



Data Quality and Bias in the Coin Hoards of the Roman Empire Database: A Logistic Regression Analysis

Philip Rademacher, University of Düsseldorf, Germany, philip.rademacher@hhu.de

The Coin Hoards of the Roman Empire database currently contains more than 18,000 coin hoards from all areas of the Roman Empire and beyond. This large data collection encourages data-driven approaches to research the ancient world. However, there are several sources of bias in the data that make data-driven research difficult. One problem is that the hoards are recorded differently. For some hoards, archaeological information is available; for others, only numismatic; and for others, not even the exact location of the find is known. This paper uses logistic regression to analyze whether the data quality correlates with some observable hoard characteristics. It can be shown that numismatic data tends to be superior for hoards from the first century AD. In addition, there are significant differences in data quality between hoards from ancient cities and forts versus those in rural areas or outside the Roman provinces.



Introduction

Archaeologists use the term hoard to describe a find that consists of a closed group of artifacts (Aitchison 1988: 271). A coin found alone is considered a single find, while several coins found together constitute a hoard, as long as it is clear that the coins were deposited together at the same time. This definition, originally formulated by Montelius (1903), was motivated by the idea that a hoard contains more information than a single find. When several coins are found together, there is a high probability that they were deliberately deposited. The location, the items contained and the way they are deposited are visible results of human actions and decisions (Oras 2013: 68).

Ancient coin hoards are of primary relevance for economic and social history because the ancient literature on these topics has survived poorly. The Coin Hoards of the Roman Empire database (hereafter: CHRE) aims to systematically record hoards from the Roman Imperial period (CHRE 2023). This large database is an advantage for historians because not only does it facilitate access to hoards as a source, but it also allows them to use the data as a collection for historical research.

Some research papers utilize coin hoard data to address economic, demographic or cultural questions. The hoarded coins can provide insights into economic topics like money supply or circulation (e.g. van Heesch 2011; Swan 2020; Mairat et al. 2022: 237–333). Hoards can also shed light on demographics: a strong temporal and spatial concentration of hoard finds has long been connected to wars or crises. The idea behind this is intuitive; hoards that were formed as a store of value could not be recovered by their owner due to the owner's death in a war or forced displacement. Hoard data can therefore be analysed as an additional indicator for unrest and demographic trends (e.g. Turchin and Scheidel 2009; De Callataj 2017). If the archaeological context of a hoard is known, one can analyse cultural aspects of hoarding, such as systematic preferences in the choice of the deposition place (e.g. Bland et al. 2020: 71–135).

Larger datasets, like the CHRE database, can provide the basis for a wide range of quantitative research topics, in which distributions are analysed. Regarding the study of coin hoards, these could be the regional and temporal distribution of hoards or specific coins. However, quantitative research questions are often affected by biased or incomplete data. There are several sources of bias in archaeological datasets that make data-driven research difficult or even impossible.

The aim of this paper is to analyse to what extent the quality of numismatic, archaeological context, and findspot information is related to spatial and temporal hoard characteristics. For example, it is considered whether urban hoards more often have better numismatic or archaeological context information than rural hoards. The purpose of this analysis is to discern if an information disparity exists among the hoard

finds that follows a measurable pattern. Such a disparity can be called a ‘recording bias’ because it is caused by unequal documentation of coin hoard finds.

The Problem of Bias

When researching archaeological data, one must be aware of a fundamental aspect: databases that collect artifacts are always incomplete, as they can only record the artifacts that are known to them. The CHRE database thus represents only a small portion of the original population of all deposited hoards. It is likely that many hoards are not included in the database due to reasons such as being recovered in antiquity, being destroyed, not being found yet, or not being scientifically documented (cf. Robbins 2013).

However, a bias does not mean that we only know a subset of the statistical population. Rather, the issue is that the database is not a representative subset. In other words, the database does not contain a random sample of the original population of hoards. We must expect that hoards with certain characteristics are under- or overrepresented in the database.

A brief look at the geographical distribution of the hoards in the CHRE database already shows that the density of coin hoards differs greatly between countries. Luxembourg and England have a higher density of hoard finds per square kilometer than any other country. Does this provide any insights into ancient hoarding practices? Does this suggest that the Romans hoarded more in what is now known as Luxembourg? Perhaps, but it is likely that such differences in hoard densities reflect an imbalance in the intensities of searching and recording. In regions where more excavation is undertaken (e.g. in England due to the widespread use of metal detectors), more hoards can be found, and in regions where more hoards are scientifically documented (e.g. Luxembourg), more hoards are recorded in the database (cf. Bland 2020).

Thus, the distribution of hoards is biased and does not purely reflect ancient patterns, because hoards with certain characteristics are overrepresented in the database. In this example, the certain characteristic is the location; hoards deposited in the area of present-day England or Luxembourg could be more likely to be included in the database.

Discussing and analysing bias is also relevant in datasets other than the CHRE database. Bias has become a major concern in initiatives that generate and curate extensive archaeological and numismatic datasets. One example is the Rural Settlement of Roman Britain project, which explores regional patterns of archaeological excavations and considers underlying biases: ‘Another major challenge with Big Data projects is to recognise factors that might bias the data recovered and to evaluate the degree to which patterning is a true reflection of past activities rather than a product of modern ones’ (Smith et al. 2016: xx).

Another example is the Portable Antiquities Scheme (PAS), whose user guide addresses biases in detail. Alongside theoretical explanations, it provides five case studies illustrating various data-driven methods for detecting regional biases (Robbins 2014). Similarly, the project ‘Framing the Late Antique and Early Medieval Economy’ (FLAME), which also examines coin finds (including but not limited to hoards), dedicates a section of its user manual to discuss biases (FLAME 2024). This project addresses general biases and includes essays analysing regional biases. They describe factors like institutional and legal frameworks or research traditions in the country and assess their impact on the database. For instance,

‘...what is distinct and unusual in England and Wales is the very large published body of finds (hoards and single coins) made by metal detector users. These reflect modern legal and administrative arrangements, not necessarily a uniquely rich monetary economy in centuries past. Were the same regulations in place in France, Italy or Turkey, it is likely that Britain would look far less unusual’ (Naismith 2021: 4).

This paper, while closely related to biases in archaeological data, diverges from the broader topic as it does not focus on biases that render the CHRE database a non-representative sample. Instead, the focus here is on internal biases that arise from the collected data and are caused by differences in data quality and availability. The analysis ignores the question of whether the CHRE database is representative and the recording bias is analysed in isolation from all previous bias levels. Formally, a smaller subset (e.g. hoards with high data quality) is compared with another larger subset (hoards in the database), testing whether the smaller subset is representative of the larger subset, but it remains unclear whether the larger subset is representative of its original population (all hoards formed during the Roman Empire). Therefore, in the next chapter, the data quality measures are presented and differences in data quality are discussed in detail.

Differences in Data Quality

In the CHRE database, hoards are described with numismatic and archaeological characteristics, as well as information on their findspot. The quality of such information varies widely between hoards and is rated within the database on a scale of one to four, where a four indicates excellent quality. In addition to this rating, the database shows whether coin data is available or not available. Regarding the hoards with coin data, a distinction is made between coin data that has already been entered and coin data that is available to be entered in the future.

The hoard from Niederzier (CHRE 17212) is an example of a well-documented find. The summary text describes the findspot, archaeological context and the coins

contained: ‘Three coins of Postumus were found in the basement area of the hypocaust heating room inside the main building of a villa rustica located north-east of Niederzier, in the “München Busch” wooded area’. The coins are described in detail and referenced with a RIC number. Therefore, the numismatic rating is ‘4-Excellent’, while the other categories are rated with ‘3-Good’. A counter-example is the hoard of Aschach (CHRE 16548), where the summary states: ‘A hoard of unspecified coins found 1 km south of castle Aschach, known from archival sources. No further details are known’. Since there is no information about the coins or the archaeological context, these categories are rated as ‘1-Poor’. The location of the find, unlike the previous categories, is rated as ‘2-Fair’ due to the approximately known location of the find.

