

The Population of European Cities from 700 to 2000

Social and Economic History

Eltjo Buringh

Utrecht University, Utrecht, Netherlands

e.buringh@uu.nl

Abstract

This article presents an expanded dataset of the historical urban population in Europe, *European urban population, 700–2000* (<https://www.doi.org/10.17026/dans-xzy-u62q>). This dataset contains new and improved estimates of the urban population (in thousands of inhabitants) between the years 700 and 2000 in 2,262 European settlements, including European cities with more than 100,000 inhabitants. The dataset is based on previous historical demographic sources that have been critically assessed and systematically complemented with new population estimates for additional time windows, deriving from either quantitative sources or proxies. Missing data are covered by city-specific and time-specific imputations. The applied time windows are whole centuries before 1500 and half a century afterwards. The article discusses the robustness checks that have been performed to validate the reliability of the imputed numerical results.

Keywords

Europe – cities – population – inhabitants – middle-ages – (early)-modern period

Online publication date: 3-9-2021

- Related data set “European urban population, 700–2000” with DOI <https://www.doi.org/10.17026/dans-xzy-u62q> in repository “DANS”

1. Introduction

Historical data of city populations have been extensively used in the past by large numbers of scholars of diverse disciplines to speak on issues of long-term (economic) development, regional divergence, and even to quantify the influence of religious and political changes. Therefore, there will be a wide interest in an improved and expanded dataset of historical urban populations in Europe. This new dataset comprises a longer time series than any currently existing. New too, is its incorporation of up to now unused historical proxy information serving as a basis for very early population estimates. Missing population data have, for the first time, been imputed in a systematic and city-specific way. Furthermore, the medieval first and second nature characteristics of each city (which in the Middle Ages were often quite important for a city's long-run economic development) have been characterized at the city level. And last but not least, the whole structure of the new dataset allows its future users to easily filter out those population data or other data that do not meet their criteria of inclusion. This new dataset, therefore, has a lot more to offer than currently existing sets.

More than three decades after the seminal publication *The population of European cities from 800 to 1850* by Bairoch et al. (1988), it surely is time for an update. After the fall of the 'iron curtain' in 1989 with the later ensuing changes in the political landscape of Eastern Europe and all the new historical information that has become available since, a thorough revision of their work seems more than necessary. In this updated dataset of city sizes, I collected additional population data at several time windows not covered in Bairoch et al. (1988), namely, the years 700, 1100, 1550, 1650, 1900, 1950, and 2000. I also included, together with data about their previous demographic development, all European cities with over 100,000 inhabitants in the year 2000, and similarly all capital cities in that year. The final database comprises 2,262 European settlements, most of which have become cities. Figure 1 presents the distribution of settlements in the year 700 that met the inclusion criteria. Figure 2 shows how the more than two thousand current towns are located on the map of Europe. The population size estimates in the database are given in whole numbers of thousands of inhabitants. As always in historical demography, these estimates are of course less accurate when going back further in time and when they are not based on actual census data.

An update of Bairoch's dataset has now become feasible because urban archaeological, historical and demographic research published in the past decades has considerably enlarged our knowledge base. Furthermore, *Wikipedia* (2012–2019) has made existing (but previously often obscure) local historical information electronically accessible. This means that all in all, there now are

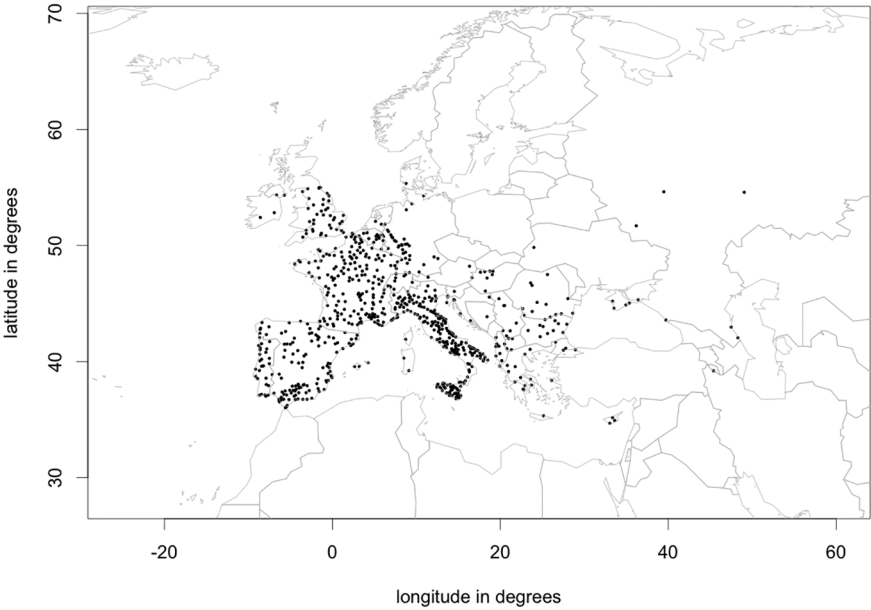


FIGURE 1 Location of settlements with one thousand or more inhabitants in the year 700

