers, etc.

[24]Certain specific *renditions*, even for *generic* names associated with the NT family, could also be included in this disqualifier list.

lengthy; for our purposes, it will suffice for this list to consist of all names other than those already included here.

Next, we need to deal with the fact that even if the ossuary for a candidate on our lists were found, we have no way of knowing a priori by which *rendition* their name would appear. Our paradigm for measuring "surprisingness" will allow us to handle this problem in an effective way, but will require an a priori assignment of a measure of "surprisingness" to any name rendition that might occur. It will be more convenient to deal with a reciprocal form of "surprisingness"; this will be a measure of "relevance and rareness" which we will call the "RR value." "Relevance" will refer to membership in an a priori list of tomb candidate name renditions. The RR value of a datum, or of a subset of data, will often be the same as the frequency of occurrence of its "relevant" components under independent random sampling from the onomasticon, but there will be essential exceptions to this. (The complete definition is somewhat involved and will be detailed further below.) For the sake of definiteness, we define "surprisingness" as $-\log(\text{RR value})$, or alternately as $1/(\text{RR value})$.

The way in which we shall assign "RR values" to name renditions of NT persons on our a priori candidates lists is via *prespecified* nested classes of sets of name renditions in which the innermost class(es) represent the most "relevant" but "rarest" (i.e., the most specific but appropriate) renditions of that person's name, and the outermost classes include the less rare renditions still considered relevant for that person. These classes are compiled in conjunction with the totality of the information in Ilan (which includes the Talpiyot names). Collections of outermost sets of such nested classes may themselves constitute a part of a partition of the generic name category from which they derive, as may happen when the generic name applies to more than one NT individual. This occurs in particular with the generic name Mariam which here can refer to three different "intimates" of the NT family, and with the generic name Joseph which can refer to two such individuals. Nothing here is intended to prevent the same rendition category from applying to more than one person.

To now become specific, for Mary Magdalene, we initially allow a nested class of renditions consisting of the following three "appropriate" and decreasingly rare sets: (a) the set consisting only of the rendition Mariamenou [$\eta$] Mara; (b) the set consisting of all versions of Mariamne, including the one in (a); and (c) the set consisting of all Mariams,[25] including those in (b). Upon consulting Table 4, we observe that no (nonfictitious) rendition of Mariam appearing in category (b) and not also in category (a) occurs among sources restricted to ossuaries. Since only ossuary-based sources ultimately figure in the analysis, our categories (a) and (b) here actually become identical; we are thus left with only two nested rendition

[25]It is possible that for Mary Magdalene only the renditions in (a) and (b) are relevant and that the remaining Mariams in (c) are not. However, the results of our analysis will not depend upon whether or not we include (c) here since it becomes included upon considering Mary.

categories for Mary Magdalene. Now, each such renditions set will have an a priori RR value associated with it, and when an observed rendition of a relevant name is encountered, the RR value associated with it will be that of the rarest set to which it belongs. The specific measure of "rareness and relevance" associated with such a set will be defined below; typically it will be the relative frequency of that set within the onomasticon, but certain exceptions to this will be permitted.

Continuing in this way, for (the mother) Mary we allow two classes, namely (a) all versions of Marya; and (b) all Mariams,[26] including those in (a). For the (possible) sisters Mariam, Salome, and Joanna we have only the generic name sets for each since none is known by any rarer rendition; this applies to Martha as well.

For the men, we must be mindful that the name of the father on the generational ossuary plays a different role than the other male names. In any case, for Jesus,[27] as well as for James and Cleopas, we have again basically only their generic name categories, while for Joses we have (a) all renditions consistent with Joses (as at the bottom of Table 5); and also (b) the generic Joseph set. As for the father on the generational ossuary two additional persons are relevant for us, one being Joseph, the father of Jesus, and the other being Jacob, the father of this Joseph—but the latter relevant primarily because he is also the possible father of Cleopas, and relevant only if the generational ossuary were to read "Cleopas son of Jacob." For Joseph and Jacob we again have only their generic name classes associated with them, as neither appears to have been known by rarer renditions. The RR values assigned to these renditions will be context-dependent owing to configurational considerations induced by the presence of the generational ossuary.

**11. A proposed method for analysis: preamble.** We turn now to develop our approach for the inference problem at hand. Because application of a classical hypothesis testing framework in the present context is not straightforward, we consider an approach centering generally on the "surprisingness" of observations, and of how frequently—under a random sampling protocol from the onomasticon—a cluster of observations of equal or greater "surprisingness" would arise. The idea is to try to circumvent specifying aspects of an alternative hypothesis "inessential" to the problem. Broadly put, "surprisingness" is related (inversely) to 'relevance and rareness" in observations (referred to as "RR" values), with "relevance" referring generally to association of the data with what might be expected to occur in a tomb of the NT family, and "rareness" connected with, but not identical to, a relative frequency associated with those data.

[26]Conceivably, one could argue here to omit the broader class (b) and allow only (a). However, owing to the presence on our list of a sister whose name might be Mariam, this decision again is inconsequential.

[27]Strictly speaking, Jesus was known *only* by the version Yeshua of the generic Yehoshua. However, the full and formal Yehoshua is *never* used in Second Temple documentary texts [T. Ilan, private communication] and for this reason we allow only the generic name category here.

An approach based on "surprisingness"—or "RR" value—possesses some useful features: First, it provides a more "natural" method for specifying relative probabilities for clusters of names under the alternative. It also permits us to deal effectively with the fact that names of "relevant" persons can present in more or less rare renditions; such renditions may be nested and different "RR" values assigned to them. Furthermore, it leads us, in a natural way, to recognize that under the *alternative* hypothesis, the probabilities associated with any given set of names are not invariant under configurational rearrangements of those names; it also provides an intuitively natural way to encode subtle features of the probability structure arising out of the complex of family interrelationships. The method is also useful in helping distinguish between those aspects of the alternative that are of an a priori nature from those that are a posteriori; in particular, it allows us to more easily recognize that the test procedures can themselves be allowed to depend upon certain aspects of the observed tomb *configuration*. Last, but not least, the method affords us the convenience to ignore names whose evidentiary values are regarded as negligible, even though many such names would be viewed as not inconsistent with a NT tombsite. Such features make the method easier to implement than a carefully crafted likelihood ratio test which requires a precise specification of an $H_1$-probability structure. Any seemingly "incorrect" specifications of the alternative hypothesis will only result in some modest losses of power.[28] We see it as not disadvantageous to make that sacrifice, viewing it as partial payment toward any inadvertent *post hoc* indebtedness in the inference.

Returning to our discussion on measuring surprisingness, if we were to mirror a standard hypothesis testing setup, $H_0$ might be the assertion that the observed configuration of names arose by purely random draws from the onomasticon; an alternative "$H_1$" would be an opposite of $H_0$ relevant to the "NT hypothesis" that the tombsite is that of the NT family. A "sample space" would consist of all possible drawings from the onomasticon, subject to the conditioning of there being two women, and four men, two of whom are in father-son generational alignment. Some modest "realism" restrictions on points in the sample space may also be required; specifically, within a small tomb, the exact name renditions of deceased persons ought to differ. We next need to order the points in the sample space "along an $H_0$—$H_1$ continuum." (This occurs naturally in the classical setup once $H_1$ is specified fully.) It cannot be entirely unambiguous as to how such an ordering should be defined; loosely put, we want to order points on the basis of how "convincingly" they reflect what one might expect to find in a NT family tombsite. Among name clusters not inconsistent with "$H_1$" this ordering could be on the basis of the probability, under purely random sampling, of prespecified aspects of the cluster that most convincingly "allude" to "$H_1$." Thus, for example,

[28] This occurs because the presence in the sample of nonrelevant names, and the absence of relevant ones is not fully "optimized" for, although such "mathematical" optimality here is more apparent than real.

the presence of a name such as Matya should not disqualify a cluster (since it is not inconsistent with the genealogy) although its probability contribution might be discounted (e.g., set to unity)—for we have, after all, no idea who this may be—while a more "rare but relevant" name such as Yoseh should have its probability accounted for in the computation. The "probability," or "RR value," resulting from a computation of this type (with the familial and other adjustments to be discussed below) will be used to measure "surprise"; smaller RR values $\Rightarrow$ greater surprise. A "tail area" for assessing "evidence" against $H_0$ would then be based on the overall probability, under the $H_0$-sampling, of the set of points in the sample space whose RR values are less than or equal to that of the observed outcome (i.e., which are as or more "surprising"). If this "tail area" is sufficiently small, we may then consider to invoke the standard logic and conclude that either we have witnessed an event of rare chance, or the null hypothesis must be untrue. We are being cautious not to use the term "$p$-value" here; a more careful discussion of the interpretation of a small tail area will be undertaken in Section 14. For an appropriate definition of "surprise"—which must be specified a priori—a key computational question then becomes: *What is the probability that a* (*permissible*) *random sample of two female and three*[29] *male ossuaries, configured as at Talpiyot, contains a cluster of names which* (*relative to this* $H_0$ *and* "$H_1$" *setup*) *is as or more* "*surprising*" *than the cluster found?*

It is perhaps worth remarking that if we proceeded classically and carried out a LR test on the basis of a priori hypotheses such as APH 1–APH 8, then if a Wald's $\chi^2$ type of approximation were applicable we would need the probabilities (under $H_0$ and $H_1$) only for the observed data point, that is, only for the names and configuration observed. But whether such a test is carried out exactly via enumeration, or only approximately via a Wald's approximation, the out-of-sample names will matter only as to their number and their probabilities, the actual names themselves will not matter; and in turn, their number and their probabilities are required only for determining the distribution of the LR test statistic under $H_0$. The imprecision in this assertion pertains primarily to matters concerning the tomb configuration and familial interrelationships among the names. But if one already includes within the alternative those names that are "configurationally active," thereby accounting for their contribution to the overall "$H_1$" probability structure, the inclusion of additional names becomes essentially straightforward, and our assertion then holds more precisely. This opens the possibility that seemingly quite different versions of "$H_1$" could lead to essentially similar test results. As long as two versions of "$H_1$" were not particularly opposed to any of the names in-sample, but otherwise had (possibly quite) different sets of out-of-sample candidates, although approximately the same in number and with comparable "$H_1$" probabilities, then the results of the tests should be similar. The robustness of any procedure—specifically

[29]The Yehuda ossuary is initially being excluded in our analysis.

to the "$H_1$" specification—could then presumably be checked by allowing for different numbers of out-of-sample names, and different probabilities for them, with the actual names themselves not mattering. Robustness to *moderate* variations in the "$H_1$"-probabilities of in-sample (as well as out-of-sample) names could presumably also be checked in this way, although the same cannot be said for *contentious* "$H_1$"-disagreements concerning any of the in-sample names.[30] It is to be understood, throughout our discussions, that all versions of "$H_1$" require a broad category of "Other" for all of the essentially "uninformative" names that could occur but are not otherwise considered to be inconsistent with "$H_1$."

One final point arises from the fact that even if a "person" in the tomb is on our "$H_1$" a priori list, we do not know what rendering of their name will occur, and in particular how "relevant and rare" that rendering will be. If a name version is rare, this would be evident from Ilan's lexicon. However *rare names are not rare* and there may well be more than one possible rare rendering for any particular individual. In the end, however, the ossuary of such an individual would have been rendered in (at most) one such way. Hence the "rare names are not rare" concern does not apply so much to any particular individual's many potential rare renderings, as it does to the case that too many *persons,* each having rare name forms, are all considered to be likely candidates under "$H_1$." Accounting for this requires that we carry out the analysis allowing for more rather than fewer possible candidates having rare names.

**12. A proposed method for analysis: the RR method.** With the background of the previous section behind us, we may now describe in further detail our proposed paradigm based on "surprisingness."

Because our inference is conditional on the observed configuration, our procedure may depend on that aspect of the observed data (although not on any other). For the sample of the two women, we (initially) consider the case that "$H_1$" allows selection from a list of persons which consists of Mary Magdalene, Mary, Mariam, Salome, and "Other," together with their corresponding name rendition classes as defined in Section 10. In numerical experimentation, this list may be reduced, and/or augmented by Joanna, Martha, Woman(1), Woman(2), etc., where "Woman(i)" is considered to be a "relevant" (out-of sample) person whose name is

[30]In the present context, one of the more contentious "$H_1$"-disagreements centers around the inclusion of Mary Magdalene as an "$H_1$"-candidate. We note that—*on purely technical grounds*—this contentiousness makes her an ideal "$H_1$"-candidate for *some* "hypothesis test." In any case, it may be that the contentiousness of Mary Magdalene as an "$H_1$"-candidate has arisen because some interpret this as intending to imply that she was a spouse to Jesus although no such assumption is made here. Sensitivities to this issue appear to have been heightened due to a recent fictional account; see Ehrman (2004). Another source for this contentiousness possibly arises from the tradition that regards Mary Magdalene as a "sinner"; the earliest historical accounts, however, do not corroborate that view.

left unstated.[31] This list is intended to reflect APH 1–APH 8. The category "Other" groups together all other female names, in particular those considered not to be informative. Any selection of a female name from the above list is "relevant," except for "Other." The "RR" value for each of the names on this list is typically, but not invariably, the probability of the rarest rendition category (among our pre-defined categories) of the observed version of that name under random sampling from the onomasticon; the name category "Other" is discounted by being assigned an RR value of 1. The RR measure (relevance and rareness) for a set of two womens' names is defined as the product of their individual RR values. The sampling of these womens' names is carried out by drawing independently from the onomasticon, except that we do not allow any name *rendition* to occur twice.

Turning next to the men, the list of persons under "$H_1$" is taken initially to consist of Joseph (as a father), Jesus, Joses, James, and "Other," to be augmented in numerical experimentation by Cleopas, Male(1), Male(2), etc., with corresponding rendition classes again as given in Section 10. The RR values for each of these name renditions are computed from their onomasticon frequencies, except for the uninformative category "Other" which is assigned an RR value of 1 and is otherwise treated as before. With these conventions, male names are selected under random sampling from the onomasticon, and assigned to the two singleton male ossuary slots, and the two slots on the generational ossuary. This random sampling for the men is restricted by realism requirements to ensure that "no man dies twice," and that a father and son have different names. The RR value (relevance and rareness) for the sample of the four male names is then defined as the product of the RR values of the individual names, except for adjustments deriving from NT familial relations detailed below. The RR value for the combined male and female sample has yet to be defined; for the moment may we take it to be the product of the RR value for the females and the RR value for the males.

The exceptions to the $H_0$-sampling may be summarized as follows. For the women, we do not permit any name rendition to occur twice. Likewise for the men, for any configured set of four male names we do not permit the name renditions of the two singleton males to be identical (unless both are "Other") as one would (again) not expect two inscribed ossuaries to have been left indistinguishable in a small tombsite. Furthermore, we do not permit the father and son name renditions to be the same (unless both are "Other"). And finally, we do not permit the name rendition of the son to also be that of one of a singleton male (unless "Other"), the idea being again that "a person cannot die twice."

We next indicate the nature of some of the definitional adjustments to the RR value imposed by "$H_1$." As it happens, these involve only the names for the males,

[31] The actual names will be unimportant except for those in-sample; only the number of such name categories and their probabilities or RR values will actually matter. This will also be the case for some of the configurational aspects operative in the case of the men.

and often involve the father in the generational ossuary. Typical among such restrictions and adjustments are the following. If the father is "Other," then the RR value for the generational ossuary is set to 1 regardless of the son's name, for we then do not know who that son may have been and therefore discount it. Next, the father of that pairing is not permitted to be Yeshua[32]; in that case we set the RR value for the pairing to 1, or even to $\infty$, the net effect being about the same. If Yosef is the father and the son is "Other," and if that Yosef cannot be the biblical brother by virtue of there also being a Yosef in the tomb, the RR for the generational ossuary is set to 1, since we again do not know who this Yosef is. These considerations are far from complete; a complete set of restrictions and adjustments of this type will be detailed in the following section.

Finally, the RR value for the entire sample is defined as the product of the RR value for the females and the RR value for the males but possibly with exceptions of the following type: We may consider requiring the name Yeshua to appear as either the son in the generational pairing or as one of the singleton males; the idea here is that nothing beats the "surprisingness" of the *ne plus ultra* name Yeshua—appearing in a consistent manner—in a tombsite being gauged for having belonged to the NT family. Nevertheless, the inferences do need to be checked for robustness to requirements of this nature. As long as the definition of "surprise" (or "RR" value) is specified fully and a priori, the resulting approximate "tail area" will essentially be "valid"; all that is still required would be to determine the distribution of the "RR" values under the null hypothesis.

**13. A statistical analysis.** In this section we summarize a statistical analysis of the Talpiyot tomb data based on the paradigm developed in the previous two sections. Our analysis, however, is predicated upon a particular set of assumptions. Statistical analysis often follows from factual direction by subject matter expertise—in this instance from specialists in the history of early Christianity, in ancient scripts and carvings, and so on. The assumptions A.1–A.9 under which we carried out our analysis[33] are by no means universally agreed upon. Furthermore, the failure of any one of them can be expected to impact significantly upon the results of the analysis. We begin by itemizing these nine assumptions.

- **A.1:** We assume the "physical facts" to be correct: that the Talpiyot burial cave was found and provenanced properly, that it had remained essentially undisturbed since antiquity, and that no ossuaries were moved into or out of the tomb between the time the burials took place and the time in 1980 when the tomb was excavated.

[32]Having one Yeshua in the tomb as a father is "problematical" enough; a second is not being permitted.

[33]These assumptions were proposed by S. Jacobovici, except for A.6 and A.9 which are due to the author.

- **A.2:** We assume that if any ossuaries bearing inscriptions were removed from the tomb they were removed *haphazardly* and with no intent to mislead "in the direction" of "$H_1$"—that is, without regard to inscriptions that may have been inconsistent with "$H_1$."
- **A.3:** We assume that the historical and genealogical information relied upon here is adequately accurate. In particular, we assume that the most appropriate rendition of the name for the mother is Marya, for the father is either Yehosef or Yosef, and that those for the siblings are as given in the NT, with the second brother's (Yoseh's) most appropriate name rendition being as in Mark 6:3 of the NT.
- **A.4:** We assume that the ossuary inscribed "Yehuda son of Yeshua" can be explained and may be disregarded in our analysis. (We shall revisit this point in Section 14.)
- **A.5:** We assume the approximate validity of the demographic estimates for Jerusalem, in particular for the number of Jewish adults deceased within the relevant time spans, for the number of ossuary burials that took place, and for their inscription rates.
- **A.6:** We assume that the lexicon of Ilan (2002) provides a sample of names of persons from the relevant era sufficiently representative for our purposes, and that our implementation for their frequencies is appropriate.
- **A.7:** We assume that the full inscription Mariamenou [$\eta$] Mara refers to a single individual and represents the most appropriate specific appellation for Mary Magdalene from among those known; we further assume that this inscription is sufficiently distinctive that it could only have applied to very few and/or very particular individuals within the generic Mariam name category. Our specific implementation of this assumption will be of the type to assume that essentially at most one out of every 74 Mariams could legitimately have been rendered in this way, and that Mary Magdalene was among those who could.
- **A.8:** We assume that the inscription of the father "Yehosef" on the "Yeshua ossuary" and the inscription "Yoseh" on that individual's ossuary were meant to distinguish among two different persons.
- **A.9:** We assume that on a *time cross-sectional* basis, the assignment of names is adequately approximated by independent sampling; thus, for instance, that fathers called Yehosef would name a son Yeshua with about the same incidence as occurs in the general population, and so on. (See also Section 14.)

We turn now to our analysis, stressing again that it is predicated upon *all* of the hypotheses APH 1–APH 8 and the assumptions A.1–A.9. We compare "surprisingness" (or rather "RR" values) for Talpiyot-like configurations of names, when sampled randomly from Ilan's onomasticon, with the corresponding values for the arrangement actually observed; these computations were based on complete enumeration over the onomasticon.

Our *baseline computation* involves sampling from the womens' name rendition categories

$$\text{MM}, \quad \text{Marya}, \quad \text{Mariam}, \quad \text{Salome} \quad \text{and} \quad \text{Other},$$

with relative frequencies

$$\frac{74 \times (1/44)}{317} = \frac{1.68}{317}, \qquad \frac{74 \times (13/44)}{317} = \frac{21.86}{317}, \qquad \frac{(74 - 1.68 - 21.86)}{317},$$
$$\frac{61}{317} \quad \text{and} \quad \frac{317 - 74 - 61}{317},$$

and assigning to these renditions the "RR" values

$$\frac{1.68}{317}, \qquad \frac{21.86}{317}, \qquad \frac{74}{317}, \qquad \frac{61}{317} \quad \text{and} \quad 1,$$

respectively; here MM stands for "Mariamenou [$\eta$] Mara" (or equivalently for our data, just Mariamne). The frequencies for MM and Marya were discussed in Section 7; the frequency for Mariam is based on the complement in the set of generic Mariams[34] after the MMs and the Maryas are removed. The RR values assigned to the name categories are the same as their corresponding assigned frequencies, but with several exceptions: The RR value for a Mariam who is not an MM or a Marya, is based on the frequency of the entire generic class; the rationale for this is that this is now a very common rendition of a very common name, and while it is consistent with the NT genealogy, it carries reduced evidentiary value. Also, the name category "Other" is assigned an RR value of 1; higher RR values still could be assigned to any womens' names thought to invalidate the find although we did not implement such an invalidation set—the impact of this being, of course, conservative.

The mens' name rendition categories for our *baseline computation* are

$$\text{Yosef}, \quad \text{Yeshua}, \quad \text{Joses}, \quad \text{James} \quad \text{and} \quad \text{Other},$$

with relative frequencies

$$\frac{(221 - 33.63)}{2509}, \qquad \frac{101}{2509}, \qquad \frac{221 \times (7/46)}{2509} = \frac{33.63}{2509}, \qquad \frac{43}{2509}$$

and

$$\frac{2509 - 221 - 101 - 43}{2509},$$

and RR values

$$\frac{221}{2509}, \qquad \frac{101}{2509}, \qquad \frac{33.63}{2509}, \qquad \frac{43}{2509} \quad \text{and} \quad 1,$$

[34] "Mariam" is being used in two senses here: as the generic name category, and as the "other" Mariams after the specialized ones are removed. This will also occur with the name Joseph. The intended meanings should be clear from the context.

respectively. The category Other is again assigned an RR value of 1. The frequency (as well as the RR value) for Joses was discussed in Section 7. The RR value for Yosef is based (initially at least), on the full generic Joseph count—again on the grounds that it is now a most ordinary rendition, although for the renditions of Yosef the situation will actually be more involved since they could refer to either the brother or to the father; we shall need to revisit such issues below.

If, in numerical experimentation, any of our baseline name renditions are removed from our a priori lists, the adjustments required to the frequencies and RR values of the remaining ones are the natural ones. And if any names such as Joanna, Martha, Cleopas are added to that list, the frequencies and RR values associated with them will be based on Ilan's (nonfictitious persons) counts, namely $12/317$, $21/317$, $7/2509$ (and so on), respectively. Updates to the frequencies for the categories of "Other" women and/or "Other" men are also the obvious ones.

To further appreciate the nature of the complications that may arise consider, for example, finding a Cleopas son of James ossuary. Should such a James be viewed as being the biblical brother with a hitherto unknown son? Or should this James be viewed as being the biblical grandfather? We are obliged to establish rules for differentiating among such possibilities.

The reader will hardly fail to notice—as our definition of "surprisingness" and "RR" value takes shape—the many judgement calls involved in their definition. Our choices are meant to mirror the intent that RR should essentially measure the probability contribution only for those aspects of the find that are considered *relevant* and *knowable* for the NT family; however what is important is that these judgement calls all be of an a priori nature and this we are attempting to do on a best efforts basis. Experimentation appears to confirm that "sensible" variations in the definitions do not make a great difference to the results of our computations—as long as one is operating within the same set of a priori hypotheses and assumptions, namely APH 1–APH 8 and A.1–A.9.