This study analyses differences in data quality in four countries: Germany, the Netherlands, Portugal and Spain. Modern state boundaries are used to define study areas because it was expected that differences in the recording of hoards were closely related to modern institutions. The four countries were selected since they contained more than 200 hoards at the time of data acquisition, all of which were presented as complete entries within the database.

Table 1 provides a quantitative summary of how many hoards in the study area have good numismatic, archaeological or findspot information. In the analysis, I classify the quality of data as ‘good’ if the numismatic rating equals or exceeds ‘3-Good’. In addition to numismatic rating, I also test whether the presence of coin data correlates with certain hoard characteristics, regardless of the data quality. For context and findspot rating, the threshold value was set to ‘2-Fair’, because the overall data quality is lower. The lower threshold value is a technical decision to have enough hoards in the minority class (i.e. with high data quality).

	Germany		Netherlands	
	Quantity	%	Quantity	%
Hoards	855		222	
Numismatic Rating ≥ 3	393	46%	107	48%
Coin Data available	553	65%	110	50%
Context Rating ≥ 2	196	23%	35	16%
Findspot Rating ≥ 2	535	63%	76	34%
	Portugal		Spain	
	Quantity	%	Quantity	%
Hoards	262		461	
Numismatic Rating ≥ 3	86	33%	167	36%
Coin Data available	119	45%	218	47%
Context Rating ≥ 2	27	10%	102	22%
Findspot Rating ≥ 2	62	24%	176	38%

Table 1: Overview of hoards with good data quality.

In each country, approximately 50% of the hoards have data on their coins, which is also often rated as good. The presence of coin data does not necessarily indicate that the numismatic rating is good or excellent. The summary on the hoard of Imsbach (CHRE 4169), found in 1846, illustrates this: ‘A plowman on his farm towards Imsbach uncovered a heavy stone slab under which he found an urn with two handles. It contained approx. 4000 coins (Diocletian – Constantine I, AD 317). 223 coins are listed in FMRD [Fundmünzen der Römischen Zeit in Deutschland]’. Although coin data is available, the rating is ‘2-Fair’ since only a small portion of the coins are known.

Unlike the numismatic rating, information regarding the context or findspot is more often rated as poor. There are two possible reasons for the lower findspot and context ratings. Many hoards were not discovered by professional archaeologists, leading to a lower quality of information. Another reason might be that hoards were considered primarily of numismatic interest for a long time, increasingly drawing archaeological attention from the 1960s (Guest 2015: 104).

Describing Hoards

To answer the question of whether the information disparity among the hoard finds follows a pattern, we need hoard characteristics that are known for all hoards in our dataset, regardless of their data quality or coin data availability. It would, for example, be impossible to determine whether hoards with coins of Nero lack coin-level data more frequently than other hoards, since answering this question would require coin-level information for all hoards, which is not currently available. The Nero example thereby shows that certain biases in the data can never be detected; if a recording bias is detected in this analysis, it is only based on observable characteristics.

Spatial Location

Since almost all hoards in the CHRE database are georeferenced, the ancient environment of the deposition sites can be described using the Digital Atlas of the Roman Empire (DARE 2017). The DARE has been maintained by Lund University since 2012 and offers a map of the Roman Empire that contains ancient cities, settlements, buildings and roads.

Three spatial characteristics are constructed with the DARE, namely whether a hoard is located within the Roman Empire, near cities, or near military fortifications. Hoards within 1.5 miles of an ancient settlement (e.g. capital, colonia, municipium or vicus) or within one mile of a military fortification (e.g. legionary fortress or castrum) are classified as urban or military, respectively.¹ The remaining hoards are considered rural. The imperial border and proximity to military fortifications are only relevant in the analysis of Germany and the Netherlands.

The choice of the hoard categories (rural, urban and military) has economic implications. In research, fortifications and cities are considered monetized places where new coins enter circulation (von Reden 2015: 156). If — in certain areas — coin data is more often available for hoards from such monetized sites, this would bias the comparison between different areas. Such spatial biases have not yet been analysed, as they can not be visualized with CHRE data, but only by linking hoard data to the DARE.

Closing year

In addition to the coordinates, the closing year (or *terminus post quem*) is also known for most hoards in the CHRE database, regardless of their coin data availability. The closing year is determined by the latest coin of a hoard and describes the earliest point in time when the hoard could have been deposited. To enable a better comparison between hoards, the closing years were grouped into centuries for the analysis.

It is worth noting that the closing year is a characteristic that is widely available even when there is no coin data. This is because the coins were often described by enumerating the emperors depicted: ‘According to a note published in 1830, a young boy found 12 denarii. Only the names of the emperors are known. Coins range from Hadrian to Septimius Severus’ (CHRE 2959). However, the closing year can sometimes be imprecise if parts of the hoard were lost, or coins were unidentifiable. For the analysis, it is assumed that such cases are exceptions, and that the absence of some coins does not often change the century in which the closing year lies.

Table 2 shows how the coin hoards are distributed across centuries in the study areas and how many hoards were found near ancient cities, forts, and within the empire. Due

	Germany	Netherlands	Portugal	Spain
Closing Century				
First	10%	22%	22%	19%
Second	20%	27%	10%	12%
Third	34%	29%	19%	29%
Fourth	30%	13%	33%	32%
Fifth	6%	9%	16%	8%
Spatial Location				
Fort	22%	23%	0%	0%
City	34%	22%	35%	42%
within Empire	64%	70%	100%	100%
Coverage	85%	82%	89%	93%

Table 2: Distribution of the hoard characteristics used in the logistic regression model.

to the use of temporal and spatial characteristics, not all hoards could be included in the analysis, because many hoards lack information on the closing year. The coverage ranges from 82% in the Netherlands to 93% in Spain.

Empirical Strategy

The analysis in this paper is inspired by sample selection models used in social and economic research, where non-random samples are a frequent problem (for a comprehensive overview refer to Winship and Mare 1992; Cameron and Trivedi 2005: 546–553; Wooldridge 2009: 606–613). The goal of this paper is to analyse how the quality of hoard data correlates with specific hoard characteristics. To identify how data quality is related to those features, a binary regression model was estimated. This research setting is close to the first step of the Heckman correction (Heckman 1979). The regression model was used to estimate the probability that a hoard has high data quality, depending on the hoard’s characteristics.

Compared to descriptive or univariate methods, the application of a multivariate method, like binary regression, has one main advantage. It enables the joint analysis of multiple hoard characteristics, whereby the model disentangles overlapping effects and identifies the characteristics most closely related to data quality. In contrast, when using univariate methods, one characteristic is always analysed in isolation from the others.

This can be illustrated using a fictional dataset of 1,200 coin hoards. For simplicity, the analysis focuses on two characteristics: the hoard location (near an ancient city or rural) and the *terminus post quem* (first or second century AD). **Table 3** presents the distribution of high-quality hoard data based on these characteristics. The 1,200 hoards are distributed equally across time and space, with 600 hoards for every characteristic. However, the spatial distribution within the two centuries is completely opposite: in the first century AD, the majority of hoards are urban, while in the second century AD, rural hoards predominate.

