MYLES LAVAN

Greek names and freed status in Roman Italy:

Why ancient historians can't ignore statistics

Abstract: The rule of thumb that an individual with a Greek name can be assumed to be of freed status is widely used in the social history of Roman Italy. This paper shows that it is based on a logical fallacy and must be abandoned. There is no valid way to use names to impute the status of individuals without knowing the mix of statuses in the population. The paper goes on to show that it possible to make inferences based on onomastics but that it requires a formal statistical model of the relationship between names and status. The method is illustrated by application to the lists of members of *collegia* from Roman Ostia.

Keywords: statistics; liberti; manumission; legal status; names.

Ancient historians are often sceptical of the application of "statistics" in a broad sense – not just the apparatus of inferential statistics but all formal quantitative analysis – to their material. There are certainly good reasons for caution. The significant biases that distort surviving datasets, especially the mass of epitaphs, can easily lead the unwary astray. But that does not mean that historians can afford to forego the intellectual discipline of mathematical reasoning. What might seem to be commendable caution can lead to errors at least as serious as over-confidence in quantification. This paper presents a cautionary example of the potential pitfalls of *not* taking a statistical approach. I will show that a rule of thumb that is widely used in the social history of Roman Italy is based on a fundamental error in reasoning. That error has proliferated unchecked precisely because of a widespread reluctance to resort to formal quantification when handling a complex question that clearly involves probabilities.

One major difficulty in the social history of Roman Italy is the precipitous decline in the use of the formal markers of status – filiation and libertination – in epigraphy of the imperial period.[1] Historians have to resort to other methods to distinguish between freed and freeborn among the so-called *incerti* in the abundant epigraphic evidence. The most important is a rule of thumb that a free person of indeterminate status who has a Greek *cognomen* can be assumed to be freed.[2] This use of names as an index of status is grounded in the observation that persons of freed status are much more likely to bear a Greek *cognomen* than

This research was funded by the Leverhulme Trust. I am grateful to the editors and readers for their comments and suggestions.

[1] A full Roman name, as was required for example in declarations for the citizen census (CRAWFORD 1996 no. 24, 146), distinguishes the freed from the freeborn by the use of libertination and filiation respectively. In the Republican period, names were often given in full on epitaphs and in many other epigraphic genres, facilitating the social analysis of these texts. Over the course of the first century CE, however, it became increasingly common to abbreviate names, omitting any or all of *praenomen*, tribe and crucially filiation/libertination. LILY ROSS TAYLOR has sketched the contours of this development for epitaphs in the city of Rome by noting the decline in the prevalence of filiation/libertination from around 93 % in epitaphs dated to the Republican period to around 66 % in the predominantly Julio-Claudian epitaphs in the necropolis outside the Porta Salaria, to around 20 % in the epitaphs in the necropoleis under St Peters and at Isola Sacra in Portus, most of which date to the second century or later (TAYLOR 1961: 118–20).

[2] In cases where the *incerti* have a single name only and may include enslaved persons, a Greek name is taken to indicate either freed or enslaved status. To simplify the presentation, I focus on the case where the *incerti* are known to be free (either from the context or because they have a *gentilicium*). But all the same objections apply to the interpretation of a single Greek name as implying that the bearer was either freed or enslaved. They also apply to a variant of the rule of thumb that treats a Greek name as a sign of freed status *or freed descent*, as I show at the end of this paper.

are the freeborn. In one form or another, the rule of thumb underpins much of what passes for knowledge about the role of freed persons in Italian society in the imperial period. Reservations have occasionally been expressed on empirical grounds. Several scholars have pointed out that the rule is not infallible: there are freeborn persons with Greek *cognomina*, and freed slaves with Latin ones.[3] But its best advocates have always cautioned that the rule is only reliable for gauging the overall composition of a group, not for determining the status of individual persons. It is a "statistical" tool, as several have remarked.[4]

The first goal of this paper is to show that there is a more profound problem that renders the rule invalid even when applied to groups. Part I shows that the rule of thumb entails a basic error of reasoning. Even if the freed were much more likely to have a Greek *cognomen* than the freeborn, it does not necessarily follow that most persons with a Greek *cognomen* were freed – as the rule implicitly assumes. The likelihood that a person with a Greek *cognomen* was of freed status would depend on the prevalence of freed status in the population in question. To assume that persons with Greek names are freed in order to gauge the prevalence of freed status is to engage in circular reasoning.

My purpose is not, however, to reject the use of onomastics to study the presence of freed slaves. The second goal is to show that it is possible to proceed directly from the overall onomastic profile of a group to an estimate of the prevalence of freed status in the group, without making any assumptions about individuals, as long as there is independent evidence for the status-specific prevalences of Greek *cognomina* among the freed and freeborn members of the group. But an entirely different method is required, involving a formal mathematical model of the relationship between names and status. Precisely because the significance of names is a matter of "statistics", a statistical approach is required if we are to make valid inferences. Part II of the paper explains this alternative method and illustrates its application to one especially well-documented case, the *collegiati* of Ostia in the late second and early third centuries CE. I will argue that we can arrive at a rough estimate for the prevalence of freed status – around 21 % – but that any individual *collegiatus* with a Greek *cognomen* is as likely to be freeborn as freed.

The method requires good data to produce reliable results. It also requires an overtly mathematical approach that might discomfit some ancient historians. So it is worth insisting at the outset that this is the only way to draw any conclusions based on onomastics. If the data seem inadequate, or if one is unwilling to venture a calculation, then one cannot legitimately conclude anything at all about the status of individual persons. The only valid alternative is *aporia*.

This is part of a larger project to study the scale of manumission in Roman Italy. This paper is uncompromisingly mathematical in its approach: the freed appear only as a statistic – some proportion of the population – while the enslaved are almost entirely absent. This is necessary because statistical analysis is often the only way we can hope to see the freed in our evidence. In the absence of formal markers of status, it may be impossible to detect them as individuals; their presence can only be inferred from the onomastic profile of the whole. We see, as it were, a shadow cast by a mass of individuals who are otherwise hidden from us. Measuring their presence is a crucial prerequisite to writing an adequate history of manumission – or indeed of slavery. Without a proper sense of scale, we cannot hope to understand the importance of freed men and women in Roman society or to understand the experience of the enslaved.

[3] For example, the wide-ranging discussion in BRUUN 2013 concludes by questioning the categorical validity of the rule: "I believe that we are far from justified in claiming that a Greek *cognomen* in each and every case is a safe indication of slave or freedman status" (35).

[4] See especially ANDREAU 1974: 152 ("une vérité statistique"), DUTHOY 1989: 198 ("Ce critère n'a qu'une valeur statistique et ne peut donc servir pour régler un cas individuel isolé."), MOURITSEN 2011: 126 ("a statistical tool"; "the status of any single individual cannot be ascertained solely on onomastic criteria.").

For now, however, my focus is on the problem of method and on the indispensability of statistical reasoning in ancient history.

I. THE CONVENTIONAL APPROACH

Names and juridical status

It is now well-established for a wide range of geographic, chronological and social contexts in late republican and early imperial Italy that persons of enslaved or freed status were much more likely to have a *cognomen* of non-Latin origin, usually Greek, than were the freeborn. See Table 1, which reports what I term the status-specific prevalences of non-Latin *cognomina* among the freed/enslaved and the freeborn respectively for nine different datasets.[5] All show a pronounced difference in the status-specific prevalences. The explanation is that enslaved persons were often given Greek names. The distinctiveness of slave onomastics is a visible legacy of one of the many techniques of symbolic as well as physical violence that underpinned the system of slavery: the master's control of the slave's name. As KYLE HARPER has observed, this act of existential domination was a symbol of mastery.[6]

Table 1. Empirical evidence for the differential use of non-Latin cognomina across Italy

		Prevalence of non-Latin *cognomina* (%)				
		Freed or enslaved		Freeborn		
Region	Dataset	%	n=	%	n=	Notes
Rome	All evidence (SOLIN)	69	27,509	12	2,650	SOLIN 2001 and SOLIN 1971. The data for freed/enslaved is based on an exhaustive study of Roman inscriptions and literary texts. The data for freeborn is based on all inscriptions in *CIL* VI, excluding persons of senatorial rank.
	Sample of epitaphs (KAJANTO)	72	635	23	176	KAJANTO 1968. Data based on *sepulcrales* in the two supplementary volumes *CIL* VI, 4, 2–3; the figure for freeborn excludes senators and soldiers. See also n. 19.
	magistri vicorum, 136 CE (*CIL* VI 975)	56	219	29	34	My analysis.*
Latium & Campania	Ostia: dedicators and dedicatees of tomb buildings	63	433	21	111	My analysis of data in MOURITSEN 2004, based on certain cases only (see n. 31) and excluding slaves.*
	Herculaneum: Album	53	284	21	56	My analysis of the names as read by CAMODECA 2008b.*
South	Beneventum: all inscriptions	55	80	9	53	KAJANTO 1968.
	Canusium: epitaphs	57	~200	14	~140	CHELOTTI et al. 1990.
North	All inscriptions from six cities	54	382	14	381	KAJANTO 1968 (Mediolanum, Aquileia, Verona, Cremona, Florentia, Parma).
	All inscriptions from fourteen cities	46	1,374	6	461	Calculated from data in MANSUELLI 1962 (a composite dataset with data of uneven quality), covering Verona, Ateste and twelve others.

[5] The table draws heavily on MOURITSEN 2013: 123–31, the most important treatment of the subject to date. Names that are neither Latin nor Greek in origin are very rare in Italian epigraphy, but also seem slightly more common among slaves and freedmen. HEIKKI SOLIN observed rates of 1.8 % and 1.1 % among slaves/freedmen and freeborn respectively at Rome (SOLIN 2001: 309 for freed/slaves and SOLIN 1971: 124 for the freeborn). Hence they are usually combined with Greek names into a larger "non-Latin" category, as in Table 1.

