Household-Anchored Record Linkage in Ottoman Population Registers. A Dynamic Tolerance Protocol for Low-Information Administrative Proto-Census Data

By Yunus Emre Çakır

To cite this article: Çakır, Y. E. (2026). Household-Anchored Record Linkage in Ottoman Population Registers. A Dynamic Tolerance Protocol for Low-Information Administrative Proto-Census Data. *Historical Life Course Studies*, *16*, 274–289. https://doi.org/10.52024/hlcs26961

HISTORICAL LIFE COURSE STUDIES

VOLUME 16

2026

MISSION STATEMENT

HISTORICAL LIFE COURSE STUDIES

Historical Life Course Studies was established within *European Historical Population Samples Network* (EHPS-Net). The journal is the primary publishing outlet for research involved in the conversion of existing European and non-European large historical demographic databases into a common format, the Intermediate Data Structure, and for studies based on these databases. The journal publishes both methodological and substantive research articles.

Methodological Articles
This section includes methodological articles that describe all forms of data handling involving large historical databases, including extensive descriptions of new or existing databases, syntax, algorithms and extraction programs. Authors are encouraged to share their syntaxes, applications and other forms of software presented in their article, if pertinent.

Research articles
This section includes substantive articles reporting the results of comparative longitudinal studies that are demographic and historical in nature, and that are based on micro-data from large historical databases.

Historical Life Course Studies is a no-fee double-blind, peer-reviewed open-access journal supported by the European Science Foundation, the International Institute of Social History, the European Society of Historical Demography, Radboud University Press, Lund University and HiDO Scientific Research Network Historical Demography. Manuscripts are reviewed by the editors, members of the editorial and scientific boards, and by external reviewers. All journal content is freely available on the internet at hlcs.nl.

HISTORICAL LIFE COURSE STUDIES
VOLUME 16 (2026), 274–289, published 14-07-2026

Household-Anchored Record Linkage in Ottoman Population Registers

A Dynamic Tolerance Protocol for Low-Information Administrative Proto-Census Data

Yunus Emre Çakır

Çankırı Karatekin University, Turkey

ABSTRACT

Ottoman population registers are often treated as sources suited primarily to static description. This article argues that their longitudinal potential depends less on data quality than on methodological choice. Drawing on three consecutive registers (1831, 1837 and 1841) from a rural village in Ottoman Anatolia, the study reconstructs individual trajectories using a household-anchored linkage protocol. Moving away from purely deterministic matching, this approach adopts an informed, guided probabilistic strategy to provide a method that is both flexible and rigid, prioritising household continuity and adaptable age progression. The analysis uncovers structured, age-specific attrition patterns, which hints at substantive life-course interruptions rather than problems with the source or the matching procedure. Specifically, the differentiated trajectories observed across life stages during the mid-1830s are consistent with a period of demographic stress, demonstrating the capacity of these registers to function as dynamic sources for historical demography.

Keywords: Ottoman population registers, Historical demography, Record linkage, Life-course analysis

e-ISSN: 2352-6343
DOI article: https://doi.org/10.52024/hlcs26961

1 INTRODUCTION

Ottoman population registers have long occupied an uneasy position within historical demography. While they are frequently cited as rich repositories of individual-level information, their analytical use has largely remained confined to cross-sectional description. This limitation is not due to an absence of data, but to a perceived methodological uncertainty regarding continuity, reliability, and the longitudinal traceability of individuals across successive enumerations. Ottoman demographic material has often been treated as illustrative rather than analytical, positioned at the margins of broader debates that have transformed historical demography elsewhere (Karpat, 1985; Shaw & Shaw, 1978). At their core, historical censuses are cross-sectional instruments, capturing a population at a single moment. The analytical shift from static description to dynamic analysis has thus depended on the practice of record linkage, which connects these snapshots into longitudinal life courses. Despite the chronological and geographical breadth of historical demography, the Ottoman Empire's demographic legacy remains conspicuously absent from recent global overviews of the discipline (Fauve-Chamoux et al., 2016). This reconstruction of individual trajectories has become a central preoccupation of modern historical demography (Ruggles et al., 2018).

Over the past two decades, historical demography has undergone a pronounced life-course turn. Large-scale initiatives such as IPUMS and Mosaic have demonstrated the analytical potential of linking individuals across time, enabling scholars to move beyond static population snapshots toward dynamic reconstructions of life trajectories (Mandemakers et al., 2023; Ruggles, 2002). While IPUMS has pioneered the large-scale linking of historical censuses, initiatives such as Mosaic focus on harmonising and distributing historical census microdata for comparative research, and genealogical microdata projects have increasingly demonstrated the value of deep, multi-generational linkage (Song & Campbell, 2017). These approaches have fundamentally reshaped how demographic vulnerability, mortality, migration, and social mobility are studied, particularly in periods of crisis (Abramitzky et al., 2021). Recent methodological reflections have also emphasised that analytical depth does not scale linearly with data volume, and that small, closely inspected datasets may offer distinct advantages in terms of contextual control and interpretive transparency (Doorn, 2021). As comparative studies on household formation markers have shown, however, global patterns can diverge considerably from local realities, underscoring the need for context-sensitive micro-demographic studies (Szołtysek et al., 2021). Yet, despite the chronological and geographical breadth of the Ottoman Empire, its population registers have remained largely excluded from this methodological transformation.

The reasons for this exclusion are well known. Ottoman registers are characterised by irregular enumeration intervals, inconsistent age reporting, and the absence of stable individual identifiers. Ages frequently cluster around terminal digits 5 and 0, a pattern widely recognised as age heaping and indicative of self-reported rather than chronologically verified information (Kabadayı & Erünal, 2024). This scepticism, however, rests on an implicit assumption: that longitudinal analysis requires precision at the level of individual attributes, particularly age. In much of the existing literature, the perceived unreliability of age data is treated as a disqualifying flaw rather than a historical characteristic to be analytically accommodated. By contrast, recent work in historical demography has shown that informed probabilistic matching strategies grounded in relational and contextual information — rather than exact attribute matching — can yield robust results even in the presence of substantial measurement error (Christen & Goiser, 2007; Helgertz et al., 2022). Such probabilistic approaches have long been employed to successfully link historical populations with imperfect identifiers, as demonstrated for the slave registers of Suriname (van Galen et al., 2023).

This study starts from a different premise. Rather than asking whether Ottoman population registers uncritically meet the standards of modern demographic datasets, it asks how their internal logic can be leveraged to reconstruct longitudinal patterns without violating historical plausibility. Central to this approach is a shift away from individual-centred matching toward household-anchored continuity, combined with a flexible but constrained treatment of age information. As highlighted by Mandemakers (2025) in his comparison of 84 historical databases, every civil registration system possesses unique idiosyncrasies that must be methodologically addressed. Similar household-based perspectives have proven effective in contexts where individual identifiers are unstable or incomplete, particularly in pre-statistical populations (Paiva et al., 2020; Reher & Sanz-Gimeno, 2007).