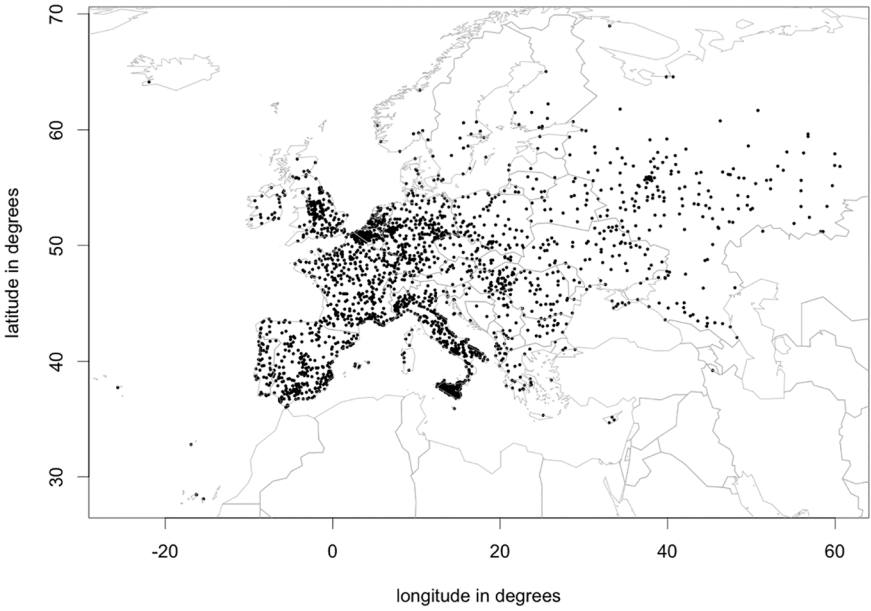


FIGURE 2 Location of settlements with one thousand or more inhabitants in the year 2000

a larger number and also more diverse sources available than back in the 1970s and 1980s when Bairoch and his coworkers collected the majority of their data. This enables me to use new proxies and also allows a motivated correction of previous anomalies in some of the historical population data.

This article is composed as follows: the first section on demographic data describes the proxies used to estimate urban population sizes. Next, it describes the following step: the imputation of the then still missing values with a universal algorithm. In the second section, I describe some robustness checks and discuss the results. These demonstrate that this new database of European city sizes stands up to scrutiny, and on a general level probably is quantitatively accurate. Annex A gives a brief description of the various units contained in the actual database and Annex B describes the geographical data that have been collected for the different European cities.¹

2. Demographic Data

- European urban population deposited at DANS – DOI: <https://www.doi.org/10.17026/dans-xzy-u62q>
- Temporal coverage: 700–2000

The most accurate numbers of urban inhabitants are those based on census data. Whenever available these were used. Data of censuses were collected for the year as close to that of the nearest time window as possible. However, in Europe, population censuses are limited to the last couple of centuries at most, and virtually non-existent before 1800. When no census data were available the currently existing population estimates generally have been used. Their sources were: Bairoch et al. (1988), Bosker et al. (2013), the *Encyclopaedia Britannica* (1910), *Lexikon des Mittelalters* (1999) and *Wikipedia* (2012–2019).²

¹ Both Annexes are available alongside the database via E. Buringh, *European urban population, 700–2000* at DANS: <https://www.doi.org/10.17026/dans-xzy-u62q>.

² Wikipedia as a source has its problems because its contents are not stable over time and its information also differs per language. During the period of data collection (2012–2019), Wikipedia has been consulted for every individual city in the database in English, German, French and Dutch, as well as in the local language of the country because such local language versions of Wikipedia often present census data of the town in question. Any historical population numbers other than actual census data, which were at times mentioned in Wikipedia, have not been trusted at first sight. Next to census data, Wikipedia has been used often for the collection of additional information, for example, when a town was first mentioned in a chronicle or historical document or to find out what its surface area was at various dates or whether it contained information on one of the other historical proxies that were used in this study. Because Wikipedia is not stable over time, the actual digital path to

I also collected information from the *Grote Winkler Prins* encyclopedia (Röst, 1990–1993), the *Great Soviet Encyclopaedia* (Prokhorov & Paradise, 1973–1982), and the *Encyclopaedia of Islam* (Gibb et al., 1975).

The updated numerical data based on proxies have been indicated in the dataset. When proxies were available the various numbers of medieval inhabitants in a city have been numerically estimated with historical proxy values (just as has been generally done in the past in the field of historical demography), such as a town's total inhabited urban surface area, the circumference of its town walls, how many towers a town wall contained or the number of its city gates, medieval tax data concerning numbers of houses or hearths, how many fighting men a town could mobilize, the amount of poll tax a town produced or the numbers of parishes in a town at a certain point in time. All such new and proxy-derived values that were quantitatively differing from previous estimates have been indicated in the database. An important proxy, whose usefulness up to now has not been fully exploited in historical demographic studies, is when a settlement gets mentioned in a chronicle or charter for the first time in history. When, for instance, was some medieval settlement recorded to have been plundered? Or when did a place get its first mint, market rights, city rights, or received its first bishop? Between 700 and 1500, such historical events often occurred for the first time in Europe when a town had a minimal size of around one thousand inhabitants. This allows inferences about its population when a city's size is unknown but such events have been recorded in a historical source.