	1 st century	2 nd century	Total
City	450/500	90/100	540/600
	90%	90%	90%
Rural	50/100	250/500	300/600
	50%	50%	50%
Total	500/600	340/600	
	83%	57%	

Table 3: Distribution of hoards with high-data quality (fictional dataset for illustration purposes). Urban hoards have a higher data quality than rural ones.

The data was generated in such a way that 90% of hoards from ancient cities exhibit high data quality, while only 50% of the rural finds meet the same data quality criteria. For example, in the first century AD, 450 out of 500 urban hoards have high data quality but only 50 out of 100 rural hoards. This difference between urban and rural hoards remains the same over time. However, since the first century AD sample contains more urban hoards (500) than the second century AD sample (100), the overall data quality appears to be higher in the first century AD (83%) compared to the second century AD (57%). Analysing the data quality in a univariate way (i.e. focusing solely on the total values of the last row and column) yields two biases. The analysis shows a difference of 40 percentage points (pp.) between urban and rural hoards, and a 26pp. difference between both centuries. However, the second bias is misleading because it is driven entirely by the different composition of the century samples and a double-count of the urban-rural effect. Comparing rural and urban hoards separately between both centuries yields no difference in data quality. The share of rural hoards with high data quality is always 50% and the share of urban hoards is always 90%.

In this example, a binary regression model would only measure one bias that is related to the location of the hoard. The estimated model would return an average marginal effect of 40pp. for the urban hoards, which means that the probability that a hoard has a high data quality is 40pp. higher for urban than for rural hoards. The average marginal effect for the second century AD (relative to the first century AD) would be zero.

In order to measure a temporal bias, the probabilities must be generally different between the centuries. The second example (**Table 4**) shows such a situation. Urban finds again have better data quality than rural finds, and finds from the second century AD have better data quality than those from the first century AD. A binary regression model would measure two significant average marginal effects: a 20pp. difference between rural and urban hoards and a 10pp. difference between hoards from the first and second centuries AD. A univariate analysis would again double count effects.

	1 st century	2 nd century	Total
City	60/100	350/500	410/600
	60%	70%	68%
Rural	200/500	50/100	250/600
	40%	50%	42%
Total	260/600	400/600	
	43%	67%	

Table 4: Distribution of hoards with high-data quality (fictional dataset for illustration purposes). Urban hoards have a higher data quality than rural ones and hoards from the second century AD have a higher data quality than those from the first century AD.

The differences in the probabilities would be 26pp. between city and rural, and 24pp. between the first and second centuries AD.

In summary, regression analysis provides a clearer understanding of the relationships between hoard characteristics and data quality, avoiding the weaknesses of univariate approaches. In the two examples above, the biases were constructed in such a way that they were visible when looking at the inner cells of the table. In contrast, real datasets are more complex, and binary regression is a helpful tool to detect, simplify and quantify biases.

Logistic Regression

In this paper, logistic regression was chosen as a binary regression model and country-specific regressions were conducted.² Each indicator — namely, the availability of coin data, as well as the data quality indicator for numismatic, archeological, and findspot information — was modeled in a separate regression.³ In the end, there are 16 regression models, one for each quality indicator and each country.

In these models, the quality indicator serves as the dependent variable (y), while the chosen hoard characteristics are the explanatory variables (X). The dependent variable is binary in nature. For instance, in the case of coin data, the quality indicator is assigned a value of zero if there is no coin data, and a value of one otherwise. When considering context and findspot ratings, the quality indicator is set to one if the rating exceeds '2-Fair' and to zero if the rating is lower. For the numismatic rating, the threshold value is '3-Good'.

Logistic regression aims to discern the relationship between the quality indicator and the hoard characteristics. Therefore, the relationship between the hoard characteristics and the data quality is parametrized with coefficients. The principle is similar to a linear regression:

$$y = \hat{\beta}_0 + \hat{\beta}_{EMP}EMP + \hat{\beta}_2C2 + \hat{\beta}_3C3 + \hat{\beta}_4C4 + \hat{\beta}_5C5 + \hat{\beta}_{CITY}City + \hat{\beta}_{FORT}Fort + \hat{u}.$$

The coefficients ($\hat{\beta}$) indicate how strongly a hoard characteristic can be related to data quality. For example, $\hat{\beta}_{EMP}$ shows whether hoards within the Roman Empire have higher data quality than those outside of the Roman provinces. The coefficients $\hat{\beta}_2 \dots \hat{\beta}_5$ show whether a century has better or worse data quality than the first century AD, which is the reference category and therefore omitted in the regression equation.⁴ The last coefficients show whether forts or cities have higher data quality than rural hoards where $\hat{\beta}_0$ is the intercept and \hat{u} denotes the residuals, which reflects the difference between model prediction \hat{y} and the true value of y . The equation above can be written in matrix notation as $y = X\hat{\beta} + \hat{u}$. The model predictions are $\hat{y} = X\hat{\beta}$.

The main difference from the linear regression model is that logistic regression applies a sigmoid function to the linear equation, ensuring that the model outcomes are probabilities bounded between zero and one. The linear equation sketched above would not be restricted to values between zero and one for \hat{y} . The logistic regression uses the cumulative distribution function of the logistic distribution (Λ) as a sigmoid function. The application of a sigmoid function means that the regression model does not have a closed form and must be estimated iteratively with maximum likelihood (Cameron and Trivedi 2005: 139–145):

$$\Pr[y = 1|X] = \Lambda(X\hat{\beta}) = \frac{e^{X\hat{\beta}}}{1 + e^{X\hat{\beta}}}$$

whereby $\lim_{X\hat{\beta} \rightarrow \infty} \Lambda(X\hat{\beta}) = 1$ and $\lim_{X\hat{\beta} \rightarrow -\infty} \Lambda(X\hat{\beta}) = 0$.

In logistic regression, we model the probability that the hoard has high data quality: $\Pr[y = 1|X]$. The estimated probabilities are conditional on the hoard characteristics, which are the explanatory variables (X) of the model. This means that hoards with identical characteristics receive the same probabilities in the model. If we model data quality only with the characteristic ‘hoard lies within an ancient city or not’, all the coin hoards deposited within ancient cities would receive the same probabilities.

Ideally, data quality is random and does not correlate with specific hoard characteristics. Consequently, the conditional probabilities of all hoards are approximately equal. There are no hoard characteristics that imply better or worse data quality. On the other hand, a bias is detected if hoards with certain characteristics have significantly higher or lower probabilities.

To detect characteristics that increase or decrease the conditional probabilities for data quality, average marginal effects (hereafter: AME) are computed. For each hoard, the difference between the prediction with and without the characteristic of interest (e.g. city) is calculated ($\tilde{x}_i\tilde{\beta}$) are all other characteristics multiplied by their coefficients. Then the average of all marginal effects is calculated (Wooldridge 2009: 577):

$$AME_j = \frac{1}{n} \sum_{i=1}^n \Lambda(\tilde{x}_i\tilde{\beta} + \beta_{city}) - \Lambda(\tilde{x}_i\tilde{\beta}),$$

test statistics with $z_{AMEj}^* = \frac{AME_j}{SE_{AMEj}}$ and a reject null hypothesis of $AME_j = 0$ if $|z_{AMEj}^*| > z_{1-\frac{\alpha}{2}}$.

To assess whether an AME is different from zero, z-tests were conducted. Insignificant effects have a large standard error (SE_{AMEj}) relative to the estimated effect (AME_j). Therefore, we cannot be sure that they are different from zero. In the z-test, the absolute value of this ratio is compared with the corresponding quantile from the

standard normal distribution (z). The quantile is determined by the error probability (α) we allow. We reject the null hypothesis that the marginal effect is zero if $|z_{(AMEj)}^*| > z_{(1-\frac{\alpha}{2})}$. In the results, the stars indicate the significance level up to which the null hypothesis can be rejected. The significance level is $1 - \alpha$.

