UNIVERSIDAD DE SALAMANCA

Departamento de Estadística

Máster en Análisis Avanzado de Datos Multivariantes y Big Data

Trabajo Fin de Máster

Modelo de Cox y análisis de supervivencia basado en datos retrospectivos

Trayectoria de reunificación de pareja entre inmigrantes en España

Aurora: Mengyao Wu

Tutor: Dr. Javier Martín Vallejo

2023

UNIVERSITY OF SALAMANCA

Department of Statistics

Master of Sciences in Advanced Multivariate Data Analysis and Big Data

Cox model and survival analysis based on retrospective data

Partner Reunification Trajectory among Immigrants in Spain

Author: Mengyao Wu

Supervisor: Dr. Javier Martín Vallejo

2023

Acknowledgements

I would like to thank my supervisor, Dr. Javier Martín Vallejo, for his helpful guidance and incredible patience in completing this thesis. Working with him over the past few months has truly been a tremendous learning experience for me. My sincere gratitude also goes to Dr. Alberto del Rey Poveda, who provided valuable insights and constructive feedback on an earlier version of this thesis. In addition, I would like to thank Dr. Peng Li for his generous support in addressing my questions and concerns regarding statistical methods during my research stay at the Max Planck Institute for Demographic Research in Rostock, where much of the work on this study was done.

This thesis was supported by the University of Salamanca through the Scholarships for University Studies 2022-2023 for the completion of the MSc Advanced Multivariate Data Analysis and Big Data. The author is grateful for the financial support received.

Abstract

The Cox Proportional Hazards model and survival analysis are both important tools for time-to-event analysis in many disciplines. In particular, they have been widely used in sociological studies to analyze the time to the occurrence of life events and to identify relevant factors associated with the hazard of the event, where it is referred to as *event history analysis* (Allison, 1982; Yamaguchi, 1991). To better understand the application of these statistical methods in the social sciences, this thesis utilizes Cox regression models and survival estimation techniques to examine migrant couples' reunification process in Spain using retrospective data from the 2018 Spanish Fertility Survey. Specifically, we constructed Kaplan-Meier survival curves of reunification by sex of pioneer migrants and analyzed differences between the curves using the log-rank test. In addition, we performed univariate and multivariate Cox proportional hazards regression analyses to identify the potential explanatory factors associated with the relative risk of reunification. Hazard ratio (HR) values and forest plots were generated to present the results of the univariate and multivariate analyses, and scaled Schoenfeld residuals were used to test the Cox model. The results showed significant differences in the couple's reunification process by the gender of the first migrant, where female primary migrants tend to reunite with their partners earlier than their male counterparts. Potential covariates of age, childbearing status prior to migration, family economic status, duration of union, as well as partner's age at arrival are identified as significant factors related to the couple's reunification process. Overall, this thesis advances our understanding of the use of advanced statistical methods in other disciplines, and the novel empirical findings of this study contribute to the literature on family dynamics among immigrant populations in Spain.

Resumen

El modelo de riesgos proporcionales de Cox y el análisis de supervivencia son herramientas importantes para el análisis del tiempo transcurrido hasta la ocurrencia de un suceso. En particular, se han utilizado en estudios sociológicos para analizar el tiempo transcurrido hasta la ocurrencia del evento y para identificar un conjunto de variables explicativas, donde se denomina *event history analysis* (Allison, 1982; Yamaguchi, 1991). Para comprender de mejor manera la aplicación de estos métodos estadísticos en las ciencias sociales, esta tesis utiliza modelos de regresión de Cox y técnicas de estimación de supervivencia para examinar el proceso de reunificación de las parejas migrantes en España utilizando datos retrospectivos de la Encuesta de Fecundidad 2018. Para desarrollar estas ideas, en primer lugar, construimos curvas de supervivencia para mostrar el tiempo transcurrido hasta la reunificación en función del sexo de los inmigrantes pioneros y analizamos las diferencias entre las curvas a través de Log-rank test. En segundo lugar, realizamos análisis de regresión de riesgos proporcionales de Cox univariantes y multivariantes para identificar los factores explicativos asociados con el riesgo relativo de reunificación. Se generaron valores de hazard ratio (HR) y *forestplot* para presentar los resultados, y utilizamos residuos de Schoenfeld para comprobar el modelo de Cox. Los resultados mostraron diferencias significativas en el proceso de reunificación de la pareja según el sexo del inmigrante, donde las inmigrantes tienden a reunirse con sus parejas en menos tiempo que los inmigrantes. Las covariantes de la edad, la situación de procreación antes de la migración, la situación económica familiar, la duración de la unión, así como la edad de la pareja a la llegada se identifican como factores significativos relacionados con el proceso de reunificación familiar. En general, esta tesis contribuye a la mejor comprensión del uso de métodos estadísticos avanzados en otras disciplinas, y los novedosos resultados empíricos de este estudio contribuyen, a su vez, a la literatura sobre la dinámica familiar entre la comunidad inmigrante en España.