[6] HARPER 2011: 226–7

* See n. 28 below for the rationale for re-analysing these datasets (to ensure comparability with my counts for the the Ostian *collegiati*) and the criteria used.

The figures in Table 1 deserve two comments. First, there is no single right way of calculating these statistics, which means that one should not expect two scholars analysing the same set of names to arrive at precisely the same results. One potential source of divergence is the criteria used to determine who is "certainly" freed or "certainly" freeborn. The identification of "certain" cases often relies not just on filiation/libertination, but also on a number of other markers that can identify a person as freed (e.g. the use of terms like *libertus*, *collibertus* or *patronus,* or *verna* or *conservus* when applied to a free person) or freeborn (the rank of decurion or *eques Romanus*). Scholars vary in the criteria they use, and occasionally apply other, less conclusive criteria as well.[7]

Scholars can also diverge in the classification and counting of *cognomina*. The category of "non-Latin" (or "Greek") names is just a heuristic device. It is an attempt – for the purpose of establishing a link between names and statuses – to approximate a distinction that seems to have been meaningful in Roman society. While many names can easily be classified, others pose problems:[8] names based on loanwords long-established in Latin (e.g. Hilarus or Leo[9]), other names of foreign origin but long established in Roman usage (e.g. Hermes or Grypus[10]) and names derived from ethnonyms (e.g. Bithynus or Numida[11]) – to name just a few. There are two underlying sources of difficulty. The first is that we are almost entirely dependent on etymology alone to predict judgements that contemporaries would have made based on a wider range of considerations including contemporary usage. The second is that we are trying to impose a binary distinction ("Latin"/"Non-Latin") on what was surely a continuum. A proper account of how names were actually experienced would have to use polythetic or "fuzzy" categories to manage the borderline cases. The binary distinction is still useful in revealing patterns, but it is important to recognise that it is a simplification and that scholars can legitimately differ in how they classify individual cases within a binary scheme. Further problems arise in counting: what to do about persons who do not have a *cognomen*, *cognomina* that take the form of a *gentilicium*, persons with multiple *cognomina*, otherwise unattested forms and fragmentary names. Again, there is no single correct solution to any of these issues.

The possibility of legitimate divergence in the criteria of classification means that calculations of the prevalence of non-Latin names in any sample are liable to vary by several percentage points depending on the scheme used. The scope for variation is small relative to the pronounced differences between the freeborn and freed/enslaved across all the datasets in Table 1. There is no question of those differences being an artefact of how names are counted. But it is a potential complication when making comparisons between datasets.

The second feature of Table 1 that deserves note is the variation in the two status-specific prevalences across the datasets. Some of this may be due to different scholars using different principles of classification. There will also be some random variation in what are often relatively small samples. But it seems clear that onomastic patterns varied significantly from context to context. The observed differences probably reflect a complex interaction of geographic, chronological and social factors. It is plausible that onomastic patterns varied from place to place. Rome has produced relatively high rates of non-Latin names among slaves and freedmen, while northern Italy shows relatively low rates (though the variation may also reflect differences

[7] See e.g. n. 16 below on KAJANTO'S idiosyncratic criteria for identifying the freeborn.
[8] SOLIN 1971: 48–85 is a thorough discussion of the difficulties.
[9] SOLIN 1971: 56 considers both Latin, but argues that there is no general rule: Mimus is Latin but Tragicus Greek (58).
[10] SOLIN 1971 considers Hermes as Greek (49), but Grypus as Latin, at least in aristocratic circles (71–2).
[11] It is conventional to classify names based on ethnonyms or toponyms from the East of the empire as Greek, but those from the west as Latin – even though they are often derived from local languages.

in the chronological distribution and social profile of the samples). A second plausible source of variation is diachronic change. It is very difficult to measure this, given the difficulty in dating most inscriptions, but we can glimpse some developments. For example, HEIKKI SOLIN inferred that Greek names became progressively more common in the freeborn population of the city of Rome in the first and especially second centuries CE.[12]

Even at a given time and place, there will have been some social variation in the status-specific prevalences. For example, the rate of non-Latin names among the freeborn seems to have declined toward the top of the social hierarchy. In Herculaneum, non-Latin names are much rarer among the curial elite (~2.5 %) than among the freeborn males listed on what appears to have been an album of *municipes* and *incolae* (21 %).[13] Similarly low rates can be observed among the magistrates and *curiales* of other cities and among the senatorial elite.[14] Conversely, non-Latin names may have been somewhat more common among the freeborn children of freed parents than among the rest of the freeborn population, though we lack a reliable comparative analysis.

There may well have been similar variation within the enslaved/freed population. It is noteworthy that, in HEIKKI SOLIN'S sample of names from the city of Rome, non-Latin *cognomina* are significantly less common among those described as home-born (*vernae*) (43 %; n=575) than in the total population (69 %; n=27,509).[15] If favoured *vernae* were more likely to be given a Latin name and were also more likely to be freed (as seems plausible), then the prevalence of non-Latin names would be somewhat lower among the freed than among the enslaved. It follows that the sub-populations of the enslaved and the freed might have had different onomastic profiles, though they are usually conflated for the purpose of onomastic analysis.

Social and chronological variation probably contribute to the surprising divergences between the three Roman samples in Table 1. Non-Latin *cognomina* are significantly less common among the freed *magistri vicorum* of 136 CE (56 %; n=219) than among the composite population of freedmen and slaves that appear in Roman inscriptions (69 % in SOLIN'S exhaustive corpus, including attestations in literary texts; 72 % in KAJANTO'S sample of Roman epitaphs). This may illustrate some divergence between the freed and the enslaved of the sort I just suggested and/or chronological change (since most of the certain *liberti* in the epigraphic record are much earlier than the second-century list).

The other striking discrepancy concerns the prevalence of non-Latin names among the freeborn: 12 % (n=2,650) in SOLIN'S analysis of inscriptions in *CIL* VI, but 23 % (n=176) in KAJANTO'S much smaller sample of epitaphs alone. KAJANTO'S figure may be inflated slightly by the erroneous inclusion of some freed persons due to an idiosyncrasy in his classification criteria.[16] Some of the residual discrepancy may be due to random variation in his relatively small sample. But it probably also reflects social variation across the two samples. SOLIN'S category of *ingenui* include soldiers, who appear to have been less likely to have a non-Latin *cognomen* than the civilian population, while KAJANTO'S excludes them. KAJANTO'S sample, being based on epitaphs alone, may also have a different social profile than SOLIN'S, which includes other types

[12] SOLIN 1971: 96–7 and 124–5.
[13] Curial elite: CAMODECA 2008a: 162. *Album*: Table 1 with WALLACE-HADRILL 2015 on the nature of the list.
[14] MOURITSEN 2011: 125–6. JONGMAN 1988: 327 observes a similar differentiation within the *ordo* of Canusium in 223 CE (*CIL* IX 338), with the prevalence of Greek names decreasing with rank from 32 % among the *pedani* to 9 % among the *quinquennalicii* (and 0 % among the senatorial patrons).
[15] Statistics from SOLIN 1971: 156 and SOLIN 2001: 309 respectively. BRUUN 2013 draws attention to the importance of this finding. On *vernae*, see further HERRMANN-OTTO 1994.
[16] KAJANTO considered *incerti* whose father had a *gentilicium* (and hence were free) to be certainly freeborn (521). The index is far from conclusive, since a freed person could have a freed or even freeborn father; their status at birth would have been determined by their mother's status, not their father's. This is likely to have led to the inclusion of some freed persons among KAJANTO'S "*ingenui*", though I suspect the numbers are relatively small

of inscription. It is notable that the rate of 29 % among the freeborn *magistri vicorum* (n=34) is much closer to that KAJANTO'S sample. To be frank, we are not yet in a position to give a conclusive explanation of the divergences. But they illustrate the caution required in handling the available onomastic data. They also demonstrate how much work remains to be done, even for the city of Rome.

Inferring the presence of freed slaves

Despite these caveats, there can be no doubt that freed persons were significantly more likely to carry a non-Latin *cognomen* than their freeborn peers across Italy in the late republican and early imperial period, even if the precise ratio varied from context to context. The question is how this insight can be used to determine the status of individual *incerti* or to gauge the prevalence of freed status in a group of *incerti*.