As a second step, the missing values between two dates of quantified proxies or censuses have been imputed with a universal, but city-specific and time-specific algorithm for each of the more than two thousand European cities included in the database. Such imputed values have also been indicated in the dataset. Later on, both steps will be carefully described and evaluated by comparing the obtained results with other historical data that were not used for the dataset. Although the various checks show that the obtained dataset of city sizes probably can be trusted at a general level, individual medieval cities may have had considerably different numbers of inhabitants from the proxy-estimated or imputed values. For more city-specific studies, the reader is therefore urged to look for other data and proxies that this study undoubtedly will have missed, as it has not been possible to be exhaustive and cover all sources for every individual city. Such currently not used other proxies may lead to a better-founded and probably more accurate estimate for the city

the information in question has not been included in the database, as an exact path to varying information would not have secured its retrieval. Therefore, in the relevant cell of the dataset, the reference is made just to Wikipedia in general.

in question.³ However, at a more general area-level or when averaged over a longer time period, the undoubtedly present deviations from historical reality in this database due to the imputation procedure or the use of proxies will more or less even out.

This study has also tried to critically examine previously published city sizes with over ten thousand inhabitants (indicated as '10 k' for short, with the use of 'k' as an indication of kilo with a value of one thousand, a standard prefix for SI units of measure). When necessary these city sizes were corrected based on new information. In Bosker et al. (2013) this has been done already for medieval cities with more than 60 k inhabitants (such as Cordoba, Palermo, and Paris) and now I also looked critically at smaller cities. The French city of Laon will be presented here as a typical example to show how my update, based on habited surface areas, has led me to adjust its previously estimated early medieval population, which was based on a different proxy (parishes), downwards considerably. According to Bairoch et al. (1988, p. 26), in the year 900 Laon had – a for the tenth century enormous number of – 28 k inhabitants. (For the next time window, a century later, they indicate 25 k inhabitants in Laon). Bairoch's source for this population size of 28 k is Chandler and Fox (1974, p. 113), who have based themselves on a French tourist guide of Laon from 1896, which indicated that in 900 Laon had 23 parishes. This number of parishes was multiplied by the average number of people in a parish. Generally, Chandler and Fox (1974) have used 3 k inhabitants per medieval parish (pp. 114, 118, 119, etc.) to estimate the population of a town; however, they used other values too, such as 1.25 k (p. 113) per parish.⁴ Though this lower proxy value was used to estimate the size of Laon, it still made it the most populous city in tenth-century France.

That this last claim is hard to believe, can be corroborated by local archaeological research reported by Gaillard and Jorrand (2006, p. 165). They indicate that around 700 AD Laon had a rather limited habited surface area of some 20 ha. For Laon's urban surface area around half a millennium later, Russell (1958) gives a value of 42 ha. If the claimed population of 28 k for Laon in 900 is believed, this then leads to urban population densities of somewhere between 1,400 to 667 inhabitants per ha, which in the early Middle Ages were

3 I am interested in keeping the database as up-to-date as possible, and hope in the future to include any improvements that will become available, for example, because new proxies have been applied in one of the cities or other currently to me unknown sources have been unearthed. Any suggestions for improvements of specific city size in a century, including the sources where to find the underlying numbers or proxies, will be accepted thankfully at e.buringh@uu.nl.

4 Nevertheless, both values of 3 k or 1.25 k per parish used by Chandler and Fox (1974) are higher than the considerably lower 0.4 k to 0.45 k inhabitants per average fourteenth-century parish in France and England suggested in a more recent study (Buringh, 2011, p. 289).

densities unheard of. (See Figure 3 for actual medieval population densities of on average one hundred inhabitants per hectare: 0.1 k/ha). The use of the city-size-related proxies and my calculations have motivated me to reject the previous population estimates in Laon around the year 900. This rejection can be further validated by other sources. A historical study by Lusse (1992) indicates that at the turn of the millennium there were some twenty churches in Laon and the surrounding Laonnais area, nine of those churches were located outside the town walls then protecting a habited surface of 36 ha.⁵ Possibly four of these churches were built before 1000. According to the proxies used here, this would suggest a population that might have risen to 4 k in 1000 in Laon. (The database currently estimates a value of 3 k inhabitants in Laon around 1000 with the imputation procedure.) Parisse and Leuridan (1994) corroborate this adjusted view and indicate that Laon then had one cathedral and two abbeys making its size rank (based on the number of religious institutions per city in Parisse and Leuridan 1994) end up somewhere in the lower middle bracket of French cities, and certainly not the largest in tenth-century France.

This typical example of the population of the city of Laon around the year 1000 could be elaborated with that of numerous other medieval cities where the previous values of historical demographic populations have not been taken at their face value but challenged with other and often more reliable (proxy) data. However, a full description of all these cases for all time windows would lead to a virtually unreadable pulp of words that would interest only a minute fraction of users. Therefore, the various sources of proxies have been indicated as briefly as possible in cablese (see Annex A). The second important conclusion from the Laon example is that even literature referenced historical demographic data can be superseded by probably more accurate estimates that were based on proxies and imputations. The fact that some urban population has been proxied or imputed does not necessarily make the so obtained values less reliable.