Contents

List of Figures

List of Tables

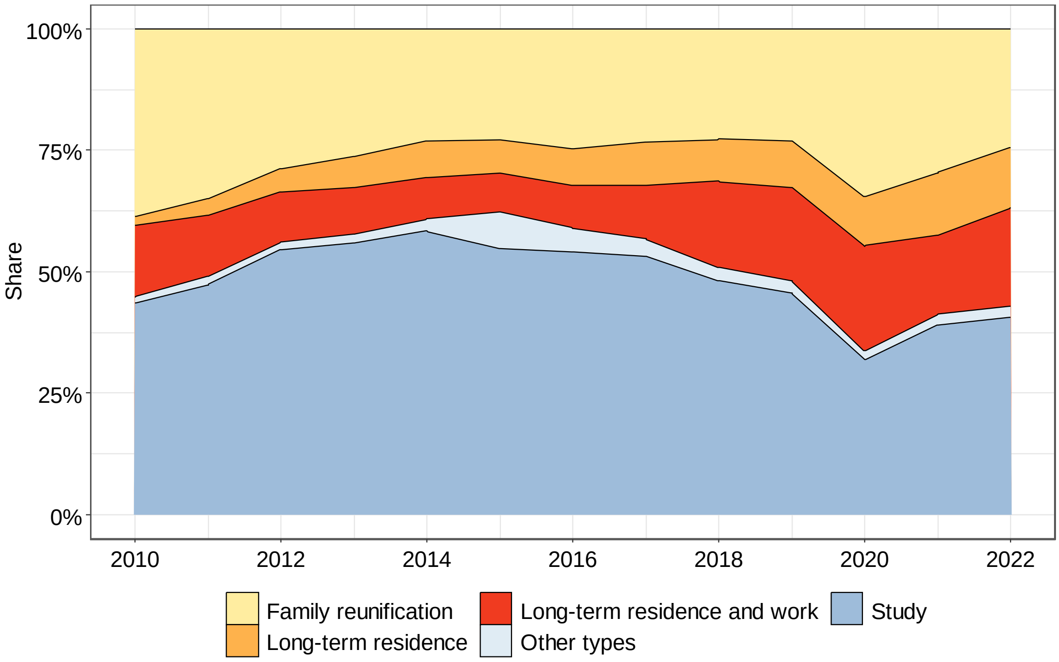
1

Introduction

Family reunification, which refers to the process by which migrants bring close relatives (i.e., spouses and children) to the country of immigration (Beauchemin et al., 2015), has gained prominence in European countries due to growing concerns about government policies and the integration process of reunited family members (Galeano & Gerber, 2023). Spain, as one of the newer immigration countries in Southern Europe since the 2000s, has also experienced significant family-related migration in recent years (Eremenko & González-Ferrer, 2018). According to the Ministry of Inclusion, Social Security and Migration (**Figure 1**), family reunification has become one of the main reasons for long-term visa applications, increasing from 23,044 in 2014 to 45,959 in 2022, representing 24.4% of the total number of residence permits granted in the country.

Family reunification, on the other side, is also a gendered process (Bonizzoni, 2015; Del Boca, 2002; Toma & Vause, 2014). In contrast to the traditional pattern of couple reunification, in which female dependents mostly followed the male breadwinner and were therefore treated as a secondary form of migration (Cerrutti & Douglas S. Massey, 2001; Kanaiaupuni, 2000), in recent years with women increasingly becoming key to the financial stability of their fa

Figure 1. Relative distribution of issued long-term visas in Spain, 2010-2022



Source: Author's calculations based on data from Ministry of Inclusion, Social Security and Migration.

-milies, migration flows have become more diversified in terms of the gender pioneer migrants (Estévez-Abe & Caponio, 2022; Kreyenfeld, Diehl, Kroh, & Giesecke, 2021). Indeed, Spain experienced a large increase in female-led migrations flows recently, in which men entering family-related migration as spouses or fiancés (as shown in **Table 1**), especially from societies with more equitable gender norms, such as Latin American countries (Bueno & Vidal-Coso, 2018; Del Rey & Grande, 2017; Heering, van der Erf, & van Wissen, 2004; Hervías Parejo, 2015).

Table 1. Issued Family reunification visas by gender in Spain, 2014-2022

	2014	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Male	8,554	10,033	11,200	12,027	13,544	15,351	13,067	15,835	17,241
Female	14,490	16,734	18,983	20,310	22,673	25,185	22,490	27,033	28,718
Total	23,044	26,767	30,183	32,337	36,217	40,536	35,557	42,868	45,959

Source: Author's calculations based on data from Ministry of Inclusion, Social Security and Migration.

Although men and women may experience reunification in very different ways (Kanaiaupuni, 2000; Morris, 2015), gender remains largely absent from important discussions about family reunification and integration in the EU context, and no systematic large-scale study has offered a gendered exploration in relation to family migration trajectories, especially in Southern European countries.

To fill this gap, the aims of this thesis are twofold: first, to study the process of couple reunification by the gender of pioneer migrants (male/female-led migration) by controlling for other covariates through Kaplan-Meier estimates; and second, to identify the factors that explain the timing of reunification of both partners in Spain by using Cox proportional hazards model, a multivariate model that allows one to predict time to event (in this case reunification) based on a set of predictor variables. This technique is particularly relevant in social science research as it allows the inclusion of both time-varying variables and censored time data, which greatly facilitates the empirical study of transitions over the life course, such as the transition to emancipation (Rey & Stanek, 2022), partnership formation and dissolution (González-Ferrer, Hannemann, & Castro-Martín, 2016), birth progressions (Lübke, 2015), and migration (Eremenko & González-Ferrer, 2018).

To develop these ideas, this thesis is organized as follows Chapter 2 reviews the relevant literature on family reunification by identifying the main factors related to the probability and timing of spousal reunification. Chapter 3 describes the data selection process and provides an overview of the sample. Chapter 4 presents our empirical analysis using Kaplan-Meier (KM) estimates and a Cox regression model. Additionally, a diagnostic test for the Cox model is provided using Schoenfeld residuals. In Chapter 5, we discuss the results of the analyses performed and provide closing remarks. With these analyses, our study aims to construct a Cox model to predict the relative risk or relative speed of spousal reunification and the results may contribute to a better understanding of reunification dynamics among immigrant groups and provide important implications for public policy on family reunification in Spain.