In addition to the "realism"-based sampling restrictions outlined in the last section, the computations in our *baseline* case involve a series of 14 configuration-related familial adjustments to the RR values whose interactions with each other can be a bit complicated. These were devised on the basis of what is believed known of the genealogy of the NT family and of our relative expectations of how one may have thought such names might or ought to be configured in a NT tomb. The parameters proposed below were all selected on the basis of appearing to be reasonable a priori choices, but the sensitivity of the computations to these choices was nevertheless checked to gauge their influence. For the *baseline case*, we now itemize the complete set of adjustments to the RR values as implemented in our "R" computing code [Feuerverger (2008)]:

- If the father is Yeshua, the RR value for the generational ossuary is set to 1.
- If the father is Other, the son's RR value does not count (i.e., is set to 1).
- If the father also appears as one of the singletons, his name is not counted twice toward the RR value.

- If the two singleton males are Yosef and Yoseh,[35] then under "$H_1$" we do not know who Yosef is and therefore set his RR value to 1.
- If Yoseh is the father then the RR value for the son is set to 1 since the biblical brother Yoseh did not have a son whose name we know. However, since it was not uncommon for sons to be named after close blood relatives we shall allow the particular names Yeshua, Yosef, James, and Cleopas for the son,[36] but in those cases we discount the RR value for those son's names by multiplying by 5.
- Likewise, if Cleopas is the father then the RR value for the son is set to 1, however, we shall allow the particular names Yosef, James, and Yosa for the son but in those cases we discount the RR value for those son's names by multiplying by 5.
- If Yoseh is the father, and a Yosef appears as a singleton, then we do not know who that Yosef is (even though this name is not considered to be invalidating) and so we assign to that Yosef an RR value of 1.

In respect of the next four points (with Yosef being the father in each), we bear in mind that the name Yosef can refer to either the biblical brother or to the biblical father, unless Yoseh is the name of the son or a singleton male, in which case Yosef can only refer to the biblical father or to someone we don't know; we must therefore make RR value adjustments to account for the resulting scenarios:

- If Yosef is the father but is not also a singleton male, and if Yoseh is either the son or a singleton—thereby ruling out that Yosef is referring to the biblical brother—then the RR value for the generational ossuary is set to 1, unless the son is either Yeshua, Yoseh or James, in which case the generational ossuary receives its "full" RR value.
- If the father is Yosef and is not also a singleton male, and if a Yoseh does not also appear in the tomb—thereby making it possible that Yosef refers to either the biblical father, the biblical brother, or to someone we don't know—then the RR value for the generational ossuary is set to 1, unless the son is either Yeshua or James, in which case the generational ossuary receives its "full" RR value.
- If the father is Yosef and he is also a singleton male, and a Yoseh does not appear in the tomb then he can only refer to the biblical brother or to someone we don't know. In either case we do not know the name of the son. For our baseline case we allow the son to be either Yeshua or James but multiply that son's RR value by 5, and apply the usual RR value for the name Yosef.

[35] Note that the case where Yoseh is the son and Yosef is a singleton will get handled (*q.v.*) by the fact that if Yoseh is the son of anyone other than Yosef then he cannot be the biblical brother. The reverse case where Yosef is the son and Yoseh is a singleton will get handled (*q.v.*) by the fact that Yosef will then be an unknown person.

[36] These four names correspond to persons believed to have died prior to the year 70 CE.

- If the father is Yosef, and Cleopas is the son, and if a Yoseh is nowhere in the tomb, then regardless of whether or not Yosef is also a singleton, we treat him as referring to the biblical brother. The RR value for the generational ossuary is then the product of the RR values for Yosef and Cleopas except multiplied by 5 since that son's name was not known.

In respect of the next two points (James being the father in both), we bear in mind that the name James can refer to either the biblical brother or to the biblical father of Yosef and Cleopas; we must therefore make RR value adjustments to account for the resulting scenarios:

- If James is the father and is also one of the singletons, then under "$H_1$" he can only refer to the biblical brother or to someone we don't know and cannot refer to the biblical grandfather Jacob. In this case we permit the son to be either Yoseh, Yeshua, or Yosef, or even Cleopas, but we multiply the son's RR value by 5.
- If James is the father and not also one of the singletons, then he can be referring to either the biblical grandfather or to the biblical brother. In that case, if Cleopas is the son the generational ossuary is given its full RR value, but if the son is Yoseh, Yosef or Yeshua, the son's rarity is multiplied by 5.

And one final adjustment:

- If Yeshua is the son, and Yosef is the father, then in the baseline case we apply a "bonus" factor to this "prize" case by dividing the RR value by 1.2.

In numerical experimentation, the "downweighting" factor of 5 for "unknown sons" was varied and we also could entirely disallow RR contributions for the names of such sons. We also could omit the 1.2 bonus factor for the Jesus son of Joseph combination. Further, we could also require that a Yeshua must appear in the tomb before it could be considered to be as "surprising" as that at Talpiyot. Experimentation confirms, however, that the results of the computations are not unduly influenced by modest variations in such specifications for the definition of the RR values as long as such rules are selected in a generally reasonable way.

We turn finally to the results of our computations which are based on exact enumeration over Ilan's onomasticon. There are, firstly, a total of $317^2 \times 2509^4 = 3.982 \times 10^{18}$ possible samples (of persons) that can be drawn from the onomasticon (if order is allowed to matter); of these, $3.608 \times 10^{18}$ pass our "reality" requirements—that is, approximately 90.6% of drawn samples are "valid." For the Talpiyot tombsite, the RR values are computed as

$$\frac{74 \times (1/44)}{317} \times \frac{74 \times (13/44)}{317}$$

for the women,

$$\frac{221 \times (7/46)}{2509} \times 1$$

for the singleton men, and

$$\frac{101}{2509} \times \frac{221}{2509} \Big/ 1.2$$

for the generational ossuary, with the RR value for the overall find then being the product ($1.451 \times 10^{-8}$) of these three RR values; this computation takes into account all of our baseline rules including the 1.2 bonus factor for the prized Jesus son of Joseph pairing. Next, for our baseline context, we found that $1.981 \times 10^{12}$ of the "valid" samples have an RR value less than or equal to that of the Talpiyot tomb—that is, are considered to be as or more "surprising" than the Talpiyot find; the proportion of these is $5.491 \times 10^{-7}$, or about 1/1,821,000. Multiplying this proportion by 1,100, that is, by the estimated maximum number of Talpiyot-like tombsites that can be formed from all inscribed ossuaries that had been produced in that region and in that era—gives 0.0006041, or about 1/1,655. The interpretation of such a "tail area" is discussed in Section 14.

One intuitive explanation for this (baseline) result is as follows. The names of the four males can be arranged in 12 different configurations—4 choices for father, then 3 for son, the other two being singletons whose order does not matter. In Talpiyot the 4 male names which occur there were found in their unique "best" configuration. Loosely put, this contributes a factor of about 1/12 to the tail probability. When combined with the "rareness and relevance" of the Mariamenou inscription these largely counteract that we are looking at the best of 1,100 tombsites. The remaining names are not equally rare but they are nevertheless relevant ones and random sampling over the onomasticon does not beat them too easily, particularly when NT familial relationships are properly accounted for.

We next examine the sensitivity of this computation to the various parameter choices, restrictions, candidate lists, and so on, underlying the baseline case. (We do not, however, deviate here from any of the assumptions A.1–A.6.) The questions at issue here concern how far we can push the "$H_1$" specification before the results become meaningless. This "stress testing" work involves: (1) Adding additional candidate names to "$H_1$," and/or removing names; (2) Changing the probabilities or RR values for names in "$H_1$"; (3) Changing the numerical values of parameters; (4) Adding or dropping various "$H_1$" restrictions and/or configurational bonuses; and (5) Combinations of the above. To prevent this high-dimensional task from becoming unwieldy, we carry out such steps one at a time, as well as in judicious combinations.

The following tail areas are obtained under the indicated "single condition" changes from the baseline case:

- Require that Yeshua be in the tomb before it can be considered to be more surprising than that at Talpiyot: 0.000552.
- Remove the bonus factor of 1.2 for the Yeshua/Yehosef generational pairing: 0.000726.

- Reduce the rarity adjustment factor (of 5) for unknown sons by half: 0.000696.
- Double the rarity adjustment factor for unknown sons: 0.000604.
- Do not count unknown sons (set their RR value to 1): 0.000597.
- Remove Salome: 0.000367.
- Add Joanna: 0.00111.
- Add Martha: 0.00103.
- Add Cleopas: 0.00267 [worst case[37]].
- Reduce the frequency and RR-value for MM by half: 0.000181.
- Double the frequency and RR-value for MM: 0.000953.
- Reduce the frequency and RR-value for Yoseh by half: 0.000323.
- Double the frequency and RR-value for Yoseh: 0.00131.
- Allow the father on the generational ossuary to be named Yeshua: 0.000697.

The following results are obtained under the indicated "multiple condition" changes from the baseline case:

- Add Joanna and Martha: 0.00159.
- Add Joanna and Cleopas: 0.00463.
- Add Martha and Cleopas: 0.00429.
- Add Joanna, Martha and Cleopas: 0.00669 [worst case].
- Double the frequency and RR-values for MM and Yoseh: 0.00220.

In the next group of results, Joanna, Martha, and Cleopas are all included, this being the "worst" of the cases computed above.

- Remove bonus factor for the Yeshua/Yehosef generational pair: 0.00752 [worst case].
- Require that Yeshua be in the tomb before it can be considered to be more surprising than Talpiyot: 0.00380.
- Remove bonus factor for Yeshua/Yehosef generational pair but require that Yeshua be in the tomb: 0.00415.

In the next group of results, Joanna, Martha, and Cleopas are all included, and no bonus factor is used for the Yeshua/Yosef pairing; this is again the "worst" of the cases considered above.

- Do not allow the RR value for unknown sons to count: 0.00635.
- Reduce the rarity adjustment factor (of 5) for unknown sons by half: 0.00871 [worst case].
- Double the rarity adjustment factor for unknown sons: 0.00678.

[37] Adding Cleopas results in the greatest deterioration in "tail area" among "single condition" changes. Here, as well as in each block of results below, we indicate the "worst case" within the block. Shortly, we pursue "steepest ascent" based on such "worst case" results.

In the next group of results, Joanna, Martha, and Cleopas are all included, no bonus factor is used for the Yeshua/Yosef pairing and the RR adjustment factor for unknown sons is reduced by half. (This is the "worst" of the cases considered above.)

- Reduce MM (frequency and) RR-value by half: 0.00410.
- Double MM RR-value: 0.0193.
- Reduce Yoseh RR-value by half: 0.00414.
- Double Yoseh RR-value: 0.0173.
- Double MM and Yoseh RR-values: 0.0353 [worst case].

In the next group of results, Joanna and Cleopas are included, but Martha is excluded; no bonus factor is used for the Yeshua/Yosef pairing, and the RR adjustment factor for unknown sons is reduced by half.

- For the case just stated: 0.00594.
- Reduce MM (frequency and) RR-value by half: 0.00274.
- Double MM RR-value: 0.0132.
- Reduce Yoseh RR-value by half: 0.00281.
- Double Yoseh RR-value: 0.0116.
- Double MM and Yoseh RR-values: 0.0238 [worst case].

In our last group of results, Joanna and Cleopas are included, but Martha is excluded; no bonus factor is used for the Yeshua/Yosef pairing, and unknown sons are not counted toward the RR value.

- For the case just stated: 0.00423.
- Reduce MM RR-value by half: 0.00199.
- Double MM RR-value: 0.00944.
- Reduce Yoseh RR-value by half: 0.00190.
- Double Yoseh RR-value: 0.00836.
- Double MM and Yoseh RR-values: 0.0169 [worst case].

**14. Discussion and concluding remarks.** We begin with some remarks on our computations. In some respects, the results are driven by the conditioning on the observed configuration of the inscribed ossuaries in the tomb, and their number is fortuitously close to being "optimal" for "allowing detection." With more inscriptions the combinatorial growth of possibilities dilutes power and with fewer inscriptions the premium on "rareness" diminishes. (Fortuitous "relevant" rarenesses among the renditions which occurred also play a critical role.) However, even with this seemingly ideal number of inscribed ossuaries our "tail areas" become "not significant" if the set of a priori candidates for a NT tombsite and their sets of name renditions (rare ones, in particular) become too large. This also occurs if these lists exclude certain in-sample names and renditions, in particular the rare (and controversial) "MM."

A number of simplifications were used to bound computational labour. We have, first, not implemented a list of names which invalidate a find. However, doing so would only invalidate some of the samples under $H_0$ hence further reducing our "tail areas" since the Talpiyot site contains no such names; therefore the effect of that simplification is conservative. In fact, even within the generic names among our candidates, there occur renditions for them that also belong on our list of invalid names, or should at least be treated as "Other" so far as their contribution to RR value is concerned. The effects of our not having done so are again conservative since (1) the frequencies for the relevant names are then higher than they really should be, (2) because some of these renditions do not then wind up on an "invalid candidates" list, and (3) because these renditions are wrongly assigned "legitimate" RR values in cases when they should have been treated as "Other." A second labour-saving approximation involved not concerning ourselves unduly with the possibility of drawing identical name renditions (for the two women, or the two singleton men, or the father and son) when those names arose from the "Other" names categories; needless to say this should hardly impact on the results.

Certain additional items of "evidence" or "data" that may carry "information" relevant (in varying degrees) to our problem have not been incorporated into our analysis because such observations do not typically correspond to a priori hypotheses; the question of if, and precisely how, such information can be quantified in a *formal* statistical analysis is therefore problematical. The items of this type of which we are aware are: (1) The untypical carving of the circle and upward pointing gable on the entrance wall of the tomb; (2) The rightward leaning "cross" at the head of the Yeshua ossuary inscription which might be thought more distinctive than a mason's mark (although its meaning, if any, is not known); (3) The proximity of the tombsite to the Temple; (4) The unusually high proportion (6/10) of ossuaries bearing inscriptions; (5) The languages used on the inscriptions, and in particular the use of Greek script on Ossuary #1; (6) The fact that these ossuaries are all of adult size; (7) Purported mitochondrial DNA evidence suggesting that Yeshua and Mariamenou were not "maternally" related; (8) The alignment of the three names Yehosef, Yeshua, and Yehuda which appear on the two generationally sequenced father-son ossuaries ("A son of B son of C") being the *only* one among the six possible arrangements for those names that does not *immediately* invalidate the find; (9) Purported electron microscopy tests which suggest that the spectral element signature of the patina of the James ossuary matches to the Talpiyot tomb; and finally, (10) The relative absence of archeological features which could be used to help further rule out the possibility of this being the NT tombsite. Two further points also bear noting here. The first is that on a priori grounds, the sisters (Mariam and Salome, say) are perhaps less likely to occur in a NT tombsite due to the possibility that they may have been married and hence been with families of their own. The second is that if the disputed James ossuary were to prove authentic, then James could no longer be an a priori candidate. (A related consideration arises if James was buried at the place of his execution.) Needless to say, if any

of these out-of-sample names were "removed" from our a priori lists, or otherwise "downweighted," our "tail areas" would all decrease.

Let us next consider the impact of some of the assumptions. First, as concerns assumption A.8 (that Yoseh and Yehosef do not refer to the same person), the situation is somewhat subtle. While reasonable arguments may be advanced in favor of this assumption, if we were to choose to carry out an analysis without it, the probability structure under $H_0$ could then no longer be approximated by independence. Specifically, the drawings of the father and of the singletons would then become dependent in a way which cannot be specified in an obvious manner so that the combined RR value for Yoseh and the father Yehosef could then not be approximated by ordinary multiplication. One could, however, carry out analyses under two eventualities—the first (as we have done) under the assumption that these persons differ, and the second under the assumption that they are in fact the same. In the latter case, the father Yehosef in the generational ossuary would then become regarded as being the biblical brother (with only Yoseh, and not Yehosef, contributing toward the RR value), and the son Yeshua would then not count toward the RR value (or might count but in only a diminished way). Thus overall, without assumption A.8, the computations would not result in "significance."

Curiously, assumption A.4—regarding the Yehuda son of Yeshua ossuary—involves less computational complexity than at first seems since our analyses may in fact be carried out allowing for the presence of a full "generationally aligned" sequence "A son of B son of C." Because the NT genealogy has no *known* father-and-son pair with *both* dying between 30 CE and 70 CE, the youngest of this aligned trio—namely "A"—would *never* contribute toward the RR value. Hence the results of such analyses would actually be identical to those already carried out. A quite different conclusion would be reached, however, if the presence of this ossuary in the tomb was permitted to count "negatively," that is, in the direction of invalidating the find.

Concerning our specialized independence assumption A.9, a referee has argued that if the population of Jerusalem consisted of a small number of large clans, each sharing only a few ancestors, it could lead to name clustering, and the longitudinal dependences would then result in cross-sectional dependence as well. Of course, the cross-sectional approximate independence is ultimately a judgement call which we would have preferred to avoid, except that doing so would then limit the power of statistical procedures that can be devised. The data base for "assessing" this assumption more broadly (for the era in question) is limited, but it is not null. The series of "begats" in the NT are one potential data source which could be studied. More usefully, Ilan's (2002) compilation allows us to reconstruct some name matchings. Thus, of the 23 entries of (generic) Yeshua derived from ossuaries, 13 are matched with the name of either a son or a father, with two of these being matched with both a father as well as a son. (One further entry is matched with a Salome, presumably a wife or sister.) From that data a slight tendency may be discerned for fathers called Yeshua to also name their sons Yeshua, but little else

of significance is in evidence. Of the 45 entries of (generic) Yosef derived from ossuaries, 32 are matched with the name of either a son or a father, with one of these being matched with both a father and a son. [In two cases a daughter is mentioned (both times Martha). In another case a twin is mentioned (Eleazer), and in a related case two sons are mentioned (Eleazar and Joseph).] Two of these 32 cases indicate a son to be Yeshua (one corresponding to Talpiyot); none show Joseph as being a son of Yeshua. There appears to be a significant tendency for the sons and the fathers of (these ossuary-derived) Yosefs to have such rather unusual names as Shabi, Yoezer, Kallon, Agra, Benaiah, and so on. The impact of this on our analysis is conservative since the direction of the dependence implied only renders the Talpiyot observations more rare.

The last assumption we discuss here is A.7 concerning the name of Mary Magdalene. This assumption was suggested to us under the rationale outlined in Section 6 and it is the case that without the "rareness and relevance" of the Mariamenou [$\eta$] Mara inscription our test procedures would not prove "significant." Having no germane historical expertise, the author worked under this assumption, but the question may fairly be put as to whether or not it arose under the influence of the data. For inferences to be valid, the renditions for Mary Magdalene (particularly the most specialized ones) must, of course, be specified a priori. As this point will no doubt be argued by others it is unnecessary for us to belabour it here; however we offer two comments. First, our analysis does indeed assume the name of Mary Magdalene to have been either Mariamne or Mariamen (or a closely related rendition), a point legitimately subject to corroboration—or otherwise—by historical scholars. Should such scholarship ultimately prove inconclusive, an approach along the following lines may perhaps be considered: We have at our disposal a list of some 80 Mariams of the era whose actual name renditions are known to us; this includes the two Mariams from the Talpiyot find. If now we sought to categorize these 80 renditions according to the degree to which they appear to be appropriate ones for Mary Magdalene then it might well be that the rendition Mariamenou [$\eta$] Mara would be the one selected as being the most so. Here again, it would be the remarkable character of that rendition that would lead us to offer it that consideration. A separate issue is whether or not Mary Magdalene's candidature is legitimately a priori; while the logic behind the hypothesis APH 5 of Section 10 is "best efforts"-based, it is not incontestable.

The issues arising from the remaining assumptions, as well as their impacts on the analysis are more straightforward. We only remark, yet again, that all of the assumptions must be met for our "tail areas" to be meaningful.

Finally, concerning the (disputed) ossuary of James, it has been speculated that it might actually provenance to the Talpiyot site. On the basis of the currently available evidence the author does not believe any such claim to have been established, but its impact on the computations can nevertheless be described. First, with that ossuary included the statistical "significance" of the find would strengthen substantially even though the number of ossuaries conditioned upon would also have increased. No additional "RR" value would accrue for the common father, although

some modest contribution might accrue on account of two patronymic ossuaries then likely being brothers. As for the (disputed) "brother of Jesus" component of the inscription, no further "RR" value would accrue from the repeated mention of Jesus. Of course, the mere mention of *that* particular name, *and in this way*, would obviously be considered to be sufficiently remarkable that any further statistical efforts would be rendered unnecessary.

Let us finally turn to the question of how one may interpret the "tail areas" computed in the preceding section, that is, the proportions ("under $H_0$") of obtaining "surprisingness" values as great as at Talpiyot. The issues here are not straightforward. Suppose, for the sake of this discussion, that agreement has been reached with respect to *all* of the hypotheses, assumptions, and conditions upon which our computations were carried out; we shall hereafter collectively refer to these as our *provisos*. Using our "baseline" case for purposes of illustration, our computations suggest that a clustering of names as "surprising"—that is, "as relevant and as rare"—as those at Talpiyot occurs (approximately) once per 1,821,000 tombs under random sampling from the onomasticon. This number is considerably greater than the number of persons—let alone families—that died during the relevant span.[38]

We are, in fact, now in a position to carry out a particular *hypothesis test*: Here $H_0$ is the hypothesis that *all* 1,100 tombs in the vicinity of Jerusalem arose under random assignment of names, and $H_1$ is the hypothesis that *one unspecified one among these* 1,100 *tombs is that of the NT family.* The test statistic we shall use for this purpose is the lowest $H_0$-tail area for the RR values of the 1,100 tombs. A $p$-value for this test is *bounded above*[39] by the probability that one among these 1,100 tombs would have an RR value corresponding to an $H_0$*-tail area* less than or equal to $1/1{,}821{,}000$; this probability bound is $1/1{,}655$. *We therefore conclude*, *subject to the stated provisos*, *that there exists a NT tombsite*, *and furthermore that it is one of the* 1,100 *tombs in the vicinity of Jerusalem.* This is the first step in our inference, although it may be bypassed if we are prepared to accept the stated conclusion.

Interestingly—if counterintuitively—we cannot as an *immediate* next step conclude from this that the tomb at East Talpiyot must be that one. Our finding does, however, permit us to *objectively* assign a probability of $1/1{,}100$ of being that of

[38]Hence if, for example, the entire population could be divided into 10,000 Talpiyot-size tombs, the probability is $1/182$ (under random assignment) that another family would have matched this tail area, and $1/1{,}655$ that such a family would have occurred among the 1,100 existing tombs. Of course, larger families could have better odds that some *deliberately selected* subset of their names might be deemed to be as "surprising."

[39]It is *bounded above* because not all existing tombs have as yet been "measured," and one or more among them could conceivably provide a still lower tail areas. The fact that not all tombs were configured identically complicates our arguments, however such conditioning is accepted statistical practice.

the NT family to any randomly selected one among these existing 1,100 tombs. Constructing a formal hypothesis test for whether or not the East Talpiyot tomb is actually that one is however not straightforward[40]—a price we pay for the absence of a probability model (for RR values) under the "NT hypothesis." We are thus faced with the situation that we know (with $p = 1/1{,}655$) that one of the 1,100 tombs in the vicinity of Jerusalem is the NT family tombsite, and furthermore know that this knowledge was derived from an (extreme) RR tail area measurement which occurred *at a single tombsite*. And yet we cannot *immediately* conclude from this that this one tombsite must be that of the NT family. We do however know that the NT tombsite is either *the one* at East Talpiyot or *one of the others* among the 1,100 tombs in the vicinity of Jerusalem; unless a "type 1" error has occurred in our "first step," no other options are available.[41]

The second step in our inference involves the Bayes formula

$$\frac{P(A|B)}{P(\overline{A}|B)} = \frac{P(A)}{P(\overline{A})} \times \frac{P(B|A)}{P(B|\overline{A})}$$

for updating prior odds by a likelihood ratio. Here $A$ is the event that the Talpiyot tomb is that of the NT family, and $\overline{A}$ is the event that it is not. The conditioning event $B$ can be chosen in more than one way here. The "natural" choice—where $B$ is the event of obtaining the specific cluster of names found at Talpiyot—is awkward to work with. We shall condition instead on the event that the $H_0$-tail area of the tomb being examined is less than or equal to that which occurred at Talpiyot. In proceeding, the following notation will be useful. Let $n_1$ be the number of tombs in the vicinity of Jerusalem that have already been excavated; that number[42] is approximately 100. Let $n_2$ be the number of tombs—approximately 1,100—that exist in the vicinity of Jerusalem. Let $n_3$ be the number of tombs (of "Talpiyot size") that could have been formed had the *entire* population of Jewish adults been buried in tombs with inscribed ossuaries; that number is somewhat less than 10,000. Let $q$ be the $H_0$-tail area of the RR statistic for the Talpiyot tomb according our baseline, or to any other "case" being considered; the order of magnitude of $q$ is about $10^{-6}$. In this notation, the $p$-value for our test at step one is

[40]There are analogies between our problem and one arising in "DNA matching" where a probability $P(A|B)$ is computed, although $P(B|A)$ is the one desired. In our application, what has been computed is the probability of obtaining an equally "surprising" cluster of names *given* that the tomb is not that of the NT family while what is desired is the probability that this is the NT family tomb *given* that the cluster of names is so surprising. Some considerations that apply in such DNA studies therefore carry over to our problem. However our problem differs from the DNA one in that the DNA profile of the "accused party" is fully known, while the a priori profile for the NT tombsite is not.