3. Two-step Procedure of Size Updates

3.1. *Proxies*

First of all, I used several proxies concerning the specific study period to estimate missing population numbers. Per habited urban surface of one hectare,

5 I could not find a copy of the actual 1896 source of Chandler and Fox (1974) in any Dutch library, so I could not check whether it contained the numbers of parishes in the Laonnais area in general or only those in the city of Laon itself. Based on Lusse (1992), I would suspect that Chandler and Fox's source could have meant the Laonnais area.

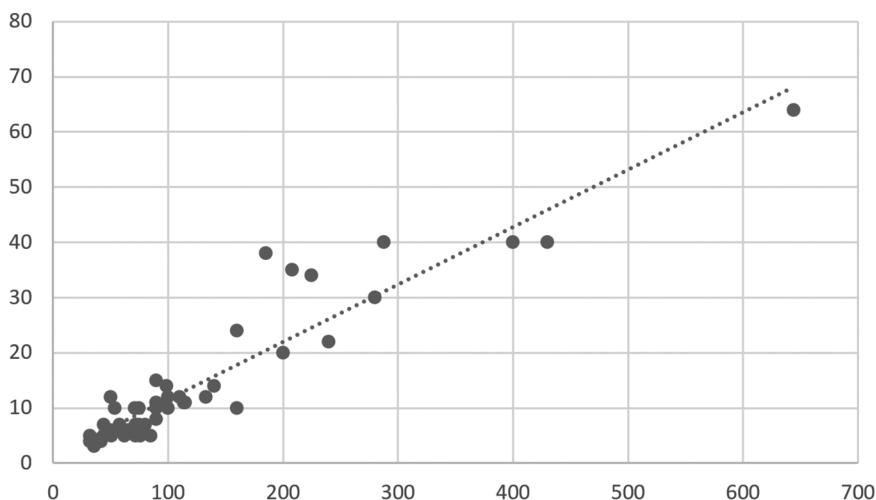


FIGURE 3 Calibration of habited surface area (in ha, x-axis) and population (in thousands, y-axis) in 57 medieval European towns based on Bairoch et al. (1988)

MAIN SOURCES FOR SURFACE AREAS: RUSSELL (1958, 1972), *LEXIKON DES MITTELALTERS* (1999), AND CHANDLER AND FOX (1974).

I have assumed a medieval habitation of a round figure of 100 persons (0.1 k/ha). Numerically such a value is corroborated by a linear relationship found between the medieval urban population in Bairoch et al. (1988) and the average habited urban surface areas, resulting in an R-square of 0.88 in 57 European towns (see Figure 3 for the calibration data).

Of course, this number of 0.1 k/ha is just an average that seems to be right in general, but that may have had rather different values in individual cities and periods (see the substantial scatter of a factor of two around the regression line in Figure 3). Therefore, one should consider the range of population densities in medieval cities to span half this value in considerably less densely populated towns (circa fifty persons per hectare, or 0.05 k/ha) to around double the average value of 0.1 k/ha. For instance, for medieval Islamic cities in southern Europe sometimes higher densities of up to 0.15–0.2 k/ha may be warranted. When other densities than 0.1 k/ha have been used for population estimates based on a surface proxy—for example, in a large and densely populated city as medieval Paris a value of 0.125 k/ha was used—, this has been indicated in the source column in the dataset.

The values of the other proxies that have been used in this database, such as the average number of persons per parish (0.4 k to 0.45 k/parish), the number of persons per household or hearth (4 to 5.5 /household-hearth), or the fraction with which the number of potential male soldiers has been multiplied to find the average number of urban inhabitants (somewhere around 5 to 6 /

soldier) have been indicated, whenever relevant for the calculations. Another source of information was medieval and early-modern travel accounts, which at times compared sizes of cities or described the miserable state of a city after it had been conquered and razed to the ground. Such eyewitness accounts can sometimes be highly informative (see, e.g., van Bavel et al., 2018).

Next to the use of proxies to estimate population numbers at a specific date, I also specifically collected information (placed to the nearest time window) on the first time a city got mentioned, for instance, in a medieval chronicle, or because it became a bishopric, obtained its mint, officially got city rights or perhaps was granted market rights. Of course, if a specific city size (or proxy) was mentioned when one of the above events happened, this value was recorded in the database. Only when no other size-related information was found than the first mentioning of a town, the number of 1 k inhabitants was recorded as a default value for the town of the time window in question. The underlying assumption is that such events generally started to happen when the population in a settlement had reached some minimal size, possibly somewhere between roughly five hundred to fourteen hundred inhabitants, which rounded off would become 1 k inhabitants in the database.⁶ Also when a town was mentioned for the first time as having been ravaged by either Vikings, Magyars, or Mongols (and having no other size information) it was given a default value of 1 k, because such plunder would not have been worthwhile to record for posterity if it would not have had enough inhabitants to make such an event worthwhile for the attackers (one could argue that in some sense this is similar to when a town gets mentioned for the first time).