2

Literature Review on Spousal Reunification

Although the legal and financial constraints imposed by immigration policies in receiving nations largely influence the capacity of family members to migrate, the trajectory of spousal reunification has been shown to be a dual selection process (Battistella, 1995; Beauchemin et al., 2015), in which migrants, on the other side, also exercise their agency to comply with legal rules through self-selection (Bledsoe & Sow, 2011; Jasso & Rosenzweig, 1986). Recent studies have shown that individual factors, such as the sociodemographic characteristics of the first migrant, the spouse, and the household/union, also can mediate the pace of spousal reunification among migrants (Kraus, Sauer, & Wenzel, 2019; Toma & Vause, 2013).

First, an emerging strand of the family migration literature has highlighted the importance of cultural origin in explaining differences in the likelihood of spousal reunification (González-Ferrer et al., 2014; Mato Díaz & Miyar Busto, 2017). In particular, reunification tends to be more selective among couples from less gender-egalitarian countries (di Belgiojoso & Terzera, 2018; Lahav, 1997). Findings from a recent study in European countries, for example, show that reunification in both destination and origin countries is relatively rare for Senegalese couples due to the relatively high prevalence of polygamous unions (Baizán, Beauchemin, & González-Ferrer, 2014). In contrast, for migrants from more gender-egalitarian societies, spousal reunification in destination countries appears to be relatively common and rapid (Bonizzoni, 2015). For instance, a recent study in Spain found that immigrants from Eastern European countries and non-Andean Latin American countries were more likely to rejoin their partners than immigrants from Sub-Saharan, Asian, and Andean Latin American countries (Requena & Sánchez-Domínguez, 2011)

Second, the educational level and labor status of migrant couples, which are related to their ability to afford their migration trajectories to destination countries (Mato Díaz & Miyar Busto, 2017), also significantly affect the likelihood of family reunification (González-Ferrer, 2007). In particular, women with higher levels of education are more likely to join their husbands in destination countries than less educated women (González-Ferrer, 2007; Kanaiaupuni, 2000; Toma & Vause, 2013), as they have greater employment prospects and agency to realize their migration project rather than being left behind (Gupta, 2003). Similarly, migrant men with tertiary education are more likely to bring their partners to the destination country than those with primary education, as they have greater economic stability to afford the costs of family life in the host country (González-Ferrer, 2007; Mato Díaz & Miyar Busto, 2017). In addition, factors at the couple level, such as childbearing status, the household's economic situation, as well as the duration of the union, are also associated with the family reunification process, according to recent studies (Caarls & Mazzucato, 2015; González-Ferrer, 2007; Kraus et al., 2019),

Nevertheless, despite these findings from existing studies, there is no consensus regarding the relationship between the gender of the primary migrant and the risk of couple reunification in family studies (di Belgiojoso & Terzera, 2018). Although migration policies have long been constructed within a male-dominated framework (Hervías Parejo, 2015), where women are considered particularly disadvantaged in exercising their right to reunification due to the vulnerable labor status of female migrants in receiving countries (Kofman, 2004), recent empirical studies have shown that when women act as pioneers of family migration, the propensity of couples to reunite significantly increases and the duration of reunification is also shorter compared to male pioneer migrants (González-Ferrer, 2013). These mixed results add another layer of complexity to the study of family reunification. Therefore, by applying survival analysis, this study aims to provide additional insights to the literature on partnership dynamics through its methodological innovations as well as novel empirical findings.