The AME is interpreted relative to the reference category. The reference group consists of the characteristics that were omitted as variables from the regression equation. For example, the AME of the characteristic ‘city’ shows how the probability for coin data changes on average if a coin hoard was deposited in an urban environment instead of a rural one. If the effect is positive, coin hoards from ancient cities are more likely to be recorded at the coin level than rural hoards. A negative effect indicates the opposite.

The empirical analysis was carried out in Python using the following additional libraries: Pandas and Numpy for data preparation in general, and Geopandas and Sklearn for the creation and manipulation of spatial variables. The logistic regression was estimated with Statsmodels. The diagrams presented in the following section were created with Plotly, while the map was created with Matplotlib.

Results

Table 5 shows the AMEs for Portugal and Spain. The significance level is highlighted with stars depending on the significance level. Effects without at least one star are not significant, which means they are not necessarily different from zero. While the tables contain significant and insignificant effects, only significant ones are interpreted. In addition to the stars, a significant effect is highlighted with a grey background.

Portugal	Coin Level Data	Numismatic Rating	Context Rating	Findspot Rating
2 nd century	-0.170	-0.173**	-0.063	-0.116
3 rd century	0.105	-0.048	0.064	-0.085
4 th century	-0.079	-0.181***	0.020	-0.053
5 th century	-0.306***	-0.308***	0.028	-0.062
City	0.145**	0.152**	0.167***	0.396***
Spain	Coin Level Data	Numismatic Rating	Context Rating	Findspot Rating
2 nd century	-0.200**	-0.115	-0.002	-0.015
3 rd century	-0.120*	-0.116*	0.105	0.110*
4 th century	-0.080	-0.155***	0.013	0.039
5 th century	-0.295***	-0.261***	-0.042	0.009
City	0.103**	0.102**	0.202***	0.378***

Table 5: AMEs for Portugal and Spain. Significant effects are highlighted with a grey background. Significance levels are indicated with stars: 90% (*), 95% (**) and 99% (***).

In the regression models for Portugal and Spain, the characteristics ‘rural’ and ‘first century AD’ were omitted and form the reference group. All AMEs are interpreted relative to the reference group. For example, the AME of 0.110 for the third century AD in the findspot rating model for Spain indicates that the probability of a hoard from the third century AD having good findspot data is 11pp. higher than for those of the first century AD. The marginal effect expresses a change in percentage points; for example, if the probability in the reference group is on average 40%, the marginal effect would increase the probability to 51%.

In both countries, the data quality of hoards found near or in ancient cities is significantly better compared to rural hoards. Urban hoards frequently have more accurate coin data, along with detailed context and findspot information. The disparity between urban and rural hoards is very large in the findspot data. The probability of having good findspot information increases by an average of 38pp. in Spain and 40pp. in Portugal, if the hoard was deposited in an urban area. Simultaneously, the availability of coin data increases on average by 14pp. in Portugal and 10pp. in Spain. Rural hoards have poorer data quality.

This bias in favor of ancient cities could have several reasons. One plausible explanation could be the frequent proximity of research facilities to these ancient cities, given that many modern cities are built upon the foundations of their ancient counterparts. This proximity may lead to more consistent scientific documentation of the hoards (see also Robbins 2014: 39). Empirical evidence from the CHRE database supports this hypothesis. While many rural finds were discovered by amateurs and lack information on their whereabouts, a large share of urban hoards has this information. Many urban hoards were handed over to museums or other academic institutions (**Figure 1**).

In Spain, however, there could be an additional explanation for the better archaeological context data of urban hoards. The discovery years of rural and urban hoards are distributed differently, with a large proportion of urban hoards discovered after the 1970s, while many rural hoards were discovered in the first half of the twentieth century or earlier. The CHRE data shows that hoards discovered earlier were often only recorded numismatically and lack precise findspot or archaeological context information. The archaeological investigation of hoards began in the 1960s and has increased since then (Guest 2015: 104). Roughly 60% of the hoards discovered in Spain in the last twenty years have archaeological context and findspot information, while at the beginning of the twentieth century, the ratio was the other way around (**Figure 2**). Therefore, the data quality of urban hoards in Spain could be better due to their late discovery (**Figure 3**).

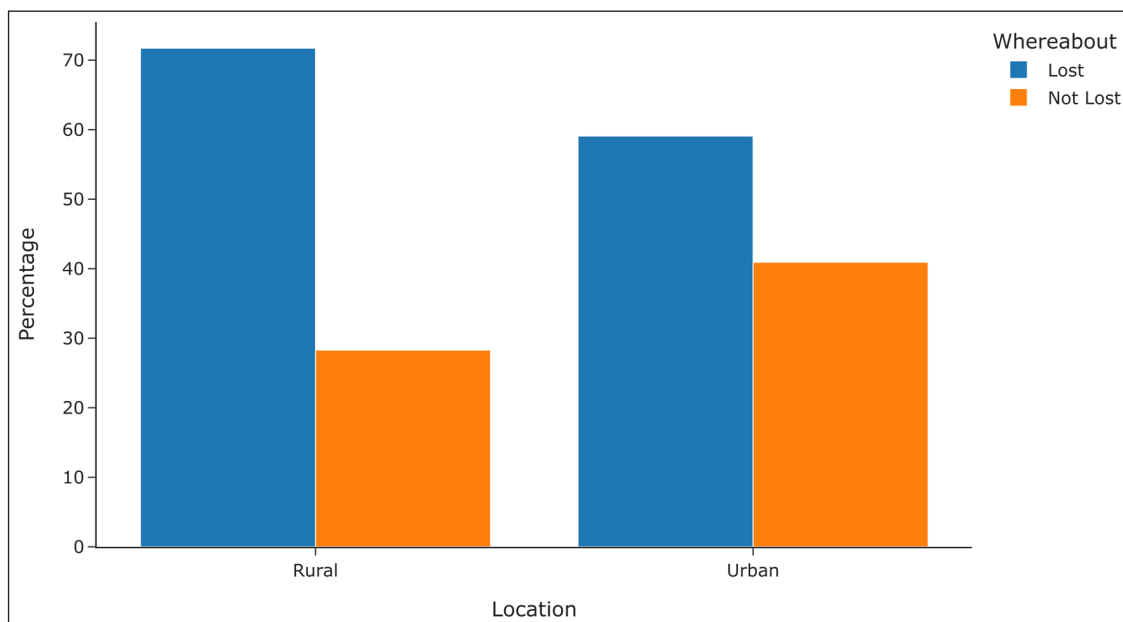


Figure 1: Share of rural and urban hoards with and without information about their whereabouts (Spain and Portugal). The shares of each group on the x-axis add up to 100%. Source: Author.