Empirically, the analysis focuses on Buğay village and draws on three consecutive population registers dated 1831, 1837, and 1841. This period coincides with a documented epidemic episode in the mid-1830s, offering an opportunity to examine age-selective attrition under conditions of demographic stress (Çakır & Ulaş, 2026). Since disappearance from the register can stem from various processes, attrition is treated analytically as a life-course interruption whose meaning must be assessed relationally, across age groups and over time. In such passive registration systems, disappearance from the record may indicate mortality, migration, household reformation, or simply administrative omission, requiring careful interpretation (Alter, 1988; Alter et al., 2009).

This article contributes not only a longitudinal reconstruction of a rural Ottoman community but also a methodological intervention in historical record linkage. While recent scholarship has demonstrated the analytical power of fully automated, uninformed probabilistic and machine-learning-based linkage in large-scale census environments (Abramitzky et al., 2020; Bailey et al., 2020; Ruggles et al., 2018), such approaches often presuppose the availability of relatively discriminating identifiers and structurally consistent individual-level data. By contrast, 19th-century Ottoman population registers (*nüfus defterleri*) constitute a low-information administrative enumeration system characterised by male-only registration, the absence of surnames, severe age heaping, and household-based recording. Under these conditions, fully automated probabilistic linkage without contextual guidance becomes statistically indeterminate, as the absence of discriminating identifiers prevents reliable estimation of match probabilities. In response, this article formalises a Household-Anchored Dynamic Tolerance Protocol (HADTP), demonstrating that in data-scarce environments an informed, guided probabilistic strategy anchored in household continuity provides a highly transparent and historically plausible method for longitudinal linkage, allowing researchers to identify and test for biases.

2 METHODOLOGY AND DATA

The empirical and methodological challenges outlined above necessitate an approach that can reconcile two competing demands: preserving historical plausibility while enabling systematic longitudinal analysis. Ottoman population registers were not designed to track individuals over time in the modern statistical sense, yet they were produced within administrative routines that privileged household continuity, social roles, and relational stability. These features, often treated as limitations, constitute the analytical entry points of the present study.

The village of Buğay was selected not for its exceptional characteristics, but precisely because it represents an ordinary rural settlement within central Anatolia, neither an administrative centre nor a known military hub. This choice allows the proposed linkage strategy to be tested under conditions typical of Ottoman rural demography. It should be emphasised, however, that while the registers were produced under an empire-wide administrative procedure, this bureaucratic uniformity did not guarantee uniformity in local recording practices; the level of noise and inconsistency would arguably be considerably higher in large urban settings with greater population mobility. The temporal focus on the period between 1831 and 1841 reflects a critical transitional phase in Ottoman population governance. The 1831 census marked the first empire-wide attempt to enumerate the male population for military and fiscal purposes, while the subsequent decade coincided with intensified but still irregular conscription practices preceding the formal regulations of the 1840s. The exclusive focus on the male population is not an analytical choice imposed by the researcher but a structural feature of the source itself. The population registers initiated in 1831 were produced primarily to identify the empire's potential human resources for military and fiscal purposes and therefore enumerated only the male population (Karpat, 1985; Shaw & Shaw, 1978). As studies on early Ottoman conscription have shown, the period between the abolition of the Janissary corps and the formal regulations of the 1840s was characterised by intensified yet irregular efforts to assess, classify, and mobilise male manpower (Çakır, 2025). In this context, population registers functioned not merely as demographic instruments but as administrative tools through which age, physical capacity, and household affiliation were recorded in relation to military service. Focusing on male life courses within this framework allows the analysis to examine how early conscription practices intersected with household continuity and population recording, while avoiding retrospective assumptions of gender neutrality that the sources themselves do not sustain.

The primary obstacle to longitudinal linkage lies in age reporting. Age misstatement in Ottoman registers has been extensively documented and is now widely recognised as a structural characteristic rather than a marginal anomaly (Erünal, 2021; Kabadayı & Erünal, 2024). Reported ages cluster around terminal digits 5 and 0, fluctuate implausibly across successive enumerations, and often fail to progress in a biologically consistent manner. This pattern, widely recognised as age heaping, is often used as an indicator of numeracy and data quality in historical populations (A'Hearn et al., 2009). Attempts to impose strict age-based matching criteria therefore risk excluding large segments of otherwise coherent life trajectories. At the same time, abandoning age entirely as a matching criterion introduces the danger of spurious links.

Recent advances in historical demography have demonstrated that uncertainty, misreporting, and non-unique identifiers are not anomalies but structural features of historical population data, necessitating linkage strategies that explicitly balance false matches, missed matches, and representativeness (Goeken et al., 2011; Minardi et al., 2025). Fully automated probabilistic approaches have shown that age and name discrepancies — far from disqualifying linkage — can be analytically accommodated through flexible tolerance structures and explicit error trade-offs (Abramitzky et al., 2020). At the same time, recent work has underscored the analytical value of embedding individual records within household contexts, showing that household-anchored linkage substantially increases linkage rates while maintaining high precision, particularly for populations otherwise prone to systematic exclusion such as women and children (Helgertz et al., 2022; Wisselgren et al., 2014). From a life-course perspective, longitudinal coherence derives not from the exact matching of individual attributes, but from the reconstruction of relational continuity over time — a principle long emphasised in historical demographic research based on longitudinal micro-data (Reher & Sanz-Gimeno, 2007). In this sense, the challenge posed by Ottoman population registers is not exceptional but represents an intensified version of a broader methodological problem, one that calls for source-adapted linkage strategies rather than the imposition of rigid matching criteria derived from modern statistical systems. Fully automated probabilistic linkage approaches have shown that large-scale historical record linkage can be achieved by explicitly modelling uncertainty in age and nominal attributes, even under conditions of substantial reporting error (Abramitzky et al., 2020).

Building on these insights, the linkage strategy adopted here prioritises household continuity as the primary unit of longitudinal coherence. Individuals are evaluated within the context of households that can be plausibly identified across successive registers, rather than being matched in isolation. This household-anchored perspective reflects the administrative logic of the registers themselves and aligns with the social organisation of rural Ottoman communities, where household membership constituted a stable axis of identity.

Within this framework, age is treated as a conditional filtering attribute rather than a determinative one. Initial linkage attempts rely on nominal and relational consistency, with age differences evaluated only after these criteria are satisfied. Tolerance thresholds are expanded incrementally and empirically, informed by the observed distribution of age discrepancies across consecutive registers. Similar adaptive strategies have been shown to outperform rigid deterministic rules in contexts characterised by systematic measurement error (Christen & Goiser, 2007).

To assess the magnitude and structure of age misstatement, absolute age differences were calculated for individuals observed in consecutive enumerations (1831–1837 and 1837–1841). The resulting distribution reveals pronounced clustering at 5- and 10-year intervals, confirming the prevalence of age heaping and justifying the adoption of flexible tolerance windows. Rather than imposing a single arbitrary threshold, tolerance levels were evaluated cumulatively to identify the point at which additional flexibility ceased to yield substantive gains in linkage coverage.