It is regularly asserted that *incerti* with a non-Latin *cognomen* can be assumed to be of freed (or enslaved) status.[17] Thus a non-Latin name has become the most important of a number of markers used to identify free persons of unspecified status as freed slaves. Others include having one of a sub-set of Latin *cognomina* identified as characteristically "servile", serving as an Augustalis (an office supposedly monopolised or at least dominated by freedmen), being enrolled in the *tribus Palatina*, or having a spouse with the same *gentilicium* (taken to be characteristic of persons freed from the same household).[18]

These rules of thumb also underpin almost all attempts to date to assess the presence of freed persons in a sample of *incerti*. These attempts generally take one of two forms. The first and more common merely reports the prevalence of non-Latin *cognomina*, and sometimes some of the other supposed indices, without any explicit interpretation – as if the meaning were obvious. The implicit assumption seems to be that the prevalence of non-Latin *cognomina* is a minimum estimate for the freed presence (a minimum because there will have been other ex-slaves who did not have a non-Latin *cognomen*). The reasoning is made explicit in studies of the second type, which do venture a formal quantification of the freed presence. The usual approach is to define a set of indices of freed and freeborn status respectively. Freed *incerti* are identified as such by possession of a non-Latin *cognomen* and some selection of other markers; freeborn *incerti* are identified by a different set of indices such as military service or holding a magistracy (both of which normally required free birth). The prevalence of freed status is then estimated by comparing the number of persons identified as freed to those identified as freeborn (usually ignoring the *incerti* who cannot be assigned a status).[19]

There are two fundamental problems with this approach. The first is a vulnerability to selection bias. There is good reason to believe that the freed are more likely to be identifiable as such than the freeborn, and hence will be over-represented among persons who can be assigned a status. For example, there are a wide range of relational terms can conclusively identify a free person as freed: *patronus*, *libertus*, *collibertus*, *verna*, *conservus*, etc. But there are no corresponding terms that can identify a person as freeborn since terms such as *pater*, *mater* and *filius* were demonstrably used of those born in slavery as well as those born free. There is a similar asymmetry among the usual indirect indices of freed status. No one would dream of identifying an *incertus* as freeborn merely because the *cognomen* is Latin

This is a major problem for estimates that focus on a subset of persons who can be assigned a status, discounting the rest. That said, the problem could be addressed by keeping the indeterminate cases in the

[17] See e.g. ABRAMENKO 1993: 17, CAMODECA 2008b: 91 and MOURITSEN 2005: 41. MOURITSEN 2011: 126 is more cautious: "it represents a statistical tool and that the status of any single individual cannot be ascertained solely on onomastic criteria."

[18] MOURITSEN 2013: 126–7 is the best discussion of these indices of freed status.

[19] See e.g. TAYLOR 1961 on *incerti* in Roman funerary epigraphy, JONGMAN 1988: 271 on witnesses' names on the Iucundus tablets from Pompeii, ABRAMENKO 1993: 16–21 on *Augustales* and *seviri* in Italy and the provinces and MOURITSEN 2004 on dedicators and dedicatees of tomb buildings at Ostia.

denominator and treating the resulting estimate as a minimum. But there is a second and fatal weakness in the traditional approach. As I will show, assuming that persons with non-Latin *cognomina* are freed in order to estimate the prevalence of freed status involves circular reasoning. Exactly the same problem applies to most of the other putative indices of status, such as having a supposedly "servile" Latin *cognomen* or a spouse with the same *nomen*, since they depend on similar probabilistic reasoning. But I will focus on the example of non-Latin *cognomina* since they are the keystone of the conventional method.

Conditional probabilities in ancient history

It is impossible to understand and hence avoid the error in the traditional approach without a modicum of probability theory and some mathematical notation. To simplify the presentation, I will henceforth use "Greek" as a short-hand for "non-Latin" when referring to *cognomina*. In mathematical notation, *p(X)* denotes a probability. The prevalence of any quality, such as freed status, in any group can be described in terms of probability. *p(Freed)*, the probability that a randomly selected member of a group will be freed, is nothing other than the prevalence of freed status in the group in question. *p(Freeborn)* is the prevalence of freeborn status. If the group in question is composed of free persons only,

$$p(Freed) = 1 - p(Freeborn)$$

In other words, the two probabilities or prevalences always sum to 100 %.

The prevalence of Greek and Latin *cognomina* can also be described as probabilities, *p(Greek)* and *p(Latin)*, which also sum to 100 %.

The question at hand requires the additional concept of conditional probability. A conditional probability is a probability that takes the form "the probability that A is true given that B is true". In mathematical notation, this is represented as *p(A|B)*. The status-specific prevalences of Greek *cognomina* can be described as conditional probabilities. *p(Greek|Freed)* is the probability that a randomly selected person will have a Greek name *if we observe that they are freed*. That probability is just the prevalence of Greek *cognomina* among the sub-group who are of freed status. Similarly *p(Greek|Freeborn)* is just the prevalence of Greek names among the freeborn.

The situation we face is one in which we can observe *p(Greek)*, the prevalence of Greek *cognomina*, in a particular sample and we want to infer *p(Freed)*, the prevalence of freed status, in that sample. Considerable empirical work has demonstrated that freed persons are much more likely than freeborn persons to have a Greek *cognomen*. Expressed in terms of conditional probabilities, this means that *p(Greek|Freed)* is usually much higher than *p(Greek|Freeborn)*. To be confident in identifying someone who has a Greek *cognomen* as freed, however, we would need to know a different conditional probability, namely *p(Freed|Greek)*. This is the probability that someone who has a Greek *cognomen* is of freed status – or, in other words, the prevalence of freed status among the sub-set of persons with Greek *cognomina*. The two conditional probabilities are entirely distinct. *p(A|B)* and *p(B|A)* are termed inverse probabilities. Research into the psychology of probabilistic reasoning has shown that people often confuse conditional probabilities, implicitly assuming that p(A|B) and p(B|A) must have similar magnitudes – though they need not. This tendency has been called "confusion of the inverse".[20] The fact that *p(Greek|Freed)* is high does not in itself prove that that *p(Freed|Greek)* is also high.

[20] O'HAGAN et al. 2006: 44–5.

The relationship between p(A | B) and its inverse p(B | A) is given by Bayes' rule, one of the most important theorems in probability, given here in a simple and expanded form:[21]

$$\mathrm{p}(\mathrm{A} \mid \mathrm{B}) = \frac{\mathrm{p}(\mathrm{B} \mid \mathrm{A}) \cdot \mathrm{p}(A)}{\mathrm{p}(B)} = \frac{\mathrm{p}(\mathrm{B} \mid \mathrm{A}) \cdot \mathrm{p}(A)}{\mathrm{p}(\mathrm{B} \mid \mathrm{A}) \cdot \mathrm{p}(A) + \mathrm{p}(\mathrm{B} \mid \mathrm{A}') \cdot \mathrm{p}(A')}$$

Applied to the question at hand, the expanded this gives:

$$\mathrm{p}(\mathrm{Freed} \mid \mathrm{Greek}) = \frac{\mathrm{p}(\mathrm{Greek} \mid \mathrm{Freed}) \cdot \mathbf{p}(\mathbf{Freed})}{\mathrm{p}(\mathrm{Greek})}$$

$$= \frac{\mathrm{p}(\mathrm{Greek} \mid \mathrm{Freed}) \cdot \mathbf{p}(\mathbf{Freed})}{\mathrm{p}(\mathrm{Greek} \mid \mathrm{Freed}) \cdot \mathbf{p}(\mathbf{Freed}) + \mathrm{p}(\mathrm{Greek} \mid \mathrm{Freeborn}) \cdot \left(1 - \mathbf{p}(\mathbf{Freed})\right)}$$

I present this formula merely to observe that, as a matter of mathematical logic, it is impossible to proceed from *p(Greek | Freed)* and *p(Greek | Freeborn)* to *p(Freed | Greek)* without also knowing *p(Freed)* – the prevalence of freed status in the sample (shown in bold above). There is no valid chain of reasoning that moves from the observation that Greek *cognomina* are much more common among the freed than the freeborn to the conclusion that persons with Greek *cognomina* are likely to be freed without knowledge of the prevalence of freed status. The traditional approach is logically indefensible.

Since Bayes' rule can seem opaque on first acquaintance, it may be useful to illustrate the problem with an example. Consider two hypothetical populations, A and B. Both populations are characterised by a sharp onomastic divide of the sort observed in peninsular Italy: the prevalence of Greek *cognomina* is 60 % among freed persons and just 20 % among the freeborn. The hypothetical populations differ in the prevalence of freed status. Freed slaves make up 90 % of Population A and only 10 % of Population B.

Table 2. Hypothetical population A

	Persons by *cognomen* type			Prevalence of Greek *cognomina*	Corresponding probability
	Greek	Latin	Total		
Freed	540	360	900	60 %	*p(Greek | Freed)*
Freeborn	20	80	100	20 %	*p(Greek | Freeborn)*
Total	560	440	1000	56 %	*p(Greek)*
Prevalence of freed status	96 %	82 %	90 %		
Corresponding probability	*p(Freed | Greek)*	*p(Freed | Latin)*	*p(Freed)*		

Table 2 shows the population structure for Population A and calculates the relevant prevalences. Because we have perfect knowledge, we can observe *p(Freed | Greek)* directly. It is just the prevalence of freed status among persons with a Greek *cognomen*, i.e. 540/560 or 96 %. It can also be calculated using the expanded form Bayes' rule, in order to illustrate its application:

$$\mathrm{p}(\mathrm{Freed} \mid \mathrm{Greek}) = \frac{\mathrm{p}(\mathrm{Greek} \mid \mathrm{Freed}) \cdot \mathrm{p}(\mathrm{Freed})}{\mathrm{p}(\mathrm{Greek} \mid \mathrm{Freed}) \cdot \mathrm{p}(\mathrm{Freed}) + \mathrm{p}(\mathrm{Greek} \mid \mathrm{Freeborn}) \cdot (1 - \mathrm{p}(\mathrm{Freed}))}$$

[21] In the expanded form, *p(A′)* is the probability that A is false. Hence *p(A′) = 1– p(A)*. In the present case, *p(Freed′) = p(Freeborn)*. The expanded form distinguishes the two conditions under which one could have a Greek *cognomen*: being freed or being freeborn.

$$= \frac{60\,\% \cdot 90\,\%}{60\,\% \cdot 90\,\% + 20\,\%.10\,\%} = 96\,\%$$

The relatively high value for *p(Freed|Greek)* means that a person in Population A who has a Greek *cognomen* is indeed likely to be freed.

The situation is very different in Population B (Table 3). Because the freeborn are significantly more numerous, the minority who have Greek *cognomina* outnumber the majority of freedmen who have a Greek *cognomen*. Hence *p(Freed|Greek)* is just 25 %. Only a quarter of persons with Greek *cognomina* are freed. In this population, it would be a serious mistake to treat a Greek *cognomen* as an index of freed status.