[41]We shall not consider here the possibility that the foregoing arguments (as well as some others below) may be repeated using the 100 tombs already excavated in lieu of the 1,100 "in existence."

[42]The ossuary-sourced listings in Ilan also divide up into approximately 100 Talpiyot-like configurations.

$p = n_2 q$, while our odds-updating formula becomes

$$\frac{P(A|B)}{P(\overline{A}|B)} = \frac{1}{(n_2 - 1)} \times \frac{\theta}{q} = \frac{\theta}{(n_2 - 1)q},$$

where $\theta \equiv P(B|A)$ is the probability that a NT family tomb would consist of a cluster of names as surprising (based on our RR approach) as that at Talpiyot. Some readers may believe that $\theta = 1$, or in that order of magnitude; for them the inference process will now be completed. A similar remark applies to readers prepared to at least believe that $\theta$ is not terribly small.

Readers who prefer not to assume that $\theta$ is not very small may consider, as a third step, to obtain a lower confidence bound for $\theta$. Among the $n_2$ existing tombs, that of the NT family has probability $\theta$ of "attaining $q$" while the probability that one among the $n_2 - 1$ others does is given by $(n_2 - 1)q$ since their tail areas are uniformly distributed. Hence the probability that the tail area value of $q$ will be attained in the group of all $n_2$ existing tombs is given by

$$\tau \equiv \theta + (n_2 - 1)q - (n_2 - 1)q\theta = \theta[1 - (n_2 - 1)q] + (n_2 - 1)q.$$

This in fact is the probability of a Bernoulli event. A decidedly conservative $100(1-\alpha)\%$ lower confidence bound for $\tau$ is given by 0 if the "$q$-event" is not attained, and by $\alpha$ if (as in our case) it is. Solving $\tau \geq \alpha$ then gives the $100(1-\alpha)\%$ lower confidence bound

$$\theta \geq \frac{\alpha}{1 - (n_2 - 1)q} - \frac{(n_2 - 1)q}{1 - (n_2 - 1)q}$$

for $\theta$, from which we obtain the confidence bound

$$\frac{P(A|B)}{P(\overline{A}|B)} \geq \frac{\alpha - \beta}{\beta(1 - \beta)}, \qquad \text{where } \beta \equiv (n_2 - 1)q,$$

for the odds ratio; for small $\beta$, this bound is approximately $(\alpha/\beta) - 1$. For illustration, in our baseline case, $n_2 = 1{,}100$, and $q = 1/1{,}821{,}000$; if $\alpha = 0.05$ or 0.01, the lower confidence bound for $\theta$ is 0.0494 or 0.0094, and in turn the lower confidence bound for $P(A|B)/P(\overline{A}|B)$ will be 81.90 or 15.58, respectively. If we had assumed instead that $\theta = 1$, 0.5, or 0.1, then using the value $\theta/\beta$ we would have obtained odds ratios of 1657, 828 and 167, respectively. These results are, of course, all dependent upon our provisos.

To summarize now, in this paper we have conveyed an interesting data set and have provided some background essential for its interpretation. We have also proposed a paradigm intended to deal with the purely statistical questions such data pose—that based on "surprisingness," or the "RR" (relevance and rareness) measure. Although related to classical methods, this paradigm differs from them in a number of ways. In practice, there are probably few real-data-based analyses of consequence on controversial issues which do not lend themselves to counterargumentation. The results of our analysis could be challenged on the basis of the

methodology applied or the assumptions on which it was based. We hope that the statistical methodology itself will not be found unduly controversial. As concerns the assumptions, the situation is different; while we have provided a rationale for each, they are not unassailable. Furthermore, arguments could be mounted to the effect that no a priori lists of persons and name renditions could ever be legitimately assembled after the fact. The influence of the Mariamenou [$\eta$] Mara inscription in the analysis particularly flags it as a "target."

If the assumptions A.1–A.9 under which our computations have been carried out are accepted, and if an a priori list of NT tomb candidates, together with an a priori set of name renditions for them were accepted as well, and further, if the list of candidates contained at least those key persons which the Talpiyot inscriptions seemingly allude to, then our computations strongly suggest that the possibility that the Talpiyot tomb is that of the NT family merits serious consideration. Subject to the stated provisos, our numerical experiments also suggest that this conclusion is robust to *moderate* variations in the specifications of the lists of candidates and name rendition categories. It is also reasonably robust with respect to variations in the relative frequencies for these name renditions and with respect to "reasonable" variations in the components of our definition of "surprise" (or "RR" value).

Even if statistical significance of the "RR" value of the Talpiyot tomb were accepted as fact, nothing in the purely statistical aspects of our analysis *directly* addresses such questions as whether or not Jesus and Mary Magdalene might have been married, or whether or not they may have had a son; certainly other possible explanations exist as well. Further, statistical significance only establishes that either the null hypothesis must be false, or we have observed an event of rare chance; either of these are possibilities.

Among the various assumptions made, perhaps the one that most "drives" our analysis in the direction of "significance" is the extraordinary inscription Mariamenou [$\eta$] Mara. It has been speculated that Mary Magdalene was a principal driving force in the movement founded by Jesus but was later vilified in the course of patriarchal power struggles. While we are in no position to weigh in on any such theories, what we can say is that from a purely statistical point of view, this much is true: It is the presence in this burial cave of the ossuary of Mariamenou [$\eta$] Mara, and the mysteries concerning the identity of the woman known as Mary Magdalene, that hold the key for the degree to which statistical analysis will ultimately play a substantive role in determining whether or not the burial cave at East Talpiyot happens to be that of the family of Jesus of Nazareth.

**Acknowledgments.** For helpful discussions and other assistance the author is indebted to David Andrews, Nicole Austin, James Charlesworth, Radu Craiu, Tom DiCiccio, Laurel Duquette, Grace Feuerverger, Steve Fienberg, Camil Fuchs, Don Fraser, Shimon Gibson, Felix Golubev, Itay Heled, Tyler Howard, Tal Ilan, Simcha Jacobovici, Marek Kanter, Georges Monette, Hadas Moshonov, Radford

Neal, Nancy Reid, Ben Reiser, James Tabor, Robert Tibshirani, the staff at Associated Producers Ltd., the editors of this Journal and the referees, and certain others who prefer to remain anonymous. Due to confidentiality agreements the author was obliged to respect, not all those named were aware of the nature of this work; it goes without saying that the author alone is responsible for the contents of this paper. I particularly wish to reiterate my indebtedness to Simcha Jacobovici for bringing this extraordinary data set to my attention, for sharing his extensive knowledge base regarding this archeological find, and for facilitating scholarly contacts; the assumptions under which our analysis was carried out were proposed by him. I also thank Associated Producers for permission to reproduce the images of the six inscribed ossuaries. Special thanks to Tyler Howard for his astute suggestions and highly conscientious work in checking my original S-PLUS code and converting it into R code. I also take this opportunity to reaffirm my indebtedness to David Andrews for the many incisive insights he has generously shared with me over the years on the subtleties of statistically significant applications; no finer statistician have I ever known. Likewise, S. Fienberg and G. Monette were most gracious in supplying invaluable comments and suggestions on an earlier draft. Last, but not least, I wish to thank Professor James Tabor for having so generously shared with me his wealth of knowledge concerning historical matters of the New Testament era, and for tirelessly responding to many queries, particularly during the process in which the a priori hypotheses of Section 10 were being formulated.

## SUPPLEMENTARY MATERIAL

**Computing code for "Statistical analysis of an archeological find"** (doi: 10.1214/08-AOAS99supp; .txt). This file contains the R computing code used to produce the results in this paper. The code is self-explanatory and is easily modified to generate the reported results. It may also be modified to account for different assumption sets to enter into the "RR" (relevance and rareness) computations.

## REFERENCES

ANDREWS D. F. A. and FEUERVERGER, A. (2005). The statistical significance of Palm Beach County. In *Statistical Modeling and Analysis for Complex Data Problems* (P. Duchesne and B. Rémillard, eds.) 17–40. *GERAD 25th Aniv. SER.* **1**. Springer, New York. MR2189529

BOVON, F. (2002). Mary Magdalene in the *Acts of Philip.* In *Which Mary?—The Marys of Early Christian Tradition* (F. Stanley Jones, ed.) 77–89. *Society of Biblical Literature Symposium Series* **19**. Society of Biblical Literature, Atlanta.

BROSHI, M. (1977). How many people lived in Jerusalem (from the time of Herod until the destruction of the Second Temple)? ETMOL **14**. (In Hebrew.)

BROSHI, M. (1978). Estimating the population of ancient Jerusalem. *Biblical Archeology Review* June **4** 10–15.

DIACONIS, P. and MOSTELLER, F. (1989). Methods for studying coincidences. *J. Amer. Statist. Assoc.* **84** 853–861. MR1134485

EHRMAN, B. D. (2004). *Truth and Fiction in The Da Vinci Code.* Oxford Univ. Press, Oxford.

FLAVIUS, J. (1943). *Antiquities of the Jews.* (A. Wikgren, transl.) Harvard Univ. Press.

FEUERVERGER, A. (2008). Supplement to "Statistical analysis of an archeological find." DOI: 10.1214/07-AOAS99SUPP.

FUCHS, C. (2004). Demography, literacy and names distribution in ancient Jerusalem—How many James/Jacob son of Joseph, brother of Jesus were there? *Polish J. of Biblical Research* 1–30.

HACHLILI, R. (1994). The change in burial customs in the late Second Temple period. In *Tombs and Burial Customs in the Ancient Land of Israel Yitzhak Ben Tzvi Memorial and the Society for Research of the Land of Israel and its Antiquities* 173–189.

HACHLILI, R. (2005) *Jewish Funerary Customs*, *Practices and Rites in the Second Temple Period.* Brill, Boston.

ILAN, T. (2002). *Lexicon of Jewish Names in Late Antiquity, Part 1*: *Palestine 330 BCE–200 CE*. Mohr Siebeck, Tubingen.

KLONER, A. (1996). A tomb with inscribed ossuaries in east Talpiyot, Jerusalem. *Atiquot* **XXIX** 15–22.

LEVINE, L. I. (2002). *Jerusalem*: *Portrait of the City in the Second Temple Period (538 BCE–70 CE)*. Jewish Publication Society, Philadelphia.

RAHMANI, L. Y. (1994). *A Catalogue of Jewish Ossuaries in the Collections of the State of Israel*. Israel Antiquities Authority, Israel Academy of Sciences and Humanities, Jerusalem.

SHANKS, H. and WITHERINGTON, B. (2003). *The Brother of Jesus: The Dramatic Story and Meaning of the First Archeological Link to Jesus and His Family.* HarperCollins, New York.

TABOR, J. D. (2006). *The Jesus Dynasty.* Simon & Schuster, New York.

WILKINSON, J. (1974). Ancient Jerusalem: Its water supply and population. *Palest. Explor. Q.* **106** 33–51.

DEPARTMENT OF STATISTICS
RM. 6009, SIDNEY SMITH BLDG.
100 ST. GEORGE STREET
UNIVERSITY OF TORONTO
TORONTO, ONTARIO
CANADA M5S 3G3
E-MAIL: andrey@utstat.toronto.edu

*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 55–56
DOI: 10.1214/08-AOAS99A
Main article DOI: 10.1214/08-AOAS99


# DISCUSSION OF: STATISTICAL ANALYSIS OF AN ARCHEOLOGICAL FIND

BY STEPHEN M. STIGLER

*University of Chicago*

Statistics is the science of uncertainty, and it should be capable of helping to address even hard to quantify problems. Indeed, the very attempt to quantify may itself shed light and understanding, and can often lead to better articulation of even qualitative evidential arguments. Yet, when statistical ideas are used in areas where wide segments of the population hold strongly divided passionate views, areas such religion or politics, the entry of statistics into the discussion is seldom accorded a warm and friendly reception. Instead, the greeting is at best extraordinarily skeptical, with quibbling over minor points that would be passed by silently in less-contentious studies, and with inhospitality to even the best of intentions. At worst, the intruder is burned at the stake or removed from the rolls of the employed, although such extremes are rarer these days than they were at the time of Giordano Bruno and Galileo.

Is this resistance rational? Do questions like that confronted in Andrey Feuerverger's painstakingly honest study of an archeological find, questions involving broad public knowledge and wide publicity, require a different standard of proof than run-of-the mill scientific questions? I think they may well, for several reasons.

1. The very wide public attention to the area, even before the discovery of the evidence, changes the way we think of the evidence. For example, the temptations to persons of unknown identity (even in the distant past) to fraudulently manufacture evidence must be considered, and the weighing of potential forms of fraud in any modeling context is a highly vexing question.

2. Even aside from any possible fraud, the conditions surrounding the arrival of the evidence can legitimately raise questions that would never arise in more mundane investigations. For example, we are told that, "No information is available regarding the placement of the various ossuaries among the *kokhim*." But the names involved in this case are so universally recognized that it might be argued that the absence of information is in this case informative, as the dog who did not bark was to Sherlock Holmes. One might believe that had the ossuaries been arrayed together in a meaningful order, this would with some probability have been noted, and the lack of such notation suggests they were not.

3. Francis Galton issued a caution in 1863 for those dealing with small data sets with uncertain generating mechanisms: "Exercising the right of occasional

Received December 2007.

suppression and slight modification, it is truly absurd to see how plastic a limited number of observations become, in the hands of men with preconceived ideas" (*Meteorographica*, London: Macmillan, 1863, page 5). Since occasional suppression and slight modification can be a part of sound statistical analysis, it is easy to overlook this potential bias, for it will not always be obviously present or consciously operating in a deceptive way.

I commend Andrey Feuerverger for undertaking this investigation. That it may be greeted skeptically is no reflection upon him, only upon the nature of the question he considers. Some of the assumptions he forthrightly makes, such as the independent assignment of names in families, may not survive later scrutiny. But in the face of all these difficulties, his carefully qualified analysis reminds us that addressing a question is not the same as resolving it, and that issues of wide general interest where prior opinions are sharply divided present novel problems of statistical formulation. I look forward to the ensuing dialogue, which will hopefully have greater focus because of the pains Feuerverger has taken to frame and present the issues.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637
USA
E-MAIL: stigler@uchicago.edu

*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 57–65
DOI: 10.1214/08-AOAS99D
Main article DOI: 10.1214/08-AOAS99


# DISCUSSION OF: STATISTICAL ANALYSIS OF AN ARCHEOLOGICAL FIND

BY CAMIL FUCHS

*Tel Aviv University*

**1. Introduction.** The starting points of Feuerverger's paper are both exciting and promising: A scientific puzzle of major importance is settled by a novel statistical approach. The puzzle is related to the re-analyzed inscriptions on the ossuaries from an ancient tomb from Jerusalem unearthed in 1980. The new analysis, also documented in a book [Jacobovici and Pellegrino (2007)] and a documentary movie [Cameron (2007)], claims that the inscriptions indicate that this may be the burial site of the New Testament (NT) family. Undoubtedly, if validated, a discovery with potential to stir major interest both in academic as well as in religious circles. At this point, the statistical methodology is called to settle the controversy and a new statistical approach is developed to handle the intricacies of the complex problem.

The results presented in the paper seem to justify the prior excitement. In terms of the new approach, the defined level of "surprisingness" for the cluster of names in the tomb is found to be very high, that is, under the specified provisos, there is a very low probability that a random sample of such ossuaries contains a cluster of names which is more surprising than the cluster found. Furthermore, when the probabilities related to the level of surprisingness are translated into the classical terms of conditional odds ratios, the odds that the Talpiot tomb is that of the NT family are also found to be very high.

It seems like the statistical methodology succeeded in settling the controversy, and the verdict is in favor of the tomb being the NT family tomb. In the process, a new approach was developed to settle cases in which judgment has to be rendered on whether or not a multiple characteristics event is or is not a result of random draws.

On a personal note, I confess that I would have been very pleased to be able to conclude my discussion with two positive statements: (a) that I found the results convincing and we can second Prof. Feuerverger's claim that the tomb is most likely that of the NT family, and (b) that the new approach is preferable to the existing methods in deciding whether the tested object is the special one.

Unfortunately, to anticipate the findings detailed below, despite the initial excitement and the personal preferences, I find myself in disagreement with the results and the conclusions. As for the new approach, it may evolve and prove

Received October 2007.

beneficial, although not necessarily preferable to existing methods. I believe its properties have yet to be investigated.

**2. The statistical analysis.** Let us first briefly review the relevant statistical features in Feuerverger's approach and their application to the particular data set. The justifications of the above-mentioned contentions are presented in this context.

The analogue of a null hypothesis $H_0$ is defined to be the assertion that the observed configuration of names (on the ossuaries in the tombsite) arose by purely random draws from the onomasticon. The alternative $H_1$ is presented as "an opposite of $H_0$ relevant to the "NT hypothesis" that the tombsite is that of the NT family." An intermediate formulation (with weaker $H_1$) is also presented, with $H_0$ being the assertion that all possible tombs comparable to that of Talpiot arouse under random assignment of names and $H_1$ is the event that among the such possible tombs, one *unspecified* tomb is that of the NT family. With respect to the intermediate $H_1$ and for various prior-like probabilities, Feuerverger assesses from the $H_0$-tail area the odds ratios of the event that the Talpiot tombsite is that of the NT family.

The data from the Talpiot tomb includes six inscribed ossuaries with the following inscriptions:
#1:$Μαριαμηνου$ [$\eta$] $Μαρα$, #2: יהודה בר ישוע, #3: מתיה, #4: ישוע בר יוסף, #5: יוסה, #6: מריה
transliterated as:
#1: Mariamene [$\eta$] Mara, #2: Yehuda son of Yeshua, #3: Matya, #4: Yeshua son of Yoseph, #5: Yoseh, #6: Marya.

At least some of the names are reminiscent of the names related to the NT family. As a first step in determining how significant or (in terms of the proposed approach) how "surprising" is this find, one has to assess how common were those names in the vicinity of Jerusalem in the late Second Temple period. Table 1 presents the frequencies and the relative frequencies of the generic names out of the total compiled male and female nonfictitious names from ossuary and non-ossuary sources [Ilan (2002)]. Furthermore, the table also presents the frequencies and relative frequencies of the relevant renditions of Mary/Mariam and Yoseph from ossuary sources.

Under the proposed approach, the data analysis conditions on both the number of inscribed ossuaries and their gender distribution, as well as on the generational sequence in two of the four male ossuaries. However, the basic analysis deals only with the inscriptions from five ossuaries, with the Yehuda son of Yeshua ossuary being discarded.

Now, the new approach defines "an a priori defined" measure of "surprisingness" related to the $H_0$–$H_1$ continuum. The "surprisingness" value of a particular configuration increases as the configuration is in some respect closer to $H_1$. The reciprocal form of the "surprisingness" value is defined as "relevance and rareness" (RR value). "Relevance" refers to membership in an a priori list of candidates for

TABLE 1
*Frequencies of the named inscribed in the Talpiot ossuary*

| | All sources | | | Ossuary sources | |
|---|---|---|---|---|---|
| **Generic name** | **Frequency** | **Relative frequency** | **Renditions** | **Frequency** | **Relative frequency** |
| | | | **Female** | | |
| Mary/Mariam | 74 | 0.233 | Mariamene | 1 | 0.023 |
| | | | Marya | 13 | 0.295 |
| Females— Total all sources | 317 | | Total ossuaries— Females named Mariam | 44 | |
| | | | **Male** | | |
| Yehuda | 171 | 0.068 | | | |
| Yeshua | 101 | 0.040 | | | |
| Matya/Mattityahu | 62 | 0.026 | | | |
| Yoseph | 221 | 0.088 | Yoseh | 7 | 0.152 |
| Males— Total all sources | 2509 | | Total ossuaries— Males named Joseph | 46 | |

inclusion in an NT tombsite, and "rareness" is defined relative to an a priori list of nested possible name renditions for each such candidate. The initial relevant lists are supposed to include names which are reasonable to assume that they have potential to be found in a NT family tomb, based on a set of a-priori formulated hypotheses. The relevant lists have to reflect those hypotheses. In addition, the relevant lists are also allowed to include unrelated names, defined as "Other," as possibly belonging to persons about whom there are no records. The population and the sample are stratified, and separate a priori lists of tomb candidate name renditions are compiled by gender.

In the analysis of the Talpiot data the following assumingly a priori lists of tomb candidate name renditions for men and women are presented:
Men: Yoseph, Yeshua, Yoseh, James and "Other"
Women: Mary Magdalene (denoted MM or Mariamene), Marya, Mariam, Salome and "Other"

Thus the Matya from ossuary #3 is considered as "Other" (one of those possibly belonging to persons about whom there are no records), and Mariamene [$\eta$] Mara is added to the women's list as being "the most specific appellation to Mary Magdalene from among those known." As can be seen from Table 1, this is the only such exact rendition of Mariam among the recorded names.

The RR value of a datum or of a subset of data is defined as the adjusted relative frequency of occurrence of the components under independent random sampling from the onomasticon. The RR for a generic name is its relative frequency, while the RR value for a particular rendition of a generic name is computed as a product

of the name's overall relative frequency and the relative frequency from ossuaries sources of the particular rendition within the generic name. For some particular configurations, quite complex (and relatively reasonable) definitional adjustments imposed by $H_1$ are used in the computation of the RR values. In particular, a prized bonus is applied when Yoseph is the father and Yeshua is the son with the RR-value being divided by 1.2.

Under the suggested approach, the names defined as "Other" receive an RR value of 1, and thus have no effect on the product which yields the RR value for the entire cluster. As expected, and as illustrated below, a sample's RR value is critically affected by the two major features of the approach: the definition of the a priori list and the value given to names defined as "Other."

Table 2 presents the RR values for the cluster of names found in the Talpiot tombsite. We can see that Matya is assigned an RR of 1, while the ossuary #2 is discarded (with its two names, Yehuda and Yeshua, but the name Yeshua does appear in the table from ossuary #4).

The product of the individual RR-values yields $1.74 \times 10^{-8}$. Following the division by the prized bonus factor of 1.2, the RR-value for the cluster is $1.45 \times 10^{-8}$. Clusters with a similar configuration (i.e,. two female and three male ossuaries, where one male ossuary has two men in father–son generational alignment) and with a lower RR value are considered to be more "surprising" than the studied tombsite. Out of the $n_1$ and $n_2$ male and female persons in the population, the total possible number of such samples is $n_1^4 \cdot n_2^2$ and the total number of valid samples (which pass pre-specified "reality" requirements) is $\beta n_1^4 \cdot n_2^2$ with $\beta < 1$. In this case, $n_1 = 2509$, $n_2 = 317$ and Feuerverger found that $\beta = 0.906$, yielding $\beta n_1^4 \cdot n_2^2 = 1.981 \cdot 10^{12}$. Among them a proportion of $5.89 \times 10^{-7}$, or about $1/1{,}821{,}000$ have an RR value lower than $1.45 \times 10^{-8}$. The size of the estimated population who could have been interred in ossuaries includes about 4,400 males and 2,200 females. Dividing those values into the studied configuration of 4 male and 2 female inscriptions we obtain an estimate of 1,100 potential "trials" with which the Talpiot tombsite has to be compared.

The $p$-value for testing the alternative that among the comparable possible tombs one *unspecified* tomb is that of the NT family is assessed by the probability that at least one among the 1,100 would have an $H_0$-tail area less or equal to $5.89 \times 10^{-7}$. This probability is bounded above by $1/1{,}655$. For the Bayes-type computation of the posterior probability that this is indeed the NT family tombsite, Feuerverger defines by $\theta$ the (prior) probability that an NT family tomb would consist of a cluster of with an RR value as surprising as that at Talpiot. For $\theta = 1, 0.5$ and 0.1, the posterior probabilities are 0.9994, 0.9988 and 0.9940, respectively.