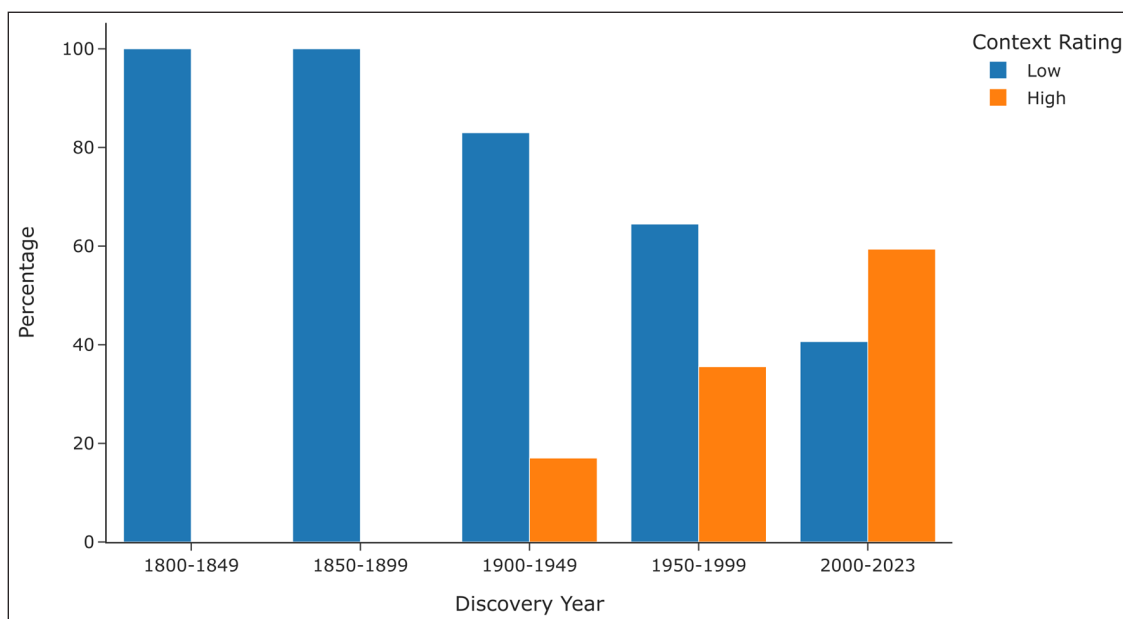


Figure 2: Share of Hoards in Spain with high and low archaeological context rating relative to their discovery year. Hoards found in Spain after 1800 are included. The Discovery Years are grouped in 50-year steps. The shares of each group on the x-axis add up to 100%. Source: Author.

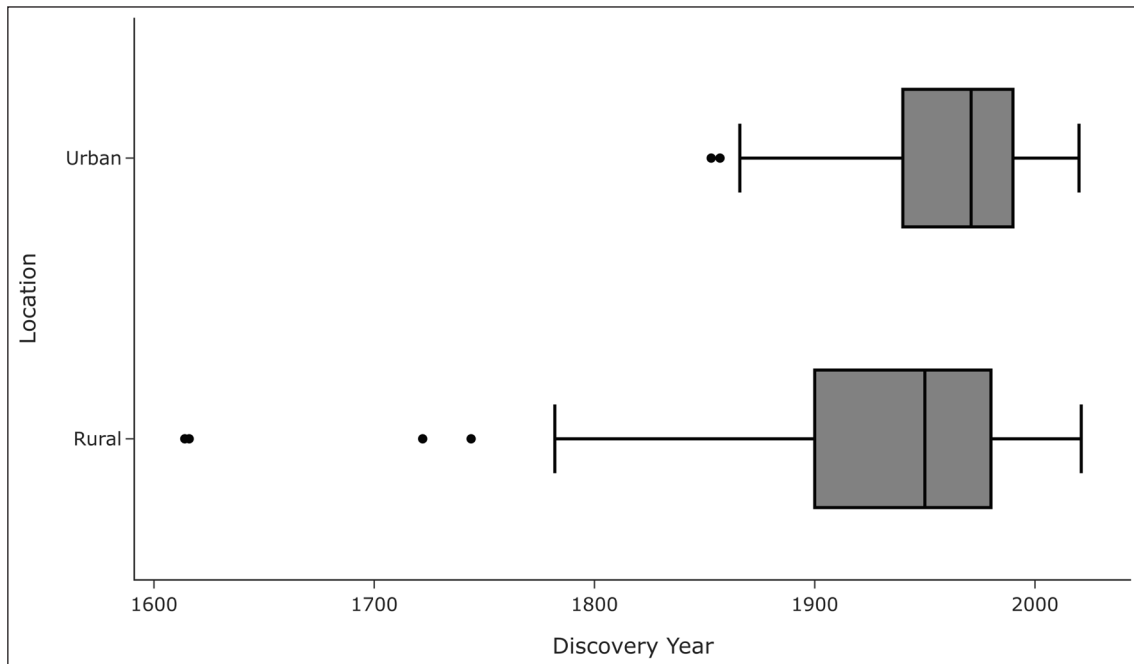


Figure 3: Distribution of discovery years by urban and rural hoards in Spain. The boxplots show that many urban hoards were found later. Source: Author.

In addition to the higher data quality of urban hoards, there are also some temporal differences in both countries. The temporal effects are interpreted relative to the first century AD. To illustrate the interpretation, consider the coin-data availability of Spain: **Figure 4** shows the temporal distribution of hoards with and without coin-level data. If the coin data availability in Spain is random, it is reasonable to expect that the hoards with coin data have a similar temporal distribution as the hoards without. From the figure, it becomes clear that the first and fourth centuries AD are overrepresented among hoards with coin data, as their proportion in the subset of hoards with coin data is larger than in the other subset. Of all hoards without coin data, 14% have a *terminus post quem* in the first century AD, while of all hoards with coin data, 23% are dated to the first century AD. Based on the same logic, the remaining centuries are underrepresented. For example, the fifth century AD has a share of 11% among the hoards without coin data, but only a share of 4% among the hoards with coin data.

Consequently, hoards from the first century AD have a high probability for coin data. Changing the closing century to the second, third or fifth centuries AD would decrease the probability of coin data. Therefore, the AMEs of these centuries are negative. The fourth century AD is equally overrepresented as the first century AD, leading to no significant marginal effect in the regression results. The results of the regression can be easily compared with the histogram. However, this histogram (Figure 4) does not take into account the location of the hoard, while regression analysis does.

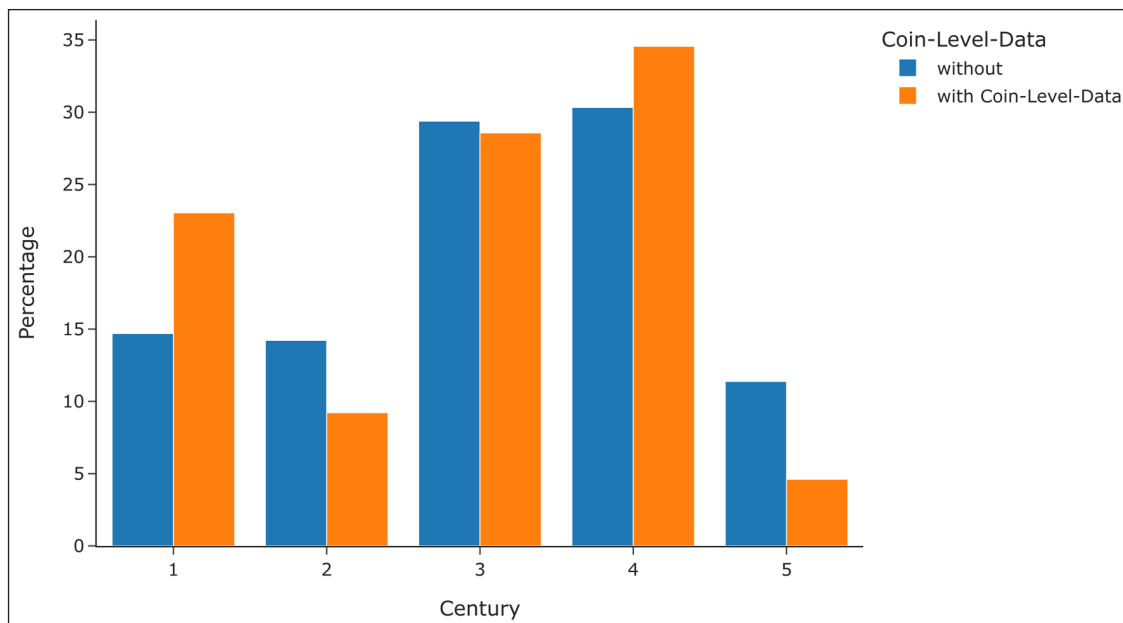


Figure 4: Distribution of hoards with and without coin-level data in different centuries in Spain. The shares of each color group add up to 100%. Source: Author.