Table 3. Hypothetical population B

	Persons by *cognomen* type			Prevalence of Greek *cognomina*	Corresponding probability
	Greek	Latin	Total		
Freed	60	40	100	60 %	*p(Greek|Freed)*
Freeborn	180	720	900	20 %	*p(Greek|Freeborn)*
Total	240	760	1000	24 %	*p(Greek)*
Prevalence of freed status	25 %	5 %	10 %		
Corresponding probability	*p(Freed|Greek)*	*p(Freed|Latin)*	*p(Freed)*		

The point of these two hypothetical examples is just to illustrate the more general point revealed by Bayes' rule. Even if Greek names are much more common among the freed than among the freeborn, one cannot be confident that a Greek *cognomen* is a good index of freed status – i.e. have a high value for *p(Freed|Greek)* – without knowing the prevalence of freed status in the relevant group. It is entirely possible that most free persons with Greek names were freeborn.

II. A MODEL-BASED APPROACH

It is possible to estimate the prevalence of freed status on the basis of onomastic data, but an entirely different method is required. The conventional approach involves first imputing the status of individual persons and then calculating the proportion who were probably freed. All such calculations are invalidated by the circular reasoning that is unavoidable in the first step. There is no valid way to impute individual status without knowing the prevalence of freed status in the relevant population. But it is possible to proceed directly from the prevalence of Greek *cognomina* to an inference about the prevalence of freed status without making any assumptions about individuals. If we know how common Greek *cognomina* are among freed and freeborn persons respectively, then the proportion of Greek *cognomina* in a sample of persons of mixed status is a direct index of the prevalence of freed status in that sample.

This is an example of a more general inference problem that has been termed "compositional analysis":[22] "The problem of apportioning composite material to original sources is a well-established yet difficult problem. … We do not observe directly the proportions of each source comprising the mixture samples. Rather, we obtain measurements of … features of the … samples which we call markers. Our task is to infer the relative proportions of each source within a mixture sample using these marker measurements. We use the term compositional analysis to describe methods of estimating these proportions. For our method, we require marker data not only on the mixture samples but also on "pure" samples from each source." In the present case, there are two "sources" (freed and freeborn), one "marker" (the prevalence of Greek *cognomina*) and just three samples (one "mixture sample" and two "pure samples"). It is a relatively

[22] BREWER et al. 2005: 19–20.

simple instance of the species, yet it is still highly complex. I am not aware of any analytical method to compute an expected value or confidence interval for the mixture proportion (i.e. the prevalence of freed status) given the statistics of the three samples. But it is possible to derive a simple linear model that will produce a best estimate of the prevalence of freed status in a sample, given the prevalence of Greek *cognomina* in the sample and in two representative reference samples composed of freed and freeborn persons respectively.

A simple model

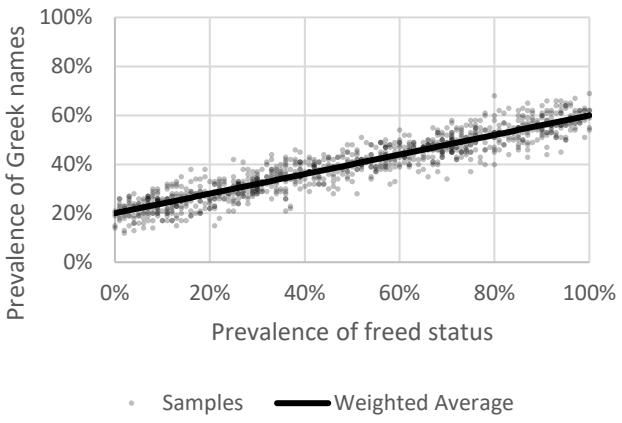
The model-based approach is most easily apprehended by starting with the reverse problem. If we knew the prevalence of freed status in some group but not the prevalence of Greek names, it would be possible to predict the latter if we knew the status-specific prevalences. Take a hypothetical population in which 60 % of freed and 20 % of freeborn persons have a Greek *cognomen*, assuming for the moment that we have perfect knowledge of those parameters. What proportion of Greek *cognomina* would we expect to find in a sample composed entirely of freed persons? The answer is obvious – around 60 %. Conversely, we would expect a sample composed entirely of freeborn persons to show a proportion around 20 %. What about a sample in which freed and freeborn were at parity? Since both groups are represented equally, we would expect the average of the respective prevalences, namely 40 %. To generalise, we would expect the prevalence of Greek names in a mixed sample to be the weighted average of the status-specific prevalences.

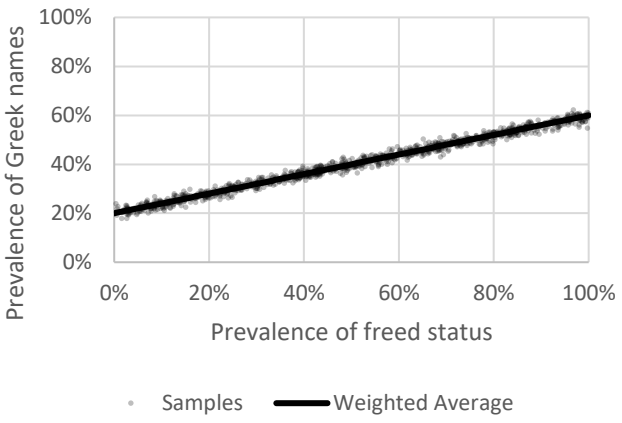
The actual proportion of Greek names would of course vary from sample to sample due to random variation. Figures 1a and 1b illustrate this with simulated data, for samples of 100 and 1,000 persons respectively. Each chart shows the result of a simulation which generated 1,000 random samples of the required size. Each dot represents a sample, showing the prevalence of freed status on the X-axis and that of Greek *cognomina* on the Y-axis. The former was assigned randomly, all values between 0 and 100 % being equally likely. This determined the number of freed and freeborn persons in the sample. Names were then assigned randomly, with each freed person having a 60 % chance of having a Greek *cognomen* and each freeborn person a 20 % chance.[23] The overall prevalence of Greek *cognomina* was then measured, and shown on the Y-axis. The two charts also show a black line representing the weighted average of the status-specific prevalences of Greek names in the population (60 % and 20 %) as the weight of the freed population increases from 0 % to 100 %.

Figure 1. 1,000 samples of 100 persons (Fig. 1a) / 1,000 persons (Fig. 1b) with a randomly determined prevalence of freed status and the required numbers of persons drawn randomly from a population in which 60% of freed and 20% of freeborn persons have a Greek name

a. Samples of 100 persons *b. Samples of 1,000 persons*

[23] This is modelled with two binomial distributions with the appropriate probabilities and number of trials.



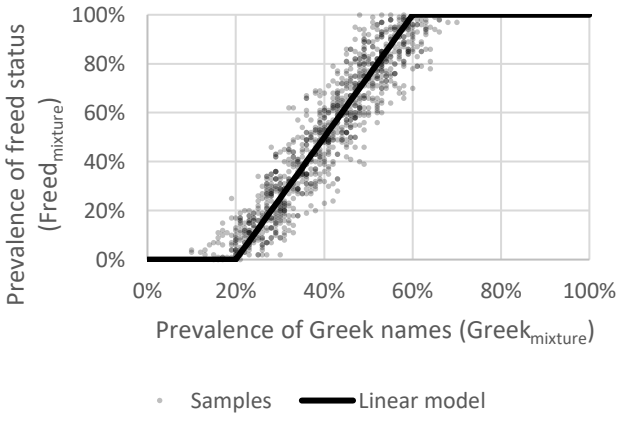


The charts illustrate that the weighted average of the status-specific prevalences provides a simple linear model that could be used to predict the prevalence of Greek names if we knew the prevalence of freed status. They also illustrate that the prediction would always entail a margin of error. The vertical gap between a dot and the black line is the error that we would make if we used the model to estimate the prevalence of Greek names in that particular sample. The errors arise because of random variation in the status-specific prevalences *within* samples, which can diverge from the values of 60% and 20% in the larger population. The scope for divergence and hence the margin of error depends on the sample size, as can be seen by comparing Figures 1a and 1b. The larger the sample, the more precise the prediction can be.

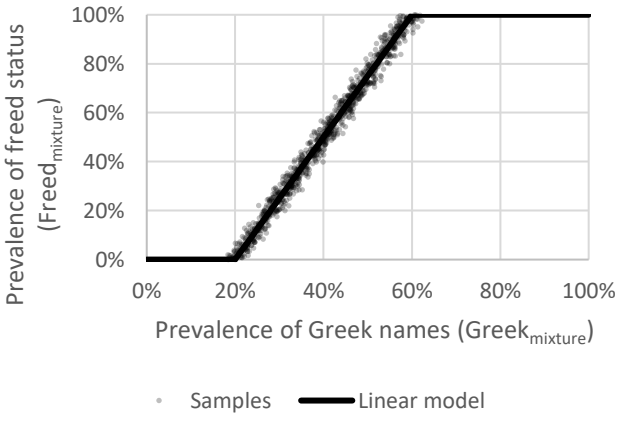
The problem at hand involves the opposite direction of inference: predicting the prevalence of freed status on the basis of the observed prevalence of Greek names. This is more complicated. Figures 2a and 2b present the same simulated data, but with the axes inverted. The X-axis shows what we observe, the prevalence of Greek names; the Y-axis shows what we want to predict, the prevalence of freed status. Again, there is a clear and approximately linear relationship between the two quantities. The simulated samples cluster around a line that is the inverse of the weighted-average function in Figures 1 and 2.