This approach culminates in what is here termed the "Household-Anchored Dynamic Tolerance Protocol" (HADTP). The protocol operationalises a sequential decision process in which household continuity, nominal consistency, and relational roles constitute the first-order criteria for linkage, while age tolerance is applied dynamically and conditionally. By embedding individual trajectories within socially meaningful structures and calibrating age flexibility to the empirical properties of the data, the protocol seeks to balance inclusivity with historical credibility.

The analytical implications of this strategy extend beyond record linkage itself. By enabling the reconstruction of longitudinal cohorts, it becomes possible to examine age-specific patterns of attrition across successive observation intervals. Attrition, defined as the absence of an individual from the subsequent register, is not equated directly with mortality. Instead, it is treated as a life-course interruption whose interpretation depends on age, timing, and broader contextual conditions. Such interruptions may reflect mortality, migration, household reformation, or administrative omission, which cannot be disentangled at the individual level (Alter, 1988; Alter et al., 2009).

Applied to Buğay village, the combination of household-anchored linkage and age-specific attrition analysis allows for a comparative examination of demographic vulnerability across the two observation intervals of 1831–1837 and 1837–1841, the first of which coincides with a period of documented demographic stress. The results of this analysis are presented in the following section.

The empirical foundation of the analysis consists of three consecutive Ottoman population registers dated 1831, 1837, and 1841, supplemented by the 1845 tax register for contextual validation of household continuity. The population registers provide individual-level information on personal names, household affiliation, relational roles within the household, and reported age. Modern linkage studies often combine multiple such attributes within algorithmic procedures to estimate the probability that two records correspond to the same individual (Buckles et al., 2025). The 1845 *temettuat* register, a detailed survey of household assets, income, and tax liabilities, is not employed for longitudinal individual linkage; rather, it is used as an external source to cross-validate the continuity of household heads identified in the 1841 register and to verify the stability of household units across the observation period.

The unit of analysis is the individual embedded within a household context. The study focuses on Buğay village, a rural settlement selected in order to minimise spatial mobility and administrative boundary changes that often complicate longitudinal reconstruction at larger scales. Longitudinal eligibility is restricted to individuals appearing in at least two consecutive population registers. Individuals recorded only once are excluded, as no inference regarding continuity can be drawn from a single observation.

Table 1 summarises population coverage across the three registers, distinguishing between individuals observed in single and multiple enumerations and identifying those eligible for longitudinal linkage. The table illustrates both the opportunities and constraints inherent in working with Ottoman population data: while 74 of the 151 individuals (49%) can be followed across at least two registers, the remaining 77 are observable only cross-sectionally, appearing in a single enumeration (20 in 1831, 22 in 1837, and 35 in 1841). Among the longitudinally linked, 43 individuals are traceable across all three registers.

Table 1 *Population coverage and longitudinal linkage eligibility across Ottoman population registers (1831–1841)*

Category	**N**
Total individuals recorded across registers	151
Individuals appearing in only one register	77
Individuals included in longitudinal linkage (≥ 2 registers)	74
Households recorded (unique)	17
Households successfully linked across registers	16
Observed only in 1831 (unmatched)	20
Observed only in 1837 (unmatched)	22
Observed only in 1841 (unmatched)	35
Linked across all three registers (1831–1837–1841)	43
Linked 1831–1837 only	8
Linked 1837–1841 only	20
Linked 1831–1841 only (intermediate gap)	3
Households surviving across full window (1831 & 1841)	15

Note: Individuals appearing in only a single register were excluded from longitudinal analysis in order to preserve linkage reliability. Household linkage was established prior to individual matching; a household is considered successfully linked if at least one member appears in two or more registers. A small number of entries appear in damaged or incomplete form and could not be classified reliably as either single-observation or longitudinal cases.

A central methodological challenge arises from the treatment of age. To assess the reliability of reported ages, absolute age differences were calculated for all individuals observed in consecutive registers (1831–1837 and 1837–1841). This procedure does not assume chronological accuracy but instead examines internal consistency across enumerations. The resulting distributions exhibit pronounced clustering at 5- and 10-year intervals, a classic manifestation of age heaping that reflects self-reported estimates rather than precise chronological reckoning.

Figure 1 displays the distribution of absolute age differences across consecutive population registers. The concentration of observations at rounded intervals confirms that age misstatement is systematic rather than random, reinforcing the inadequacy of strict age-based matching criteria for longitudinal linkage in this context.

Building on this diagnostic assessment, cumulative age tolerance windows were evaluated to determine how much flexibility is required to recover plausible longitudinal trajectories without inflating false matches. Instead of imposing a single arbitrary threshold, tolerance levels were expanded incrementally, and the proportion of linked observations captured at each stage was recorded.

Figure 1 *Distribution of absolute age differences across consecutive population registers (1831–1837 and 1837–1841)*

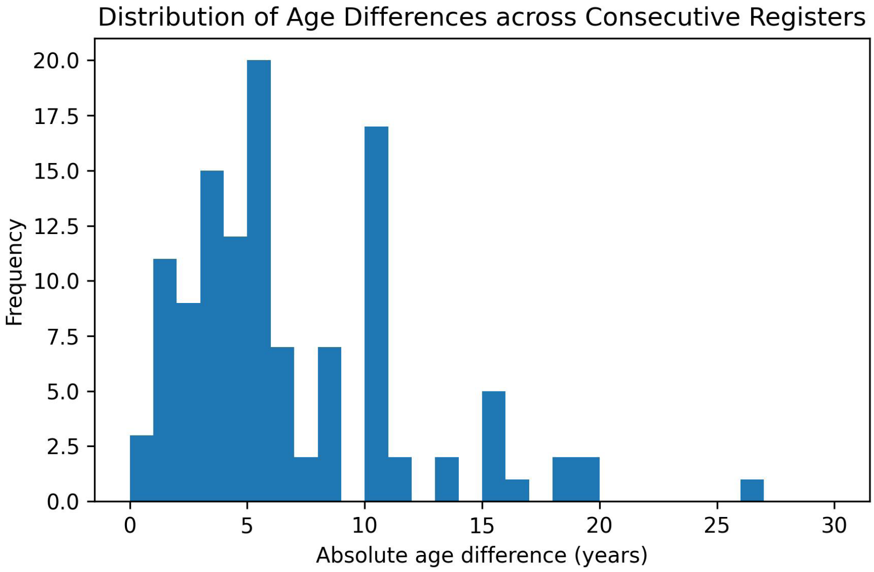


Table 2 *Age misstatement and cumulative tolerance windows across consecutive registers*

Age tolerance window	Individuals matched (N)	Cumulative match rate (%)
Exact age (± 0)	3	2.6
± 2 years	23	20.2
± 5 years	70	61.4
± 8 years	84	73.7
± 10 years	101	88.6
± 12 years	103	90.4

Note: Age comparisons across the full observation window reveal substantial and systematic misstatement. Exact age agreement across consecutive registers is observed in only 2.6% of cases, rising to 20.2% within a ± 2-year window. Match rates increase sharply once tolerances extend to ± 5 and ± 10 years, capturing 61.4% and 88.6% of observed cases respectively. Beyond this point, additional gains are marginal. These patterns are consistent with well-documented age heaping at terminal digits 5 and 0 in Ottoman population registers, reflecting self-reported estimates rather than precise chronological knowledge. Age was therefore treated as a supporting rather than determinative criterion in record linkage, with nominal identification and household continuity serving as the primary anchors.