For the temporal AMEs of Portugal and Spain, no country-specific explanations were found, but after analysing the results for Germany and the Netherlands, a cross-country explanation is given. **Table 6** shows the average marginal effects for Germany and the Netherlands. In contrast to Portugal and Spain, urban hoards do not have a higher data quality overall. While the findspot information is indeed more precise for urban hoards and in Germany the context data is also better, there is no effect on the availability or quality of numismatic data in this country/area. Interestingly enough, in the Netherlands, there is a negative effect. The likelihood of having good numismatic information decreases by an average of 15pp., when the hoard was deposited in the neighborhood of an ancient city. Neither country shows evidence that coin hoards found near ancient cities have better numismatic data. However, in Germany, this is the case for hoard finds located near ancient forts.

The border of the Roman Empire was also considered in the models of Germany and the Netherlands and revealed strong effects in both countries. In Germany, hoards within the empire have better data quality compared to those deposited outside of the provincial borders. In the Netherlands, at least for numismatic data, it is the other way around. Hoards within the empire have a 20pp. lower probability of coin data. This effect is shown on **Figure 5**, which demonstrates that most hoards without coin data are located in the south of the Netherlands.

Germany	Coin Level Data	Numismatic Rating	Context Rating	Findspot Rating
2 nd century	0.015	-0.087	-0.069	-0.088
3 rd century	0.050	-0.044	-0.105**	-0.040
4 th century	0.026	-0.120*	-0.123***	-0.137**
5 th century	-0.015	-0.189**	-0.150***	-0.194**
Within Empire	0.156***	0.164***	0.126***	0.156***
City	0.035	0.053	0.128***	0.113***
Fort	0.091**	0.138***	0.035	0.162***
Netherlands	Coin Level Data	Numismatic Rating	Context Rating	Findspot Rating
2 nd century	-0.202**	-0.114	0.010	0.027
3 rd century	-0.229**	-0.176*	-0.021	-0.017
4 th century	-0.210*	-0.158	-0.029	-0.079
5 th century	0.028	-0.111	0.040	-0.020
Within Empire	-0.209***	-0.216***	0.120**	0.083
City	-0.128	-0.154*	-0.058	0.265***
Fort	-0.002	0.008	0.023	0.266***

Table 6: AMEs for Germany and Netherlands. Significant effects are highlighted with a grey background. Significance levels are indicated with stars: 90% (*), 95% (**), and 99% (***).

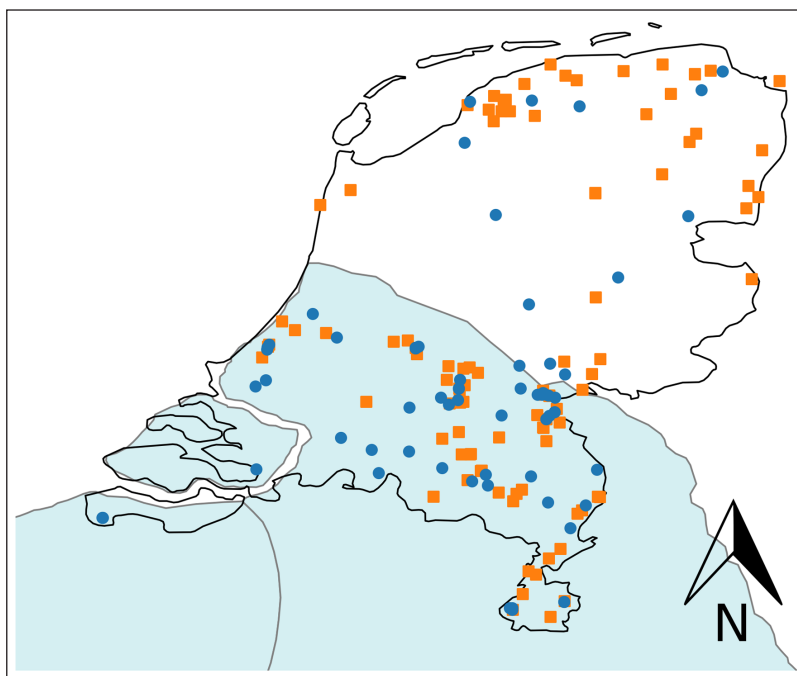


Figure 5: Hoards with and without coin data in the Netherlands. Orange squares indicate the availability of coin data and blue points indicate the non-availability. The light blue background shows the borders of the Roman Empire. Source: Author.

In the Netherlands, the border effect can be explained by the existence of the *Studien zu Fundmünzen der Antike* monograph series. These volumes have systematically collected and processed numismatic information on coin finds in various countries. However, for the Netherlands, volumes are mainly available for the northern part (except Volume 3, which covers the city of Nijmegen (Radnóti-Alföldi 2002)). The hoards in the north have been predominantly recorded in the database using these volumes. In contrast, data from the south was derived from multiple sources, which seems to have an impact on numismatic data quality.

Regarding the temporal effects, it is noticeable across all four countries that they predominantly occur in models for coin-level data or numismatic rating and, if significant, they are negative. In all countries, the first century AD hoards have the highest coin data availability and the best quality of numismatic data.

An explanation for the bias towards the first century AD could lie in *The Roman Imperial Coinage* (RIC) volume series. Good numismatic data is characterized by clearly identified coins, whose identification is often achieved through the assignment of RIC numbers. However, the coins were cataloged in different volumes depending on the year of issue. The first volume, cataloging all coins up to AD 69, was published in 1923 (RIC I), while the last volume for the fifth century AD was not published until 1994 (RIC XX). Therefore, coins from first century AD hoards found after 1923 could be immediately referenced. In contrast, for a hoard from the fifth century AD found in the 1970s, only coins from earlier centuries could be assigned RIC numbers. These hoards could only receive RIC numbers for all coins afterwards, assuming the whereabouts of the coins are known. The RIC publication year can therefore be seen as an indicator of scientific development in the field of numismatics.

Figure 6 shows the percentage of hoards that had a discovery year later than the publication year of the corresponding RIC volume. The figure is consistent with the marginal effects. We see that only a small share of hoards from the fifth century were discovered after 1994, when the RIC volume was published. The share lies between 23% (in Portugal) and 50% (Netherlands). In contrast, the first century AD has the highest share, which lies between 60% (Germany) and 86% (Spain). Many hoards from the first century AD could potentially be directly referenced with RIC numbers.

The figure shows another interesting point: the (scientific) recording of hoards seems to have a longer tradition in Germany because there are many more hoards with a finding year before the publication of the RIC volumes than in other countries. This observation could also be the reason why in Germany numismatic data quality differs less between the centuries.