3

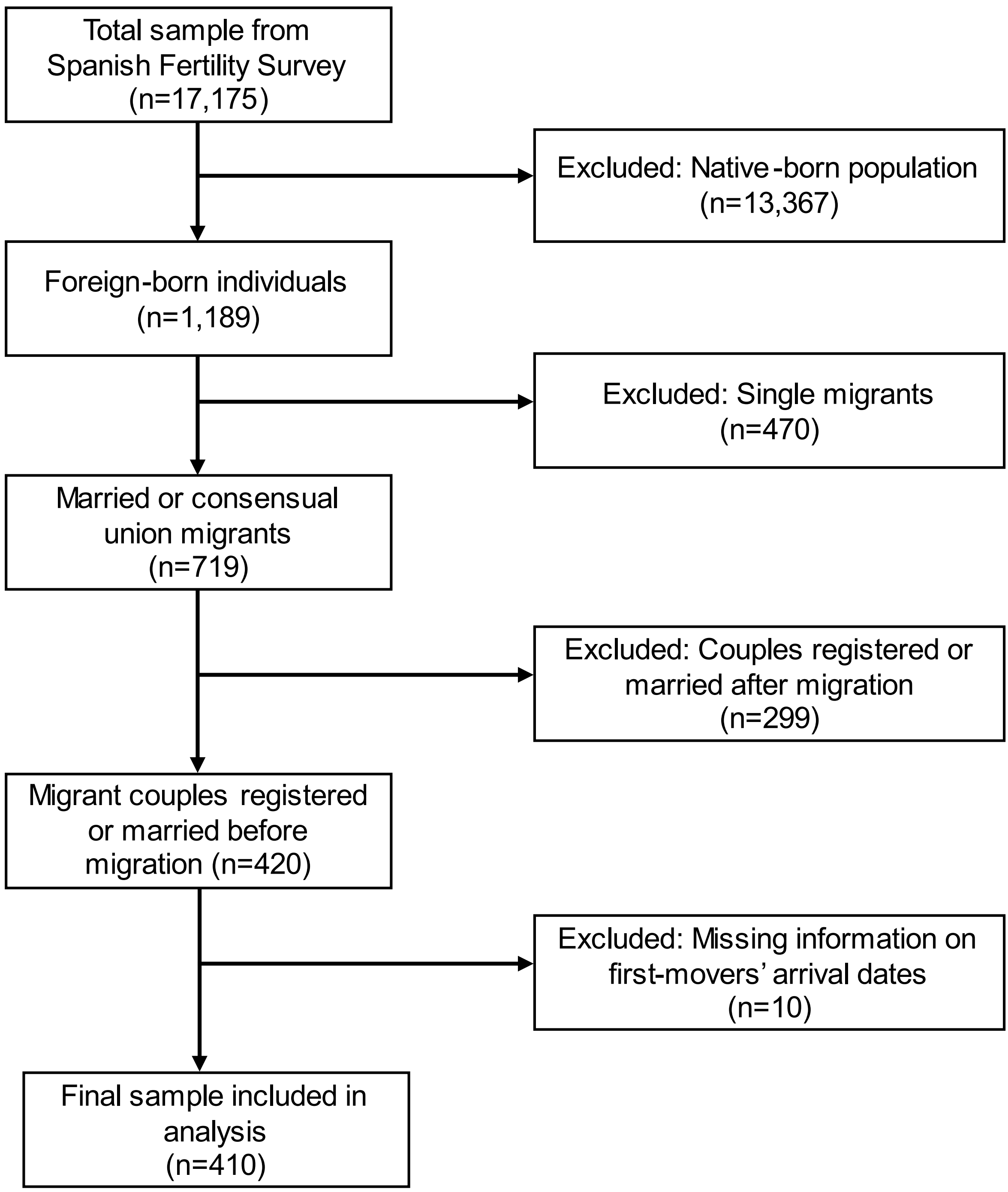
Data and sample

In this study, we use the 2018 Spanish Fertility Survey (EF-2018), a nationally representative survey conducted by the Spanish National Statistics Institute (INE). This dataset contains detailed retrospective information at the individual and household level, including dates of birth, marriage, and migration of primary migrants and their partners. In our study, such retrospective life events are converted into person-month records from the respondent's birth to the time of the survey, allowing for survival analysis. In addition, information on typical individual characteristics such as age, income, and marital status is also available.

3.1 Sample selection

We used selection criteria to construct the sample. First, since the aim of this study is to examine the process of spousal reunification, we restrict the analysis to immigrant groups, thus excluding natives (N= 13,367) from our sample. The initial samples included 1,190 immigrants in Spain. Second, we excluded individuals who reported not being involved in a marital union or cohabitation (N= 470). Third, we excluded couples who were registered or married after migration (N= 299), as well as those couples with missing information on the date of arrival of the first migrants (N= 10). In total, 410 cases who met the eligibility criteria were (39.5%) were included in this study. **Figure 2** showed the flow chart for selection included in this study, of which 230 are males pioneer migrants (60.5%) and 150 are female primary migrants of study population.

Figure 2. Flow chart for selection of study population



Source: Author's own elaboration.

3.2 Sample Characteristics

Our main independent variable of interest is the gender of the pioneer migrants, which is categorized into male and female in the dataset. We include several explanatory variables in the models. First, we include a categorical variable for the pioneer migrant's highest level of education and categorized into "low-medium" (which includes: less than primary education, primary education, first stage of secondary education, second stage of secondary education, and non-higher post-secondary education) or "high" level (which includes vocational training, university diplomas, master's degrees, specializations in health sciences, and doctoral education). In addition, note that this categorical variable was also created for the educational level of the migrant partner for the variable "partner's educational level" in our sample.

Second, we control for the migrant's region of origin in five categories in the dataset: Western Europe, Africa, Latin America, Central or Eastern Europe, and Asia. Third, marital status has two categories: married and in a consensual union (pareja de hecho). Fourth, the legal status variable of primary migrant is classified into two levels: with or without acquired Spanish citizenship. Another important covariate to control is couple's household economic situation by the variable "family economic hardship" to indicate reported financial hardship of migrant couples. Finally, we also include variables such as migrant's age at arrival (<40 years and ≥40 years), partner's age at arrival (<40 years and ≥40 years), and duration of union (≤10 years and >10 years). For simplicity, we coded these continuous variables into categorical variables with a small number of levels. This can also facilitate the stratified tests that we will discuss in the following chapter.

Table 2. Descriptive statistics for total sample.