Figure 2. 1,000 samples of 100 persons (Fig. 3) / 1,000 persons (Fig. 4) with a randomly determined prevalence of freed status and the required numbers of persons drawn randomly from a population in which 60% of freed and 20% of freeborn persons have a Greek name

a. Samples of 100 persons



b. Samples of 1,000 persons



This line provides the simplest mathematical model for predicting the prevalence of freed status on the basis of the onomastic data.[24] If $Freed_{mixture}$ is the (uncertain) prevalence of freed status in the mixed sample, $Greek_{mixture}$ is the (observed) prevalence of Greek *cognomina* in the mixed sample, and $Greek_{freed}$ and $Greek_{freeborn}$ are the (observed) status-specific prevalences of Greek *cognomina* in the population from which the sample is drawn:

$$Prediction(Freed_{mixture}) = \frac{Greek_{mixture} - Greek_{freeborn}}{Greek_{freed} - Greek_{freeborn}}$$

for $Greek_{freed} \leq Greek_{mixture} \leq Greek_{freeborn}$; if $Greek_{mixture} < Greek_{freeborn}$, $Prediction(Freed_{mixture}) = 0$ %; if $Greek_{mixture} > Greek_{freed}$, $Prediction\ (Freed_{mixture}) = 100$ %

Note that a complete linear function would predict impossible values when the observed prevalence of Greek names is below 20 % or above 60 %. Hence the need for discontinuities at those points and a z-shaped or piecewise-linear model.

Unlike the conventional method, this calculation makes no assumptions about the status of individual persons. Hence it avoids the circular reasoning entailed in first assuming that persons with a Greek *cognomen* are freed and then estimating the prevalence of freed status. It uses all the data, thus avoiding the problem of selection bias that arises when only a subset of the *incerti* are assigned a status. Finally, the calculation is complete in itself. There is no need to add in the prevalence of other indices of freed status, as is usually done. Indeed, it would be an error to do so. I will return to the question of how additional information could be incorporated.

HENRIK MOURITSEN once described Greek *cognomina* as a "yardstick" for the composition of a group of people.[25] The metaphor is apt. The proportion of Greek names can provide a direct measure of the prevalence of freed status on a scale from zero to 100 %. There is no need to make inferences about individuals first. But there is no single, universal yardstick. It is clear that the status-specific prevalences of Greek names ($Greek_{freed}$ and $Greek_{freeborn}$) varied from context to context. Hence each context requires its own yardstick based on the onomastic pattern in the particular group in question. One cannot simply compare proportions of Greek names across different contexts. For example, an observed proportion of 50 % Greek names might imply a very different prevalence of freed status in northern Italy than it would in the city of Rome (cf Table 1).

The simple model requires three caveats. First, the linear structure is a simplification that leads to bias when the predicted prevalence approaches zero or 100 %, because the scope for error becomes asymmetric. The problem can be observed visually in Figures 2a and 2b. As before, the vertical gap between a dot and the black line is the error that we would make if we used the model to estimate the prevalence of freed status. It is fairly obvious that the model is an unbiased predictor of $Freed_{mixture}$ for middling values of $Greek_{mixture}$: the dots are evenly distributed above and below the black line. As the predicted value approaches 0 % and 100 %, however, the distribution becomes asymmetric. This asymmetry makes the linear model a biased

[24] I am certainly not the first to think along these lines. HENRIK MOURITSEN applied essentially the same calculation in an early note on the prevalence of freedmen in inscriptions from the city of Rome, estimating it at around 82 % (MOURITSEN 1988: 194 n. 228). It is a great pity that he reverted to the traditional approach in his later work (MOURITSEN 2004, MOURITSEN 2011: 123–31), even to the extent of preferring the logically invalid estimate in TAYLOR 1961 to his own, better-founded calculation here (MOURITSEN 2011: 127). Essentially the same calculation was also applied by VERBOVEN 2011: 90–3 to several different datasets (though for some reason a different and erroneous method was used to estimate the presence of freedmen among the Ostian *collegiati*). The simple linear model should be fairly obvious to anyone who considers the problem. What is surprising is that no one has seen that this – or something more complex – is the only valid way to interpret the onomastic data.

[25] MOURITSEN 2011: 126.

estimator at extreme values. A completely unbiased estimator would be an asymptotic function of $Greek_{mixture}$. In other words, it would curve as it the predicted value approaches zero or 100 %. But this would require a far more complex statistical model. Since the bias only becomes significant at extreme values and/or in small samples, the linear model is probably adequate for what will only ever be rough estimates.

Second, it does not contain a measure of the margin of error. Moreover, the problem of random variation is larger than I suggested in the thought experiment above. We can never know the status-specific prevalences of Greek *cognomina* in the population. We can only estimate them from other samples. Hence the precision of the estimate will depend not just on the size of the sample of *incerti*, but also on the sizes of the two samples from which the status-specific prevalences are estimated. Ideally, we would compute a confidence interval for $Freed_{mixture}$ given the statistics of the three samples, but I am not aware of any analytical method to do so. As noted earlier, the statistics of compositional analysis problems are highly complex.

The margin of error could in principle be reduced by incorporating additional evidence from one or more independent indices. The most obvious candidate is the prevalence of so-called "servile" Latin *cognomina*. Scholars have long supposed that there is a sub-set of Latin *cognomina* that are typical of slaves – names such as Hilarus, Faustus, Fortunatus and Ianuarius – and can, like Greek *cognomina*, function as an index of freed/enslaved status. All the same objections apply to the notion of using these names to infer the status of individuals *before* establishing the composition of the population. Yet the overall prevalence of such *cognomina* could in principle provide another yardstick for the mix of freed and freeborn. There would be no question of adding the prevalences of Greek and these "servile" Latin *cognomina*, as is usually done. Rather the two prevalences would provide the basis for two independent estimates of the prevalence of freed status. Combining them would require a more complex statistical model, since they would never deserve exactly the same weight in the analysis. The effect would be to refine the estimate and reduce the margin of error around it. Unfortunately, however, the Latin *cognomina* commonly identified as "servile" do not in fact appear to be reliably more common among the freed than among the freeborn (see Appendix). It is possible that future work will be able to identify a set of Latin names that can provide a reliable additional index of the freed presence. For the moment, however, Greek names must stand alone.

The third and final caveat is that the accuracy of the estimate will also depend on the representativeness of the reference populations used to estimate the status-specific prevalences. The best available reference data will often be from a different period, a different social context or even a different city than the *incerti*. An injudicious choice could easily introduce significant error into the estimate.

Applying the model

These caveats should caution against applying the method to small samples, or where there is no good reference data. In such cases, it may be advisable to retreat to a position of total ignorance and forgo making any inferences about the composition of the group or the status of individuals. But there are datasets where the samples are large enough, and the reference populations representative enough, for the method to be expected to produce useful results.

One such example is *collegiati* of Ostia. In the second and early third centuries CE, Ostia's many commercial *collegia* had a habit of commissioning monumental lists of their members. Numerous fragments survive.[26] The lists usually begin with the names of patrons, who were often of high rank and hence not representative of the wider population, and officers, who might also be unrepresentative. But the vast majority of the names are ordinary members (*plebei*). Five particularly large examples, from four different *collegia*, each

[26] See especially *CIL* XIV 246ff, 4560ff and 5354ff.

preserve the names of hundreds of members (Table 4).[27] Another six smaller lists from these and other commercial *collegia* bring the total number of named *plebei* to around 1,400 – an extraodinrarily large sample by the standards of Roman history.

Bearing in mind my earlier remarks about the scope for divergence in the classification and counting of names, I have re-analysed the *collegiati* and all other datasets to which I will compare them in order to ensure consistency.[28] The five larger lists are remarkably consistent in their onomastic profile, with the prevalence of non-Latin *cognomina* among *plebei* varying narrowly within the range 27-31 %. The other smaller lists, taken together, show a similar profile, with a prevalence of 33 %. The overall prevalence in the combined corpus of 1,365 persons with *classifiable* cognomen is 30 %. The consistent representation of non-Latin *cognomina* across several independent datasets suggests that they are representative of a larger population from which the various *collegia* drew their members and in which the prevalence of non-Latin *cognomina* was around 30 %.

Table 4. Lists of members of collegia at Ostia

Collegium	Date (CE)	Persons with non-Latin *cognomina*	Persons with Latin *cognomina*	Prevalence of non-Latin *cognomina*	n=
corporati qui pecuniam ad ampliandum templum contulerunt (*CIL* XIV 246)	140	57	126	31 %	183
lenuncularii tabularii auxiliares Ostienses (*CIL* XIV 250)	152	37	88	30 %	125
lenuncularii tabularii auxiliares Ostienses (*CIL* XIV 251)	192	76	182	29 %	258
fabri tignuarii (*CIL* XIV 4569)	198	96	222	30 %	318
fabri navales Portenses (*CIL* XIV 256)	Late 2nd / early 3rd c.	86	230	27 %	316
Six other lists from commercial *collegia* (*CIL* XIV 252, 4564, 4565, 4573; Bloch 1953 280)	160s-210s	54	111	33 %	165
Total		406	959	29.7 %	1,365

To interpret this figure in terms of the prevalence of freed status requires onomastic data from a reference population of known status. Given the scope for migration to have produced a distinctive onomastic pattern in this major port city, it is essential to have reference data specific to Ostia.[29] We have no data

[27] On these *collegia*, see further MEIGGS 1973: 311–36, TRAN 2006 and ROHDE 2012. On the commercial character of the *corporati qui pecuniam ad ampliandum templum contulerunt*, see must recently BRUUN 2016b: 362.