In a nutshell, the exposition above reviews the basics of the new proposed approach as applied to the specific data set.

**3. The a priori hypotheses.** As emphasized, the foundation for the analyses is a set of "hypotheses, assumptions and conditions upon which the computations

TABLE 2
*RR-values for the cluster of names in the Talpiot tombsite*

| | | | All sources | | Ossuary sources | RR |
|---|---|---|---|---|---|---|
| **Ossuary** | **Name on ossuary** | **Generic name** | **Relative frequency** | **Renditions** | **Relative frequency** | |
| | | | **Female** | | | |
| #1 | Mariamene | Mary/Mariam | 0.233 | Mariamene | 0.023 | $0.0053 = 0.23 \cdot 0.02$ |
| #6 | Marya | | | Marya | 0.295 | $0.0690 = 0.23 \cdot 0.30$ |
| | | | **Male** | | | |
| #2 | Yehuda | Yehuda | 0.068 | | | Discarded |
| #4 | Yeshua | Yeshua | 0.040 | | | 0.0403 |
| #3 | Matya | Matya/Mattityahu | 0.026 | | | 1 |
| #4 | Yoseph | Yoseph | 0.088 | | | 0.0881 |
| #5 | Yoseh | Yoseph | 0.088 | Yoseh | 0.152 | $0.0134 = 0.09 \cdot 0.15$ |

are carried out, collectively named as the provisos." The paper presents detailed explanations and justifications for the features in the provisos and performs sensitivity analysis under some variations of the provisos. We concentrate here on the basic provisos and the resulting conclusions. Undoubtedly, the posterior probabilities are impressive and seem to suggest that this is indeed the tombsite of the NT family.

However, are the provisos reasonable? And more importantly, were the provisos specified and were the analyses carried out according to the stated premises of the new approach? And if not, what is the likely effect of the deviations from those premises?

The a priori nature of the provisos is among the most important premises of the new approach. In this context, let us revisit first the issue of the female names contained in the presumably a priori list of candidates.

3.1. *The female names in the a priori list of candidates.* The list of potential candidates includes the names Mariam and Salome "commonly believed to be" Jesus' sisters, Marya (Jesus' mother), and Mary Magdalene. The addition of Mary Magdalene is explained by the fact that Mary Magdalene was "present at the burial ritual." The contention that Mary Magdalene's ossuary is presumed to be that inscribed as Mariamene [$\eta$] Mara is justified by stating that Mariamne is "the most specific appellation to Mary Magdalene from among those known." But it is difficult to avoid the feeling that in a truly a priori compiled list, the probability of adding persons whose relation was only that they were "present at the burial ritual" and had no familial relationship, were likely to be quite low. (The issue of possible

familial relationship is discussed, but the addition of the name is not based on it.) Moreover, the addition of the particular rendition of the name to the list gives a clear impression that after observing the data, the list was biased in favor of $H_1$.

Furthermore, since the particular rendition is in the relevant list, the inscription Mariamenou [$\eta$] Mara is now presented as being a unique rendition of Mariam both from ossuary as well as from nonossuary sources. The assigned RR value to that name is $1.68/317$, with the *largest* effect on the overall RR value. Clearly, if there is evidence that the elegantly rendered ossuary inscribed Mariamenou [$\eta$] Mara is indeed the ossuary of the Mary Magdalene, the finding is sensational by itself. But if we only use the statistical evidence, the fact that the effect on the overall result of the inscription Mariamenou [$\eta$] Mara (whose presence on the list is at least more ambiguous than the other names) is problematic, to say the least. Were Mariamenou [$\eta$] Mara treated as "Other," the overall RR value would have been 188 times higher, with the corresponding effect on the computed $p$-value.

The effect of the inscription Mariamene [$\eta$] Mara also illustrates a further significant deviation from the initial a priori definition of "surprise" relative to $H_1$. If the alternative $H_1$ is that this tombsite is that of the NT family, the "surprisingness" should indeed be assessed with respect to $H_1$ and not (only) with respect to the frequency table of the names. To illustrate this point consider a changed configuration of only the three male inscriptions, from (Yeshua son of Yoseph, Yoseh and Matya) to (Yoseh son of Matya, Jacob and Yoseph). Note that there is no Yeshua, and Yoseh is the son of an arbitrary Matya. Although a priori the changed configuration is by no means a serious candidate for being the NT family tombsite, under the suggested method the new configuration would have had a lower RR value than the actual one, that is, a higher "surprise."

3.2. "*Other*" *and disqualifying names.* Now let us address other features of the presumably a priori selected relevant lists. The relevant lists are allowed to include any number of names defined as "Other" as possibly belonging to persons about whom we have no records, with individual RR value of 1. Using this rule, the author computes the overall RR values as a product of the RR values of only four out of the six inscribed ossuaries (!). The ossuaries inscribed as Yehuda son of Yeshua (#2) and Matya (#3), although discussed at length, contribute nothing to the computation of the overall RR value. Following the rules set up by the suggested approach, this procedure is at least questionable. A set of rules which weigh positively (i.e., with a coefficient less than 1) names expected under $H_1$, but does not weigh negatively names which are unexpected under $H_1$, is likely to bias in favor of $H_1$.

Also, and continuing the previous point, it is mentioned that "... the list of persons (but not necessarily *names*) that would *disqualify* the tombsite as belonging to the NT family includes Joseph, Simon, and Yehuda" (as the persons' death did not occur in the relevant period of time, but the names may belong to other persons about whom we have no records). But if, say, an ossuary inscribed "Simon" would

have been found in that tombsite (say, instead of that of "Matya") how could we have known whether it belongs to "that" Simon (brother of Jesus) or not? According to the "surprisingness" approach, we would have ignored that inscription, as belonging to "Other" (as belonging to a person about whom we have no records) and set the relevant coefficient to 1. The calculated $p$-value would have been exactly as in the present case. How can one thus judge the relevance to $H_1$ and render judgment about disqualifying? The overall impression is that the inevitable exposure to the data affected the definition of the provisos in favor of $H_1$.

**4. Another analysis.** I mentioned above that the inclusion of MM in the relevant list has a substantial effect on the overall results and conclusions. We can get an idea of the order of magnitude of that effect by comparing the results presented in Feurverger's paper with those yielded by another Bayesian analysis performed on the same data by Kilty and Elliot (2007). They consider the name Mariamene [$\eta$] Mara as irrelevant, and treated it identically to the names on the ossuaries inscribed Yehuda son of Yeshua, and Matya. Their computation is based on a listing of 32 scenarios of combinations of names one might expect to find in a NT family tombsite, based on Jesus' brothers and mother. All the scenarios have to include the Yeshua son of Yoseph (in any rendition), and are assumed to be equally probable. The a posteriori probability that this is indeed the tombsite of the NT family given the data is estimated by Kilty and Elliot as 0.487, very different from the values of well above 0.994, deduced from the odds ratios mentioned in Feuerverger's article.

The comparison between Kilty and Elliot's results and the a posteriori probabilities computed by Feuerverger illustrates the effect of the inclusion of Mariamene [$\eta$] Mara in Feuerverger's list. Obviously, other analyses of this data set are possible and indeed some are presented in articles posted on the internet. I refer to Kilty and Elliot's article, since unlike others, they mention that they agree in principle with Feuerverger's conclusions and their intention in writing the article was to show that the cluster of name is "hardly what a person should expect to find randomly." They further state that their figure is "quite comparable to Feuerverger's conclusion even though the two are done from very different standpoints." The statement seems to be inaccurate, probably based on fragmentary information of Feuerverger's results.

**5. Some final remarks.** Feuerverger emphasizes the provisos for the calculations, and mentions that the conclusion and the measure of surprisingness are based on a particular—but not uncontested—set of assumptions. He mentions that "as long as the definition of surprise is specified fully and a priori, the resulting approximate "tail area" will essentially be valid." It is difficult to accept that in this case, the elements of the new approach which are mentioned in the paper that have to be a priori specified (the hypothesis for the problem, the measure of surprisingness, the list of possible candidates, and the lists of nested possible name rendition

for each candidate), have indeed been so specified. The final sentence in the paper candidly, and in my opinion very correctly, points to the weakest link in the foundation of the entire exposition and conclusions: "It is the presence in this burial cave of the ossuary of Mariamenou [$\eta$] Mara, and the mysteries concerning the identity of the woman known as Mary Magdalene, that hold the key for the degree to which statistical analysis will ultimately play a substantive role in determining whether or not the burial cave at East Talpiot happens to be that of the family of Jesus of Nazareth."

Let me re-phrase this sentence: "If the ossuary inscribed Mariamenou [$\eta$] Mara is indeed the ossuary of the Mary Magdalene from the New Testament, then, given the other names inscribed on the other ossuaries and the assumptions presented in the paper, we can state with a very high degree of confidence that that is the tombsite of the NT family."

I agree to such a statement. The only problem is that no statistical expertise is necessary to reach such a conclusion. If indeed, an ossuary *proven* to be that of Mary Magdalene was to be found, and in the same tombsite were also to be found ossuaries inscribed as Yeshua son of Yoseph, Yoseh and Marya, it is unlikely that the archeologists and the historians would appeal to statisticians for help. In such a case, as mentioned, the ossuary of the Mary Magdalene would have been by itself an important historical relic.

On the other hand, if we don't have that level of confidence regarding the Mary Magdalene ossuary, we have to rely on statistical analysis. Unfortunately, in my opinion, the stated principles of setting the assumptions were not followed, both in the presumably a priori compilation of the relevant lists as well as in the definition of the RR values (which allows discarding data which may point toward $H_0$ and assigns "surprisingness" values based the rareness of name frequencies rather than the actual closeness to $H_1$). The resulting effect on the conclusions reached is dramatic. Indeed, the narrator in the movie [Cameron (2007)] announced that Feuerverger's model concludes that "there is only one chance in 600 that the Talpiot tomb is not the Jesus family tomb, if Mary Magdalene can be linked to Mariamene." Later, in an interview on the *Scientific American* website [Mims (2007)], Feuerverger is quoted as saying that "I did permit the number one in 600 to be used in the film. I'm prepared to stand behind that but on the understanding that these numbers were calculated based on assumptions that *I was asked to use*," a statement far removed from the rigorous demand of a priori assumptions. [On his webpage, Feuerverger (2007) mentions that the quotations in the interview are "sufficiently accurate to be considered fair".]

In spite of the fact that, in my opinion, the analysis of the "surprisingness" based on the configuration of names failed to yield the stated conclusions, I refrain in this article from passing judgment on the subject matter issue of whether or not this is the tombsite of the NT family.

Furthermore, notwithstanding the reservations from the analyses applied to the discussed data, I applaud the bold initiative taken in the discussed paper to develop

a new approach to tackle a problem characterized by a degree of complexity that precludes the straightforward application of the classical hypothesis framework. The general problem of rending judgment on whether a multiple characteristics observation represents the pursued specific entity or it is just the result from random draws is interesting and intriguing. Cases of disputed paternity and DNA matching come to mind in this context. Unlike the Talpiot case, in those cases a standard for comparison is available. The new approach and concepts of "surprisingness," "relevance" and "rareness" may evolve and prove beneficial in cases in which there is no such standard exists.

Classical methods, usually based on Bayesian analysis are available for those cases, but their application may be difficult in complex situations. If the new approach is to be applied, its performance needs to be compared to existing methods in situations in which it is known whether the null hypothesis (or the analogous null hypothesis) is correct. I think that the features of the approach still need to be investigated theoretically or by simulations under various conditions of complexity. In any case, the assumptions have to be pre-specified to ensure valid results and a valid comparison.

## REFERENCES

JACOBOVICI, S. and PELLEGRINO, C. (2007). *The Jesus Family Tomb: The Discovery, the Investigation, and the Evidence That Could Change History*. HarperCollins, San Francisco.

CAMERON, J. (2007). *The Lost Tomb of Jesus*. Discovery Channel, March 4, 2007.

ILAN, T. (2002). *Lexicon of Jewish Names in Late Antiquity*, *Part 1*: *Palestine 330 BCE–200 CE*. Mohr Siebeck, Tubingen.

KILTY, K. T. and ELLIOT, M. (2007). Probability, statistics and the Talpiot tomb. Available at http://www.lccc.wy.edu/Index.aspx?page=547.

MIMS, C. (2007). Q&A with the statistician who calculated the odds that this tomb belonged to Jesus—An interview with the professor who estimated the probability that a particular tomb could have been the final resting place of a family other than that of Jesus. *Scientific American Website* http://sciam.com/article.cfm?articleID=13C42878-E7F2-99DF-3B6D16A9656A12FF.

FEUERVERGER, A. (2007). Dear statistical colleagues. Text page. Available at http://fisher.utstat.toronto.edu/andrey/OfficeHrs.txt.

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH
RAYMOND AND BEVERLEY SACKLER FACULTY OF EXACT SCIENCES
TEL AVIV UNIVERSITY
RAMAT AVIV, TEL AVIV
ISRAEL 69978
E-MAIL: fuchs@post.tau.ac.il

*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 66–73
DOI: 10.1214/08-AOAS99E
Main article DOI: 10.1214/08-AOAS99


# DISCUSSION OF: STATISTICAL ANALYSIS OF AN ARCHAEOLOGICAL FIND

BY DONALD L. BENTLEY

*Pomona College*

I begin this discussion by quoting a Mosaic law. This is not one that can be found in the Torah, but I know it to be authentic because I heard it from the mouth of Moses himself. The law is, "Statistics is the umpire of the sciences." This law was told me by Lincoln Moses, one of the top applied statisticians of the twentieth century, a real craftsman with data and a master of the application of statistics and statistical reasoning.

To appreciate this law, one needs to make the distinction between theoretical and applied statistics which I would like to illustrate with an example. A number of years ago, Joe Gani gave a talk to a group of statisticians. During the talk, which included a discussion of Fisher's work on predicting the number of species of butterflies in Malaysia, he made an aside remark that perhaps the same method could be used to determine the number of words in a person's vocabulary. The problem, posed in a general mathematical context, resulted in the well-known paper by Efron and Thisted (1976) titled "Estimating the number of unknown species: How many words did Shakespeare know?" And while the authors used the vocabulary framework for the structure of their research, there was no specific application in mind. The intended audience was the statistics community, not a group of Shakespearean scholars. I consider this an example of theoretical statistics.

Later, Thisted happened upon an article reporting that a newly discovered poem could well have been written by Shakespeare. Thisted and Efron (1987) set about modifying their previous results so the assumptions and methodology met the requirements necessary for applying them to the question of whether there was reason to believe that this poem had not been written by Shakespeare. The statistics now served as a tool for the primary purpose, not an end in itself. This is applied statistics. I believe the Mosaic Law, "Statistics is the umpire of the sciences," is directed toward statisticians working on applied problems. Because the problem which motivated the paper under discussion is an application, it seems appropriate to consider how well the statistics served as umpire in the research.[1]

Received December 2007; revised January 2008.

[1]The results of the research were first released by Discovery Channel at a press conference held at the New York City library on February 26th and then to the general public on March 4th, 2007 in the broadcast of their documentary "The Lost Tomb of Jesus," produced by James Cameron and directed by Simcha Jacobovici.

I believe there are five phases of a research project in which an applied statistician should be intimately involved, and the properly prepared statistician will feel comfortable participating in each of them [Bentley (1996)]. They are: (1) determining the question, (2) designing the experiment, (3) gathering and validating the data, (4) analyzing the data, and (5) communicating the results. There is a need for statistical reasoning to be applied at each phase of the project, not just during the analysis of the data. The statistician should be involved from the very beginning, which is the forming of the question the research team will attempt to answer. It is the statistician's responsibility to make sure everyone on the team understands the question and is in agreement that it is the appropriate question for the project. Moreover, equally as important is that the statistician makes sure everyone is in agreement with the assumptions behind the question. This includes making sure these assumptions form a consistent set. There should be ample evidence that they can be reasonably accepted by the community of scholars in the field of the application. The statistician, though not an expert in the substantive field, should still feel comfortable with the arguments being used by the research team members to justify the assumptions. And the assumptions should not preclude being able to answer the agreed-upon question; they must be consistent with the question. Keep in mind that a good umpire is not required to be highly skilled in playing the game, but must have a good knowledge of the rules by which the game is to be played and have had enough exposure to the game to be able to detect when the rules are being violated.

The abstract of the paper under discussion states, "An approach is proposed for measuring the 'surprisingness' of the observed outcome relative to a 'hypothesis' that the tombsite belonged to the NT family." This is followed in Section 1, the Introduction and Summary, with the statement, "Since names such as Yehosef, Marya, Yeshua, etc., were not uncommon during the era in which such burials took place, the task of assessing whether or not these ossuaries might be those of the New Testament (NT) family is not straight forward. [...] One purpose of this article is to contribute toward such efforts by developing statistical methods for assessing evidence for and against a 'hypothesis' that this tomb belonged to the family of the historical Jesus."

The analysis addresses this question under a given set of assumptions. Therefore we can only accept the conclusion if we are willing to accept the assumptions. As umpires, we need to know exactly what these assumptions are and the consequences these assumptions impose upon any inferences that are drawn from the analysis. Several times in the paper, the author refers to a "historical viewpoint" and "historical assumptions." I was unable to find a definition of "historical," but based on the assumptions and arguments presented in the text, I have been led to believe that by historical viewpoint is meant a strictly literal translation of the Gospels with one exception. The final paragraph of Section 1 states, "We remark that in assessing the evidence in any way, it is essential to adopt a strictly *historical* viewpoint, and thus to set aside considerations that a NT tombsite cannot exist. In

fact, Jewish ritual observances prevalent at the time are entirely consistent with the possible existence of such a tomb." This wording is equivalent to assuming a New Testament family tomb might exist, that it is possible there could be a family tomb which might even contain the remains of Jesus of Nazareth.

However, this is not the assumption that was used in the analysis. Instead, the analysis is conditioned on the assumption that, with probability one, there did exist such a tomb. And even beyond that is the assumption that this Jesus family tomb was in the vicinity of Jerusalem with probability one. In Section 14, the author states, "We are in fact now in a position to carry out a particular *hypothesis test*: Here $H_0$ is the hypothesis that *all* 1,100 tombs in the vicinity of Jerusalem arose under random assignments of names, and $H_1$ is the hypothesis that *one unspecified one among these 1,100 tombs is that of the NT family.*" However, the calculations do not seem to allow for an *a priori* probability greater than zero that there does not exist a family tomb in Jerusalem.

This raises a question. Even if there were a tomb for the family of Jesus of Nazareth, why would it be in Jerusalem with probability one? Why would it not be in the Galilee around Nazareth or Capernaum which was the base of Jesus' ministry? James Tabor is acknowledged in the paper as being a New Testament expert and in particular for his involvement in formulating the "A Priori Hypotheses," the assumptions that were used in the analysis. In checking Tabor's new book, *The Jesus Dynasty*, for an answer to this question one finds a picture on page 239 of Tabor kneeling on the Tsaft grave which is located in the Galilee, near Capernaum [Tabor (2006)]. In the associated text, Tabor describes a rabbinic tradition which identifies the grave as the tomb of Jesus of Nazareth, and Tabor presents an argument, based on the Gospels, as to why Jesus' body might have been returned to that area for burial. From this discussion, it appears that in 2006 Tabor did not believe the assumption that Jesus was buried in a family tomb in Jerusalem with probability one. However, he was willing to accept that assumption in establishing the "A Priori Hypotheses" for this analysis.

Another assumption which is not stated explicitly in the paper is incorporated into the estimate of the numbers of ossuaries and tomb sites in the Jerusalem area. Section 8 begins, "We require estimates of the size of the relevant population of Jerusalem and of the number of ossuary burials that took place overall." Based upon an estimate of the number of *residents* in Jerusalem during the period of ossuary burials, the researchers came up with an estimate of about 6,600 inscribed ossuaries which, when assuming a configuration of four males and two females per tombsite as in the Talpiyot site, led to an estimate of 1,100 tombsites in the Jerusalem area which is the figure used in the analysis. But this estimate excludes those persons who were not residents of Jerusalem yet might have been buried there in family tombs. Jesus of Nazareth, for example, was not included, nor was Mary of Magdala, nor was Joseph from Arimathaea whom the author argues arranged for the tomb in which Jesus was first buried.

There are two important factors the author ignored in estimating the number of ossuary burials that took place in the Jerusalem vicinity. First, it was sacred tradition that Jews make a pilgrimage to the Temple in Jerusalem three times a year, at the three important Festivals. One of these was the Passover when Jesus was crucified. Based on extra-biblical sources, scholars estimate that the number of people in Jerusalem increased by between fourfold and tenfold during these Festival periods. In other words, the number of people in Jerusalem at the time of Jesus' death would have been between 120,000 and 300,000, not 30,000. In fact, E. P. Sanders (1992) estimates the number to have been between 300,000 and 500,000. Further, many Jews then, even as today, desired to be buried in Jerusalem. There would have been nothing to prevent a family from bringing the bones of a family member from afar to place in an ossuary in a family tomb in Jerusalem, just as the bones of Joseph were carried out of Egypt. These omissions cause the assumptions used in calculating the number of tombsites in Jerusalem to be inconsistent with other assumptions used in the statistical analysis. In particular, assumptions A.5 and A.7 of the paper are inconsistent.

Among the greatest sources of controversy arising from *The Lost Tomb of Jesus* documentary is the Mariamene inscription, and the ossuary's "relevance" to Mary of Magdala.[2] At issue is whether the inscription refers to one or two persons. Assumption A.7 of the paper states, "We assume that the full inscription Mariamenou [$\eta$] Mara refers to a single individual and represents the most appropriate specific appellation for Mary Magdalene amongst those known. . . ." The impact this assumption has on the results of the analysis is acknowledged by the author in the paper's final paragraph: "Among the various assumptions made, perhaps the one that most "drives" our analysis in the direction of 'significance' is the extraordinary inscription Mariamenou [$\eta$] Mara." In describing Ossuary #1, the author states, "Rahmani (1994), pages 14, 222, reads the inscription as follows: "The stroke between the $\upsilon$ of the first and the $\mu$ of the second name probably represents an $\eta$, standing here for the usual $\eta\ \kappa\alpha\iota\ldots$ used in the case of double names. . . " and he posits that the second name is a contracted form [not a contraction] of 'Martha' leading to the reading 'Mariamene [diminutive] who is also called Mara'." The author then adds that it is Rahmani's reading which was adopted in determining the relevance of the ossuary for the analysis with the justification that it "was accepted by Kloner (1996) and has been corroborated by others [without reference] in the field." It should be noted that Kloner is not an epigrapher but rather one of the first archaeologists into the tomb in 1980.

In determining how to evaluate the inscriptions of the Yoseh and Marya ossuaries the author rejected Rahmani's interpretation. Near the end of Section 2 the author states, "Rahmani surmised that the similarities between ossuaries #5

[2]See Bovon (2007) concerning his opinion of the inappropriateness of the appellation of Mariamenou for Mary Magdalene in the first century CE, as compared to the impression given of his expert opinion in the documentary. Also more generally, Shanks (2007).

[Yoseh] and #6 [Marya] and their inscriptions, both coming from the same tomb, may indicate that Yoseh and Marya were the parents of Yeshua and the grandparents of Yehuda." In the attached footnote 5 the author adds, "If this interpretation is correct, the tombsite cannot be that of the NT family. However Rahmani does not follow up with any explanation for the messy nature of the inscription of Ossuary #4." Rahmani's suggested interpretation is treated by the author in assumption A.4 which states, "We assume that the ossuary inscribed 'Yehuda son of Yeshua' can be explained and may be discarded in our analysis." The umpire in me has to ask a question. Why should we be willing to accept with probability one Rahmani's interpretation of the "Mariamenou" inscription which supports the desired conclusion of the researchers although Rahmani only claims it is "*probable*," yet be willing to reject with probability one an interpretation which Rahmani states "*may be indicated*" when the interpretation would invalidate their theory? It should be pointed out that these decisions were made a posteriori by those who formulated the eight "A Priori Hypotheses" and the nine "Assumptions," after they had seen the data.