3.2. *Imputation*

The second step of the two-step process is to impute the missing time window-values between two dated numbers of inhabitants in a town. In Bairoch et al. (1988), as well as for cities for which I used proxies for their size classification, actual population estimates at quite a number of other time windows are missing due to a lack of data or proxies. It is, however, very well possible to impute those missing values without any further assumptions than that of a uniform growth rate between the two whole-century dates for which there are quantified

6 Actual historical confirmation for this default value of 1 k is found with Brussels and Montpellier in the 1000s; Marienberg in the 1300s; Stroud, Londonderry, Helsinki, and Montemaggiore in the 1600s; Hagen in the 1700s; Tampere and Sterlitamak in the 1800s; and, finally, Wolfsburg, Elista, and Latina in the 1900s. All these cities got mentioned for the first time in historical sources when they had around one thousand inhabitants.

values of a city size.⁷ In three consecutive centuries, for instance, two proxy- or census-based values for a certain city may have been found with one missing in between: “a. b”, in which “a” is a number and “b” is a different integer number in thousands of inhabitants. For two or more missing values in consecutive centuries this may have been: “a . . b”; or “a . . . b”, etc. The assumption of one continuous growth rate in the time period from a century with “a” inhabitants to the century with “b” inhabitants means that the missing values can be simply imputed for the geometrically averaged growth rate for a specific city between the two dates in question. For one missing value, this procedure then leads to:

$$(b/a)^{0.5} \cdot a.$$

For two consecutive missing values, it leads to:

$$(b/a)^{0.333} \cdot a \text{ for the first, and } ((b/a)^{0.333})^2 \cdot a \text{ for the second.}$$

For n consecutive missing values, it leads to respectively:

$$(b/a)^{1/(n+1)} \cdot a; ((b/a)^{1/(n+1)})^2 \cdot a; \dots \text{ and } ((b/a)^{1/(n+1)})^n \cdot a.$$

Whenever the year 1400 was involved in this imputation, a slightly different procedure was followed. During the Black Death (whose first and most devastating and then also economically most influential bout ravaged from 1347–1351 in Europe) approximately one third of the inhabitants of all age groups and wealth classes died.⁸ To find the numbers of inhabitants in 1400 (if there was

7 That such normally distributed city-specific growth rates were a characteristic symptom of the growth of cities in the past is affirmed by the general use of log-normal distributions to describe the prevalence of historical city sizes (see, e.g., Eeckhout [2004], and the following debate his paper generated). As a rule, log-normal distributions are the result of natural growth processes that are driven by small normally distributed percentage changes. In this imputation procedure, by letting the numerical population data (between size “a” and “b”) decide on the actual value of a city-specific growth rate over time, eventually also similar log-normal size distributions of medieval city sizes may be expected to arise. A limited sample (N=39) of the estimated growth rates derived from the imputation procedure and calculated as a yearly growth percentage of the urban population, also showed to be distributed nearly normally.

8 Ryan Johansson (2003, p. 225). Though the plague paid recurrent visits to the Latin West (and the Middle East) in the following centuries, its devastating shock effect on demography and society was less severe later on, as parts of the population then had managed to develop some resistance. Therefore such later bouts of the plague had a less devastating influence on the economy as a whole and are not included in this algorithm. Cohn (2002) describes that

no other numerical size information for the city, of course), I used a fraction of two thirds of the numbers of inhabitants in 1300. This loss of life is based on the negative influence of the Black Death on the European population. If a city's population in 1400 was unknown, the numbers of its inhabitants in 1300 were multiplied by $2/3$ to obtain an estimate of its size in 1400. For this, I used the imputed "1300-value" (found with the above algorithm) to calculate its post-Black Death "1400-value".

With the now calculated 1400-value, I started the imputation process anew with the 1400-value as the new number "a" and the original value "b" as "b", for a second and separate application of the algorithm for the missing values.

In practice, this simple algorithm turns out to be a powerful tool for imputing missing values. It is completely city-specific and time period-specific, as it only uses the actual data for the city in question for the time period when the population changed from "a" to "b". The algorithm is solely based on the available numerical information of the city in question, and this algorithm does not need any other assumptions concerning a city's demographic or economic development. Because all missing numbers of these more than two thousand European cities have been imputed in the same way, this imputation procedure will not influence future mutual comparisons between (sub)samples of cities in specific centuries or areas contained in the database.

3.3. *Special Cases*

The database spans the period 700 to 1500 in steps of one century and from 1500 to 2000 in steps of half a century. However, to calculate the population development in the various cities, I started collecting information from around 500 AD, the first turn of a century after the demise of the Roman Empire. Because the quantitative influence of the Justinian plague in 541–42 AD is not known, the population values obtained for 500 and 600 AD will not have any real historical meaning, and these two centuries have been omitted from the final database, which therefore starts with the numbers of inhabitants estimated around 700 AD.