[28] I define a non-Latin *cognomen* as one that does not appear in the catalogues of Latin *gentilicia* and *cognomina* in SOLIN AND SALOMIES 1994. *Cognomina* that take the form of a *gentilicium* are counted as Latin (see KAJANTO 1963: 20–1). Persons with more than one *cognomen* are counted as non-Latin if any of their *cognomina* are non-Latin. Persons without a *cognomen* are excluded from the calculation. Names are only counted and classified if the editor has restored a complete name. Since there are more short letter combinations that can identify a Greek name (e.g. Ph-, Th-, X-, Z-) than a Latin name, classifying fragmentary names will tend to over-estimate the prevalence of non-Latin names. The boundary between names that can be restored in full and those that cannot is obviously open to debate in some cases. The decision to defer to the editor is intended to remove the scope for my judgement to introduce systematic bias.

[29] On the scale of immigration to Ostia, see MEIGGS 1973: 214-6 and ROHDE 2012: 91 and 262. As it happens, however, the status-specific prevalence of Greek names among the freeborn in the Ostian epitaphs, discussed below, is not very different from that in the Roman epitaphs or on the Herculaneum album (Table 1).

specific to the second- and third-century *collegiati*, but we do have a large corpus of Ostian epitaphs which includes the names of many persons of known status, both freed and freeborn. The chronological distribution of the epitaphs overlap with the *collegia* lists but extend somewhat earlier and later, from the middle of the first century CE into the fourth.[30] The names of dedicators and dedicatees of tomb buildings have been carefully catalogued by HENRIK MOURITSEN (Table 5). Among the dedicators and dedicatees of the Ostian tombs, 63 % of the certainly freed have a non-Latin *cognomen* (n=433), compared to 21 % of the certainly freeborn (n=111).

Table 5. Dedicators and dedicatees of tomb buildings at Ostia[31]

	Prevalence of non-Latin *cognomina*	n=
Enslaved (?)	12.5 %	8
Freed	62.8 %	433
Freeborn	20.7 %	111
Incerti	52.3 %	533
Total	52.9 %	1085

The question is whether the epitaphs are likely to be representative of the status-specific prevalences of Greek names among the *collegiati*. There are two potential sources of error. One is significant diachronic change, given the imperfect chronological overlap between the two samples. The second possible issue is social variation between the population visible in the epitaphs and that visible in the *collegia*.

The best tests of the representativeness of the epitaphs are two independent datasets that are roughly contemporary with the membership lists and should be composed exclusively of slaves and freedmen (Table 6). A list of the members of Ostia's *familia publica* includes the names of 22 slaves and 59 free persons, all or most of whom must have been freed. The prevalence of non-Latin *cognomina* is 63 % overall and also 63 % among the 35 who were certainly freed public slaves (those with the *nomen* Ostiensis).[32] A series of fragments of *fasti* for the town's *seviri Augustales,* who are generally assumed to have been almost exclusively of freed status, preserve the *cognomina* of 76 office-holders from the late second and early third centuries; 63 % have a non-Latin *cognomen*.[33] The close correspondence with the epitaphs (where 63 % of the freed have a non-Latin *cognomen*) suggests that the epitaph data is a good guide to the onomastics of freed slaves in the collegial class in the late second and early third centuries. Unfortunately, there is no comparable

[30] MOURITSEN 2005: 40.

[31] To ensure consistency with the analysis of Ostian *collegiati*, I have recalculated the prevalences from the catalogue of names in MOURITSEN'S appendix using the principles set out in n. 28. My samples are slightly smaller than his because I only count certain cases and because I distinguish the certainly freed from the small number of persons MOURITSEN 2004: 286 identifies as slaves. But the computed prevalences are almost identical: MOURITSEN 2004: 286 reports 63 % (n=486) for *liberti*/*servi* and 21 % (n=129) for *ingenui*.

[32] *CIL* XIV 255. See BRUUN 2008 on the list and its date (542 n. 19).

[33] *CIL* XIV 4560–2 (on which see BRUUN 2016a). The situation is somewhat different in the fragments grouped as *CIL* XIV 4563, generally identified as an *album* rather than *fasti* and recently dated to the early second century by BRUUN 2016a. 43 % of 83 *cognomina* are non-Latin. On the face of it, this might suggest a significantly higher representation of the freeborn at that time (considerably earlier than the *fasti*, on BRUUN'S reasoning). But it is striking that most of the Latin *cognomina* are concentrated on a single fragment, 4563,5 (as noted by MOURITSEN 2007: 289 n. 10). There is a marked disjunction in prevalence of non-Latin *cognomina* between that fragment (29 %, n=48 – like that of the commercial *collegia*) and the rest (63 %, n=35 – almost identical to that observed on the *fasti*). The relationship of that fragment to the rest deserves further study.

evidence with which to test the representativeness of the freeborn in the epitaphs.[34] This is the assumption that requires most caution.

Table 6. Lists of members of two other collegia at Ostia

	collegium	Date (CE)	Persons with non-Latin *cognomina*	Persons with Latin *cognomina*	Prevalence of non-Latin *cognomina*	n=
familia publica	*familia publica* (*CIL* XIV 255)	Late 2nd / early 3rd c.	51	30	63 %	81
	Slaves		16	6	73 %	22
	Ostiens-es/ -ii		22	13	63 %	35
	Other free(d)		13	11	59 %	24
seviri Augustales	"Fasti" (*CIL* XIV 4560-2)	Mostly 190s-220s	76	45	63 %	121
	"Album" (*CIL* XIV 4563)	Early 2nd c.	36	47	43 %	83
	"Album" fragment 5		14	34	29 %	48
	"Album" excluding fragment 5		22	13	63 %	35

If we accept that the epitaphs are representative of the status-specific prevalences of Greek names among the *collegiati*, the simple linear model can be used to estimate the prevalence of freed status among them:

$$Prediction(Freed_{mixture}) = \frac{Greek_{mixture} - Greek_{freeborn}}{Greek_{freed} - Greek_{freeborn}} = \frac{29.6\,\% - 20.7\,\%}{62.8\,\% - 20.7\,\%} = 21.4\,\%$$

Names and individual status

It is only *after* estimating the overall prevalence of freed status in a group that it becomes possible to assess the likelihood that an individual person with a non-Latin *cognomen* was a freed slave – and hence determine whether a Greek name is a useful index of individual status in this context. The required probability, *p(freed|Greek)*, can be estimated by the expanded form of Bayes' rule:[35]

$$\mathrm{p}(\mathrm{Freed} \mid \mathrm{Greek}) = = \frac{\mathrm{p}(\mathrm{Greek} \mid \mathrm{Freed}) \cdot \mathrm{p}(\mathrm{Freed})}{\mathrm{p}(\mathrm{Greek} \mid \mathrm{Freed}) \cdot \mathrm{p}(\mathrm{Freed}) + \mathrm{p}(\mathrm{Greek} \mid \mathrm{Freeborn}) \cdot \big(1 - \mathrm{p}(\mathrm{Freed})\big)}$$

$$= \frac{62.8\,\% \cdot 21.4\,\%}{62.8\,\% \cdot 21.4\,\% + 20.7\,\% \cdot 78.6\,\%} = 45.3\,\%$$

[34] For what it is worth, the prevalence of non-Latin *cognomina* (18 %) is very similar to that observed among the freeborn on the *album* of Herculaneum (21 %, n=53). An even higher proportion can be observed among the *ingenui* on the list of the Roman *magistri vicorum* in 136 CE (29 %, n=34), though the sample is very small.

[35] *p(freed)* here is *prediction(*$Freed_{mixcutre}$*)* from the previous formula, i.e. the estimated prevalence of freed status in the sample; *p(Greek|Freed)* and *p(Greek|Freeborn)*, the status-specific prevalences in the sample, are estimated from the reference data, $Greek_{freed}$ and $Greek_{freeborn}$ in the previous formula.

In case the calculation by way of Bayes' rule seems too compressed, I also set out the implied structure of the population in the same form as for the two hypothetical populations in Tables 2 and 3. Table 7 shows the predicted distribution of names and statuses in a group of 1000 *collegiati*:

Table 7. A model of the structure of the population of collegiati

	Persons by *cognomen* type			Prevalence of Greek *cognomina*	Corresponding probability
	Greek	Latin	Total		
Freed	135	80	214	62.8 %	*p(Greek | Freed)*
Freeborn	163	623	786	20.7 %	*p(Greek | Freeborn)*
Total	297	703	1000	29.7 %	*p(Greek)*
Prevalence of freed status	45.3 %	11.4 %	21.4 %		
Corresponding probability	*p(Freed | Greek)*	*p(Freed | Latin)*	*p(Freed)*		

My best estimate of the proportion of *collegiati* with non-Latin names who were of freed status is 45 %. In other words, a person of unknown status who happens to have a non-Latin name would be about as likely to be freeborn as to be freed. The evidence suggests that a non-Latin name is useless as an index of status for the Ostian *collegiati*.

This illustrates the independence of two questions that are usually conflated. The first question is whether we can use onomastic evidence to estimate the prevalence of freed persons in a group. The answer to that question is a guarded "yes" – provided we have an adequate reference population and are cognizant of the scope for error. The second question is whether we can use a Greek name to determine the status of an individual. The answer to that question is "not necessarily": it depends on the status-specific prevalences of Greek names *and* the prevalence of freed status in that particular context. In the case of the Ostian *collegiati*, I can venture an estimate for the overall presence of freedmen, but I cannot identify individual freedmen. Hence the shadow metaphor that I adduced at the start of this paper. The freed slaves make their presence felt in the overall prevalence of non-Latin names. We cannot see them as individuals, but we can see their collective impact on the onomastics of the group as a whole.