	Total Sample (N=410)	Male (N=178)	Female (N=232)
Age (years)			
<40	196 (47.8%)	75 (42.1%)	115 (49.6%)
⩾40	214 (52.2%)	103 (57.9%)	117 (50.4%)
Educational level			
Low-Medium	269 (65.6%)	130 (73.0%)	139 (59.9%)
High	141 (34.4%)	48 (27.0%)	93 (40.1%)
Region of origin			
Western Europe	34 (8.3%)	14 (7.9%)	20 (8.6%)
Africa	72 (17.6%)	34 (19.1%)	38 (16.4%)
Latin America	188 (45.9%)	80 (44.9%)	108 (46.6%)
Asia	13 (3.2%)	6 (3.4%)	7 (3.0%)
Central or Eastern Europe	103 (25.1%)	44 (24.7%)	59 (25.4%)
Age at arrival (years)			
<40	260 (63.4%)	136 (76.4%)	124 (53.4%)
⩾40	150 (36.6%)	42 (23.6%)	108 (46.6%)
Acquired citizenship			
No	288 (70.2%)	124 (69.7%)	164 (70.7%)
Yes	122 (29.8%)	54 (30.3%)	68 (29.3%)
Partnership status			
Married	368 (89.8%)	162 (91.0%)	206 (88.8%)
Consensual union	42 (10.2%)	16 (9.0%)	26 (11.2%)
Family economic difficulty			
Yes	253 (61.7%)	110 (61.8%)	143 (61.6%)
No	157 (38.3%)	68 (38.2%)	89 (38.4%)
Duration of union			
⩽10 years	71 (17.3%)	24 (13.5%)	47 (20.3%)
>10 years	339 (82.7%)	154 (86.5%)	185 (79.7%)
Childbearing status before migration			
Without child	312 (76.1%)	149 (83.7%)	163 (70.3%)
With chid(ren)	98 (23.9%)	29 (16.3%)	69 (29.7%)
Partner's educational level			
Low-Medium	278 (67.8%)	123 (69.1%)	155 (66.8%)
High	132 (32.2%)	55 (30.9%)	77 (33.2%)
Partner's age at arrival (years)			
<40	260 (63.4%)	136 (76.4%)	124 (53.4%)
⩾40	150 (36.6%)	42 (23.6%)	108 (46.5%)

Source: Author's calculation based on data from the Spanish National Institute of Statistics (INE)

Table 2 summarizes the socioeconomic characteristics of our sample. The total sample size is 410, with 178 male-led migrant couples and 232 female-led family migration. The overall median age of the participants is 41.7 years (SD = 8.29), with slightly higher averages for males (44.5 years) compared to females (39.9 years). The educational level shows a majority with “low-medium” education (65.6%) compared to 34.4% with “high” education. The region of origin varies: Latin America (45.9%), Central/Eastern Europe (25.1%), and Africa (17.6%) have a significant share, while Western Europe (8.3%) and Asia (3.2%) have a relatively small share in the sample. 63.4% of all participants arrived before the age of 40, and 36.6% arrived at or after the age of 40. The acquisition of citizenship is split, with 29.8% holding citizenship and 70.2% not holding citizenship. In terms of partnership status, 89.8% of individuals are “married”, while 10.2% are in a “consensual union”. About 61.7% of the participants have family economic difficulties. “Duration of union” reflects that 82.7% have been in a union for more than 10 years. In terms of childbearing status, the majority of participants (76.1%) had no children, while 23.9% had one or more children prior to migration. Partner-related characteristics show that 67.8% of individuals have partners with a “low-medium” level of education and 32.2% have partners with a “high” level of education. The age of the partner at arrival shows a distribution of 63.4% below 40 years, and 36.6% at or above 40 years.

4

Statistical analysis and results

In this section, we present the analysis process by introducing the survival estimation techniques. Specifically, the following analyses are performed: (1) presenting the structure of the data set for the current study, (2) calculating the survival probability and hazard probability for the overall spousal reunification process, (3) comparing the survival differences between two gender groups using Kaplan-Meier survival curves and the log-rank test, (4) decomposing the effect of primary migrant's gender on reunification using stratified log-rank test, (5) fitting reunification data into univariate Cox model, (6) constructing multivariable Cox regression model using eligible covariates, and (7) testing the validity of the Cox model using Schoenfeld residuals. Note that all statistical and descriptive analyses were performed by using statistical software programs: R (version 4.2.3) with the 'survival' package (Therneau & Lumley, 2017). And p-value <0.05 was considered statistically significant by two-sided hypothesis tests.

4.1 Survival analysis

Survival analysis is a technique that widely used in population biomedical research (Collett, 1994), in which the outcome variable of interest is time until an event occurs (Kleinbaum & Klein, 2008). As such, *time* refers to the years, months, weeks, or days from the start of an individual's follow-up to the occurrence of an event. *Event* refers to death, relapse from remission, or any specified experience of interest that may occur to an individual. In the current study, we consider the event of interest to be "spousal reunification in Spain" and the endpoint to be the time to reunification. Therefore, our outcome variable is "time in months from the arrival of the pioneer migrants in Spain to the migration of the partner". Specifically, we consider that couples enter the risk of reunification at the time of the forerunner's arrival in Spain and are followed until partner's migration or until the survey date.

4.2 Censoring

One of the characteristics of survival analysis is *censoring*, which occurs when the starting or ending events are not precisely observed (Moore, 2016). In our study, couples who did not experience spousal reunification at the end of the survey (December 2018) are treated as *right-censored*. Therefore, the **dependent variable** in this study is defined as a binary variable that coded as '1' if spousal reunification happens, and '0' otherwise. Our data may be represented in tabular form as shown in **Table 3**.

Table 3. Structure of data set for current study

Subject	Start of observation	End of observation	Survtime (months)	Status
1	November 1998	February 2018	231	0
2	December 2005	February 2018	146	0
3	June 2016	December 2018	30	1
4	February 2016	August 2016	6	0
5	April 1997	December 2018	260	1
6	June 2014	December 2015	10	0
…	…	…	…	…

Source: Author's own elaboration.