Following the first mentioning of a city, all towns with Roman roots indicated in Talbert (2000) have been assigned a value of 1 k inhabitants in 500 AD (if there was no other size information, of course). For towns with Greek or Phoenician roots (as indicated in Talbert [2000] or *Wikipedia* [2012–2019]) I followed a slightly different course and assumed that the imputed value of their population in the year 900 will have been applicable too to the two

"mortalities declined rapidly after the first bout of plague in 1347–1351, and by the end of the century it had become largely a disease of children" (p. 238).

earlier centuries (700 and 800), as such very ancient towns will probably have reached a population plateau at an earlier date than the Roman towns that were generally founded later. The 87 towns involved are indicated in the database. The towns that Jedin et al. (1987) classified in 600 AD as having then been a metropole in the church hierarchy of the western church or as having then been a patriarchate in the eastern church, have all been assumed to have had 4 k inhabitants in 600 AD (if there was no other size information). The 21 towns involved are indicated in the database. Needless to say that for all ancient Roman and pre-Roman towns the imputation procedures for missing values have been the same as those used for the later medieval cities.

4. Some Robustness Checks of the Results and Discussion

The above-described two-step process of the use of proxies followed by imputation of still missing values of population numbers has produced a database of city sizes. Question is, whether the so obtained numerical results make sense. Therefore, I compared subsamples of the constructed database with several independent proxies for historical population numbers, which were not used to estimate the values in the database. For the very early period in the eighth century, I compared the numbers of churches and cathedrals in towns in western Europe with their population sizes in 700 and 800 AD. For England, there is numismatic information on the distribution of finds of stray coins that were minted around the year 1000. And of course, I compared the imputed values of medieval town sizes with their surface areas (as far as these values have not had a prior use in the proxy-derived calculation of the medieval town sizes).

4.1. *Basilicas and Cathedrals in the Eighth Century*

The numbers of churches in a town can be seen as a proxy for the numbers of inhabitants in medieval Christian cities, similar to the number of parishes, which at times has been used as a proxy for this estimation. In Gauthier and Picard (1986–2014), the numbers of cathedrals and basilicas in western European towns around 750 AD have been indicated on 118 late-antique/early-medieval town plans; these maps were drawn on a scale of 1:12,500 and based on local archaeological research. Jedin et al. (1987) present similar town plans on which churches of post-classical Rome, Ravenna, and Milan have been indicated. All churches on these maps and dating from before 750 AD have been counted and the numbers of churches per town have been correlated with a town's population number in the database (when a town was missing from the database but found in Gauthier and Picard [1986–2014], it was classified

as having zero inhabitants in the database). The number of churches in the eighth century can explain 55% of the variance in the population numbers in 700 AD and 60% of the variance in the population numbers in 800 AD ($N = 121$). These 121 towns concern churches in the larger eighth-century cities, spread over western Europe and located in current-day France, Switzerland, Germany, Belgium, Luxembourg, the Netherlands, and Italy.

Though the correlation is not perfect, it still is high enough to warrant the inclusion of the new city size estimates from 700 AD into the database. The somewhat higher correlation in the database for the population sizes in the year 800 suggests that these later figures possibly are slightly more accurate than those approximated for the previous century. That both correlations are lower than those in the other robustness checks might also have been caused by “noise” in the maps in Gauthier and Picard (1986–2014). Some of the maps did not indicate any early medieval churches built before 750, while certainly a number of these towns (Nîmes, Nice) were known to have been a bishopric or an archbishopric (Eauze) in 600 AD (Jedin et al., 1987, pp. 22–23), so at least one basilica/cathedral (and probably even some more churches) would have been expected there.⁹ A second reason for finding a lower correlation is the fact that uncertain churches have sometimes been left out of the maps (such as in Trier), while for other cities they have been indicated on the maps (such as in Metz). Because I counted all indicated churches on the maps, this also adds extra noise to the data.

4.2. *English Coins and Urban Population in 1000*

It may be expected that the number of stray coins found in a city has some direct relationship with the number of individuals that have once used these coins. As a second check, I, therefore looked at English data on coin finds reported by Palliser (2000). Surviving numbers of coins minted between 973 and 1066 have been associated with the numbers of urban inhabitants in the year 1000 in England as estimated in the database with the two-step procedure. This shows that for the 40 towns for which additional numismatic information could be found, differences in the numbers of local coins can explain 75% of the variance in English population numbers. Also here I assumed that towns not present in the database had a population of zero inhabitants in 1000. This correlation is a further confirmation of the accuracy of the two-step procedure followed here to estimate medieval population sizes.

9 Possibly a lack of actual local archaeological excavations has led to the maps where these churches were missing.

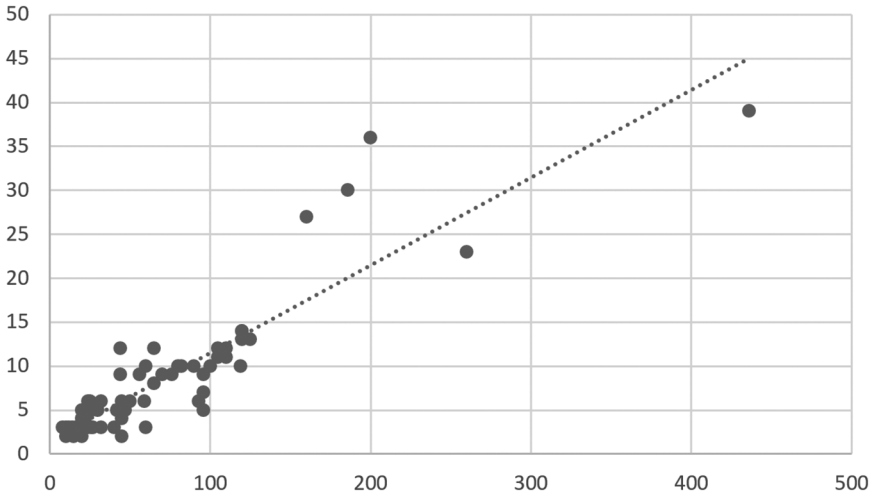


FIGURE 4 Correlation between habited surface area (in ha, x-axis) and population (in 000s, y-axis) in 65 medieval European towns determined with the method of imputation
MAIN SOURCES FOR SURFACE AREAS: RUSSELL (1958, 1972), *LEXIKON DES MITTELALTERS* (1999), AND CHANDLER AND FOX (1974).