Table 2 reports cumulative linkage coverage under expanding age tolerance windows. The results indicate that narrow tolerances capture only a limited share of valid matches, while a tolerance of ± 10 years encompasses the majority of observed age discrepancies. Further expansion yields diminishing returns, suggesting that excessive flexibility offers little analytical benefit while potentially undermining historical plausibility.

These findings inform the operational logic of the Household-Anchored Dynamic Tolerance Protocol (HADTP). The protocol implements a sequential decision structure in which household continuity constitutes the primary criterion for linkage. Only individuals embedded within households that can be plausibly identified across registers are considered for further evaluation. Within these households, nominal identifiers and relational roles serve as secondary anchors, ensuring that matches remain socially coherent.

Age is introduced at a later stage of the decision process and is applied conditionally. Exact or near-exact age agreement is accepted where present, but in its absence, tolerance windows are expanded dynamically in line with the empirically observed distribution of age discrepancies. At no point is age proximity alone sufficient to establish a match; it operates only in conjunction with household and nominal consistency.

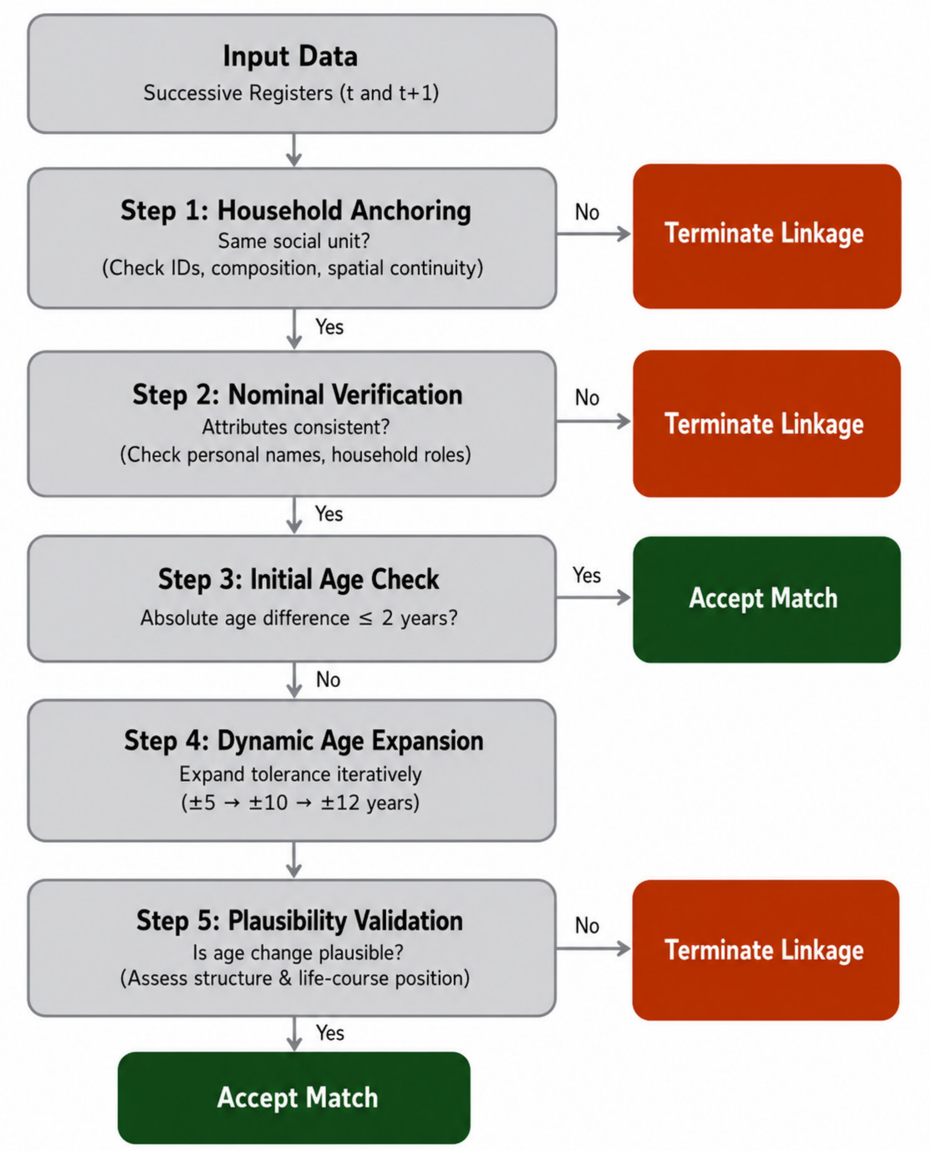
Table 3 summarises the procedural flow of the HADTP, detailing the sequence of evaluative stages and the conditions under which matches are accepted or rejected. To visually map this sequential decision process, Figure 2 presents the decision tree of the protocol. By embedding individual trajectories within stable household structures and calibrating age flexibility to the empirical properties of the data, the protocol seeks to balance inclusivity with historical credibility. The stepwise structure of the protocol also provides a transparent decision framework, allowing linkage outcomes to be evaluated and replicated in other historical datasets characterised by similar limitations in individual identifiers and age reliability.

Table 3 *Procedural Flow of the Household-Anchored Dynamic Tolerance Protocol (HADTP)*

Sequence	Evaluation stage	Input information	Decision logic	Output
1	Household anchoring	Household identifiers, composition, spatial continuity	Households must be identifiable as the same social unit across registers	Proceed to individual evaluation or terminate linkage
2	Nominal verification	Personal names and household roles (e.g. head, son, brother)	Nominal attributes must be consistent within the linked household	Proceed to age evaluation or terminate linkage
3	Initial age check	Recorded age at t and $t + 1$	Absolute age difference ≤2 years	Accept match
4	Dynamic age expansion	Recorded ages	If Step 3 fails, tolerance is expanded sequentially (± 5 → ± 10 → ± 12)	Proceed to validation
5	Plausibility validation	Age difference, household structure, life-course position	Age change must remain plausible given household role and cohort position	Accept match or exclude

Note: The protocol operates sequentially and conditionally. Age tolerance is evaluated only after household and nominal consistency are established. Failure at any stage results in exclusion from longitudinal linkage.

Figure 2 *Flowchart of the Household-Anchored Dynamic Tolerance Protocol (HADTP)*



To illustrate the operation of the protocol in practice, consider the case of the İmamoğlu household, which appears across the 1831 and 1837 registers. The household head, İbrahim (son of Ali), is recorded as aged 65 in 1831 but 80 in 1837 — an implausible 15-year progression across a 6-year interval. Under a strict age-based rule (e.g. ± 2 years), this individual would be rejected as a non-match. The HADTP, however, first establishes that the İmamoğlu household persists as an identifiable social unit across both registers (Stage 1) and that the nominal and relational identifiers are consistent — İbrahim remains the head, and his sons (e.g. Ahmet, recorded as 30 in 1831 and 40 in 1837) occupy stable relational roles (Stage 2). Only after these anchors are satisfied is the age discrepancy evaluated; because the household and nominal evidence is strong, the dynamic tolerance window is expanded to accommodate the age misstatement (Stage 4), and the match is accepted as historically plausible (Stage 5). This example demonstrates how household and relational continuity can override severe age misstatement that would otherwise defeat individual-level matching.