The other side in this debate claims that the Mariamenou inscription should be interpreted to refer to two women buried at different times, one with the name Mariame and the other the name Mara. It should be noted that it was quite common to have multiple burials in a single ossuary. Stephen Pfann (2007) provides very convincing evidence for this position. He notes that the "mark" between the first and second names, which the author accepted as representing $\eta\ \kappa\alpha\iota$, is in fact just a scratch made with a different tool than the tools used for the first and second names, and in fact the two names were carved with different tools. Further, Pfann suggests the reader compare the handwriting between the letters which form Mariame and the letters which form Mara. The first letter in each group is a capital $mu(M)$, the second an $alpha(\alpha)$, and the third a $rho(\rho)$. He points out that the first and second letters of the first name are each printed with two strokes and the third, the *rho*, is one continuous stroke. On the other hand, the first two letters of the second name are each made with a single continuous stroke, while the *rho* is formed with two clearly distinct strokes. I am not an expert in ancient Greek writing, but as an umpire I would need some convincing evidence before I would be willing to accept the assumption that the first and last words in the inscription were done by the same person and at the same time. And if they were not done at the same time, most likely the Mara inscription would be the name of a second person buried in the ossuary at a later date.

Jurgen Zangenberg (2007), a biblical archaeologist at Leiden University, addressed the use of the assumption that the Mara on the Mariamene inscription should cause it to be read as Mariamenou the Master with a rhetorical question, asking why one should assume that "Mara" here must be an honorary title unless one wishes to "prove" that which is already assumed to be known. What he is saying is that it is possible to prove anything that you assume to be true. It is the job

of the statistician, the umpire, to make sure the assumptions do not provide for this type of circular logic.

The examples I have given deal only with the first three of the five points where an applied statistician should have contributed to this research project: the point of determining the question the research was to answer and then making sure the assumptions used in the analysis were valid and also were consistent with the question, designing the experiment to gather appropriate data, and then collecting and validating the data. The concerns I raised above about the assumptions are just a few from many that I had as I read through the paper. The author, in numerous places throughout the paper, makes comments such as the following which appears in Section 1, the Introduction and Summary. "Our computations were carried out under a specific set of assumptions. . . ." And in the first paragraph of Section 13, A Statistical Analysis, we find the statement, "The assumptions A.1–A.9 under which we carried out our analysis are by no means universally agreed upon. Furthermore, the failure of any one of them can be expected to impact significantly upon the results of the analysis." There is a footnote to this statement which warrants particular attention. The footnote states, "These assumptions were proposed by S. Jacobovici, except for A.6 & A.9 which are due to the author."

Simcha Jacobovici is coauthor of the book, *The Jesus Family Tomb*[3] [Jacobovici and Pellegrino (2007)], as well as executive director of the Discovery Channel's documentary, *The Lost Tomb of Jesus*. At the press conference held on February 26, 2007 at the New York public library, Jane Ruth, the president and general manager of the Discovery Channel, referred to the subject of the documentary as, "what might be one of the most important archaeological finds in human history." After introductory remarks by Ms. Ruth and the documentary's producer, James Cameron, the podium was turned over to Jacobovici to provide the facts about the find. Following his presentation and prior to allowing questions from the media that were directed to the panel of experts, Jacobovici made the following statement. "Before I turn it over to the experts, because I have to say again, I'm not going to say I'm not an expert. I've seen a lot of internet buzz on this. I am an expert. My expertise is investigative journalism. I'm not an archaeologist. I'm not a DNA expert. I'm not a statistician. I'm a filmmaker and a journalist."

During the questioning by the media, the statistician who is the author of the paper under discussion had the opportunity to speak. He began, "The obvious needs

[3]It should be noted that on page 114 of *The Jesus Family Tomb* there is a footnote that reads, "As of this writing, Feuerverger's paper has been submitted to a leading American statistical journal and is being peer-reviewed." It must be emphasized that this peer review is being performed by statisticians who are not experts in archaeology, biblical history, or epigraphy. This review is therefore restricted to the statistical reasoning involved in the research, and should not be used as a reference for assuming the validity of the assumptions which served as the bases for the statistical analysis. The review of archaeological and biblical issues must be performed by the appropriate subject matter experts.

to be stated; that I'm not a biblical scholar, not a historical scholar. I'm just a numbers guy.... As a statistician, I do the calculations based on assumptions given to me *by the subject matter experts*, *in this case the historical biblical scholars.*"

The above quotes from the press conference raise two concerns. First, the responsibility of the statistician as an umpire of the science cannot be fulfilled by just accepting a given set of assumptions. They must be checked to make sure they meet certain standards as identified at the beginning of this discussion. But beyond that, the author did not even satisfy his own standards of using assumptions provided by "subject matter experts." By Jacobovici's own admission, his areas of expertise are filmmaking and journalism, and not the substantive fields of the complex history and archaeology of first century Judaism. Yet, as stated in both footnote 32 and the acknowledgments of the paper, it was Jacobovici who provided the assumptions which served as the basis for the statistical analysis.

Unfortunately, the lack of acceptance by the archaeological community, not to mention biblical scholars, of many of the assumptions used in this analysis is being recognized as a problem attributable to the field of statistics. As an applied statistician who is attempting to introduce more statistical reasoning into biblical archaeology, perhaps the most distressing comments I have read concerning the *Jesus Family Tomb* project were made by Sandra Scham (2007) in an article titled "The 'Jesus Tomb' on TV" which appeared in *Archaeology*, a journal she edits which is a publication of the highly regarded Archaeological Institute of America. Noting the reception this project has received among archaeologists and biblical scholars, she wrote, "At one time archaeologists loved statistics, happily performing complex regression and cluster analyses on our data and spitting out conclusions from our computers that, likely, proved the conjectures we had begun with. In the last two decades, however, we have begun to question these facile validations of our common sense. The problem is with the data. The methods may be perfectly suited to a world in which a representative sample, normal distribution, or even an idea of what the population in question might be, is possible. Archaeological evidence is precisely the opposite. We do not, in point of fact, know any of these things. In the words of one former statistically enthralled antiquarian, 'Even when the odds were good, we knew the goods were odd.' "

In my role as an umpire of "The Statistical Analysis of an Archaeological Find," I find myself in agreement with Ms. Scham.

## REFERENCES

BENTLEY, D. (1996). My first day's lectures: Past and present. In *Education in a Research University* (K. Arrow, R. Cottle, B. C. Eaves and I. Olkin, eds.) 217–229. Stanford Univ. Press, Stanford, CA.

BOVON, F. (2007). The tomb of Jesus. *SBL Forum*. Available at http://www.sbl-site.org/Article.aspx?ArticleID-656.

EFRON, B. and THISTED, R. (1976). Estimating the number of unknown species: How many words did Shakespeare know? *Biometrika* **63** 435–447.

JACOBOVICI, S. and PELLEGRINO, C. (2007). *The Jesus Family Tomb: The Discovery, the Investigation, and the Evidence That Could Change History.* HarperCollins, San Francisco, CA.

KLONER, A. (1996). A tomb with inscribed ossuaries in east Talpiyot, Jerusalem. *Atiquot* **XXIX** 15–22.

PFANN, S. (2007). Mary Magdalene has left the room: A suggested reading of ossuary CJO 701. *Near Eastern Archaeology* **69** 130–131.

RAHMANI, L. Y. (1994). *A Catalogue of Jewish Ossuaries in the Collections of the State of Israel*. Israel Antiquities Authority, Israel Academy of Sciences and Humanities, Jerusalem.

SANDERS, E. P. (1992). *Judaism: Practice and Belief 63 BCE–66 CE.* Trinity Press, Philadelphia.

SCHAM, S. (2007). The "Jesus Tomb" on TV. *Archaeology* 3 Mar. 2007. Available at http://www.archaeology.org/online/reviews/jesustomb2.html.

SHANKS, H. (2007). "The Tomb of Jesus"—My Take. *Biblical Archaeology Review* **33** 4, 72.

TABOR, J. (2006). *The Jesus Dynasty: The Hidden History of Jesus, His Royal Family, and the Birth of Christianity*. Simon and Schuster, New York.

THISTED, R. and EFRON, B. (1987). Did Shakespeare write a newly discovered poem? *Biometrika* **74** 445–455. MR0909350

ZANGENBERG, J. (2007). Yeshua aus Talpiyot und Jesus von Nazareth. *Familie Welt und Umwelt der Bibel* **44** 2–7.

DEPARTMENT OF MATHEMATICS
POMONA COLLEGE
CLAREMONT, CALIFORNIA 91711
USA
E-MAIL: dbentley@pomona.edu

*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 74–76
DOI: 10.1214/08-AOAS99H
Main article DOI: 10.1214/08-AOAS99


# DISCUSSION OF: STATISTICAL ANALYSIS OF AN ARCHEOLOGICAL FIND—SKEPTICAL COUNTING CHALLENGES TO AN ARCHAEOLOGICAL FIND

BY SHEILA M. BIRD

*MRC Biostatistics Unit*

The New Testament (NT) tomb in East Talpiyot, Jerusalem was discovered around Easter in 1980. Its surveyors at the time included Amos Kloner, whose 1980 PhD thesis was entitled "Tombs and Burials in the Second Temple Period," a topic on which he continued to publish for at least the next 15–20 years. Why did such a scholar not seize avidly the apparent historical opportunity that fell to his lot?

The tomb's excavator, Yosef Gath of the Department of Antiquities and Museums, died (date not specified) of heart failure not long after completing his work at the site. Upon completion of salvage excavations, "such bone material as remained was reburied" in accordance with Jewish ritual law. How much bone material remained? I assume that the orthodox rabbinate properly records reburials? Coincidentally, the NT tomb was discovered just as Sir Alec Jeffreys (1978–84, in Leicester, UK) was discovering DNA fingerprinting [see http://genome.wellcome.ac.uk/doc_wtd020877.html and Jeffreys, Wilson and Thein (1985)]. Some DNA analysis has been essayed, which Feuerverger side-steps. Shimon Gibson's archaeological drawings at the time of excavation indicated 10 ossuaries.

Ossuaries from the NT tomb were taken into the State of Israel Collections, but not until 1996 was it realized that records of the Israel Antiquities Authority (IAA) show only nine as having been received by it. Counting them all out and counting them all in, as famously reported by a UK journalist in the Falklands War, was inexplicably lax.

According to a 1994-published interpretation by authority Rahmani, and endorsed in 1996 by Kloner, six were found to have such Hebrew inscriptions as "Marya," "Yoseh," "Yeshua son of Yehosef," "Yehuda son of Yeshua," "Matya"... or Greek inscription of "Marmamene [diminutive] who is also called Mara." Attributions of authority are notoriously fickle: Rahmani had also interpreted Mary and Joseph as the parents of Yeshua and grandparents of Yehuda. Feuerverger argues that, if Rahmani is correct in this interpretation, then the tomb-site cannot be that of the NT family. The heretical alternative (which ancient religious authorities may have disavowed, or been unaware of) of Yeshua's having had a son by Mara is not admitted as a scientific (prior) consideration.

Received January 2008.

Rahmani's interpretation of the ossuaries' inscriptions is clearly a valid reason for the NT tomb's having not roused in the 1980s such titanic excitement as has since been engendered (http://www.theherald.co.uk/features/features/display.var.1226604.0.0.php).

As a practical statistician, my first set of sceptical questions therefore relates to the exact chronology of the tomb's discovery and excavation, the reburial of bone material (and its subsequent retrieval for DNA analysis), the registration(s) of ossuaries and deciphering of inscriptions, and the time-trail of interpretations of those inscriptions versus the publication of said interpretations.

Let me illustrate chronology by a controversy in the UK press in early January 2008 (see http://media.newscientist.com/data/pdf/press/2637/263711.pdf and http://www.guardian.co.uk/science/2008/jan/03/medicalresearch.agriculture) which surrounds the publication in December 2007 of a case-study that was submitted to Archives in Neurology [Mead et al. (2007)], an American journal, in February 2006. It concerns a 39-year old woman who died in 2000, 14 months after clinical onset of disease that was ascribed to sporadic CJD (despite atypical findings at post-mortem). Of particular note were: (a) that she was valine homozygous at codon 129 of the prion protein, and (b) that molecular analysis of cerebellar tissue demonstrated a novel $PrP^{Sc}$ type similar to that seen in vCJD. The authors reported that transmission studies were underway. This lady, were she the first clinical case of vCJD in a patient who is *not* methionine homozygous at codon 129 of the prion protein, would be as important as a first as was human-to-human, blood-borne transmission of vCJD, which merited parliamentary announcement in UK. Mysterious, therefore, were the up-to-seven-year delay in publication, failure to cite when transmission studies in mice had begun, and the authors' apparent caution that this was, in fact, *not* vCJD. Only a limited post-mortem had been permitted so that lymphoid tissue, such as from spleen and appendix, were not available for testing. The patient had a tonsillectomy but at a date and hospital unspecified; and some of the molecular techniques used were relatively recent. Transmission studies had been underway for some time so that preliminary results from them may indeed have underpinned the authors' caution. I recount this cautionary tale for two reasons: first, to illustrate that statisticians may need a hinterland of subject-matter knowledge to identify the critical questions to ask before proceeding to inference ... and, secondly, because it would be epidemiologically shocking if, for seven years, UK had overlooked vCJD in a clinical case who was valine–valine and, accordingly, the time-trail might point to pathological or molecular lacunae that needed to be plugged in UK's, European and world-wide CJD surveillance.

Let me end with the other conundrum: the missing or stolen ossuary from the NT tomb—an archaeological, if not criminal, travesty. Was an ossuary inscribed "James son of Joseph brother of Jesus" and in the possession of a private Israeli antiquities collector under prosecution for alleged forgery of *part* of said inscription from the NT tomb? Feuerverger notes that, due to the Sabbath, the NT tomb

was left open from Friday afternoon to Sunday morning in the four-day period of 28–31 March 1980. He speculates that investigating archaeologists were unlikely to have missed a seventh inscription (even prior to their having been "cleaned up") on the 10 ossuaries they'd located. Thus, if the "James" ossuary indeed came from the NT tomb, it would have to have been an 11th that the investigating archaeologists had somehow overlooked. That conveniently leaves the "missing" 10th ossuary as uninscribed. This line of argument is flimsy, but so too is it extraordinary to me that such antiquities were: (a) left open, (b) inaccurately curated, and (c) long under-rated as potentially newsworthy. . . unless scholars had indeed posed critical questions, and deployed DNA or other scientific techniques, that have unveiled more context than the problem posited, somewhat mysteriously, to investigator Feuerverger to cast statistical light on. Know thine enemy (bias).

## REFERENCES

JEFFREYS, A. J., WILSON, V. and THEIN, S. L. (1985). Individual-specific "fingerprints" of human DNA. *Nature* **316** 76–79.

MEAD, S., JOINER, S., DESBRUSLAIS, M., BECK, J. A., O'DONOGHUE, M., LANTOS, P., WADSWORTH, J. D. F. and COLLINGE, J. (2007). Creutzfeldt–Jakob disease, prion protein gene codon 129VV, and a novel $PrP^{Sc}$ type in a young British woman. *Archives of Neurology* **64** 1780–1784.

MRC BIOSTATISTICS UNIT
INSTITUTE OF PUBLIC HEALTH
UNIVERSITY FORVIE SITE
ROBINSON WAY
CAMBRIDGE CB2 2SR
UNITED KINGDOM
E-MAIL: sheila.bird@mrc-bsu.cam.ac.uk

*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 77–83
DOI: 10.1214/08-AOAS99C
Main article DOI: 10.1214/08-AOAS99


# DISCUSSION OF: STATISTICAL ANALYSIS OF AN ARCHEOLOGICAL FIND

BY HOLGER HÖFLING[1] AND LARRY WASSERMAN[2]

*Stanford University and Carnegie Mellon University*

*There are no small coincidences and big coincidences! There are only coincidences!*

From "The Statue" episode of *Seinfeld*.

**1. Introduction.** Andrey Feuerverger has undertaken a serious challenge. The subject matter is controversial and finding a sensible way to formulate the problem in a rigorous statistical manner is difficult.

The paper is notable for its thoroughness. We have rarely seen a paper on an applied problem that provides so much background material. Most importantly, the author is very careful to document all his assumptions and to remind the reader that the conclusion is sensitive to these assumptions. He resists the temptation to present his results in a sensationalistic way. Rather, he conveys his analysis in a dispassionate, understated tone. Nonetheless, he could still end up on *Oprah*.

We are trying to assess the probability of a hypothesis when the hypothesis is formed after seeing the data. This is a notoriously difficult problem. As Feuerverger notes, coincidences are common. But just how common?

One response—the nihilistic approach—is to say that it is impossible and stop there. We have much sympathy with the nihilists in a problem like this. Perhaps the scientifically honorable path is to say that any answer is misleading so it is better to provide no answer. But ultimately this is unsatisfying and we accept the author's approach to provide an analysis with many caveats.

The question may be framed formally as follows. We observe an outcome $x$—a tomb with interesting names—and we want to know: is this outcome surprising? One way to quantify surprisingness is to perform the following steps:

1. Construct a sample space $\mathcal{X}$ that contains $x$.
2. Identify all the outcomes $A$ that would have been considered surprising if they had been observed.
3. Construct an appropriate null distribution $P_0$.
4. Compute the $p$-value $p = P_0(A)$.

Received September 2007.

[1]Supported by an Albion Walter Hewlett Stanford Graduate Fellowship.

[2]Supported by NSF Grant CCF-06-25879. Thanks to Rob Tibshirani and Isa Verdinelli for helpful comments.

The most difficult step is identifying the set $A$ of interesting outcomes. It is explicitly counterfactual to ask if an outcome would have been surprising if it had occurred, knowing that it did not occur.

**2. Feuerverger's approach.** What the author has proposed is both interesting and reasonable. Numerous judgement calls have to be made but they have been carefully documented. Our summary of Feuerverger's method is this: The sample space is chosen to be sets of names on ossuaries, subject to some restrictions. The null measure is essentially random sampling from an onomasticon. The author defines a statistic (RR) that maps sets of names into products of numbers. These numbers are essentially sample proportions, modified to take into account various nuances such as surprisingness of versions of names. The result is a very small $p$-value suggesting that the find is indeed surprising.

The 'Mariamenou $\eta$ Mara' inscription has a very big effect on Feuerverger's RR statistics. An explanation for this is that the RR statistic becomes more significant if broad name categories are being subdivided into special name renditions, even if the particular name renditions are not relevant. The following example illustrates this point:

> A population has three names $A$, $B$ and $C$ each with frequency $1/3$. $A$ has 2 name renditions $A_1$ ($1/3$ of $A$) and $A_2$ ($2/3$ of $A$). Our family has two members named $A$ and $B$, and $A_1$ and $A_2$ are both relevant. The uncovered tomb has one inscription $A_1$. When only considering broad name categories, we have $RR(A) = 1/3$, $RR(B) = 1/3$ and $RR(C) = 0$. When the null is random drawing from the population, the $p$-value is then $2/3$.
>
> When taking name renditions into account, $RR(A_1) = 1/9$, $RR(A_2) = 2/9$, $RR(B) = 1/3$ and $RR(C) = 0$ giving $p$-value of $1/9$. The $p$-value decreased although both name renditions were considered relevant. The change in $p$-value can be even more substantial in more complicated cases.

In this comment, we present a Frequentist and a Bayesian approach that do not have this problem and yield quite different results.

**3. A different approach.** We would like to consider a different way of defining the basic event $A$. Our approach is more expansive and, as a result, more conservative. Instead of asking "What is the probability of getting this set of names?" we ask "What is the probability of getting some interesting set of names if one looks at several tombs?"

Let $\mathcal{X}$ be all name sets. Examples of sample points in $\mathcal{X}$ are

$$x = \{\text{Salome}\},$$

$$x = \{\text{Levi, Hanan, Simon, Mariam}\},$$

$$x = \{\text{Joseph, Jesus, Sarah}\},$$

and so on. Define a list of target names $S$. The list should include all names that will spark interest. We take this to be either the big set

$$S = \{\text{Mariam, Mary, Salome, James, Joseph, Joanna, Martha}\}$$

or the small set

$$S = \{\text{Mariam, Mary, Salome, James, Joseph}\}.$$

The name "Jesus" is not included because we will treat it separately. We assume that a tomb would have triggered interest if its name set $B$ has sufficient overlap with $S$. We lump together different version of names since interested observers would surely argue that a tomb is interesting if there is any way at all of matching the found names to potentially interesting names. Denote the name sets in the tombs by $B_1, \ldots, B_N$. Say that $B_i$ is interesting if

$$|B_i \cap S| \geq 3 \quad \text{and} \quad \text{"Jesus"} \in B_i.$$

We denote the probability of this event by $\pi_i$. Assuming independence of name assignments in and across tombs, the $p$-value is

$$p = 1 - \prod_{i=1}^{N} (1 - q(n_i, \pi_i))$$

where $n_i$ is the number of ossuaries in tomb $B_i$,

$$q(n_i, \pi) = p_J \mathbb{P}(Y_i \geq 3), \qquad Y_i \sim \text{Binomial}(n_i - 1, \nu),$$

$\nu$ is the probability that a single name drawn at random is in $S$ and $\pi_J$ is the probability of drawing the name "Jesus." We do not take $\pi_J$ to be the probability of drawing "Jesus son of Joseph" because the tomb could have been considered interesting if it had only said "Jesus."

For our calculations we take $N = 100$, $n_i = 6$. The number 100 comes from the fact that there are about 1000 tombs but only 10 percent have been excavated. Hence $\pi_i = \pi$ does not vary with $i$. We consider two possibilities for the male-femail ratio: (i) equal or (ii) unequal as represented by the onomasticon. For example, in case $S$ is equal to the first (big) choice, the male/female ratio is equal we get

$$\nu = \frac{1}{2}\left(\frac{231 + 103 + 45}{2509}\right) + \frac{1}{2}\left(\frac{81 + 63 + 21 + 12}{317}\right) = 0.3547.$$

The value of $\pi$ and $p$ for the different combinations of assumptions is as follows:

| $S$ | m/f ratio | $\pi$ | $p$-value |
|---|---|---|---|
| big | equal | 0.005 | 0.393 |
| big | not equal | 0.002 | 0.183 |
| small | equal | 0.003 | 0.290 |
| small | not equal | 0.002 | 0.158 |

We reiterate that we have not treated name variations as special. But the calculation is invariant under splitting names into subcategories since we are finding the probability of a set of interesting names, not a particular name. We also ignored family structure. We now consider two variations. We consider replacing "Jesus" with "Jesus son of Joseph" by multiplying these two probabilities. We also consider taking $N = 1000$ to reflect the unobserved tombs. The results are:

| | $N = 100$ | $N = 1000$ |
|---|---|---|
| Jesus | 0.16 | 0.82 |
| Jesus son of Joseph | 0.01 | 0.13 |

There is one case where the $p$-value is small. But the lack of robustness of this result does not make us confident in reporting a small $p$-value.

We conclude that the observed event is not rare at all. The chance that an observer would find a tomb that could be said to contain interesting target names is large. This is due to the fact that the interesting names are common and that the many tombs provide many opportunities for apparent surprises.

**4. Bayesian analysis.** Now we consider a Bayesian analysis of the problem. We need to compute

$$P(\theta = 1|x) = \frac{P(x|\theta = 1)P(\theta = 1)}{P(x|\theta = 1)P(\theta = 1) + P(x|\theta = 0)P(\theta = 0)},$$

where $x$ denotes the data, $\theta = 1$ that the tomb is from the NT family and $\theta = 0$ that the tomb is from the normal population.

In the frequentist approach, a partial ordering has to be defined on the space of all outcomes. Feuerverger does this using the RR statistic and the approach described above uses intersection of name sets. However, discerning the exact ordering on the space of outcomes may be hard or people might not agree with it. The advantage of the Bayesian approach is that the alternative distribution only has to be defined at the point $x$ and no ordering on the space of possible outcomes is needed.

4.1. *Posterior probability.* Let us introduce a little more notation at this point. Let $c$ be the configuration of a tomb, $g$ be its genealogy, $n = (n_1, \ldots, n_K)$ the broad name categories and $r = (r_1, \ldots, r_K)$ the particular name renditions. Assuming that every name rendition only depends on $\theta$ and its broad name category, we can write

$$P(x|\theta) = P(c, g|\theta)P(n|g, c, \theta)\prod_{i=1}^{K} P(r_i|n_i, \theta).$$

**Simplifying assumptions:** To make the computations easier, we make two more assumptions:

1. The configuration and genealogy we expect to see in the NT family tomb is not different from the rest of the population, that is, $P(c, g|\theta = NT) = P(c, g|\theta = P)$.
2. The particular name renditions we expect to see in the NT family tomb are no different than what we expect to see in the rest of the population, that is, $P(r_i|n_i, \theta = NT) = P(r_i|n_i, \theta = P)$. This assumption will be relaxed later.