4.2 The Survival and Hazard Functions

Two related probabilities are used to describe survival data: the *survival probability* and the *hazard probability*. The *survival probability*, also known as the survival function *S(t)*, is the probability that an individual survives from the time origin to a specified future time *t* (Clark, Bradburn, Love, & Altman, 2003a). This function provides different values of *t* that directly describe the survival experience of a study cohort and is therefore fundamental to survival analysis (Kleinbaum & Klein, 2008). The actual survival time of an individual, *t*, can be regarded as the observed value of a variable, T, that can take any non-negative value, and we call T the random variable associated with the survival time (Collett, 1994). Formally, the survival probability can be defined as following, where value *t* ranges from 0 up to infinity:

$$\mathrm{S(t) = P\ (T > t)},\ 0 < \mathrm{t} < \infty \qquad (4.1)$$

In contrast to the *survivor function*, which focuses on not having an event, the *hazard function* focuses on the event occurring and defines the instantaneous failure rate (Moore, 2016). It can be expressed as following, where $h(t)$ equals the limit, as δ approaches zero, of a probability statement about survival, divided by δ, where δ denotes a small interval of time:

$$h(t) = \lim_{\delta \to 0} \frac{pr(t<T<t+\delta|T>t)}{\delta} \qquad (4.2)$$

Additionally, the *cumulative hazard H(t)* is also commonly used, which defined as the integral of the hazard. It can also be understood as the number of events that would be expected for each individual by time *t* if the event were a repeatable process (Kleinbaum & Klein, 2008).

Another interesting function in the survival analysis is the *cumulative incidence* function, which summarizes the cumulative probability of events occurring before time t and can be expressed as:

$$\mathrm{F(t) = P\ (T < t) = 1-}\ S(t) \qquad (4.3)$$

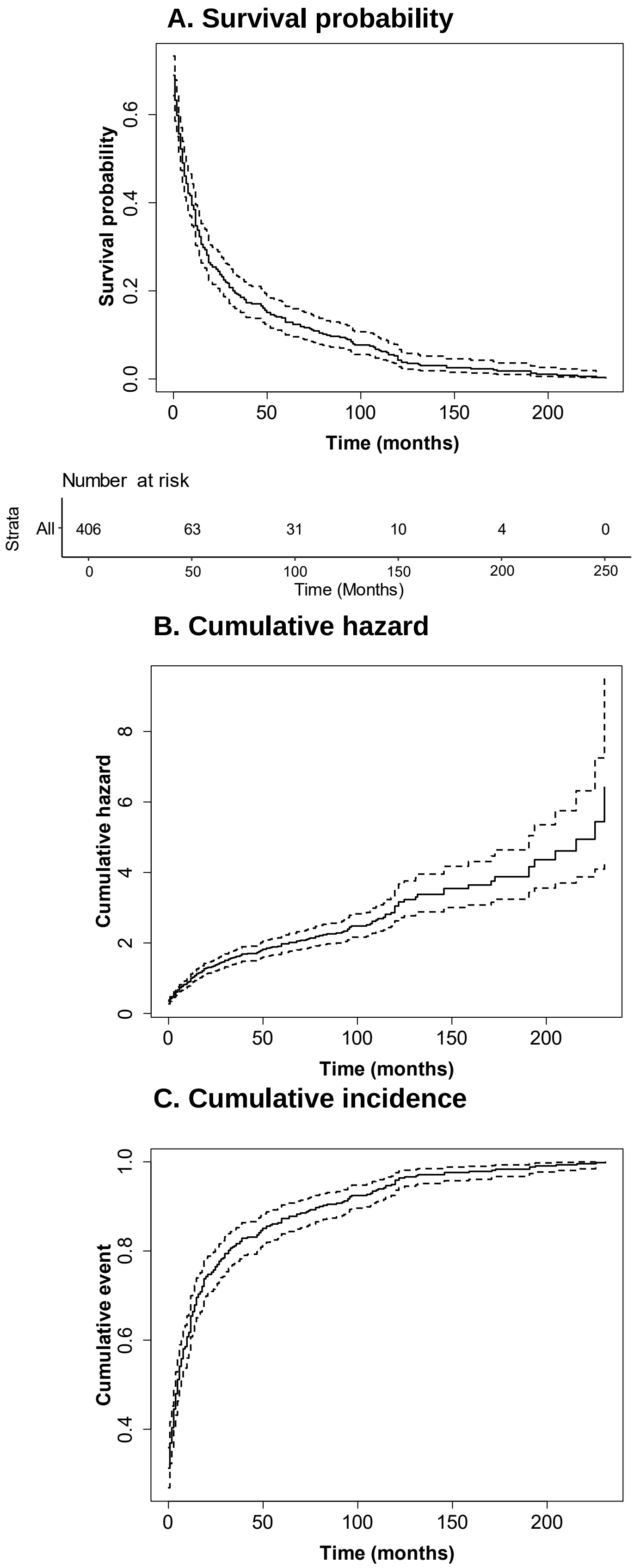


Figure 3. Survival and cumulative hazard curves for the spousal reunification study. (A) survival probability, (B) cumulative hazard function, (C) cumulative incidence curve. *Source:* Author's own elaboration.

The survival curve and its corresponding 95% confidence intervals in **Figure 3A** show the proportion of spouses residing in country of origin for each month of analysis time. First, it is important to note the high propensity for joint migration in our sample: out of the 410 observations, 126 immigrants migrated with their partners, representing 30% of all migrant couples at the time of the survey. This finding is consistent with previous studies in Germany (González-Ferrer, 2007), which suggest a preference for family migration among labor migrants. Second, as **Figure 3A** illustrates, couple reunification appears to be relatively common and rapid, especially in the first five years. The risk table below provides us with the exact number of migrant couples who remain separated by time in months. Specifically, by month 50, only 63 out of 410 spouses remain waiting in the country of origin, which indicates that about 84.6% of migrants had reunited with their partners after 4 years of separation, although a small group of migrant couples endure rather long separations, with reunification occurring almost 100 months after separation.