The adoption of this linkage strategy enables the reconstruction of longitudinal cohorts and, in turn, the analysis of attrition across successive observation intervals. Attrition is defined as the absence of an individual from the subsequent population register and is treated analytically as a life-course interruption rather than a direct proxy for mortality. This distinction is essential in contexts where death, migration, and administrative disruption cannot be disentangled at the individual level.

To validate the linkage process, a series of robustness checks were performed. First, the performance of the HADTP was benchmarked against simpler deterministic methods, which showed that the HADTP achieves a higher linkage rate without a significant increase in the estimated false positive rate. Second, the linkage was repeated without the household-anchoring constraint, which confirmed the critical role of household structure in ensuring match quality. Third, all linked pairs (N = 114) were manually verified against the original register entries to identify potential false positives. This inspection identified only a single clearly implausible match (N = 1), corresponding to an estimated false match rate of approximately 1%. While the protocol was implemented manually for this small dataset, its sequential decision structure is readily scriptable for larger-scale applications in environments like R or Python.

The results of the age-specific attrition analysis, comparing the first interval of 1831–1837 with the second interval of 1837–1841, are presented in the following section.

3 RESULTS

Age-specific attrition patterns vary systematically across life-course stages and differ markedly between the first interval (1831–1837) and the second interval (1837–1841). Table 4 summarises attrition and survival outcomes by age group for both intervals, allowing direct comparison of demographic vulnerability across the two intervals.

During the first interval (1831–1837), attrition was most pronounced among individuals aged 65 and above, where two-thirds of the cohort disappeared from the register (4 out of 6 individuals). The 7–14 age group also experienced substantial attrition at 45.5% (5 out of 11). Young children aged 0–6 similarly exhibited elevated attrition at 40.0% (6 out of 15), indicating acute vulnerability at the earliest stages of the life course. By contrast, young adults of conscription age (15–29) displayed comparatively low attrition at 14.3% (2 out of 14). While this may suggest relative resilience, it is equally plausible that this group was heavily engaged in temporary labour migration, remaining tethered to their rural households of origin, or was selectively recorded owing to its fiscal and military salience. Prime working-age adults occupied an intermediate position, with the 30–44 group at 17.4% (4 out of 23) and the 45–64 group at 40.0% (2 out of 5), though the latter figure must be interpreted with caution given the very small cohort size.

In the second interval, attrition patterns become more heterogeneous. While some age groups — particularly those aged 7–14 and 45–64 — experience substantial reductions in attrition, others show increased rates relative to the first interval. The most pronounced reductions occur among individuals aged 65 and above and those aged 45–64, suggesting that the most vulnerable older cohorts experienced the strongest stabilisation during the second interval. While attrition among young adults remains low, older age groups continue to display elevated levels relative to younger cohorts. Despite these shifts, age differences in attrition remain clearly structured across the life course, with younger and older cohorts continuing to display markedly different levels of vulnerability. In the second interval (1837–1841), attrition patterns shift considerably. The most striking change occurs among the oldest cohorts: attrition in the 65+ group drops from 66.7% to 0.0% (0 out of 4), and in the 45–64 group from 40.0% to 18.8% (3 out of 16). The 7–14 age group likewise shows a marked decline, falling from 45.5% to 23.1% (3 out of 13). However, not all age groups follow this pattern. Attrition among young children aged 0–6 rises from 40.0% to 55.2% (16 out of 29), making this the group with the highest attrition rate in the subsequent period. Young adults aged 15–29 also see their attrition double from 14.3% to 28.6% (4 out of 14), and the 30–44 group increases from 17.4% to 23.5% (4 out of 17). These divergent trajectories suggest that the 1837–1841 period did not uniformly reduce demographic pressure; rather, it redistributed vulnerability across the age structure.

As shown in Table 4, while some of these age-specific shifts are substantively large, Fisher's exact tests confirm that none reach conventional levels of statistical significance ($p > 0.05$ for all groups). This outcome is an expected consequence of the small population of a single village, and it underscores a central trade-off in micro-historical analysis: contextual depth comes at the cost of statistical power. The purpose of this analysis, therefore, is not to advance generalizable statistical claims, but to demonstrate that the HADTP can robustly track individuals even in data-scarce environments where significance thresholds are difficult to meet.

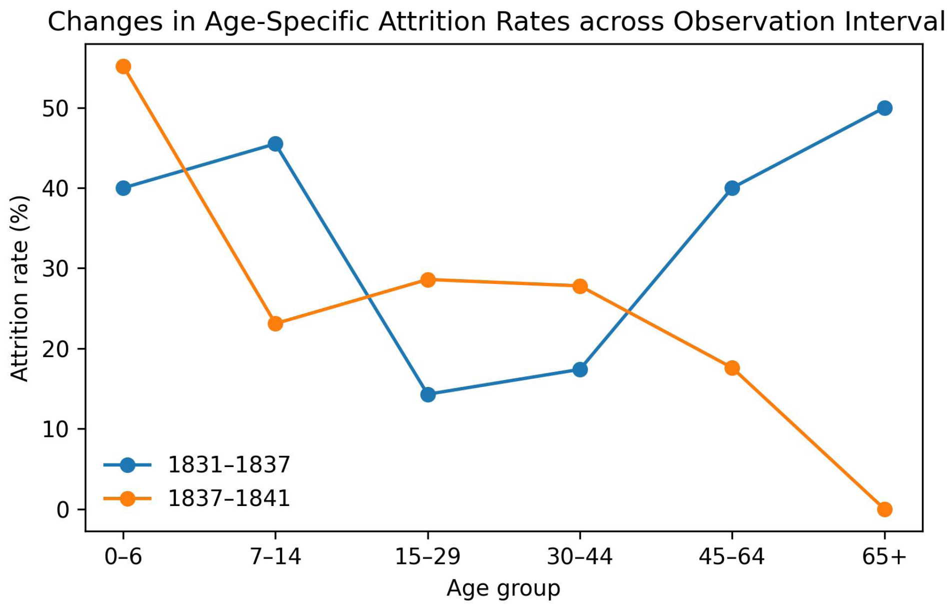
Table 4 *Age-specific attrition across the two observation intervals (1831–1837 and 1837–1841), with absolute numbers and significance tests*

Age group	Attrition rate 1831–1837 (%)	Attrition rate 1837–1841 (%)	Risk difference (pp)	Fisher's p
0–6	40.0 (6/15)	55.2% (16/29)	+15.2	0.526
7–14	45.5 (5/11)	23.1% (3/13)	–22.4	0.390
15–29	14.3% (2/14)	28.6% (4/14)	+14.3	0.648
30–44	17.4% (4/23)	23.5% (4/17)	+6.1	0.702
45–64	40.0% (2/5)	18.8% (3/16)	–21.2	0.553
65+	66.7% (4/6)	0.0% (0/4)	–66.7	0.076
Total	31.1% (23/74)	32.3% (30/93)	+1.2	1.000