4.3. *Urban Population and Imputation, between 900–1500*

In the database, there are 65 towns (between 900 and 1500) for which a medieval surface area is available that was not used as a proxy to estimate its number of inhabitants, while there is an estimate of their imputed population size in that same century. This allows the assessment of the correlation between city sizes in ha and the imputation procedure. This turns out to have an R-square of 0.79 (see Figure 4).

As might be expected, the square of the correlation coefficient is slightly lower in Figure 4 than that of the calibration in Figure 3. The still considerable value of 79% explained variance nevertheless shows that the imputation method is a versatile algorithm that produces results closely matching what would be expected. Especially the fact that the average population density (determined with the regression line in Figure 4) and based on the relationship between the surface area in ha and the imputation method numerically also comes to approximately 0.1 k/ha is quite reassuring.

4.4. *Imputations and Proxies*

Table 1 presents the number of cities in the dataset and the distribution of the two main methods used to characterize urban population numbers when they were not taken from a source that was deemed sound. The Table shows that the total number of settlements in the year 700 harbouring at least one thousand

TABLE 1 Absolute numbers of urban populations and estimation methods

Year	Total #	Imput.	Proxy	Other
700	731	0	729	2
800	872	716	149	7
900	1029	844	176	9
1000	1259	973	246	40
1100	1494	1198	285	11
1200	1771	1329	363	79
1300	1909	1298	246	365
1400	1977	1546	113	318
1500	2020	1366	72	582
1550	2031	1957	22	52
1600	2091	1169	62	860
1650	2120	2038	25	57
1700	2160	988	36	1136
1750	2180	1043	18	149
1800	2214	142	45	2027
1850	2220	193	7	2020
1900	2244	313	37	1894
1950	2248	540	2	1706
2000	2252	11	0	2241

inhabitants rises nearly threefold from a mere 731 to the number of 2,252 in the year 2000.¹⁰ In the early period, reliable numerical population data are very scarce and nearly all urban demographic estimates have been based either on proxies or on imputations. Around 1700 the majority of the urban population figures comes from other sources than proxies or imputations. This rises to some 90% for the data from 1800 onwards when census data have become more common.

The example of the number of inhabitants of Laon in 900 and 1000 has shown that its imputed and proxied population numbers are certainly no less in quality than the previously accepted quantitative data reported in various literature references. There even are some arguments in favour of the imputed and proxied values in Laon around the first millennium compared to those

10 This number is lower than 2,262 because some settlements that had over a thousand inhabitants at some point between 700 and 2000 had no inhabitants any more in the last time window or may have been absorbed into a larger agglomeration.

from previous literature references. Therefore, for the early periods covered by this study, the large fraction of such imputed or proxied values probably is something we will have to live with. Imputation helps to spread city growth evenly over the time windows with missing data. Generally, such an assumption will not be too far from the truth. However, there may have been exceptional and specific circumstances in a city when a sudden very high growth spurt should not be spread out evenly over time. Also, the reverse could be true when wars, plagues or other disasters have suddenly influenced population numbers negatively in a certain city. Due to such short-term influences, anomalies of which users should be aware may arise, as the imputed or proxied population numbers are less variable because they are averaged out more. I would like to stress that the long-term growth rates that are calculated with the imputation method are not caused by “inertia” and a consequence of the algorithms used, but are growth rates based on real numerical differences in inhabitants at two different points in time.

Even when disregarding the influence of proxies or imputations, some additional aspects have to be considered too. One is that cities by growing not only increase in population numbers but also in their spatial dimensions by incorporating the originally (non-urban) surrounding countryside into their new city area. Therefore already early-Victorian London in the nineteenth century is located on a surface area multiple times that of its Anglo-Saxon predecessor at the beginning of the study period. For London, such an urban sprawl implies that previously independent neighbouring settlements in the database such as Greenwich and Deptford eventually ended up being a part of London. When Wuppertal was created in 1929 in Germany, it encompassed the up to then-independent settlements in the dataset of Barmen, Cronenberg and Elberfeld. In Hungary, the originally separate cities of Buda and Pest eventually have grown into one: Budapest.

Furthermore, statistical agencies of individual countries have their own rules for assigning settlements into a certain urban agglomeration. Such different rules do not simplify comparisons of cities between countries. For instance, in the UK the definition for urban agglomerations classifies all settlements with more than 1,500 people on at least 20 ha into the same agglomeration when their areas are less than 200 m apart. Imposing such a strict distance criterion in a densely populated country like the Netherlands would probably have left us with only a few urban agglomerations, instead of the 60 individual Dutch cities now in the dataset.