Third, according to **Figure 3A**, we can also obtain estimates of the median survival time, which refers to the time at which the proportion of partners residing in the country of origin is 0.5. Our result shows that only 50% of the pioneer migrants remain separated with their partners 5 months after their arrival in Spain. **Figure 3B** shows the cumulative hazard for our spousal reunification data. At each time interval, “failure” probabilities are calculated as the number of couples reunited divided by the number of couples at risk of reunification. Cumulative probabilities of failure at each time scale are then calculated by multiplying all the probabilities of failure at all preceding time scales. The cumulative probability shown in **Figure 3B** indicates that the probability of spousal reunification seems to be higher in the 5 initial years of separation and decreases over time. The cumulative incidence, which refers to the incidence of spousal reunification, is also shown in **Figure 3C**, which suggests that the number of reunited couples increases over time.

4.3 Kaplan–Meier Estimate

The Kaplan-Meier survival estimate is a non-parametric estimate that makes no assumptions about the underlying distribution of the data (E.L. Kaplan, 1958). For each time interval, survival probability is calculated as the number of subjects surviving divided by the number of patients at risk. Therefore, the total probability of survival $S(t)$ to that time interval t is calculated by multiplying all the probabilities of survival at all time intervals prior to that time (Goel, Khanna, & Kishore, 2010), which can be defined as:

$$S(t) = p_1 \times p_2 \times \cdots \times p_t \tag{4.4}$$

To calculate $S(t)$ we need to estimate each of p_1, p_2, $p_3, \cdots$ and p_t. It is convenient to think of the time, t, as denoting the start of a short time interval ending at time $(t+1)$. We then use n_t as the number of partners remain separated at the start of the interval and therefore at risk of reunification during that short interval afterwards. We denote the number of partners reunifying in the short time interval just after t as d_t. The number of couples surviving (no reunifying) the interval is therefore $(n_t - d_t)$. This number in turn becomes the number starting interval $(t+1)$, which we denote by n_{t+1}. This notation enables us to write the equation as

$$p_t = 1 - \frac{d_t}{n_t} \tag{4.5}$$

where $p_t = 1$ represents at times (months) when no couple reunites, as $d_t = 0$, since the number at risk of reunification at the beginning of that month is the same as the number at risk at the end of that month. Therefore, the value of $S(t)$, the overall probability of survival (remain separated) to time t, changes only at times (months) on which at least one couple reunifies. As a consequence, we can skip over the times (months) when there is no reunification when calculating equation (4.4). Then, we can rewrite equation (4.4) by using equation:

$$S(t) = \left(1 - \frac{d_1}{n_1}\right)\left(1 - \frac{d_2}{n_2}\right)\cdots\left(1 - \frac{d_t}{n_t}\right) \qquad (4.6)$$

or more briefly as

$$S(t) = \prod_t \left(1 - \frac{d_t}{n_t}\right) \qquad (4.7)$$

The successive overall probabilities of survival, $S(1)$, $S(2)$, $\cdots$, $S(t)$, are known as the Kaplan-Meier (K-M) or product-limit estimates of survival (Machin, Cheung, & Parmar, 2006). Thus, for $S(t) = S(t-1) \times p_t$, or

$$S(t) = S(t-1)\left(1 - \frac{d_t}{n_t}\right) \qquad (4.8)$$

This result enables each successive survival probability to be obtained by successive multiplication by equation (4.5). It is necessary to specify that when $t = 0$, $S(0) = 1$, that is, all couples are assumed remain separated at time zero.

4.4 The Log-Rank Test

The log–rank test, also referred to as the Mantel-Cox test, is the most widely used method of comparing two survival curves and can easily be extended to comparisons of three or more curves (Machin et al., 2006). Indeed, it is a large-sample chi-square test that uses as its test criterion a statistic that provides an overall comparison of the Kaplan-Meier curves being compared (Kleinbaum & Klein, 2008). When only two groups are compared, the log-rank test is testing the null hypothesis that there is no difference between the populations in the probability of an event at any timepoint (Bland & Altman, 2004). Under this null hypothesis, the log-rank statistic is approximately chi-square with one degree of freedom. Thus, a *p*-value for the log-rank test is determined from tables of the chi-square distribution. Where this takes a *p*-value <0.05 then we need to refuse the null hypothesis and consider a statistically significant difference between Kaplan–Meier survival curves.

Mathematically, the test statistic is

$$\text{Log-rank test statistic} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \quad (4.9)$$

where E_1 and E_2 represent the expected number of events in each group while O_1 and O_2 are the total number of observed events in each group, respectively. The total number of expected events in a group (e.g., E_2) is the sum of expected number of events, at the time of each event in any of the group, taking both groups together. At the time of event in any group the expected number of events is the product of risk of event at that time with the total number of subjects alive at the start of the time of event in that very group. The total number of expected events in group 2 is sum of the expected events calculated at different time. The total number of expected events in the other group (i.e.,E_1) is calculated by subtracting the total number of expected events in group 2 i.e., E_2 from the total of observed events in both the groups i.e., $O_1 + O_2$.