Then the posterior odds are

$$\frac{P(\theta = 1|x)}{P(\theta = 0|x)} = \frac{P(\theta = 1)}{P(\theta = 0)} \cdot \frac{P(n|c, g, \theta = 1)}{P(n|c, g, \theta = 0)}.$$

4.2. *Distributions.* First, we define the prior distribution. Feuerverger estimates the number of tombs in the area to be about $N = 1100$. Also, let the prior probability of the NT family having a tomb at all be $t$. Then

$$P(\theta = 1) = t\frac{1}{N}, \qquad P(\theta = 0) = 1 - P(\theta = NT).$$

In order to be optimistic, we take $t = 1$ and get prior odds of

$$\frac{P(\theta = 1)}{P(\theta = 0)} = \frac{1}{1099}.$$

This prior can be thought of as a Bayesian approach to account for data snooping, that is, the potential to searching through many tombs.

For the null distribution, names are drawn randomly using the name frequencies in Ilan. Men and women are being treated separately and the list of names $n$ is treated as unordered.

When specifying the probability distribution under the alternative, it is necessary to weigh flexibility against complexity. Here we want to take the following approach: Specify a set of names from the NT family (separately for men and women) and assign each name a weight as to how likely it is to find this person in the NT family tomb. Then, the probability of a specific tomb is calculated by drawing from the nameset without replacement according to the weights. The weights can be determined in an optimistic or more conservative fashion (see Table 1).

For simplicity, the probability of being in the generational ossuary is taken to be the same for everyone, under the null as well as the alternative.[1]

**Neutral scenario:** In this case, we chose the weights in a fashion that seemed reasonable to us when we do not consider the information gathered from the tomb. Also, each name in the tomb is taken as its broad name category and it is assumed that no additional information for special name renditions is available for the NT family.

[1]This may be viewed as an oversimplification, however as the weights provide ample opportunity to fine tune prior beliefs, we do not see this as practically important.

TABLE 1
*Weights for each of the persons listed*

| Scenario | Jesus son of Joseph | James | Joses | Matthew | Judas | Others |
|---|---|---|---|---|---|---|
| Neutral | 20 | 3 | 3 | 62/2509 | 171/2509 | 3 |
| Optimistic | $\infty$ | 1 | 1 | 62/2509 | 171/2509 | 0 |

| Scenario | Marya (mother) | Mariam (sister) | Salome (sister) | Mary Magdalene | Others |
|---|---|---|---|---|---|
| Neutral | 10 | 3 | 3 | 3 | 3 |
| Optimistic | $\infty$ | 1 | 1 | 0 | 0 |

Drawing from the set is being done with probabilities proportional to the weights without replacement. The weight in the "others" category is the weight for all not listed persons.

**Neutral with special renditions:** Here, we use the same weights as in the neutral scenario, however account for the special "Mariamenou $\eta$ Mara" rendition. Each of the other inscriptions on the ossuaries is not special, so we do not make any adjustments for those. A priori, we could not have known the inscription "Mariamenou $\eta$ Mara," so how do we account for it? Under $\theta = P$, we assume that for the Marya name category, the probability of seeing a new previously unseen name is $1/80$. For $\theta = NT$, we assume that special name renditions are more likely, say $1/10$. Assuming that 'Mariamenou $\eta$ Mara' could in some way be interpreted for Maria (mother), Mariam (sister) and Maria Magdalena, this raises the odds by a factor 8 over the neutral scenario.

**Optimistic scenario:** We also wanted to explore the effect of having very optimistic assumptions which are to a large degree influenced by what has been observed in the tomb. Jesus and his mother are taken to be in the tomb for sure. For the rest of the men, the weights are equal for both brothers and set to the normal name frequency in the population for Matthew and Judas. The overall effect of this choice of weights is to effectively ignore the Matthew and Judas ossuaries, assume that one of the ossuaries is from a brother and one from a sister of Jesus and assign all eligible brothers and sisters the same weight.

4.3. *Results.* Even in the optimistic scenario, there is only about a 60% chance of the tomb belonging to the NT family. In the other two, more realistic schemes, the probability is only 22% and 3% (see Table 2). Just as Feuerverger, we also did not consider the generational part of the "Judas, son of Jesus" ossuary. Including it in the analysis would be possible; however, as prior beliefs about a possible son of Jesus are very strong, this may have overwhelmed the rest of the analysis and therefore we decided to exclude it.

TABLE 2
*Posterior probability that the Talpiyot tomb belongs to the NT family under various scenarios*

| **Scenario** | **Probability** |
|---|---|
| Neutral | 3.4% |
| Neutral—special renditions | 21.8% |
| Very optimistic | 64.1% |

**5. Conclusion.** When asked to analyze these data, we suspect that many statisticians would have said that the problem is too vague and would have stopped there. We commend Andrey Feurverger for plunging in and doing a serious analysis. Our analysis suggests that the finding does not lend support to the hypothesis that the find is indeed the tomb of the NT family. Ultimately, scholars of history and archeology will judge the validity of the claims about this find.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305
USA
E-MAIL: hhoeflin@stanford.edu

DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213
USA
E-MAIL: larry@stat.cmu.edu

*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 84–90
DOI: 10.1214/08-AOAS99G
Main article DOI: 10.1214/08-AOAS99


# DISCUSSION OF: STATISTICAL ANALYSIS OF AN ARCHAEOLOGICAL FIND

BY RANDALL INGERMANSON

*Ingermanson Communications, Inc.*

We critique the analysis by A. Feuerverger of an archaeological find that has been alleged by some to be the tomb of Jesus of Nazareth. We show that his analysis rests on six faulty assumptions that have been severely criticized by historians, archaeologists, and scholars in related disciplines. We summarize the results of an alternative computation using Bayes' theorem that estimates a probability of less than 2% that the Talpiot tomb belongs to Jesus of Nazareth.

**1. Introduction.** Andrey Feuerverger notes in his article that assumptions A.1 through A.9 are "not universally accepted." We argue that most historians and archaeologists actually disbelieve his key assumptions. (We agree with Feuerverger that the computational method he proposes can be extremely useful for difficult problems such as the Talpiot tomb.)

Assumption A.7 (the largest driver of his results) is almost universally rejected by scholars in the relevant fields. Several other assumptions are extremely dubious, and *each* of them biases the result toward $H_1$. Since all statistical biases in Feuerverger's RR values accumulate *multiplicatively*, the net effect is an enormous bias toward $H_1$.

In this article, we will look first at the most egregious problem, the "Mariamenou" inscription, which Simcha Jacobovici identified with Mary Magdalene through a long chain of reasoning that has been severely criticized by historians. In less detail, we will examine five other serious problems. By Feuerverger's own account, eliminating two of these statistical biases (the two relating to Mary Magdalene) is sufficient to destroy the statistical significance of $H_1$. But all six statistical biases should be eliminated from the baseline model of the problem.

We describe a series of calculations using Bayes' theorem that show that the probability that the tomb belongs to Jesus of Nazareth is at most about 2%, and may be much less.

**2. The primary problem: The "Mariamenou" inscription.** One of the ossuaries bears an inscription that is usually translated "Mariamenou [who is also called] Mara." Simcha Jacobovici (2007) took this to be a variant of "Mariamne"

Received November 2007; revised January 2008.
*Key words and phrases.* Jesus of Nazareth, Talpiot tomb, Bayes' theorem, statistical inference, distribution of names, onomasticon.

and interpreted it as a reference to Mary Magdalene. That is, he believed Mary Magdalene went by this name and that very few other women did. Jacobovici based his theory on the work of Dr. Francois Bovon. But Bovon (2007) immediately repudiated this interpretation of his work in a web article. The key point is this statement: "I do not believe that Mariamne is the real name of Mary of Magdalene."

Dr. Richard Bauckham (2007), a renowned expert in first-century Jewish names, has analyzed the "Mariamenou" inscription in detail in a guest blog article. His conclusions are:

(1) Grammatically, "Mariamenou" is the genitive case of the rare form "Mariamenon," a diminutive endearment deriving from the common name "Mariam."
(2) The name is not derived from "Mariamne."
(3) The name is very rare, and no other instance is found in antiquity.
(4) We have no evidence that Mary Magdalene ever went by this name.

One should ask what name Mary Magdalene went by, according to the data we have. Stephen Pfann (2007) has tabulated the references to Mary Magdalene in the various books of the New Testament, the earliest sources that mention her. She is called by the formal name "Mariam" four times and by the shorter, more intimate form "Maria" 10 times. These are the *only* names used in the New Testament to refer to Mary Magdalene.

With these facts at hand, we can answer the following question: Assuming that Mary Magdalene was actually buried in ancient Jerusalem, if one finds the inscription "Mariamenou" in that city, what is the probability that it might refer to Mary Magdalene? The answer is that the inscription is neither more nor less likely to refer to Mary Magdalene than to any other Mary of Jerusalem. (There were roughly 8500 other Marys.) This demolishes Jacobovici's theory, because "Mariamenou" simply can't be identified as "the real name" of Mary Magdalene.

In Feuerverger's article, he assigns an RR value to the Mariamenou inscription that carries an illicit factor of $(1/44)$, due to his belief that the inscription "represents the most appropriate specific appellation for Mary Magdalene from among those known." But it doesn't, and therefore this factor $(1/44)$ should be changed to 1.

This faulty assumption biases the entire calculation very strongly toward $H_1$ and is the primary driver behind the allegedly remarkable results.

**3. Five other significant problems.** In addition to the "Mariamenou" issue, there are a number of other problems in Feuerverger's work that bias the computation toward $H_1$. Each of them contributes a factor smaller than 1. The result of multiplying them all together is an enormous bias toward $H_1$. These problems are as follows:

(1) Assumption A.3 asserts that "the most appropriate rendition of the name of the mother is Marya." Note that "Marya" is the short form of the more formal name "Mariam" and is often spelled "Maria" in English. Assumption A.3 asserts that the mother of Jesus could *not* be listed as "Mariam" on her ossuary. With this assumption, Feuerverger inserts a factor of (13/44) into his RR value for the "Maria" inscription. The problem is that there is no compelling reason to believe A.3. The New Testament data compiled by Stephen Pfann (2007) shows that the mother of Jesus was called "Mariam" 13 times and "Maria" six times. So the data runs counter to Feuerverger's assumption. The mother could be called by either name. Feuerverger's factor of (13/44) is illicit and should be eliminated.

(2) Assumption A.3 likewise asserts that the short form "Yoseh" is the most appropriate rendition of the second brother of Jesus, whose formal name was "Yehosef" like his father. The New Testament refers to this brother once by the short form and once by the long form. A complicating factor here is that any randomly chosen "Yehosef son of Yehosef" would be quite likely to carry an alternative form of the name, so as to distinguish between father and son. Feuerverger inserts a factor of (7/46) into his RR value, which is too small, because it is at the minimum of the range of possible values. The correct value should lie somewhere between (7/46) and 1.

(3) An inscription "Judah son of Jesus" indicates that the Jesus buried in the tomb had a son. Jewish men of the time were very likely to be married and have children. But it is probable that Jesus had no sons. Recall that Jesus had four brothers who assumed positions of influence in the early Jesus movement. If a son also existed, he would likely have joined his uncles in a position of influence and we would have heard of him. Since we have not, we can conclude that the probability that Jesus had a son is *lower* than the probability for a randomly selected man of Jerusalem. Feuerverger's calculation fails to account for this. This inserts a bias into his computation.

(4) If the Talpiot tomb contained the *family* of Jesus of Nazareth, *would we expect Jesus to be in it*? Archaeologist Jodi Magness (2007) argued from a historical perspective that we should not. (But note James Tabor's rebuttal (2007), which argues that the tomb "should not be dismissed." We agree that it should not be dismissed, but it must stand on its merits.) Magness and Tabor at least agree that the data indicates that the body of Jesus went missing within days after the crucifixion. The earliest Jesus movement explained this by asserting that Jesus was resurrected, a claim outside the bounds of scientific investigation. If one looks for a naturalistic explanation, Magness says that much the likeliest one is that Jesus was reburied in a simple trench grave like other poor men of his time. She argues on several grounds that it is implausible that Jesus was buried in a rock-cut tomb like the one at Talpiot. Feuerverger's analysis fails to penalize $H_1$ on account of this issue, thereby introducing another source of statistical bias into his calculations.

(5) Would Mary Magdalene be buried in the family tomb of Jesus? According to Bauckham (2007), the usual practice was that only family members were buried

in a family tomb. It is *possible* that Mary Magdalene was a family member. It is even *possible* that she was married to Jesus. But we can have no certainty that she was. Most historians would estimate a probability *substantially* less than 1 for these possibilities. Feuerverger's analysis assumes that Mary Magdalene should be in the tomb and his computation achieves statistical significance *only if she is assumed to be in the tomb*. This introduces another very serious source of statistical bias into his computations.

**4. A calculation using Bayes' theorem.** It is beyond the scope of this short comment to give full details on a more correct calculation. This journal has given us space on its web site for a 29 page article that defines the statistical issues of the tomb and then describes a series of calculations we have performed. Here, we will merely summarize the results of that article [Ingermanson (2008)].

We define the two events $J$ and $T$ as follows:

$J =$ the "Jesus son of Joseph" in the Talpiot tomb refers to Jesus of Nazareth,

$T =$ the observation of the rest of the Talpiot tomb data.

We denote the negation of the event $J$ by the symbol $\sim J$.

We are interested in computing the conditional probability $P(J|T)$ using Bayes' theorem:

$$P(J|T) = \frac{P(T|J)P(J)}{P(T|J)P(J) + P(T|\sim J)P(\sim J)}.$$

Define the two ratios

$$\alpha \equiv \frac{P(\sim J)}{P(J)},$$

$$\beta \equiv \frac{P(T|\sim J)}{P(T|J)}.$$

Then our formula simplifies to

$$P(J|T) = \frac{1}{1+\alpha\beta}.$$

The results of many computations can be summarized as follows: $\alpha$ tends to be large, while $\beta$ is near 1. Therefore, $P(J|T)$ tends to be small.

We can estimate $\alpha$ very quickly. Feuerverger quotes the results of Camil Fuchs (2004) that the number of adult males who died in Jerusalem in the relevant time period was about 36420. This is overly precise, but it is reasonable in magnitude.

Assuming that 4% of men were named Jesus and 8.8% were named Joseph, we estimate the number of men named "Jesus son of Joseph" to be about 128. One of these men was Jesus of Nazareth. The other 127 are unknown to history.

Therefore, if we are given a randomly chosen man of Jerusalem named "Jesus son of Joseph," the probability that he is Jesus of Nazareth is $P(J) = 1/128$. The probability that he is not is $P(\sim J) = 127/128$. Taking the ratio, we estimate $\alpha \approx 127$. In general, if there were $N_J$ men of Jerusalem named "Jesus son of Joseph," then we have $\alpha = N_J - 1$.

The estimation of $\beta$ is much more complicated and we describe it in detail in the supplemental article [Ingermanson (2008)]. The general procedure is as follows:

We are comparing two hypotheses, $J$ and $\sim J$, using the data $T$ to distinguish between the two. For each of these two hypotheses, we imagine a statistical ensemble of tombs "similar" to the Talpiot tomb. We'll make random draws from each ensemble and tabulate the frequency of "hits" (random draws that agree with the data $T$).

We'll stipulate that each member of these two ensembles should contain an ossuary inscribed with "Jesus son of Joseph" and a second ossuary inscribed with "Judah son of Jesus." It should also contain two ossuaries bearing female names, two ossuaries bearing male names, and four uninscribed ossuaries. The distribution of names on the inscribed ossuaries must match the distribution of the names of persons living in Jerusalem in the first century, subject to the constraints of the two hypotheses.

In the case of the $\sim J$ hypothesis, there are no constraints.

In the case of the $J$ hypothesis, the only constraint is that the tomb must contain at least the names of certain members of the family of Jesus, with any remaining slots in the tomb filled with names chosen using the distribution of names in Jerusalem.

The procedure outlined above is similar in spirit to that followed by Feuerverger. Here are the primary differences in our calculations. We say that:

(1) The name of the mother of Jesus could have been inscribed as *any* form of Mary, including "Marya," "Mariam," or any other variant (even including the much-debated "Mariamenou Mara" inscription).

(2) "Judah son of Jesus" is considered less likely to appear in the tomb of Jesus of Nazareth than in the tomb of a randomly selected "Jesus son of Joseph."

(3) Jesus of Nazareth is considered less likely to be buried in a rock-cut tomb than was a randomly selected "Jesus son of Joseph."

(4) Mary Magdalene is not assumed to be in the tomb, and the "Mariamenou Mara" inscription is not assumed to be an appellation that applies to her with any higher probability than to any other Mary of Jerusalem.

(5) The probability of finding a Yoseh in the tomb is reckoned to be higher than usual, because the patriarch of the Talpiot family is named Joseph.

(6) The measure of "surprisingness" is the count of family members in the tomb, not Feuerverger's RR values. We use six different ways of defining this count.

The calculation was performed in Java using a wide variety of assumptions for the composition of a "Jesus family tomb" and using six different definitions of "surprisingness." Random draws were made in groups of 10,000, and results were tabulated.

The baseline calculation returned an estimate for the upper bound of $P(J|T)$ at about 2% (with a standard deviation of about 2%). A number of variants were tried, and the highest value found for $P(J|T)$ was 5.67%, using one assumption we consider unlikely. (The assumption that Yoseh should be exactly as rare in the Talpiot tomb as it is in tombs that do not have a patriarch named Joseph.)

We found that by tightening two assumptions, the upper bound could be substantially reduced. These are as follows.

We have assumed that the relative probability $\rho_{\text{son}}$ that Jesus had a son (as compared to other men of his time) was less than 1. That is, we defined a random variable $\rho_{\text{son}}$ uniformly distributed on the interval [0, 1]. Many historians would argue that this distribution should be strongly weighted toward zero. Doing so would strongly reduce our estimates of $P(J|T)$.

Likewise, we have assumed that the relative probability $\rho_{\text{tomb}}$ that Jesus was reburied in a rock-cut tomb (as compared to other men of his time) was also less than 1. We defined a random variable $\rho_{\text{tomb}}$ uniformly distributed on the interval [0, 1]. As noted earlier, Jodi Magness (2007) has argued strongly that $\rho_{\text{tomb}}$ should be heavily weighted toward zero. Doing so would again sharply reduce our estimates for $P(J|T)$.

We leave it to historians and archaeologists to debate such matters. We expect that their conclusions will tend to reduce our upper bound for $P(J|T)$ to be less than 2%, but it is impossible to predict how far it might drop. Such matters are irreducibly subjective.

**5. Conclusion.** Feuerverger's computation contains a number of statistical biases, each of which favors $H_1$. One of these (the "Mariamenou" inscription) introduces an illicit factor of $1/44$ to RR, which accounts for a very strong bias all by itself. But five other factors enter in with moderate statistical bias toward $H_1$, and the net effect is to create the appearance of statistical significance where none actually exists.

We have performed a series of calculations using Bayes' theorem that estimate a likely upper bound for the probability that the Talpiot tomb is the tomb of Jesus of Nazareth. This upper bound is about 2% with a standard deviation of about 2%.

**Acknowledgments.** I thank Jay Cost for reading this article and making comments. I also thank Richard Bauckham, Mark Goodacre, Gary Habermas, Michael Heiser, Stephen Pfann, and James Tabor for helpful discussions over the last several months. I take full credit for any errors.

## SUPPLEMENTARY MATERIAL

**Analysis of the Talpiot tomb using Bayes' Theorem and random variables** (doi: 10.1214/08-AOAS99GSUPP; .pdf). We analyze the Talpiot tomb, which has been alleged to be the family tomb of Jesus of Nazareth. Using Bayes' Theorem, we derive a simple function that estimates the probability that the tomb houses the remains of Jesus and his family. Unfortunately, this function cannot be evaluated exactly, because several of the key parameters are unknown. By using random variables with reasonable probability distributions, we examine the mean behavior and range of the function under a variety of conditions. We conclude that the probability is low (on the order of 2% or less) that the Talpiot tomb is the family tomb of Jesus of Nazareth.

## REFERENCES

BAUCKHAM, R. (2007). Guest article on Chris Tilling's blog on March 1, 2007. Available at http://www.christilling.de/blog/2007/03/guest-post-by-richard-bauckham.html.

BOVON, F. (2007). Article on the Society of Biblical Literature web site, posted March, 2007. Available at http://sbl-site.org/Article.aspx?ArticleId=656.

FUCHS, C. (2004). Democracy, literacy and names distribution in ancient Jerusalem—how many James/Jacob son of Joseph, brother of Jesus were there? *Polish J. Biblical Research* 1–30.

INGERMANSON, R. (2008). Supplement to "Discussion of: Statistical analysis of an archeological find." DOI: 10.1214/08-AOAS99GSUPP.

JACOBOVICI, S. and PELLEGRINO, C. (2007). *The Jesus Family Tomb*. HarperOne, SanFrancisco.

MAGNESS, J. (2007). Article on the Society of Biblical Literature web site, posted March, 2007. Available at http://sbl-site.org/Article.aspx?ArticleId=640.

PFANN, S. (2007). Available at http://www.uhl.ac/Lost_Tomb/HowDoYouSolveMaria/.

TABOR, J. (2007). Article on the Society of Biblical Literature web site, posted March, 2007. Available at http://sbl-site.org/Article.aspx?ArticleId=651.

2210 W. MAIN STREET
SUITE 107, BOX 103
BATTLE GROUND, WASHINGTON 98604
USA
E-MAIL: randy@rsingermanson.com
URL: www.Ingermanson.com

*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 91–96
DOI: 10.1214/08-AOAS99F
Main article DOI: 10.1214/08-AOAS99


# DISCUSSION OF: STATISTICAL ANALYSIS OF AN ARCHEOLOGICAL FIND

BY J. MORTERA AND P. VICARD

*Università Roma Tre*

**1. Introduction.** The paper by Feuerverger analyses interesting data on the inscriptions found on the ossuaries of a burial tomb unearthed in Jerusalem in 1980. A statistical analysis is made of the plausibility that the names inscribed on the ossuaries match those of the New Testament (NT) figures. The evidence on which the analysis is based is the distribution of names in the era when the tomb was dated. The results are based on assumptions which may drive some of the results.

Some questions immediately come to mind.

- The author assumes that a tomb of Jesus of Nazareth exists—this assumption is disputed by many people, as stated by Colin Aitken in the interview given on March 1, 2007 to *The Herald*. Moreover, even assuming the existence of a tomb of Jesus of Nazareth, why should it be located in Talpiyot and not, say, at the Sepulchre in Jerusalem or in another site or city?
- What is the uncertainty of the estimated number 1,100 of inscribed adult ossuaries? It would be important to measure the variability around that estimate.
- What implications does the statement that the Talpiyot finding is the "best of many trials" have on the results?
- Why was the DNA evidence available only for the ossuaries with the inscriptions "Yeshua son of Yhosef" and "Mariamenou e Mara?" Why was DNA not extracted from all the remains?
- Assumption A.7, which interprets the name on Ossuary #1 as being that of Mary Magdelene, is one factor that has a very strong influence on the results of the analysis since it is such a rare name. Is there no uncertainty in this interpretation?

Here we discuss further aspects of the paper and propose possible ways in which the statistical analysis could be extended.

The assumptions made by the author are based both on *anonymous* sources, such as the 4th century CE version of the Acts of Philip[1] and the NT gospels written between 65 and 100 CE. A possible way to handle the different reliability of these sources could have been that of assigning different weights to the assumptions based on historical sources and to those based on other sources, such as the apocryphal narratives.

Received January 2008; revised February 2008.

[1] Craig A. Blaising, "Philip, Apostle." In *The Encyclopedia of Early Christianity*, Everett Ferguson, ed. (New York: Garland Publishing, 1997).

Since a hypothesis such as the one investigated by the author could have an impact on the history of religion, it would be appropriate to examine other pieces of evidence. These could help explore the plausibility that the Talpiyot family configuration was so rare at that time that there could have been only one family with that configuration.