Past estimates and the notion of "freed descent"

Most work on the membership lists is that the *collegia* were dominated by freedmen and their descendants.[36] For example, TENNEY FRANK thought that "at least three-fourths" of the *lenuncularii tabularii auxiliares* of 192 were of "servile descent", a conclusion based on summing together persons with Greek names, "servile" Latin names or imperial *nomina*.[37] JANET DELAINE concluded that "at least two-thirds" of the *fabri tignuarii* of 198 had *cognomina* "which suggest that they were freedmen or of recent freed descent", an observation presumably based on adding the number of Greek and "servile" Latin names.[38] NICOLAS TRAN inferred, apparently on the same basis, that freedmen and their close descendants accounted for around half of the total sample of *collegiati*.[39]

All three scholars based their conclusion on the traditional method of treating names as an index of individual status. Rather than applying the rule of thumb in its strict form, however, they all applied a common variant of it which holds that a Greek name is an index of freed status *or freed descent* (usually left

[36] ROHDE 2012: 140-1 is a rare sceptic.
[37] FRANK 1940: 247; cf 249 and 250–1.
[38] DELAINE 2003: 727 (without any further explanation).
[39] TRAN 2006: 117 (the estimate appears to be based on summing the 25 % with Greek *cognomina* and 15 % with "servile" Latin names).

undefined). This variant might seem more guarded and hence more defensible than the pure form. But it commits the same logical fallacy.

Assume for the sake of argument that it could be demonstrated that, on some definition of "freed descent", the prevalence of Greek names was much higher among freeborn persons of freed descent than in the rest of the freeborn population (those not of freed descent, so defined). It still would not follow that most freeborn persons with a Greek name were of freed descent. That would depend on the prevalence of freed descent (so defined) in the relevant population. It would in theory be possible to estimate that by comparing the prevalence of Greek names to the status-specific prevalences among freeborn persons of freed descent and those not of freed descent (using exactly the same logic as the model for estimating the prevalence of freed status in a mixed population of freed and freeborn persons). But we would rarely if ever have the data to do so. Even on the simplest definition of freed descent – having a freed father – there would scarcely be sufficient categorizable cases (where we know the names and status of both father and child) to establish adequate reference samples. Without knowledge of the prevalence of freed descent (however defined), one cannot assume that most freeborn persons with Greek names were of freed descent.

In a different mode and with characteristic caution, HENRIK MOURITSEN limited himself to observing that the fact that the 26–30 % of the *collegiati* have Greek names implies "a substantial representation of freedmen".[40] He does not explain how he proceeds from the observation to the conclusion. But there is no logically valid route from one to the other except by means of a calculation like that presented above. One could legitimately estimate the presence of ex-slaves among *collegiati* at around 21 % and then choose to report this as "substantial". But replacing a quantitative estimate by a vague verbal descriptor defeats the purpose of the analysis. The conclusion is essentially unfalsifiable. What proportion of freedmen would not be "substantial"? The corollary of un-falsifiability is that it contains little or no information and hence precludes further analysis. Only by making numerical estimates that can be subjected to scrutiny can we hope to make any progress in understanding the scale of the freed population in this or any other population.

CONCLUSION

My inference that around 21 % is the Ostian *collegiati* were freed is just a rough estimate. Though the samples may seem large by the standards of Roman history, they are still relatively small in statistical terms particularly the sample of certainly freeborn persons from the epitaphs. Moreover, there remains significant uncertainty about the representativeness of the epitaph data for the status-specific prevalences of Greek names among the *collegiati*, again particularly for the freeborn. But the evidence base is as good as we are likely to encounter in a problem of this sort. Almost all other cases will involve smaller samples and/or reference data from a context further removed from the *incerti*.

It is perhaps surprising that the lists of Ostian *collegiati* have not figured larger in the history of manumission in the Roman world. Without making any claims about the representativeness of the *collegiati* for the whole free population of Ostia, or about the representativeness of late-second and early-third century Ostia for other Italian cities or other periods, it is hard to deny that the huge dataset gives us our best view – blurred though it may be – of the presence of freed persons in a meaningful segment of the free population of an Italian city. Wariness of formal statistical thinking has perhaps led ancient historians to neglect the importance of the few cases where we do have sufficiently large samples to draw useful conclusions.

In any case, a calculation along these lines is the only way to proceed from the onomastic evidence to any conclusions about the presence of freedmen in a group. It is logically invalid to draw any inferences about individuals before establishing the composition of the group in which they appear. We must discard the

[40] MOURITSEN 2011: 129.

rule of thumb in all its forms (including the variant that treats a Greek name as an index of freed status or freed descent) and any conclusions based on it.

My larger point is a methodological one. I hope to have illustrated the dangers of the hyper-scepticism about "statistics" that is widespread among ancient historians. It is one thing to reject statistical approaches and simultaneously foreswear any opinion on quantitative matters. Principled aporia is a perfectly reasonable response to the inadequacies of our data. But it another thing entirely to reject calculation, yet attempt to form qualitative judgements – to conclude, for example, that persons with a Greek name are "likely" to have been of freed status or that a group has a "significant" presence of freed slaves. Such inferences might seem more cautious than venturing a calculation. But guarded vagueness can hide fundamental errors of reasoning. Like many other subjects of historical analysis, the relationship between names and status is a matter of probability not certainty. Any reasoning about it must respect the laws of probability, even if one prefers to couch one's conclusions in qualitative terms. Probabilistic reasoning is hard and has many pitfalls for the unwary (not limited to the confusion of inverse probabilities).[41] Formal analysis is the best way to avoid them. As the problem of Greek names and freed status shows, the attempt to find a more cautious alternative to formal analysis can lead to errors of reasoning that would easily have been avoided by formal analysis. Historians ignore statistics at their peril.

School of Classics

University of St Andrews

St Andrews

United Kingdom

mpl2@st-andrews.ac.uk

Appendix: "servile" Latin cognomina

It has long been held that there is a subset of Latin *cognomina* that were common among slaves and freedmen but rare among the freeborn.[42] There is no consensus on a precise list, but candidates include Hilarus, Faustus, Fortunatus, Ianuarius and various other names that were demonstrably very common among slaves and freedmen.

The problem is that we do not have as much data on the status-specific prevalences of these names as we do for Greek *cognomina*. One strand of work on the subject has focused on the observation that some names were very common among slaves. With HEIKKI SOLIN'S exhaustive catalogue of the names of slaves and freedmen at Rome, it has become possible to measure the prevalence of different names in that population and to identify the most common.[43] The Latin names Felix, Primus, Hilarus, Faustus, Fortunatus and Secundus all figure in the top ten. But data on slaves alone cannot tell us anything about the prevalence of these names among the freeborn.[44] It is thus impossible to say whether they were significantly more common among slaves and freedmen than among the freeborn.

A second strand has drawn on IIRO KAJANTO'S study of Latin *cognomina*, which compiled around 130,000 names from across the western provinces.[45] KAJANTO himself was doubtful about the notion that some

[41] LINDLEY 2006 is an excellent and accessible introduction to probabilistic reasoning.
[42] See especially FRANK 1916: 692, DUTHOY 1989: 189–98, MOURITSEN 2011: 126.
[43] SOLIN 1996.
[44] So also BODEL 2003.
[45] KAJANTO 1965.

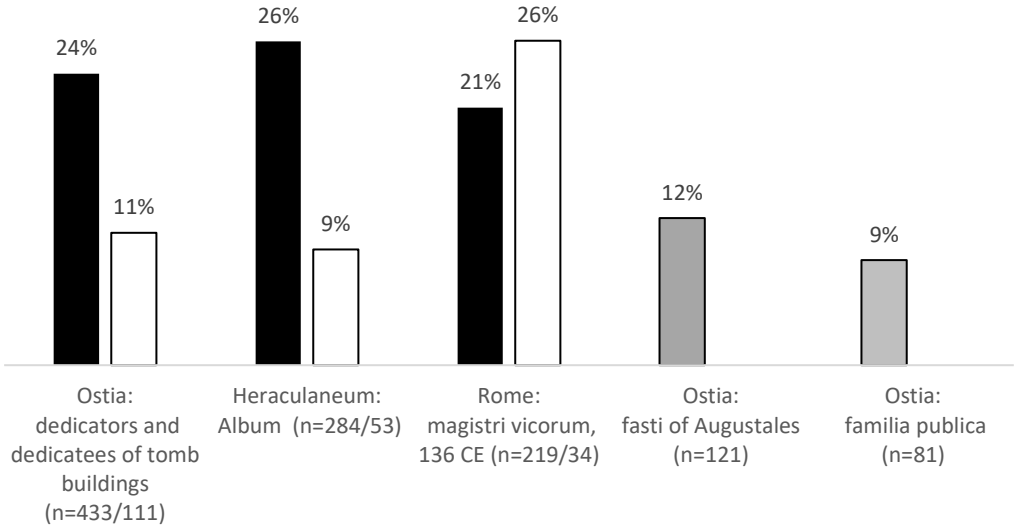
names were typical of slaves, since he observed that since many of the names that were common among slaves and freedmen were also common among the freeborn.[46] But his catalogue of *cognomina* records the number of persons attested with each *cognomen* and notes any who were certainly of freed or slave status. Later scholars have used this information to identify certain names as "servile" on the basis of a particularly high proportion of slaves/freedmen among the attested bearers. But KAJANTO does not report the number of certainly freeborn persons for each name; instead, they are combined with *incerti*. Even if he did, this would still not be enough to compute prevalences, since one would also need to know the total number of persons in each category. Given that the certainly freed often outnumber the certainly freeborn in inscriptions, the raw counts would be deceptive. ROBERT DUTHOY made some progress by cataloguing certain *ingenui* as well as certain slaves and freedmen in Roman and Italian inscriptions, but he only examined a tiny sample of 18 names.[47] And it is still impossible to interpret his data in terms of prevalences, since he has no information about the total numbers of *ingenui* and freed/enslaved persons in his corpora.