A second additional aspect concerns the fact that historical population numbers in the Middle Ages and the early-modern period could vary considerably in a very short term due to local disasters. When values have been

imputed or proxied, such temporary circumstances probably have been missed. Depending on the use one wants to make of it, one has to be aware of these limitations of the dataset.

5. Conclusion

The various checks performed on the estimated population numbers between 700 and 1500, all largely agree with what is known from other sources, and therefore I think that on a general level the results are probably accurate. However, as explained before, at individual city level medieval population estimates based on proxies can easily differ a factor of two – see the considerable scatter in the data points in Figures 3 and 4. Especially the numbers of urban inhabitants estimated for the first millennium will be the least reliable, because in this period also proxies were more scarce. Later on in the Middle Ages – in the first half of the second millennium – the various urban population estimates have been backed up with more proxies and will have therefore become more reliable. The population numbers at the windows nearer to the early-modern period probably are slightly more reliable still than those of the actual Middle Ages, while finally those of more recent periods should be considered to be the most accurate, as they are largely based on census data.

References

- Bairoch, P., Batou, J., & Chèvre, P. (1988). *La population des villes européennes, 800–1850: Banque de données et analyse sommaire des résultats = The population of European cities, 800–1850: data bank and short summary of results*. Librairie Droz.
- Bosker, M., Buringh, E., & van Zanden, J. L. (2013). From Baghdad to London: Unraveling urban development in Europe, North Africa and the Middle East, 800–1800. *The Review of Economics and Statistics*, 95(4), 1418–1437. https://www.doi.org/10.1162/REST_a_00284.
- Buringh, E. (2011). *Medieval manuscript production in the Latin West*. Brill.
- Chandler, T., & Fox, G. (1974). *3000 years of urban growth*. Academic Press.
- Cohn, S. R., Jr. (2002). *The Black Death transformed: Disease and culture in early renaissance Europe*. Arnold.
- Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 94(5), 1429–1451.
- Encyclopaedia Britannica. (1910). *Encyclopaedia Britannica* (11th ed.). 28 vols.

- Gaillard, M., & Jorrand, J.-P. (2006). Laon. In , L. Pietri , B. Beaujard , J. Biarne , R. Brulet , E. Dabrowska , M. Gaillard , C. Mériaux , & , M.-T. Raepsaet-Charlier (Eds.), *Topographie Chrétienne des cités de la Gaule, des origines au milieu du VIII^e siècle*: Vol. XIV. *Province ecclésiastique de Reims (Belgica secunda [pp.163–170])*. Impr. de Boccard.
- Gauthier, N., & Picard, J.-C. (Eds. [with several subsequent editors]). (1986–2014). *Topographie Chrétienne des cités de la Gaule, des origines au milieu du VIII^e siècle*. 16 vols. Impr. de Boccard.
- Gibb, H. A. R., Kramers, J. H., Lévi-Provençal, E., & Schacht, J. (Eds.). (1975). *Encyclopédie de l'islam*. (2nd ed.). 12 vols. Brill.
- Jedin, H., Scott Latourette, K., & Martin, J. (Eds.) (1987). *Atlas zur Kirchengeschichte: Die christlichen Kirchen in Geschichte und Gegenwart* (rev. ed. by J. Martin). Herder. *Lexikon des Mittelalters*. (1999). 9 vols. J. B. Metzler'sche Verlagsbuchhandlung.
- Lusse, J. (1992). *Naissance d'une cité, Laon et le Laonnais du Ve au Xe siècle*. Presses Universitaires de Nancy.
- Palliser, D. M. (Ed.). (2000). *The Cambridge urban history of Britain: vol. 1, 600–1540* (P. Clark , Gen. Ed.). Cambridge University Press.
- Parisse, M., & Leuridan, J. (Eds.). (1994). *Atlas de la France de l'an mil*. Picard.
- Prokhorov, A. M., & Paradise, J. (Eds.). (1973–1982). *Great Soviet encyclopedia*. 32 vols. Macmillan.
- Röst, L. C. M. (Ed.). (1990–1993). *Grote Winkler Prins. Encyclopedie in 26 delen* (9th ed.). 26 vols., with supplements 1994 and 2002. Elsevier.
- Russell, J. C. (1958). *Late ancient and medieval population*. Transactions of the American Philosophical Society, vol. 48, pt. 3.
- Russell, J. C. (1972). *Medieval regions and their cities*. Indiana University Press.
- Ryan Johansson, S. (2003). Epidemics. In J. Mokyr (Ed.), *The Oxford encyclopedia of economic history* (vol. 2, pp. 223–233). Oxford University Press.
- Talbert, R. J. A. (Ed.). (2000). *Barrington atlas of the Greek and Roman world*. Princeton University Press.
- van Bavel, B., Buringh, E., & Dijkman, J. (2018). Mills, cranes and the great divergence: the use of immovable capital goods in western Europe and the Middle East, ninth to sixteenth centuries. *The Economic History Review*, 71(1), 31–54.
- Wikipedia*. (2012, January, to 2019, August).