We will base our discussion on the following issues: in Section 2 we show how to deal with the uncertainty in name frequencies; comments on the DNA evidence are given in Section 3; the analysis of different items of evidence is given in Section 4 and Section 5 shows how an object-oriented Bayesian network (OOBN) can be structured for combining different items of evidence.

**2. Uncertain name frequencies.** In Section 5 the author gives details on the available documentation that could be used to obtain the distribution of names in the era relevant to the study. The name frequencies of three different sources are shown. Table 1 (from Table 2 in the paper) shows the relative frequency of Ilan's nonossuary and ossuary names. Category "Other" indicates all the other names having overall frequency $f_i = 1 - \sum_j f_j$.

The author tells us that "the relative frequency of female ossuaries (names) is under represented" since sometimes fathers (and occasionally husbands) were named on female ossuaries. Furthermore, the name distribution sources refer to a range in time period wider than that of the burial tomb in question. There is thus potential bias and many sources of uncertainty in the name frequency distributions. This should be appropriately accounted for, not by ad hoc adjustments, but in a fully probabilistic framework.

Thus, when analyzing the data, the name frequencies are not fixed probabilities, but empirical frequencies. These are most probably *not* a random sample from the

TABLE 1
*Frequency distribution of Jewish female names*

| **Names** | **Ilan nonossuary** | **Ilan ossuaries** |
|---|---|---|
| Mary | 0.242 | 0.228 |
| Salome | 0.161 | 0.212 |
| Shelamzon | 0.048 | 0.098 |
| Martha | 0.032 | 0.088 |
| Joanna | 0.040 | 0.036 |
| Shiphra | 0.024 | 0.047 |
| Berenice | 0.056 | 0.010 |
| Sara | 0.024 | 0.026 |
| Imma | 0.016 | 0.031 |
| Mara | 0.016 | 0.026 |
| Other | 0.339 | 0.197 |
| N. females | 317 | 193 |

population of names of the era. The uncertainty about these name frequencies can be modeled by assuming a Dirichlet prior and multinomial sampling. In Green and Mortera (2008) we show how to model uncertain frequency distributions in forensic inference in a fully probabilistic way in a Bayesian network [Cowell et al. (1999)]. Taking all uncertainties into account, in a probabilistically coherent way, would avoid those arbitrary adjustments (like multiplying by 5 or dividing by 1.2) that are made in computing the $RR$ values.

Furthermore, a very strong assumption made is that of considering independence among the names and then applying the product rule to obtain the overall $RR$ value. Also, the fact that brothers do not commonly have the same name is ignored. These dependencies as well as the fact that "in assignment of names within a family, children frequently are named as earlier 'nodes' in the family tree" can be taken into account in structuring a Bayesian network to analyze this problem.

Finally, *all* uncertainties, the name frequency distributions, the number of inscribed adult ossuaries and the relevant population size should be accounted for and modeled appropriately.

**3. DNA evidence.** The discriminatory power of DNA analysis in forensic identification is well known. Mitochondrial (mtDNA), Y-chromosome DNA and even nuclear DNA can be extracted from ancient human remains. This information is extremely important for reconstructing a probable family pedigree and establishing the sex of the owners of the bones. From this analysis one can compute the probability that the bones either belong to individuals of the same nuclear family, or to possible relatives of the family, or are from unrelated individuals. So, as stated before, why was the mtDNA of the bones found only in the ossuaries with the inscriptions "Yeshua son of Yhosef" and "Mariamenou e Mara" analyzed?

In the well-known Romanov case, mtDNA played a central role in the attempt to discover whether Anastasia, the daughter of the Tsar Nicholas II, was killed and buried with her parents [Gill et al. (1994)]. Nine skeletons unearthed in Ekaterinburg, Russia, in 1991, were tentatively identified as the remains of the last Tsar, his family and the Royal Physician and three servants. Sex testing and nuclear DNA were extracted from the bones in order to confirm that a family group was present in the grave. mtDNA (and Y-chromosome DNA) is transmitted unchanged—apart from the possibility of mutations—in the maternal (paternal) line. To verify the hypothesis that these remains were effectively from the Tsar, the Tsarina and their children, the DNA of their living descendants were analyzed, among which that of the Duke of Edinburgh. The DNA evidence supported the hypothesis that the remains were those of the Romanov family. From all the evidence—the DNA analysis, the statistical analysis and historical facts—the conclusion was reached that the nine skeletons were those of Tzar Nicolas II, the Tzarina, three of their four daughters, the court doctor and three servants. A complex statistical analysis was also made to obtain the most probable pedigree given the DNA evidence [Egeland et al. (2000)].

Although the Romanov remains are of much more recent origin than the bones found in the Jerusalem ossuaries, DNA can be extracted from ancient remains. In fact, both mtDNA and nuclear DNA has been extracted from fossils of a Neandertal man [Green et al. (2006)].

In contrast to the Romanov case, we do not have *known* descendants of the NT family. Therefore, the DNA analysis can only be used to verify the hypothesis about a specific pedigree. It can thus help to disconfirm the hypotheses that this is the NT family, but cannot be used to confirm that the hypothesis is true.

Furthermore, information on the dating and measurements taken from the ossuaries and the human remains, would be helpful to determine the age group, sex and estimated burial time of each remain.

**4. Analyzing many items of evidence.** There are many similarities in the analysis made in this paper to those commonly made in forensic identification, some of which we will illustrate here. Figure 1 shows a pictorial representation of a network for analyzing two different items of evidence pertaining to the hypotheses of interest. In this case, it is not possible to make forensic identification but it is only possible to make inference about specific pedigrees.

Let $\mathcal{E}$ denote one or more items of evidence (perhaps the totality). We need to consider how this evidence affects the comparison of the hypotheses, $H_0$ : `Tomb=NTped`, the tombsite belonged to a family with a pedigree like that of the NT family;[2] one alternative hypothesis could be $H_1$ : `Tomb≠NTped`, the tombsite does *not* belong to a family with a pedigree equal to that of the NT family. This alternative hypothesis could be formed by a number of hypotheses pertaining to each possible relationship.

When we are only comparing two hypotheses $H_0$ and $H_1$, the impact of the totality of say $k$ different elements of evidence $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_k)$, from all sources, is embodied in the *likelihood ratio*,

$$LR = P(\mathcal{E}|H_1)/P(\mathcal{E}|H_0). \tag{1}$$

When the items of evidence $\mathcal{E}_i$ for $i = 1, \ldots, k$ are conditionally independent given the hypotheses, the overall $LR$ can be computed as $LR = \prod_i LR_i$, where $LR_i = P(\mathcal{E}_i|H_1)/P(\mathcal{E}_i|H_0)$. Given the likelihood ratio, $LR_i$, based on the distribution of names (loosely, onomasticon) this can be updated with the $LR$s based on other items of evidence (e.g., all DNA profiles) and the evidence given in (1) to (10) of Section 14, to form the overall likelihood ratio.

We thus do not see the reason why the author excludes the possibility of computing a $LR$ and of using other pieces of evidence as well.

[2]The fact that no official sources contain information about Jesus from Nazareth having had sons should be appropriately considered.

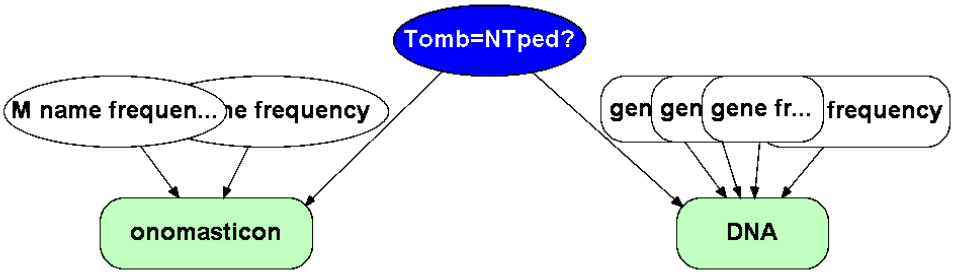


FIG. 1. *OOBN for tomb identification using onomasticon and DNA evidence.*

**5. OOBN for analyzing two or more pieces of evidence.** An object-oriented Bayesian network for analysing two or more pieces of evidence. OOBNs have shown to be an extremely versatile tool to handle different pieces of evidence relating to an identification issue; see, among others, Cowell, Lauritzen and Mortera (2007), Dawid, Mortera and Vicard (2007) and Taroni et al. (2006). A network can be built to compute the overall likelihood ratio given all the pieces of evidence.

Figure 1 shows an example of an OOBN for evaluating the weight of two pieces of identification inference: that from onomasticon together with that from DNA profiling.

In the network, the two hypotheses, described in Section 4, bearing on the pedigree of the tombsite ownership, are represented by the *true/false* states of the Boolean node `Tomb=NTped?`. The `onomasticon` node represents a complex subnetwork having as input both the Female and Male name frequencies, represented by nodes `F name frequency` and `M name frequency`, respectively. For example, the probability distribution and states of node `F name frequency` are given in Table 1. The `DNA` node represents another complex subnetwork having as input the gene frequencies represented by nodes `gene frequency`. The evidence on the tombstone names and the DNA extracted from the bones is entered in `onomasticon` and `DNA` and propagated throughout the entire network yielding, in node `Tomb=NTped?`, the overall likelihood ratio based on *all* the evidence.

We enjoyed reading the paper and writing this discussion. We recognize that Feuerverger does not have the DNA test results, but we wonder if he could facilitate access to these data so that further analysis could be made on this interesting case.

## REFERENCES

COWELL, R. G., DAWID, A. P., LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1999). *Probabilistic Networks and Expert Systems*. Springer, New York. MR1697175

COWELL, R. G., LAURITZEN, S. L. and MORTERA, J. (2007). A Gamma Bayesian network for DNA mixture analyses. *Bayesian Analysis* **2** 333–348.

DAWID, A. P., MORTERA, J. and VICARD, P. (2007). Object-oriented Bayesian networks for complex forensic DNA profiling problems. *Forensic Science International* **169** 195–205.

EGELAND, T., MOSTAD, P. F., MEVÅG, B. and STENERSEN, M. (2000). Beyond traditional paternity and identification cases: Selecting the most probable pedigree. *Forensic Science International* **110** 47–59. DOI: 10.1016/S0379-0738(00)00147-X.

GILL, P. D., IVANOV, P. L., KIMPTON, C., PIERCY, R., BENSON, N., TULLY, G., EVETT, I., HAGELBERG, E. and SULLIVAN, K. (1994). Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics* **6** 130–135.

GREEN, R. E., KRAUSE, J., PTAK, S. E., BRIGGS, A. W., RONAN, M. T., SIMONS, J. F., DU, L., EGHOLM, M., ROTHBERG, J. M., PAUNOVIC, M. and PÄÄBO, S. (2006). Analysis of one million base pairs of Neanderthal DNA. *Nature* **444** 330–336. DOI: 10.1038/nature05336.

GREEN, P. J. and MORTERA, J. (2008). Sensitivity of inference in Bayesian networks to assumptions about founders. Research report 08:04, Statistics group, Univ. Bristol. MR2271510

TARONI, F., AITKEN, C., GARBOLINO, P. and BIEDERMANN, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science. Statistics in Practice.* Wiley, Chichester. MR2271510

DIPARTIMENTO DI ECONOMIA
UNIVERSITÀ ROMA TRE
VIA SILVIO D'AMICO 77
00145 ROMA
ITALY
E-MAIL: mortera@uniroma3.it
vicard@uniroma3.it

*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 97–98
DOI: 10.1214/08-AOAS99B
Main article DOI: 10.1214/08-AOAS99


# DISCUSSION OF: BAYESIAN VIEWS OF AN ARCHAEOLOGICAL FIND

BY JOSEPH B. KADANE

*Carnegie Mellon University*

Andrey Feuerverger (2008) is to be congratulated on having given us such a careful analysis of a very interesting data set. He has obviously gone to great efforts to understand the archaeology (in several languages) and background of the tomb in question, and the literature and history surrounding it. This effort is exactly what a modern statistician should be doing in an applied problem.

Unfortunately Feuerverger is hampered, in my view, by his predisposition toward sampling theory. His technique relies on his RR-values ("relevance and rareness"), but he gives no theory of RR. Just what is it? What justifies multiplying them together? What have you got when you're done? Second, his method computes the probability of data as or more extreme than that observed were the null hypothesis true, which violates the likelihood principle because it counts as relevant data that did not occur. Finally, his method is very limited in the conclusions it permits one to draw: either the null hypothesis is false or something unusual has happened. Well, which is it? Using his paradigm, he is unable even to give a probability of which of these is the case. A great deal of effort goes into establishing a conclusion whose form does not address the question of interest, at least as I interpret it.

By contrast, a Bayesian treatment has clear-cut and simple rules. These have been worked out extensively for problems in forensic science; indeed the present problem can be so regarded. The question, as Feuerverger himself points out, is to calculate $P(B \mid A)/P(B \mid \overline{A})$ where $A$ is the event that the Talpiyot tomb is that of the $NT$ family, and $\overline{A}$ is that it is not. The event $B$ is the evidence we have, namely the specific names found in Talpiyot. $P(B \mid A)$ is probability of this tomb arising if it were the tomb of the $NT$ family. Thus it involves what other renditions of names might have been used for the persons in the $NT$ family, and the possible identities of the unidentified persons in the tomb. Similarly $P(B \mid \overline{A})$, which is essentially what he is computing from the onomasticon, is the probability of this configuration arising from some other family or group of people. While he says that this specification of $B$ is "awkward to work with," it seems to me that it leads us to address the essential questions in analyzing the Talpiyot tomb.

Höfling and Wasserman (2008) and Ingermanson (2008) in preceding comments on the paper give differing Bayesian analyses of this problem, and Mortera and

Received October 2007; revised December 2007.

Vicard (2008) stated how they would use DNA analysis in one. Should we be disturbed that the former two make different assumptions, and derive different posterior probabilities? I would argue not. The strength of the Bayesian approach is that it requires the assumptions to be stated explicitly and argued for. The acceptability of those assumptions is for each reader to judge for himself or herself. All the Bayesian argument ensures is that each writer is coherent, that is, does not contain internal contradictions in a certain technical sense. Thus the Bayesian view of probability is arguably like a language. That a sentence is in grammatical English does not require the reader to agree with it; proper grammar only helps us to understand what the writer means. Similarly, an opinion expressed in probabilistic terms is explicit, that is, a reader can understand what the writer's view is, but it is up to the writer to be persuasive to the reader. Each reader, then, needs to state the beliefs found most congenial, and to compute his or her own posterior probability accordingly.

Finally, it is obviously necessary to say something about how the statistical analysis of this data set relates to the religious beliefs of many people. Fortunately there is no contradiction between the Bayesian paradigm and such beliefs. Bayes Theorem in odds form reads, as Feuerverger points out,

$$\frac{P(A \mid B)}{P(\overline{A} \mid B)} = \frac{P(A)}{P(\overline{A})} \times \frac{P(B \mid A)}{P(B \mid \overline{A})}.$$

Here the factor $P(A)/P(\overline{A})$ is the prior odds of the event $A$. For those whose religious beliefs specify $P(A) = 0$ and $P(\overline{A}) = 1$ (i.e., there is no chance that the tomb is that of the $NT$ family), whatever the likelihood contribution [here $P(B \mid A)/P(B \mid \overline{A})$], the posterior odds of $A$ [here $P(A \mid B)/P(\overline{A} \mid B)$] are zero. This set of beliefs is coherent in the technical sense (i.e., it does not lead to sure loss), and hence is fully consistent with the Bayesian view.

## REFERENCES

FEUERVERGER, A. (2008). Statistical analysis of an archeological find. *Ann. Appl. Statist.* **2** 3–54.

HÖFLING, H. and WASSERMAN, L. (2008). Discussion of: "Statistical analysis of an archeological find" by A. Feuerverger. *Ann. Appl. Statist.* **2** 22–83.

INGERMANSON, R. (2008). Discussion of: "Statistical analysis of an archeological find" by A. Feuerverger. *Ann. Appl. Statist.* **2** 84–90.

MORTERA, J. and VICARD, P. (2008). Discussion of: "Statistical analysis of an archeological find" by A. Feuerverger. *Ann. Appl. Statist.* **2** 91–96.

DEPARTMENT OF STATISTICS
BAKER HALL 232A
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213
USA
E-MAIL: kadane@stat.cmu.edu

*The Annals of Applied Statistics*
2008, Vol. 2, No. 1, 99–112
DOI: 10.1214/08-AOAS99REJ
Main article DOI: 10.1214/08-AOAS99


# REJOINDER OF: STATISTICAL ANALYSIS OF AN ARCHEOLOGICAL FIND

BY ANDREY FEUERVERGER

*University of Toronto*

I thank all the discussants for their many critiques and comments, and for their considerable efforts. Many of the points raised are ones with which I (at least in part) agree. It therefore seems easiest to first deal with a number of points with which I don't agree.

First, Fuchs states (and Bentley appears to assume) that my analysis is documented in a book and in a movie, neither of which I have authored. In fact, it is documented *only* in my paper which references neither of these, and neither does it reference any developments which occurred subsequent to my work. Although I will need to comment on one such development below, I otherwise confine this reply to the contents of my paper and to those comments of the discussants which appear within this issue of the *Annals*. In particular, I avoid being drawn here into discussions concerning representations made elsewhere by others, or to any matters alluded to by discussants that are peripheral to the central and substantive statistical issues of the problem. Nothing in this work was ever intended to cause offence to anyone. In my view, the statistical problems here are of methodological interest, and the subject matter is one of historical and archeological significance. If this tomb is not that of the NT family (as indeed it may not be) then archaeological work could still one day unearth a tomb that is and the question of what statistics might then contribute toward such a pursuit could then become important.

I also want to say that my paper does not—as some discussants intimate—claim that the Talpiyot tomb "is most likely that of the NT family." What it tries to do is develop tools to assist subject matter experts in their work of gauging the veracity of any such claims. The function of statistics here is to help out in the difficult historical and archeological work. The critical role which historical assumptions play here means that such calls are not ours to make; and like Fuchs, I too refrain from passing judgment on the subject matter issue of whether this is or is not the NT tombsite. Of course, after the fact, it is easy to gain a sharpened appreciation for the safety of a "nihilistic" approach, one that—as Höfling and Wasserman put it—provides no answers. However, the intellectual temptations posed by a problem of this nature are surely too great to simply set aside.

Received February 2008; revised February 2008.

*Meaning of surprise.* Turning to some specific issues raised by the discussants, I think it is important to distinguish more carefully between "interesting" or "relevant" collections of names, and what I have defined as being "surprising" collections of names. If a NT tombsite actually exists, it is certainly within the realm of prior possibilities that it contains within it only the most common renditions for the names of persons who might be recognizable to us. Were this so, no purely statistical procedure would then be able to "detect" it because such collections of names would not occur rarely enough in the general population to allow any procedure at least an opportunity to attain significance—that is, we then could never know. Indeed, only if the actual burials had taken place under rarer relevant renditions of the names (and only if in a tomb of a certain size) could there ever be a chance to "detect" it statistically. In other words, some historical and archeological "good fortune" would also be required.

Both Höfling and Wasserman, as well as Fuchs, appear to misinterpret my definition of "surprisingness" and its intent. In fact Höfling and Wasserman state that "the RR statistic becomes more significant if broad name categories are being subdivided into special name renditions, even if the particular name renditions are not relevant." But that does not take into account that the specialness of a name rendition is permitted to count only if it is relevant, and only if it appears in a prespecified nested list of increasingly more specialized (i.e., "rarer and more relevant") name renditions. The rareness alone of a name rendition (even if it corresponds to a generic name category deemed to be highly relevant) is not of essence. When Höfling and Wasserman state that "interested observers would surely argue that a tomb is interesting if there is any way at all of matching the names found to potentially interesting names," they bypass the fact that such matchings will be relatively too probable to be significant if they were to occur under the most common renditions for the names. Likewise, Fuchs suggests in an example that, had the Talpiyot tomb contained a Salome in lieu of the Mariamenou [$\eta$] Mara inscription, it would have been considered still more surprising even though its RR value would then have been higher. In fact, based on the definition of surprisingness, had a Salome been found in lieu of the Mariamenou inscription, the cluster might conceivably be described as being more relevant, but (in view of how common Salome was as a name) it certainly would not have been more surprising, that is, it would not have provided a greater evidentiary value (under our provisos).

*Other misinterpretations.* Fuchs also remarks that if a Simon had been found in the tomb instead of the Matya, our RR value, and hence our tail areas, would have been unchanged even though that Simon might have been a brother of Jesus. Now one can certainly carry out analyses allowing for the brother Simon to be a candidate for a NT tomb. The reason we did not do so, however, is that Simon is presumed to have died subsequent to the time (70 CE) when the practice of ossuary burial ceased. If this is so then it is appropriate to have, as Fuchs puts it, "ignored that inscription."

Fuchs also states that assigning an RR value of 1 to names in the "Other" category means that such names will not contribute to the RR value of a cluster, the implication being that such names are then simply ignored by our procedure. This actually is not so. Such values of 1 do in fact contribute in the sense of being values much higher than would have occurred had relevant names been encountered instead. (As well, such values of 1 for "Other" also affect our null distribution.) And in any case, we have also allowed for values exceeding 1 via the device of a disqualifier list, but we did not implement such a list since no such names appeared in-sample; disregarding such names was therefore conservative.

One referee of the paper had stated that merely including the Talpiyot names within the onomasticon necessarily biases our results. I do not agree that this is the case. Any and all available names may and should in fact be used to aid in the process of constructing the prespecified categories of name renditions and nesting them according to "relevance and rareness." What is important is that this process be carried out without reference to which of the names actually originated from within the tomb.

Fuchs mentions that the RR value for the Talpiyot tomb uses only four out of the six inscribed ossuaries on account of the fact that the Matya ossuary only contributes a 1 to the RR value for the tomb. That view is not entirely correct. As noted above, the Matya ossuary renders the RR value for the tomb much higher than it would have been had a more relevant name been encountered in its place, and in turn this results in a considerably increased tail area (for the RR value) under the null distribution. The situation with respect to the sixth inscription (i.e., Yehuda bar Yeshua) will be taken up separately below; for the moment we only note that, in effect, this inscription also contributes an RR value of 1.

*A hypothesis testing issue.* Fuchs states that "A set of rules which weigh positively (i.e., with a coefficient less than 1) names expected under $H_1$, but does not weigh negatively names which are unexpected under $H_1$, is likely to yield biased results in favor of $H_1$" and that "this procedure is at least questionable." Kadane's critique of the RR measure points in a related direction. Intuitive as such remarks may seem, however, they are not entirely correct. First, as long as a test statistic is specified a priori (in particular, without making reference to the data), and as long as its distribution under the null hypothesis is specified correctly, the resulting test will be unbiased in the sense of having its stated level of significance. Essentially, only the power of the test will be at issue—a consideration that leads us to seek procedures with high ability to discriminate. Second, our allowance for a disqualification set does to some extent permit certain names to weigh negatively; the reason we did not implement such a set (as mentioned previously) was only because Matya was not considered to be a disqualifying name so that our choice not to do so was conservative. (In fact, the Matya name was viewed to be neutral.) Contrary to Fuchs' assertions, there is nothing questionable about our actual *procedure*. The matter of the Yehuda inscription will be discussed separately below.

*Post hoc inference.* Fuchs does point out correctly that "the a priori nature of the provisos is amongst the most important premises" of the analysis. He further states that "The overall impression is that the inevitable exposure to the data affected the definition of the provisos." Concerning the degree to which the provisos were truly a priori, he adds: "It is difficult to accept that ... the elements ... have indeed been so specified." On such points I am certainly sympathetic to the general nature of the concerns raised by the discussants and therefore revisit certain of the 'provisos' in the discussions below. It is indeed true here that prior exposure to the data was inevitable, and the principle is well understood that biases result if data is used when setting up an inference. In fact, the best we have been able to do was to stress the fact that the data has been seen. And we also tried, both conscientiously and hard, and on a best efforts basis, to construct our inference to be as a priori as possible in the circumstances. The extent to which we have succeeded or not in this task is one which each reader ultimately must judge for themselves. For regardless of the degree of objectivity any analyst may wish or seek to claim in carrying out an analysis under such circumstances, no convincing or irrefutable proof of such objectivity can ever be offered. This is the perennial problem of pre- versus post-hoc inference, and the present statistical problem provides a good example of it. It is also the reason why I tried to be so careful to isolate and exhibit all of the assumptions under which the analysis was carried out.