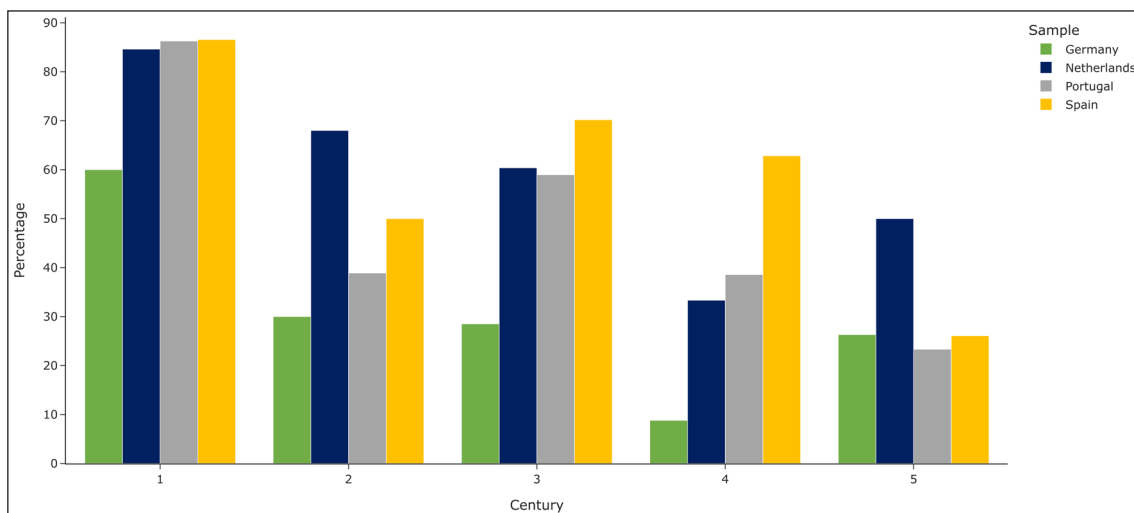


Figure 6: Share of Hoards found after the publication of the respective RIC volume. Source: Author.

To test the hypothesis of whether the existence of a RIC volume at the time of discovery impacted numismatic data quality, the regression models for numismatic rating were re-estimated and the indicator ‘RIC published’ was included in the model. This indicator is assigned a value of one if a RIC volume was available for the *terminus post quem* at the time of discovery and zero otherwise. The objective is to determine whether the temporal characteristics lose their significant negative effects in the re-estimated models while the RIC indicator becomes significant. This would demonstrate that the development of scholarship — at the time of discovery — influenced data quality and caused the temporal biases in the previous regression models.

However, in the samples used for the previous models, hoards were included for which the discovery year is not available. The share of hoards without a discovery year is 7% in Germany, 11% in Portugal and 18% in Spain. Excluding these hoards from the sample and then re-estimating the models makes the comparison of the coefficients between the models difficult, as changes in the coefficients could result from the inclusion of the new RIC variable as well as from the removal of the hoards. To enable a better comparison of the models, the following strategy was employed: for each hoard without a discovery year, the value of the indicator ‘RIC published’ was set to zero. Simultaneously, an additional control variable, ‘No discovery year available’, was included in the model. This approach allows us to determine how hoards without a discovery year are correlated with the biases that were measured in the previous models. This procedure is not applied for the Netherlands, as all hoards there have a discovery year.

Table 7 illustrates the marginal effects of the re-estimated models. To make the changes in the coefficients clearer, those that have lost significance are highlighted in grey in the upper part of the table and the new indicators (RIC and discovery year) are highlighted if they are significant. In Portugal and Germany, the coefficient of the RIC indicator is significant positive, whereas in Spain and the Netherlands, the coefficient is insignificant. In addition, in Portugal and Germany, the temporal characteristics are no longer statistically significant. In the original model, the second, fourth and fifth centuries AD in Portugal and the fourth and fifth centuries AD in Germany were negatively significant. The fact that these coefficients are now insignificant and the RIC indicator is significant suggests that the development in scholarship causes some of the temporal biases detected in the previous model.⁵

Numismatic Rating	Portugal	Spain	Germany	Netherlands
2 nd century	-0.004	-0.100	0.021	-0.129
3 rd century	0.068	-0.108*	0.071	-0.186*
4 th century	-0.045	-0.135**	0.047	-0.150
5 th century	-0.159*	-0.210***	-0.080	-0.113
Within Empire			0.145***	-0.214***
City	0.154***	0.069	0.034	-0.149*
Fort			0.117***	-0.019
RIC published	0.246***	0.075	0.314***	0.066
No discovery year available	-0.390***	-0.278***	0.166***	

Table 7: AMEs of the re-estimated models for numismatic rating (all countries). Significance levels are indicated with stars: 90% (*), 95% (**), and 99% (***).

In Spain and the Netherlands, the RIC indicator is insignificant and the other coefficients remain largely unchanged. The only exception is the characteristic ‘city’ in Spain, which was originally significant. The inclusion of the control variable ‘discovery year unknown’ results in the loss of significance for the city characteristic. This is because the majority of hoards without a known discovery year are located in rural areas and have a low numismatic rating. While in Portugal and Spain hoards without a discovery year are often those with an unclear disposition and/or without known coin data, in Germany many of them were deposited in museums and exhibit good numismatic data. It is also striking that, in the Netherlands, a discovery year was assigned to every hoard. At first glance, this can be seen as a sign of high data quality, but it might be an indication that the database is very selective and has only included well-documented hoards. This potentially means that there are potential biases at an earlier stage.

The re-estimated models show that explanations for biases can vary by country. The hypothesis that the presence of RIC volumes in the discovery year impacts the numismatic data quality can only be confirmed empirically in two out of four countries. In general, it becomes clear that it is difficult to identify the causes of biases in a data-driven way. In the specific question of whether the presence of a RIC volume at the time of discovery influences the numismatic data quality, there are two reasons for that: firstly, the RIC indicator is an approximation for developments in scholarship and a simplification of a more complex process. Secondly, identification of biases is difficult due to missing data that is required to test hypotheses, such as the discovery year.

Discussion

The biases discussed in this paper represent merely the tip of the iceberg, with three technical reasons contributing to the partial nature of the results. Firstly, the biases that have been considered here are only at the final stage. We examine a form of recording bias in isolation from preceding levels. Secondly, even if the analysis is restricted to a bias based on observable characteristics, biases based on unobservable hoard characteristics are still possible. Finally, it should be noted that the CHRE database continues to receive new entries of hoards and information. The regression results therefore are preliminary and can change in the future.

Assessing the impact of bias on coin hoard research is challenging. While we recognize various processes that introduce bias (e.g. the willingness of a finder to report a hoard, a longer history of the recording, etc.), we are not able to assess their direct effects on our knowledge. Many biases are regarded as numerical losses, in the sense that we only know a subset of the original body. The major problem is that this subset can be a selection that is not representative of the original. The situation is different, for example, with the transmission of ancient texts. We know that the transmission process was selective and therefore certain topics were only preserved in fragments. In simple terms, this is because the characteristic of a text, namely the topic, had a direct effect on its transmission. This allows us to roughly estimate which topics are missing.⁶

In the case of coin hoards, we do not know how many biases correlate with hoard characteristics and restrict the knowledge we can gain from studying hoards. The only bias that is frequently discussed are regional patterns (i.e. different densities of hoard finds across countries). For this reason, data quality and availability were analysed in relation to selected hoard characteristics in this study. The analysis contributes less to obtaining unbiased data, an objective that is, in fact, unattainable. Instead, it achieves a different objective, namely the increasing visibility of biases in archaeological and numismatic data that are typically only discussed theoretically and are not connected to certain hoard characteristics.

Conclusion

The analysis has shown that hoard data can be biased with respect to data quality and availability. There are certain hoard characteristics that imply lower or higher data quality. For example, in Portugal and Spain urban hoards more often have a high data quality. In the Netherlands, the dataset provides more insights about coin circulation outside of the Roman Empire than within. If the hoard data is used as a numismatic source (e.g. to study coin use, circulation or political communication), then in the most countries the first century AD is the best period to study.