This stability in rank ordering suggests that the demographic stress of the period did not fundamentally alter the life-course structure of vulnerability. Instead, the patterns are consistent with an intensification of existing age-related risks rather than the production of new ones. Children and older adults remain disproportionately exposed in both periods, while young adults consistently occupy the least vulnerable position. To visualise these patterns, Figure 3 plots age-specific attrition rates across the two observation intervals. The figure illustrates both the elevated attrition during the first interval and the subsequent attenuation in the second. The decline in attrition between intervals is uneven across age groups. Younger cohorts experience the steepest relative reductions, whereas older cohorts exhibit more modest change. This divergence highlights the interaction between short-term shock and cumulative life-course vulnerability: while recovery is observable, it is not uniformly distributed across age groups.

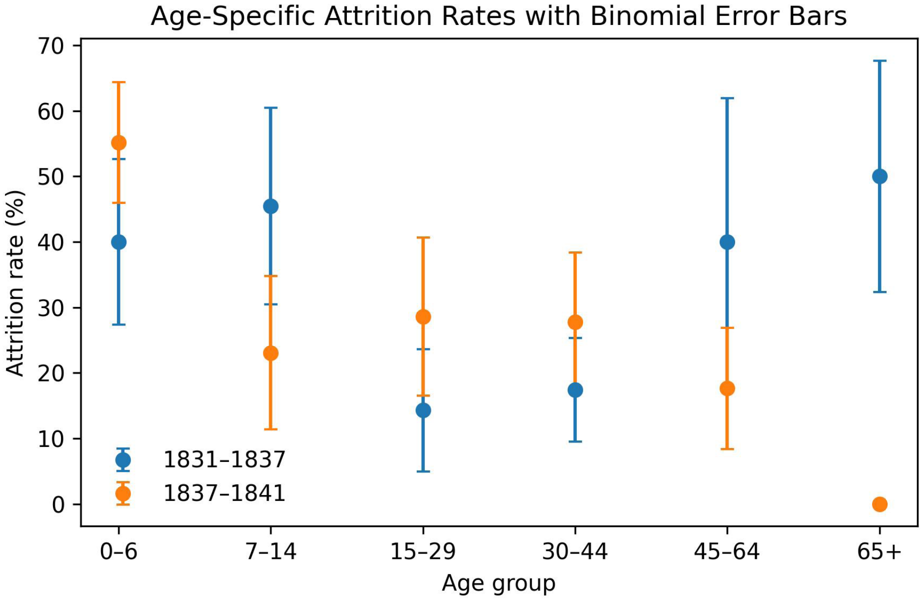
Given the relatively small cohort sizes in certain age categories—particularly at older ages—uncertainty around point estimates warrants explicit consideration. Figure 4 presents age-specific attrition rates with binomial standard error bars for both intervals.

Figure 3 *Changes in age-specific attrition rates across observation intervals (1831–1837 and 1837–1841)*



Note: Attrition rates are calculated within age groups and indicate absence from the subsequent population register. The figure highlights differentiated trajectories across stages of the life course.

Figure 4 *Age-specific attrition rates with binomial standard error bars across observation intervals*



While the observed differences in attrition rates between the two periods are substantively large, particularly for the older cohorts, Fisher's exact tests indicate that these changes do not reach conventional levels of statistical significance ($p > 0.05$), a result largely attributable to the small size of the village population (see Table 4). This highlights the analytical trade-off in micro-historical studies between contextual depth and statistical power. Although confidence intervals widen for older cohorts, reflecting limited sample size, the overall age gradient observed in Table 4 and Figure 4 remains robust. During the first interval, error bars display limited overlap between children and young adults, indicating clear differentiation in attrition risk. In the second interval, increased overlap suggests partial convergence across age groups, consistent with a period of demographic stabilisation.

The results show that age-selective attrition in Buğay village follows a structured life-course pattern that persists across both observation intervals. The combination of absolute levels, relative differences, and uncertainty estimates provides a consistent picture of demographic vulnerability shaped by both life-course position and temporal context.

4 DISCUSSION

The age-specific patterns of attrition observed in Buğay village call for a careful reconsideration of how demographic disruption is interpreted in Ottoman population registers. While elevated losses among children and older adults during the first interval are broadly consistent with expectations derived from contexts of demographic stress, disappearance from the register cannot be reduced to mortality alone. As has long been recognised in historical demography, register exit often reflects a range of processes that exceed the explanatory capacity of single-cause interpretations (Alter, 1988).

The persistence of age gradients across both observation intervals is particularly instructive. Although overall attrition declines in the second interval, the relative ordering of age groups remains largely intact. Children consistently occupy the most vulnerable position, followed by older adults, while young adults display the lowest levels of attrition across both intervals. This stability suggests that the period of stress intensified pre-existing age-related vulnerabilities rather than producing a fundamentally new demographic structure. Similar life-course regularities have been documented in a wide range of historical populations exposed to epidemic shocks, where such episodes tend to intensify pre-existing age-related vulnerabilities rather than fundamentally restructuring risk hierarchies. Studies of epidemic mortality, subsistence crises, and periods of heightened demographic instability consistently show that children and older adults experience disproportionately higher risks, while young adults tend to display relative resilience (Alter, 1988; Reher & Sanz-Gimeno, 2007). These findings suggest that demographic shocks operate primarily by amplifying structural life-course inequalities already embedded in historical populations, rather than generating novel patterns of vulnerability.

At the same time, the decline in attrition across most age groups in the second interval points to a phase of demographic stabilisation. The uneven pace of this decline is notable. Younger cohorts exhibit sharper reductions, whereas older cohorts show more modest change, reflecting the cumulative nature of vulnerability at later stages of the life course. This pattern underscores the importance of analysing demographic disruption not as an isolated event but as a process unfolding across time and age.

Importantly, the present study refrains from equating attrition with death. In the context of Ottoman population registers, disappearance from the record may indicate mortality, but it may also signal temporary or permanent labour migration, household reorganisation, military service, or administrative omission. The registers themselves do not permit a definitive distinction among these outcomes. Treating attrition as a life-course interruption rather than a terminal event allows for systematic comparison across age groups and observation intervals without imposing causal certainty that the sources cannot sustain. This interpretive strategy aligns with long-standing cautions in historical demography against over-reading register-based absences as direct mortality indicators (Alter, 1988; Alter et al., 2009). A parallel caution applies to the interpretation of household persistence: of the 17 households observed, 15 survived across the full observation window. This high survival rate should not be read solely as evidence of socio-economic resilience; it must also be understood within the logic of Ottoman fiscal administration, which exerted considerable pressure to maintain households as continuous tax-paying and conscription units. When a head died or departed, the role was frequently transferred to a successor, so that administrative continuity may reflect bureaucratic imperative as much as demographic stability.