*Mariamenou need not be Mary Magdalene.* Before addressing the critiques to my provisos, there is one further item that needs to be clarified. Fuchs intimates that the RR value assigned to the Mariamenou [$\eta$] Mara inscription means that this name must then necessarily refer to Mary Magdalene. (A related remark is also made by Bentley.) This interpretation is not correct. That RR value resulted from that version of the name having been considered, on an a priori basis, to be the most 'rare and relevant' rendition of the her name from amongst those names that we know. (But on this point, note the further discussion below.) It assumes no more than that only one woman having the generic name category of Mariam out of about every 44 such women could legitimately have been called by that rendition, and that Mary Magdalene was among those who could so be called. In particular—although that possibility was weighed into the process when deciding upon our a priori nested rendition categories—it certainly was *not* assumed that the Mara in the inscription *must* be an honorary title, only that it *might* be. For if that were assumed one must surely agree with Fuchs (and Bentley) that no statistical analysis would then be required.

We now turn our focus to some of our various provisos starting with the ones associated with Mary Magdalene.

*Mary Magdalene as a priori candidate.* With respect to Mary Magdalene, there are at least two distinct considerations. The first of these is the matter of her

inclusion on our list of a priori candidates for a NT tomb. Of course, this is primarily a historical issue, not a statistical one, and as such needs to be vetted through *dispassionate* subject matter expertise. While sensitivities surrounding this point render the scholarly work more difficult, I really do not see how one can exclude her from that list. This is in no way tantamount to any assumptions about to whom, if anyone, she may have been married. The perceptions of Mary Magdalene having been unchaste apparently originates with Pope Gregory the Great in the last decade of the sixth century and has no basis at all in the NT—a point that even the Vatican conceded in 1969. Her presence is felt both prominently and strategically throughout the NT accounts. She is the pivotal figure and primary source for the resurrection. She accompanies Jesus over substantial distances and over a substantial period of time. She even appears, from the accounts, to have been highly active in Jesus' ministry. She is present at the crucifixion, and also at the burial where (in view of the likely nature of such rituals in that era) one would expect only intimates of the family to attend. Indeed, she is also cited as having been in the vicinity of the tombsite on multiple occasions. So, on balance—and in view of the possibility that she may have been buried in the Holy Land—I really do not see how one can realistically exclude her from at least being a *candidate* for a NT tombsite. One must avoid a certain blurring of logic that can occur inadvertently here: The inclusion of Mary Magdalene on an a priori list of candidates for a NT tombsite is *not* equivalent to asserting that she must actually be found in such a tomb. It only says that she is, a priori, among the plausible *candidates*. The distinction between these must not be blurred by the occurrence of the Mariamenou [$\eta$] Mara ossuary within our data. Of course, there is no obligation on anyone's part to accept the argument we make here; if one chooses to omit Mary Magdalene as an a priori candidate then the impact of that choice is clear: no statistical analysis applied to the Talpiyot data would then attain significance.

*Names for Mary Magdalene.* With respect to the second key consideration pertaining to Mary Magdalene, the situation is more problematic. At the time I did the analysis my "due diligence" in respect of constructing an a priori nested list of name renditions for Mary Magdalene included such elements as the following:

(a) The itemization of the 80 known renditions for the generic name of Mariam as recorded in Ilan (2002).

(b) The meaning of the Aramaic word "mara"; specifically, Ilan (pp. 392 and 423) states: "Mara means 'lord, master' in Aramaic."

(c) An inference, based on (a) and (b), that of these 80 name renditions for Mariam, the extraordinary Mariamenou [$\eta$] Mara one may provide an arguably tighter "fit" to Mary Magdalene than any of the other 79.

(d) The reference by Hippolytus, around 175 CE, in his Refutations, to a particular Mariamne who was legitimately a "mara" in the Aramaic sense of that word.

(e) The unwavering opinion of Rahmani, and of some other highly regarded epigraphers (e.g., Leah Di Segni), that the full inscription on ossuary #1 was intended to refer to a single individual only.

(f) The article by Francois Bovon (2002) identifying the Mariamne in the Acts of Philip as Mary Magdalene, and identifying Philip as her brother.

And finally:

(g) Information provided to me (but note the discussions below) that Professor Bovon—a highly respected scholar and expert on this subject—was on record as having authenticated that Mariamne was most likely the actual name of Mary Magdalene.

I was, of course, also aware of the fact that the inscription Mariamenou [$\eta$] Mara had occurred within the tomb, and obviously also of the fact that such information must be disregarded when forming a priori assumptions. However there is as well a concomitant piece of information of a seemingly ancillary kind, and not entirely unrelated to our conditioning on the tomb's configuration. Namely, it is known that the Mariamenou [$\eta$] Mara inscription was rendered in Greek, but that it occurred within a tombsite containing five other inscribed ossuaries all of which were rendered in Aramaic. How and if such a piece of information may be used in forming a priori assumptions is not entirely clear to me and I leave this as a question for readers to consider. Similar issues also arise in respect of such considerations as the nature of the actual incisions and so on. I also point out in passing—although this should not be regarded as being an a priori observation—that to the best of my knowledge, Mary Magdalene is the only historical personage who was ever referred to by the generic name of Mariam combined with the Hebrew letter "nun," and that she is referred to in that way in two distinct sources (Hippolytus and the Acts of Philip).

The controversies resulting from the airing of the documentary film was a unique event in the context of any statistical problem I had ever dealt with, and went beyond what I might realistically have been able to prepare for. Scholars and others who were involved in any way were subjected to pressures that sometimes made it difficult to discern where the actual facts lay. Speaking for myself, I was interested only in the facts. The story of the crucifixion has held sway over the history of humanity for some 2000 years. It therefore seemed worthwhile to stay the course that happenstance had led me to, and to steadfastly pursue the facts to whatever would be their logical conclusion.

*Bovon's clarification*. This brings us to the subject of the clarifications subsequently issued by Professor Bovon. There is no doubt whatever now that these were not retractions in response to pressures nor were they motivated by a recognition of the possible uses which might be made of such work. In fact, Bovon's clarifications are those of a serious scholar whose remarks—having inadvertently been misinterpreted by Jacobovici—were conveyed to me out of context. To quote from Bovon's statement to the Society of Biblical Literature:

"I do not believe that Mariamne is the real name of Mary of Magdalene. Mariamne is, besides Maria or Mariam, a possible Greek equivalent, attested by Josephus, Origen, and the *Acts of Philip*, for the Semitic Myriam."

"Mariamne of the *Acts of Philip* is part of the apostolic team with Philip and Bartholomew; she teaches and baptizes. In the beginning, her faith is stronger than Philip's faith. This portrayal of Mariamne fits very well with the portrayal of Mary of Magdala in the Manichean Psalms, the Gospel of Mary, and Pistis Sophia. My interest is not historical, but on the level of literary traditions."

Without benefit of the last element, that is, (g), of the itemization above, I do not regard the assumption A.7—concerning the most appropriate name rendition for Mary Magdalene—as being equally adequately justified by the remaining elements (a) through (f) on that list. In particular, this means that we cannot (on the basis of our RR procedure) say that the Talpiyot find is statistically significant in any meaningful way. Readers who wish to form their own judgement on this should note that the germane question here is not whether or not Mariamne was the actual name of Mary Magdalene, but whether or not we are justified—on an a priori basis—to say that the rendition Mariamenou [$\eta$] Mara provides a better fit to the name of Mary Magdalene than any of the others, whilst bearing in mind that she is repeatedly referred to in the NT as having come from Migdal, and is not referred to there as Mariamne. We shall see below, however, that this matter is not yet closed.

*The Yehuda ossuary.* Now let us deal with the matter of the sixth ossuary—the admittedly problematical one inscribed Yehuda bar Yeshua. When I encountered this data set I did not at first have a clear idea of how that datum should be dealt with in an analysis and I tentatively set it aside. It would be fair to say that the apparent implications suggested by that ossuary would hardly have found any mention of or allowance for in my list of a priori assumptions for several reasons, not the least of which being that such a possibility would not ever have occurred to me. After the RR approach evolved, it became clear to me that this sixth ossuary was actually being incorporated within the computations in a particular way. As indicated in Section 14 of the paper, the analysis may in fact be carried out allowing for the presence of a generationally aligned sequence of the form "A son of B son of C" with the youngest of this trio "not counting" toward the RR value due to our lack of knowledge about *any* father-and-son pair *both* dying within the 30–70 CE timeframe. Of course, this still leaves open the question of associated a priori assumptions. If one ascribes to certain theological interpretations later placed upon the historical events, the decision is clear: the outcome observed must belong to the disqualification set, and the matter is closed. If one does not so ascribe, the situation becomes more difficult, for then one must interpret the historical records as best one can to assess the plausibility of such an outcome, and address such questions as the following: Would a union in such an instance have been sanctioned? Was it— in that era—viewed as improper to father a son? Did Jesus advocate against it for either self or followers? If there were a son, would there have been

a recognized threat to his life? We cannot answer these or other such questions on behalf of the reader. Certainly the NT does not record any union or any son (although much other information is left unprovided as well). As for the statistical analysis based on RR, what we can say is that in assigning an RR value of 1 to the sixth ossuary, our procedure in effect acts with absolute neutrality on this question.

*Some extensions.* A few methodological points seem worth noting. Instead of prespecifying nested collections of name renditions one can (for each candidate individual) preassign numerical RR or surprisingness ratings to each of the onomasticon entries under their generic name. Only comparative (not numerical) values would actually matter, and the RR computation for an encountered rendition would then be based on the "tail area" resulting within the generic name category. Since many entries in the generic collection will have identical ratings, the resulting "discreteness" of the tail areas would act much as in our nested collections approach but would allow for somewhat greater flexibility. Note also that we could allow for the existence of rare renditions as yet unknown. For any candidate individual, the rareness of any such renditions would at most be (in the order of) that of a single unique entry in the onomasticon. We should remark here that a certain amount of variability in our results is attributable to the fact that name proportions are derived from the onomasticon which itself constitutes only a sample; Mortera and Vicard propose one method for assessing such variabilities.

*Other explanations and concerns.* Although many of the discussants focus on critiques to the analysis that might have been anticipated to arise out of theological grounds, Stigler alludes to some which stem from nontheological sources. For example—although the circumstances of the find assure us that the tomb had been undisturbed for many centuries—we do know that the Talpiyot tomb had been accessed at some point in antiquity. While it seems implausible to assume undue efforts on the part of those who did so, suppose they had found there only five of our six inscribed ossuaries and "recognized" the names on them. Might they not have thought it amusing to then take one of the uninscribed ossuaries there and crudely scratch upon it the name Yeshua bar Yehosef using an implement at hand? As for Stigler's reference to Sherlock Holmes' dog who did not bark, I did independently pursue the matter of why the placement of the ossuaries among the *kochim* had not been noted and concluded that this likely had occurred only on account of the general circumstances of the find and of Gath's untimely death upon which that potentially priceless piece of information was permanently lost. It must be remembered that the archaeologists who were sent there were not statisticians, that they could hardly have anticipated the nature of the questions that would later arise from this duty, that they had limited time inside a tomb containing only seemingly typical names, and that the messy Yeshua inscription could hardly have been decipherable to them at first. In fact one could (following up on a comment made by Bird) argue equally (although for what I believe are good reasons I do not) that

the lengthy period which elapsed between the time of the tomb's discovery, and the time of the publication of its details, provides a yet contrasting instance of the dog not barking.

Stigler also raises the matter of our specialized independence assumption A.9. Our concerns, as well as our reasoning about this assumption, were discussed in Sections 7 and 14 of the paper. But in bringing this data set to the attention of the statistical community, it was understood that questions which merit further study would arise from it and the issue of cross-sectional independence is one of them. Here the question is not whether or not this assumption is true; we *know* that it is not. The question is whether the nature of that dependence affects the null distribution in an essential and nonconservative manner. I refrain from any rejoinders to Stigler's references to Bruno and Galileo finding such remarks too frightening to even contemplate.

Explanations based on coincidence should also not be overlooked; indeed, perhaps these data can be assessed under the framework of Diaconis and Mosteller (1989). Within the context of coincidence, odds of 1000 to 1 are hardly uncommon. Three "coincidences" weighed substantively in our analysis. One is the ossuary of Yeshua bar Yehosef. Another is the match to the rare name version Yoseh. And the third is the remarkable Mariamenou [$\eta$] Mara inscription. There are, however, also three further coincidences that (for reasons stated in the paper) I did not incorporate in the analysis but nevertheless seem worth noting. The first of these is the generational alignment of the three names Yehosef, Yeshua and Yehuda, with the alignment at Talpiyot being the only one among the six not immediately inconsistent with the NT family. The second is the seemingly suggestive choice, among the six ossuaries, on which the Greek script actually occurred, with the other five having been in Aramaic. And the third is the suggestive choice for which of the six ossuaries bore the messiest inscription—that choice being seemingly consistent with some theories that might be advanced to account for the empty tomb. Finally, there is yet one further coincidence: The youngest member of the generationally aligned ossuaries—namely Yehuda—has the same name as the youngest (or second youngest) brother of Jesus, with the accounts of Mark and Matthew having curiously reversed their two names.

Let us now address some specific further matters raised by the discussants.

*Höfling/Wasserman's first method.* In the "Different Approach" proposed by Höfling and Wasserman, the most essential difference actually lies in the treatment it accords to the different name versions. In particular they "lump together different versions of names" arguing "that a tomb is interesting if there is any way at all of matching the found names to potentially interesting names." Unfortunately, for common names, "interesting" will not be enough; there will be little opportunity for detection (i.e., the power will be low throughout all of the alternative) unless the renditions which occur match more specifically to the NT individuals, and if the specificness of such renditions is appropriately accounted for. A manifestation of

this is that their calculation is "invariant under splitting names into subcategories," while our calculation (which attempts to account for the degree of rareness and relevance among the possible renditions) is not. Thus, had an inscription such as (say) "Yeshua of Nazareth, son of Yehosef" occurred in the tomb, their computations would be indifferent to an essential aspect of the name. Incidentally, Höfling and Wasserman are not correct in suggesting that what I have computed is "the probability of getting this set of names."

*Bayesian notions.* Several referees argue in favor of a Bayesian approach, something I tried to avoid due to the great divergences expected amongst priors (some of which have been influenced by theological considerations). Also, I do not entirely understand Kadane's remark about violation of the likelihood principle. Kadane appears to suggest that the uncertainties in deciding between whether a null hypothesis is false or whether a rare event has been observed is merely an artifact of the frequentist approach. It seems to me, however, that no purely statistical method can ever circumvent its analogue for "type 1 error." Further, in allowing a prior to place a zero probability on a discrete event, Kadane highlights a difficulty that can arise in a purely classical Bayesian approach, unless one takes to its extreme the view that "coherence" alone must suffice. It is also not entirely clear to me how straightforward it would be to implement LR procedures of the type Mortera and Vicard advocate. The Bayesian approach proposed by Höfling and Wasserman, however, does on first glance appear to lead to results comparable to those of a frequentist approach, as long as the assumptions under which the two approaches are implemented—in particular the assumptions concerning the renditions for the relevant names—are taken to be similar. Bayesian-like ideas may of course also be used to rationally combine subjective beliefs about individual assumptions into a plausibility for the collection of all assumptions. Our approach has been, however, for the RR method to act as a measuring device, to be tuned by the investigator in accordance with his or her expert assumption set.

*Bentley.* Bentley is correct in stating that my analysis assumes that a NT tomb might exist, but I do not fully agree with him that my analysis is conditioned on the assumption that such a tomb *must* exist with probability one. Also, while it may be fair of Bentley to argue with the estimates I used for the number of tombsites in Jerusalem, I am not aware of any expert opinion suggesting that the true number of tombsites is greatly in excess of the numbers I had used. Bentley's critiques regarding the Mariamenou [$\eta$] Mara inscription are well taken and this matter has been dealt with at length in our discussions above. If, in spite of my labors, Bentley wishes to be critical of them, he is within his rights. Nonetheless —lest Bentley's comments regarding James Tabor be misconstrued—I wish to say that in my discussions with Professor Tabor I found him always to be a scholar of impeccable integrity. Some of Bentley's comments, for example his closing remarks about archeologists and archeology being now at odds with statistics and

with statisticians admittedly make for provocative and dramatic reading; unfortunately pressures of time do not permit me to enter into such debates.

*Ingermanson.* Ingermanson energetically presents "the case against" for essentially each one of the assumptions under which my computations were carried out. Although his critiques seem occasionally overzealous to me, they do provide a useful checklist of items that should be considered by anyone who seeks to arrive at a fully informed opinion about the Talpiyot tomb. Needless to say, an analysis of such data needs to be carried out under assumptions that are reasonable and defensible, even though no single assumption can ever be regarded as absolutely unassailable. Having already discussed many of such matters in my paper and throughout this reply, I do not repeat those arguments here, but instead return to two specific items. The first item concerns the treatment of the name Yoseh mentioned by Ingermanson (as well as others). I add here two additional points to those already made in the paper. First, surnames were not typical in that era and exceedingly common names (such as Yehosef) would not have provided adequate differentiation amongst individuals. In that respect, one needs to bear in mind the distinction between what we commonly refer to as being a "nickname" versus a name rendition or variant that is in itself intended to act as an actual name. (An instance of this may, e.g., have occurred in the case of the NT family.) Second, there is no singleton Yehosef inscription occurring within this tomb. Therefore such a seemingly more "formal" or more "respectful" version for Yoseh would have been available for use by the family without any risks of confusion but they chose not to use it. The second item concerns the name of the mother. I take issue with the objections Ingermanson (and others) raise regarding the a priori "rarer and more relevant" rendition of her name used in my analysis. The very earliest historical reference to her appears in Mark 6:3: "Is this not the carpenter/builder, the son of Maria, the brother of. . . " The second earliest reference also occurs in Mark when he mentions Maria as being at the cross, although whether or not this Maria ("the mother of James and Joseph") is meant to be Jesus' mother is not entirely certain. The third earliest historical reference to her appears in Matthew 1:18: ". . . when His mother Maria had been betrothed to Joseph. . . " (Luke does not use the form Maria but rather Mariam, however Luke is historically a significantly later source.)

*DNA and other evidence.* Mortera and Vicard raise the question of why DNA evidence was not collected and assessed more broadly and indicate some possible uses of such data. Bird alludes to some related matters as well. Having had reservations, such as about the risks of contamination, these were evidentiary points I decided not to pursue. I understand that it is, in any event, the case that such data cannot actually be obtained. While I appreciate the reasons behind such concerns, I also believe that Bird may be making more of the missing tenth ossuary than may actually be warranted by the facts.

*Critiques.* The critiques of the discussants encompass both the methodology and the assumptions under which it is being applied. The vigor of their remarks represents an important component of the scientific process when results seen to be controversial are being assessed. Although some of the discussants hold strong prior views on the subject matter, all critiques do nevertheless need to be considered on their own merits. So far as the methodology itself is concerned, I think I have addressed the main points that have been raised; however, the situation regarding the assumptions is necessarily different. These need to be vetted by *dispassionate* subject matter expertise. It is a curious and perhaps unique feature of this problem, however, that the body of subject matter expertise here is itself divided along very particular lines.

*A symposium.* In January 2008, at about the time I prepared this reply to discussion, I had the privilege to attend The Third Princeton Symposium on Judaism and Christian Origins held in Jerusalem. Several of the sessions at this conference were connected to matters relevant to evaluating the context of the Talpiyot tomb. Among the subject matter participants there working primarily with historical approaches, some indicated that they did not regard the Talpiyot tomb's being that of the NT family as an impossibility. All of the participants however (myself included) indicated that they did not regard that possibility as having been proven. The most interesting session there—relative to the requirements of our statistical analysis—was one on the epigraphy of the Talpiyot ossuaries during which the Mariamenou [$\eta$] Mara inscription was discussed. As might be expected, no consensus was reached in that session, but one remarkable *possibility* emerged of which few members in the audience (which consisted of nonstatisticians) grasped the immediate significance. That possibility, raised by Jonathan Price—a classical Greek epigrapher (among other qualifications)—was that this inscription had been done by one hand, that it likely referred to a single individual, and that it should likely be read as "Mariam also known as Mara," the presumption being that Rahmani misread an intended $\kappa\alpha\iota$ as an $\nu o\upsilon$ ending in the first name together with an $\eta\ \kappa\alpha\iota$, and that the $\kappa\alpha\iota$ in this instance was intended to signify a double name. (See Figure 1.) Were that the case, it seems to me that the element (g) of our "due diligence" list above could then be supplanted by one that would now be considerably stronger still. It is worth mentioning here that the classical Greek epigrapher Roger Bagnall had earlier independently arrived at a similar reading: "Mariam, also called Mara," finding Rahmani's reading to be "not acceptable," but proposing that Mara may have been intended as a short form of Maria, although "some uncertainty remains" (quoted from a June 2007 e-mail communication). Unfortunately, the (spiraling) multiplicity of readings and interpretations for that inscription, and the nature of the relative uncertainties among them, makes it difficult to give unequivocal preference to any one of the readings, and until further work and consensus establishes at least the correct reading of this inscription (let alone any correct interpretation of it) further progress along this front seems unlikely.

One possibility that had not occurred to me was raised by a participant [Claude Matlofsky] and seems worth noting here. Namely, that the Talpiyot tomb *might* also fit the profile of the family of Jesus' brother Yoseh. Under that scenario, Yoseh would have named a son after his slain brother, but the assumption A.8 about Yoseh and Yehosef being necessarily distinct individuals would have to be suspended. It bears pointing out here that statistical 'evidence' of the nature described in my paper, even if significant, cannot automatically be used to also identify the actual persons buried in the tomb, nor any of their relationships to each other; these are separate inferential problems.

*Some opinions and concluding remarks.* During the course of this work I have had occasion to meet many of the individuals involved in this matter, including Andre Lemaire who "discovered" the James ossuary, Oded Golan who owns it and who kindly permitted me a private viewing of part of his remarkable antiquities collection, Shimon Gibson and Amos Kloner both of whom (along with the late Yosef Gath) were present at the Talpiyot find in 1980, as well as with a number of other key persons. In such meetings I tried to gather information, or at least to form impressions, about some of the nonstatistical aspects relevant to the analysis. A few such observations may be worth mentioning here. First, opinion on the authenticity of the James ossuary is divided so I have no basis in forming a judgement on that matter. Either way, if the James ossuary provenanced to the Talpiyot tomb (as some have claimed) the statistical implications would be nontrivial. However, in my opinion there is at present no credible evidence to tie that ossuary to the Talpiyot tomb. Second—although no one who has witnessed first hand the intensity that can be engendered by this subject matter would deny that such an eventuality should, at the very least, be momentarily considered—the possibility of any "cover-up" of facts by the archeologists involved strikes me as being pure fiction. The dynamics for such a thing to have taken place simply were not there. Third, a story made headlines when the widow of Yosef Gath announced that her late husband knew and had told her that he had discovered the tomb of Jesus, and that he was deeply concerned about the possible repercussions of that find. Having been present at the event during which she made this statement, I found it easy enough to gather sufficient information to lead me to be concerned that this could have been an instance of "false memory syndrome"; I am therefore inclined discount that information.

A few participants at the Princeton symposium indicated that it might be worthwhile to carry out further excavations at the Talpiyot tombsite and in particular at another immediately adjacent tomb. While it is always possible that such further work might lead to more definitive answers, it is also the case that Israeli laws are very strict about matters that pertain to disturbing burial sites. Therefore, unless evidence comes to light to invalidate the Talpiyot find, this, it seems to me, is where matters are likely to rest for some time to come.

## REFERENCES

BOVON, F. (2002). Mary Magdalene in the *Acts of Philip*. In *Which Mary?—The Marys of Early Christian Tradition* (F. Stanley Jones, ed.) 77–89. *Society of Biblical Literature Symposium Series* **19**. Society of Biblical Literature, Atlanta.

DIACONIS, P. and MOSTELLER, F. (1989). Methods for studying coincidences. *J. Amer. Statist. Assoc.* **84** 853–861. MR1134485

ILAN, T. (2002). *Lexicon of Jewish Names in Late Antiquity*, *Part 1*: *Palestine 330 BCE–200 CE*. Mohr Siebeck, Tubingen.