In general, the results have shown that biases are different depending on the country and type of data considered. It is therefore not advisable to assume that the bias identified in one country also exists in surrounding countries.⁷ Different national institutions, legislations and research traditions may be responsible for the varying biases between them. These national differences imply that for quantitative research, comparing hoards from different countries can minimize the risk of being affected by a single national bias. For the same reason, it is helpful to include other archaeological material, where one can assume that the underlying biases are different from those of hoards.

Biased data is a topic that is likely to become more important in the future. The development of large databases related to history seems to catalyse quantitative research, and the range of databases is growing steadily across topics and epochs. As databases grow, quantitative research will follow and questions concerning the representativeness of the data will arise. Logistic regression as a tool for data analysis is well-suited to detect internal biases in archaeological or numismatic datasets, that have so far received less attention. There are only two requirements to apply a logistic regression: a sufficient number of observations and no (or not many) missing values. Due to the increasing amounts of data, these requirements are usually met. It is therefore to be expected that more and more types of bias will be analysed and discussed in future research.

Notes

- ¹ The distances were defined in a data-driven way. For a subset of hoards the CHRE provides information on the archaeological site class, which can e.g. be 'Military' or 'Urban'. Various distances were applied, and new locational variables constructed with DARE. With a radius of 1.5 miles for cities and one mile for forts, the best matches were achieved with the information on archaeological site class in CHRE.
- ² For more details on the math behind the method see Cameron and Trivedi 2005: 464–470; Wooldridge 2009: 575–587; Hastie et al. 2009: 119–122. In addition to logistic regression, probit regression is often used in empirical articles. The results are similar, especially when the focus is on marginal effects, as in Cameron and Trivedi 2005: 472.
- ³ Hereafter, each indicator of interest is referred to as data quality indicator. This includes not only the quality of numismatic, archaeological and findspot information but also the availability of coin data.
- ⁴ The omission of one century and a locational variable is necessary to avoid perfect multicollinearity.
- ⁵ Some of the temporal biases cannot be clearly assigned to the RIC indicator, which is why the indicator 'No discovery year available' is also significant.
- ⁶ For a discussion of lost topics, refer to Rohmann 2016: 149–235.
- ⁷ The four countries in this paper were selected due to their large number of hoards and the completeness of the information on data quality and coin-data-availability. In the future, when the database is more comprehensive, further countries can be analysed with the same research setting.

Acknowledgements

I would like to thank the anonymous reviewers and the Editor, Emily Hanscam, for their valuable comments on my paper.

Abbreviations

RIC I. The Roman Imperial Coinage, 1: Augustus to Vitellius – Mattingly, H. et al. 1923. *The Roman Imperial Coinage*. London.

RIC XX. Roman Imperial Coinage – Mattingly, H. et al. 1994. *The Roman Imperial Coinage, 10: The divided Empire and the Fall of the Western Parts AD 395–491*. London.

Competing Interests

The author has no competing interests to declare.

References

- Aitchison, Nick B. 1988. Roman wealth, native ritual: coin hoards within and beyond Roman Britain. *World Archaeology* 20(2): 270–284. DOI: <https://doi.org/10.1080/00438243.1988.9980072>
- Bland, Roger. 2020. Coin hoards in the Roman Empire: a long-range perspective. Some preliminary observations. *Journal of Ancient History and Archaeology* 7(1): 119–132. DOI: http://doi.org/10.14795/j.v7i1_SI.477
- Bland, Roger et al. 2020. *Iron Age and Roman Coin Hoards in Britain*. Oxford: Oxbow Books.
- Cameron, Adrian Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.
- CHRE. 2023. *Coin Hoards of the Roman Empire Database*. Available at: <https://chre.ashmus.ox.ac.uk/> [Last accessed: 3 December 2024].
- DARE. 2017. *Digital Atlas of the Roman Empire*. Last modified 2017. Dataset from Github. Available at: <https://github.com/klokantech/roman-empire> [Last accessed: 3 December 2024].

- De Callataÿ, François. 2017. Coin deposits and civil wars in a long-term perspective (c. 400 BC–1950 AD). *The Numismatic Chronicle* 177: 313–338.
- FLAME, 2024. *Framing the Late Antique and early Medieval Economy*. Available at: <https://coinage.princeton.edu/> [Last accessed: 18 October 2024].
- Guest, Peter. 2015. The burial, loss and recovery of Roman coin hoards in Britain and beyond: past, present and future. In: John Naylor and Roger Bland (eds). *Hoarding and the Deposition of Metalwork from the Bronze Age to the 20th Century: A British Perspective*: 101–116. BAR British Series 615. Oxford: British Archaeological Reports.
- Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer. DOI: <https://doi.org/10.1007/978-0-387-84858-7>
- Heckman, James. 1979. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society* 47(1): 153–161. DOI: <https://doi.org/10.2307/1912352>
- Mairat, Jerome, Andrew Wilson and Chris Howgego. 2022. *Coin Hoards and Hoarding in the Roman World*. Oxford: Oxford University Press.
- Montelius, Oscar. 1903. *Die typologische Methode*. Stockholm.
- Naismith, Rory. 2021. FLAME Regional Bias Series: Britain. *FLAME Regional Bias Series*. Last modified: 2021–08–31. Available at: <https://coinage.princeton.edu/wp-content/uploads/2021/09/FLAME-Bias-Britain.pdf> [Last accessed: 18 October 2024].
- Oras, Ester. 2013. Importance of terms: What is a wealth deposit? *Papers from the Institute of Archaeology* 22: 61–82. DOI: <https://doi.org/10.5334/pia.403>
- Radnóti-Alföldi, Maria. 2002. *Die Fundmünzen der römischen Zeit in den Niederlanden 3.1: Nijmegen – Kops Plateau*. Berlin: Gebr. Mann.
- Robbins, Katherine. 2013. Balancing the scales: exploring the variable effects of collection bias on data collected by the Portable Antiquities Scheme. *Landscapes* 14(1): 54–72. DOI: <https://doi.org/10.1179/1466203513Z.0000000006>
- Robbins, Katherine. 2014. *The Portable Antiquities Scheme – A Guide for Researchers*. Last modified: August 2014. Available at: <https://finds.org.uk/documents/guideforresearchers.pdf> [Last accessed: 3 December 2024].
- Rohmann, Dirk. 2016. *Christianity, Book-Burning and Censorship in Late Antiquity: Studies in Text Transmission*. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110486070>
- Smith, Alex, Martyn Allen, Tom Brindle and Michael Fulford. 2016. *The Rural Settlement of Roman Britain*. London: The Society for the Promotion of Roman Studies.
- Swan, David. 2020. *Cross-Channel hoarding in the late Iron Age and early Roman Periods (200 BC to AD 43)*. Unpublished thesis (PhD), University of Warwick.
- Turchin, Peter and Walter Scheidel. 2009. Coin hoards speak of population declines in Ancient Rome. *PNAS* 106(41): 17276–17279. DOI: <https://doi.org/10.1073/pnas.0904576106>
- van Heesch, Johan. 2011. Quantifying Roman Imperial Coinage. In: François De Callataÿ (ed.). *Quantifying Monetary Supplies in Greco-Roman Times*: 311–328. Bari: Edipuglia.

von Reden, Sitta. 2015. *Antike Wirtschaft*. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783486852622>

Winship, Christopher and Robert D. Mare. 1992. Models for sample selection bias. *Annual Review of Sociology* 18(1): 327–350. DOI: <https://doi.org/10.1146/annurev.so.18.080192.001551>

Wooldridge, Jeffrey M. 2009. *Introductory Econometrics. A Modern Approach*. Ohio: South-Western, Cengage Learning.