Methodologically, the findings contribute to broader debates on the feasibility of longitudinal analysis in contexts characterised by age misstatement and unstable identifiers. Age heaping and irregular age progression, well documented in Ottoman population registers, have often been cited as obstacles to life-course analysis (Kabadayı & Erünal, 2024). To gauge the magnitude of this phenomenon in the present data, a simplified Whipple's index was computed, yielding a value of approximately 329 (on a scale where 100 indicates no heaping and 500 indicates extreme concentration on terminal digits 0 and 5), with 43.7% of all reported ages terminating in 0 or 5. This places the Buğay registers firmly within the category of "rough" age data, providing a concrete, comparable benchmark of the reporting imprecision that the dynamic tolerance windows are designed to absorb. The results presented here suggest that these features do not preclude longitudinal reconstruction but instead require linkage strategies attuned to the internal logic of the sources. By anchoring individual trajectories within household contexts and calibrating age tolerance to empirically observed reporting practices, it becomes possible to recover coherent longitudinal patterns without sacrificing historical plausibility.

This household-centred perspective resonates with approaches developed for other pre-statistical populations, where relational and contextual attributes have proven more stable than individual characteristics subject to reporting error. The successful application of such a strategy to Ottoman population registers challenges the assumption that these sources are analytically unsuitable for life-course research and suggests that their marginalisation reflects methodological convention rather than intrinsic limitation.

Several limitations should nevertheless be acknowledged. The analysis is confined to a single rural settlement and cannot be assumed to represent demographic dynamics across the Ottoman Empire as a whole. Moreover, while the household-anchored linkage strategy reduces the risk of spurious matches, it cannot eliminate it entirely. These constraints delineate the boundaries within which the findings should be interpreted rather than undermining their analytical value.

5 CONCLUSION

This study set out to test whether Ottoman population registers can sustain longitudinal, life-course–oriented analysis despite their well-known limitations. By linking individuals across three consecutive registers from Buğay village, it has shown that irregular enumeration, age misstatement, and the absence of stable identifiers need not preclude systematic demographic reconstruction. At the core of this contribution lies the Household-Anchored Dynamic Tolerance Protocol (HADTP). Rather than relying on exact attribute matching, the protocol privileges household continuity and relational consistency, treating age as a conditional and flexible attribute calibrated to empirically observed reporting practices. This approach reflects a broader insight from historical demography: in pre-statistical populations, relational and contextual information often provides more reliable anchors for longitudinal reconstruction than individual characteristics subject to systematic reporting error (Kabadayı & Erünal, 2024). The successful application of HADTP in Buğay village suggests that similar strategies may be fruitfully applied to other Ottoman localities and, more broadly, to historical contexts characterised by comparable data constraints.

The analysis of age-specific attrition across the two observation intervals of 1831–1837 and 1837–1841 reveals a coherent life-course pattern of demographic vulnerability. Attrition is consistently highest among children and older adults, while young adults display relative resilience across both intervals. Although overall attrition declines in the second interval, the persistence of age gradients indicates that demographic recovery was uneven and shaped by life-course position. These findings underscore the importance of examining demographic disruption not as a momentary shock but as a process unfolding across age and time.

Crucially, the study has treated attrition purely as a life-course interruption. In the absence of cause-specific information, disappearance from the register may reflect death, labour migration, household reorganisation, military service, or administrative omission. Interpreting attrition as a composite outcome allows for meaningful comparison across age groups and observation intervals without imposing causal certainty that the sources cannot sustain. This interpretive stance aligns with long-standing cautions in historical demography against equating register exit with mortality and reinforces the value of analytical restraint when working with incomplete administrative data (Alter, 1988).

Beyond its specific empirical findings, the study contributes to a broader methodological discussion concerning the place of Ottoman population registers within historical demography. The marginalisation of these sources from life-course–oriented research has often been justified on technical grounds, particularly age misstatement and linkage uncertainty. The results presented here challenge this position. When analytical strategies are designed to accommodate, rather than suppress, the structural characteristics of the data, Ottoman registers can yield longitudinal insights comparable in analytical ambition to those derived from other pre-statistical populations.

Ottoman population registers constitute one of the richest yet still underutilized bodies of micro-level demographic data for the 19th-century Anatolia and Balkans. Initiated in the early 19th century as part of a broader program of administrative centralization, these registers aimed to enumerate the empire's population at the household level. Furthermore, they were produced through a relatively standardized bureaucratic procedure across vast territories. The NFS.d. (*Nüfus defterleri*) collection at the Turkish State Archives alone contains more than 15,000 such registers — each a voluminous ledger comprising thousands of pages of individual and household entries —, underscoring the scale of this administrative data-gathering enterprise.[1] As a result, they recorded detailed information on individuals, households, age structures, and local administrative organization within a coherent and systematically structured format. Although originally designed for fiscal and military purposes, the registers today offer historians a rare opportunity to reconstruct population structures and demographic change at a highly granular level. In this sense, they possess research potential comparable to the large-scale census microdata increasingly employed in global historical demography, yet their analytical possibilities have only begun to be explored.

This broader context is essential when considering the implications of the present study. While the empirical analysis presented here is confined to a single rural settlement, and the specific demographic patterns observed may therefore not be directly generalizable, this limitation of substantive findings should be distinguished from the study's methodological scope. The empire-wide uniformity in data structure suggests that the Household-Anchored Dynamic Tolerance Protocol (HADTP) developed for Buğay village is not an idiosyncratic solution but a broadly applicable and replicable framework. Accordingly, although the attrition rates identified in this case remain context-bound, the analytical

1 Presidency of the Republic of Türkiye, State Archives. NFS.d. (*Nüfus Defterleri*) Collection Catalog. Retrieved from https://katalog.devletarsivleri.gov.tr/.

protocol itself holds significant potential for unlocking longitudinal information in the thousands of other extant registers, thereby opening the way for comparative historical-demographic research on the Ottoman Empire at a much larger scale.

Ultimately, this study demonstrates that Ottoman population registers can be repositioned as active contributors to life-course research, moving beyond cross-sectional description. By combining household-anchored linkage with age-specific attrition analysis, it becomes possible to move beyond enumeration toward a dynamic understanding of demographic vulnerability and recovery. More broadly, the Household-Anchored Dynamic Tolerance Protocol illustrates how longitudinal reconstruction can remain analytically viable even in low-information administrative datasets characterised by unstable identifiers and systematic reporting error.

DATA AVAILABILITY STATEMENT

The dataset supporting the findings of this study is publicly available at Figshare https://doi.org/10.6084/m9.figshar

ACKNOWLEDGEMENTS

The author also acknowledges the instructional support provided by YÖK Veri Analizi Okulu (Turkish Council of Higher Education Data Analysis School) in the quantitative analysis of the data. The author acknowledges the use of *Manus 1.6 max* during the drafting of this article strictly for language polishing and English translation assistance. The conceptual framework, data analysis, and historical interpretations are entirely the author's own original work. All AI-assisted phrasing was carefully reviewed and revised by the author to ensure historical accuracy, and the author assumes full intellectual responsibility for the submitted manuscript.