The one exception to the lack of data on prevalence is HENRIK MOURITSEN'S analysis of the dedicators and dedicatees of tomb buildings at Ostia.[48] He took KAJANTO'S catalogue as his point of departure, identifying as "servile" any *cognomen* for which KAJANTO recorded an unusually high proportion of slaves/freedmen.[49] He then observed that these *cognomina* were indeed more common among persons who were certainly slaves or freedmen (24 %) than among those who were certainly freeborn (9 %).[50] *Prima facie* this seems to be strong evidence in support of the hypothesis. But this is just a single sample. When I first analysed the Ostian *collegiati*, I hoped to use the prevalence of these supposedly "servile" Latin names as an independent index of the prevalence of freed status. But I calculated their status-specific prevalences in various other datasets and did not find a reliably large difference between the freed and the freeborn.

Figure 3 shows the status-specific prevalences of the *cognomina* that MOURITSEN identified as "servile" for five different datasets.[51] As MOURITSEN observed, there is a striking difference in the names on Ostian tomb-buildings.[52] The names on the Herculaneum album show a similar pattern. In the list of *magistri vicorum* from Rome, however, the supposedly "servile" names are actually more common among the freeborn than the freed. Also surprising are the relatively low proportions in two other Ostian datasets that ought to be predominantly freed or enslaved: the list of members of the *familia publica* and the fasti of the *Augustales*: both show proportions similar to the freeborn named on the tomb-buildings. Some of the sample sizes are relatively small, especially for the freeborn, and there are only a few of them. But the picture looks very different to that for non-Latin *cognomina*, which shows a consistently large difference across a wide range of populations (Table 1).

Figure 3. The status-specific prevalences among the freed (black) and freeborn (white) of Latin cognomina HENRIK MOURITSEN classifies as "servile". The figures in brackets give the sample size for freed and freeborn respectively

[46] KAJANTO 1965: 133–4.
[47] DUTHOY 1989: 189–98.
[48] MOURITSEN 2004.
[49] MOURITSEN does not explain his criteria at MOURITSEN 2004: 284 n. 19, but I assume he used the same rule he describes at MOURITSEN 1988: 62 (names for which the reported proportion of slaves/freedmen is more than twice the average).
[50] MOURITSEN 2004: 286.
[51] See Tables 1 and 6 for the source of the data.
[52] Because I use a slightly different protocol to classify and count names, my figures are slightly different from his. This is a reminder that these categories are a heuristic device that we define for analytical purposes.



It may be that any patterns in the use of Latin *cognomina* were more local, or shorter-lived, than those for Greek *cognomina*. Or it may be that the difference between the status-specific prevalences was too small to provide a sensitive measure. But the problem might lie with the limitations of current notions of which particular names were characteristic of slaves. It may be that current lists conflate names that were common among slaves but rare among the freeborn with others that were common in both groups. It remains possible that the careful analysis of status-specific prevalences in much larger datasets will succeed in identifying a subset of Latin *cognomina* that consistently show a large enough difference in the status-specific prevalences to provide a useful yardstick of the prevalence of freed status. In any case, it would not necessarily follow that most free persons with a 'servile' Latin name were of freed status. As with Greek *cognomina*, that would depend on the composition of the population.

Works cited

ABRAMENKO, A., 1993, Die munizipale Mittelschicht im kaiserzeitlichen Italien: zu Einem Neuen Verständnis von Sevirat und Augustalität.

ANDREAU, J., 1974, Les affaires de Monsieur Jucundus.

BLOCH, H., 1953, Ostia. Iscrizioni rinvenute tra il 1930 e il 1939, NS VII, Ser. 8ª, 239–306.

BODEL, J., 2003, Review of Solin 1996, BMCRev 2003.01.03,

BREWER, M. J. – FILIPE, J. A. N. – ELSTON, D. A. – DAWSON, L. A. – MAYES, R. W. – SOULSBY, C. – DUNN, S. M., 2005, A hierarchical model for compositional data analysis, Journal of agricultural, biological, and environmental statistics 10.1, 19–34.

BRUUN, C., 2008, La familia publica a Ostia antica, in M. L. CALDELLI – G. L. GREGORI – S. ORLANDI (eds.), Epigrafia 2006. Atti della XIVe Rencontre sur l'Epigraphie in onore di Silvio Panciera con altri contributi di colleghi, allievi e collaboratori, 537–56.

BRUUN, C., 2013, Greek or Latin? The owner's choice of names for *vernae* in Rome, in M. GEORGE (ed.) Roman slavery and Roman material culture, 19–42.

BRUUN, C., 2016a, The date of one hundred *Augustales from Roman Ostia in CIL XIV 4563: early second century CE, ZPE 198, 257–70.

BRUUN, C., 2016b, Remembering anniversaries at Roman Ostia: The *dies natalis* of Antinous, hero and divine being, Phoenix 70, 361–80.

CAMODECA, G., 2008a, I ceti dirigenti di rango senatorio equestre e decurionale della Campania romana 1.

CAMODECA, G., 2008b, La popolazione degli ultimi decenni di Ercolano, in M. P. GUIDOBADI (ed.) Ercolano: tre secoli di scoperte, 86–103.

CHELOTTI, M. – MORIZIO, V. – SILVESTRINI, M. (eds.), 1990, Le epigrafi romane di Canosa.

CRAWFORD, M., 1996, Roman statutes.
DELAINE, J., 2003, The builders of Roman Ostia: organisation, status and society, in S. HUERTA (ed.) Proceedings of the first international congress on construction history, Madrid 20–24 January 2003, Vol. II, 723–32.
DUTHOY, R., 1989, Cognomen est omen?, in M.-M. MACTOUX – E. GENY (eds.), Mélanges Pierre Lévêque. Tome 2 : Anthropologie et société, 183–205.
FRANK, T., 1916, Race mixture in the Roman Empire, AHR 21, 689–708.
FRANK, T., 1940, An economic survey of ancient Rome. Volume V. Rome and Italy of the empire.
HARPER, K., 2011, Slavery in the late Roman world, AD 275–425.
HERRMANN-OTTO, E., 1994, Ex ancilla natus: Untersuchungen zu den « hausgeborenen » Sklaven und Sklavinnen im Westen des römischen Kaiserreiches.
JONGMAN, W., 1988, The economy and society of Pompeii.
KAJANTO, I., 1963, Onomastic studies in the early Christian inscriptions of Rome and Carthage, Acta Instituti Romani Finlandiae II.1.
KAJANTO, I., 1965, The Latin cognomina, Commentationes Humanarum Litterarum`` 36.2.
KAJANTO, I., 1968, The significance of non-Latin cognomina, Latomus 27, 517–534.
LINDLEY, D. V., 2006, Understanding uncertainty.
MANSUELLI, G., 1962, La civilisation en Italie Septentrionale après la conquête romaine (IIe–Ier siècle av. J.-C. – Ier siècle après J.-C), RA 141–78.
MEIGGS, R., 1973, Roman Ostia.
MOURITSEN, H., 1988, Elections, magistrates and municipal élite. Studies in Pompeian epigraphy.
MOURITSEN, H., 2004, Freedmen and freeborn in the necropolis of imperial Ostia, ZPE 150, 281–304.
MOURITSEN, H., 2005, Freedmen and decurions: epitaphs and social history in imperial Italy, JRS 95, 38–63.
MOURITSEN, H., 2007, *CIL* X 1403: The album from Herculaneum and the nomenclature of the Latini Iuniani, ZPE 161, 288–90.
MOURITSEN, H., 2011, The freedman in the Roman world.
MOURITSEN, H., 2013, Slavery and manumission in the Roman elite: a study of the columbaria of the Volusii and the Statilii, in M. GEORGE (ed.) Roman Slavery and Roman Material Culture, 43–86.
O'HAGAN, A. – BUCK, C. E. – DANESHKAH, A. – EISER, J. R. – GARTHWAITE, P. H. – JENKINSON, D. J. – OAKLEY, J. E. – RAKOW, T., 2006, Uncertain judgements: Eliciting experts' probabilities.
ROHDE, D., 2012, Zwischen Individuum und Stadtgemeinde: Die Integration von Collegia in Hafenstadten, Studien zur alten Geschichte 15.
SOLIN, H., 1971, Beiträge zur Kenntnis der griechischen Personennamen in Rom.
SOLIN, H., 1996, Die stadtrömischen Sklavennamen. Ein Namenbuch, Forschungen zur Antiken Sklaverei Beiheft 2.
SOLIN, H., 2001, Griechische und römische Sklavennamen: Eine vergleichende Untersuchung, in H. BELLEN – H. HEINEN (eds.), Fünfzig Jahre Forschungen zur antiken Sklaverei an der Mainzer Akademie 1950–2000 : Miscellanea zum Jubiläum, 307–330.
SOLIN, H. – SALOMIES, O., 1994, Repertorium nominum gentilium et cognominum Latinorum.
TAYLOR, L. R., 1961, Freedmen and freeborn in the epitaphs of imperial Rome, AJPh 82, 122–32.
TRAN, N., 2006, Les membres des associations romaines. Le rang social des collegiati en Italie et en Gaules, sous le haut-empire.
VERBOVEN, K., 2011, The freedman economy of Roman Italy, in S. BELL – T. R. RAMSBY (eds.), Free at last! The impact of freed slaves on the Roman empire, 88–109.
WALLACE-HADRILL, A., 2015, The album of Herculaneum: problems of status and identity, in A. B. KUHN (ed.) Social status and prestige in the Graeco-Roman World, 115–51.

Abbildungsnachweis

Fig. 1a. Author

Fig. 1b. Author

Fig 2a. Author

Fig 2b. Author

Fig 3. Author