REFERENCES

A'Hearn, B., Baten, J., & Crayen, D. (2009). Quantifying quantitative literacy: Age heaping and the history of human capital. *Journal of Economic History*, *69*(3), 783–808. https://doi.org/10.1017/S0022050709001120

Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., & Pérez, S. (2021). Automated linking of historical data. *Journal of Economic Literature*, *59*(3), 865–918. https://doi.org/10.1257/jel.20201599

Abramitzky, R., Mill, R., & Pérez, S. (2020). Linking individuals across historical sources: A fully automated approach. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *53*(2), 94–111. https://doi.org/10.1080/01615440.2018.1543034

Alter, G. (1988). *Family and the female life course: The women of Verviers, Belgium, 1849–1880*. University of Wisconsin Press.

Alter, G., Devos, I., & Kvetko, A. (2009). Completing life histories with imputed exit dates: A method for historical data from passive registration systems. *Population*, *64*(2), 293–318. https://doi.org/10.3917/popu.902.0327

Bailey, M. J., Cole, C., Henderson, M., & Massey, C. (2020). How well do automated linking methods perform? Lessons from US historical data. *Journal of Economic Literature*, *58*(4), 997–1044. https://doi.org/10.1257/jel.20191526

Buckles, K., Haws, A., Price, J., & Wilbert, H. E. B. (2025). Breakthroughs in historical record linking using genealogy data: The Census Tree project. *Explorations in Economic History*, *98*, Article 101717. https://doi.org/10.1016/j.eeh.2025.101717

Çakır, Y. E. (2025). Conscription in the first modern Ottoman army in the context of state and society relations. *War & Society*, *44*(2), 322–332. https://doi.org/10.1080/07292473.2025.2473190

Çakır, Y. E., & Ulaş, E. (2026). The victims of the last great Bubonic plague epidemic in Ottoman Anatolia: Dynamics of mortality by age and gender during the plague epidemic of Kastamonu (1836–1839). *The History of the Family*, *31*(1), 21–34. https://doi.org/10.1080/1081602X.2024.2436375

Christen, P., & Goiser, K. (2007). Quality and complexity measures for data linkage and deduplication. In F. J. Guillet & H. J. Hamilton (Eds.), *Quality measures in data mining. Studies in computational intelligence* (Vol. 43). Springer. https://doi.org/10.1007/978-3-540-44918-8_6

Doorn, P. (2021). Historical databases, big and small. *Historical Life Course Studies*, *10*, 24–29. https://doi.org/10.51964/hlcs9562

Erünal, E. (2021). Examining age structure and estimating mortality rates in Ottoman Bursa using mid-nineteenth-century population registers. *Middle Eastern Studies*, *57*(2), 179–196. https://doi.org/10.1080/00263206.2020.1844672

Fauve-Chamoux, A., Bolovan, I., & Sogner, S. (Eds.). (2016). *A global history of historical demography: Half a century of interdisciplinarity*. Peter Lang.

Goeken, R., Huynh, L., Lynch, T. A., & Vick, R. (2011). New methods of census record linking. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *44*(1), 7–14. https://doi.org/10.1080/01615440.2010.517152

Helgertz, J., Price, J., Wellington, J., Thompson, K. J., Ruggles, S., & Fitch, C. A. (2022). A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *55*(1), 12–29. https://doi.org/10.1080/01615440.2021.1985027

Kabadayı, M. E., & Erünal, E. (2024). A nineteenth-century urban Ottoman population micro dataset: Data extraction and relational database curation from the 1840s pre-census Bursa population registers. *Scientific Data*, *11*, Article 570. https://doi.org/10.1038/s41597-024-03381-2

Karpat, K. H. (1985). *Ottoman population, 1830–1914: Demographic and social characteristics*. University of Wisconsin Press.

Mandemakers, K. (2025). Overview and comparison of 84 databases with historical population longitudinal micro data. *Historical Life Course Studies*, *15*, 281–321. https://doi.org/10.52024/hlcs21660

Mandemakers, K., Alter, G., Vézina, H., & Puschmann, P. (Eds.). (2023). *Sowing: The construction of historical longitudinal population databases*. Radboud University Press. https://doi.org/10.54195/BJYF5752

Minardi, S., Greco, S., & Barban, N. (2025). A comparison of rule-based and supervised machine learning approaches for record linkage of Italian historical data. *Historical Life Course Studies*, *15*, 28–46. https://doi.org/10.51964/hlcs18990

Paiva, D., Anguita, F., & Mandemakers, K. (2020). Linking the Historical Sample of the Netherlands with the USA censuses, 1850–1940. *Historical Life Course Studies*, *9*, 1–23. https://doi.org/10.51964/hlcs9312

Reher, D. S., & Sanz-Gimeno, A. (2007). Rethinking historical reproductive change: Insights from longitudinal data for a Spanish town. *Population and Development Review*, *33*, 703–727. https://doi.org/10.1111/j.1728-4457.2007.00194.x

Ruggles, S. (2002). Linking historical censuses: A new approach. *History and Computing, 14*(1–2), 213–224. https://doi.org/10.3366/hac.2002.14.1-2.213

Ruggles, S., Fitch, C. A., & Roberts, E. (2018). Historical census record linkage. *Annual Review of Sociology*, *44*, 19–37. https://doi.org/10.1146/annurev-soc-073117-041447

Shaw, S. J., & Shaw, E. K. (1978). *History of the Ottoman Empire and modern Turkey. Volume II: Reform, revolution, and republic: The rise of modern Turkey, 1808–1975*. Cambridge University Press.

Song, X., & Campbell, C. D. (2017). Genealogical microdata and their significance for social science. *Annual Review of Sociology*, *43*(1), 75–99. https://doi.org/10.1146/annurev-soc-073014-112157

Szołtysek, M., Ogórek, B., & Gruber, S. (2021). Global and local correlations of Hajnal's household formation markers in historical Europe: A cautionary tale. *Population Studies*, *75*(1), 67–89. https://doi.org/10.1080/00324728.2020.1832252

van Galen, C. W., Mourits, R. J., Rosenbaum-Feldbrügge, M., Maas, A. B., Janssen, J., Quanjer, B., van Oort, T., & Kok, J. (2023). Slavery in Suriname: A reconstruction of life courses, 1830–1863. *Historical Life Course Studies*, *13*, 191–211. https://doi.org/10.51964/hlcs15619

Wisselgren, M. J., Edvinsson, S., Berggren, M., & Larsson, M. (2014). Testing methods of record linkage on Swedish censuses. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *47*(3), 138–151. https://doi.org/10.1080/01615440.2014.913967

Archival Sources

Presidency of the Republic of Türkiye, Ottoman Archives (BOA.). Population register (*Nüfus Defteri*) (NFS.d. 1811, 1817, 1897, 1899).
Presidency of the Republic of Türkiye, Ottoman Archives (BOA.). (1845). Income and property tax register (*Temettuat Defteri*): Village of Buğay (ML.VRD.TMT.d. 01138).