Table 4. Ranked within-gender group survival times of 15 individuals in the sample

Male group				Female group			
Time	n.risk	n.event	survival	Time	n.risk	n.event	survival
0	177	17	0.90395	0	229	109	0.52402
1	160	11	0.84181	1	120	12	0.47162
2	149	9	0.79096	2	108	5	0.44978
3	140	12	0.72316	3	103	5	0.42795
4	128	8	0.67797	4	98	6	0.40175
5	120	7	0.63842	5	92	6	0.37555
6	113	7	0.59887	6	86	5	0.35371
7	106	2	0.58757	7	81	5	0.33188
8	104	3	0.57062	8	76	6	0.30568
9	101	2	0.55932	9	70	4	0.28821
10	99	5	0.53107	10	66	3	0.27511
11	94	1	0.52542	11	63	6	0.24891
12	93	9	0.47458	12	57	1	0.24454
13	84	3	0.45763	13	56	2	0.23581
14	81	4	0.43503	14	54	6	0.20961
15	77	1	0.42938	15	48	2	0.20087

Source: Author's own elaboration.

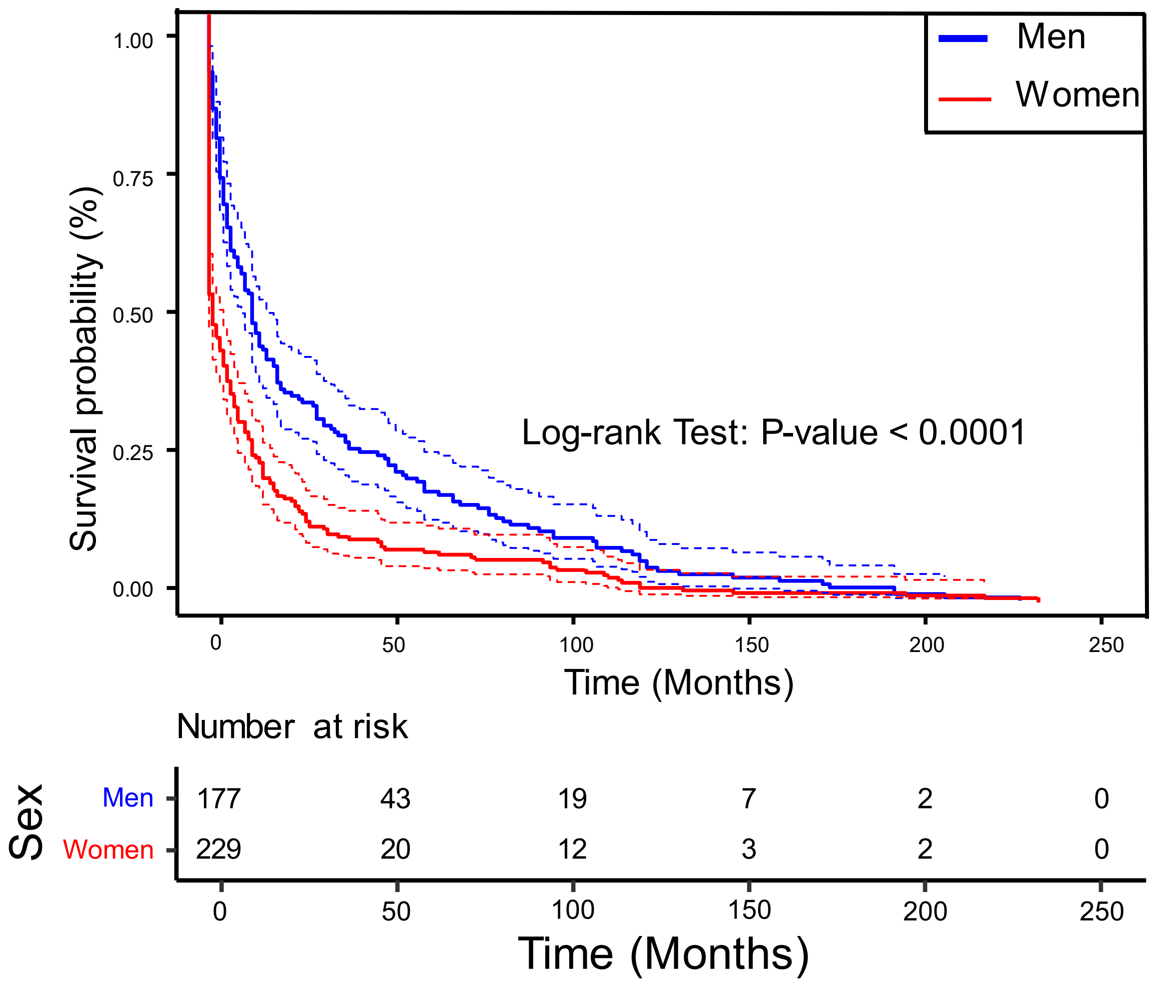


Figure 4. Kaplan-Meier estimates for the proportion of partners still residing at country of origin spousal reunification by pioneer's gender.

Source: Author's own elaboration.

The Kaplan-Meier survival curves allow us to graphically illustrate the probability of experiencing reunification by pioneer migrant's gender while accounting for censoring. Overall, the Kaplan–Meier curves in **Figure 4** show a clear difference in couple's reunification by forerunner migrant's gender, with the curve for the male group consistently lying above the curve for female group, especially at earlier times. In this regard, the log-rank test confirmed the statistically significant differences in spousal reunification (*p*-value<0.0001) between male and female pioneer couples.

This finding suggests that male-led migration families are more likely to have experienced a longer period of separation than their female counterparts. Nevertheless, the survival curves began to converge after about 50 months, suggesting that the pace of reunification tends to be similar between the two groups at later times. This finding suggests that female primary immigrants reunite even more quickly with their partners after migration when compared with male counterparts, which challenges the widely held belief that female migrants are more disadvantaged to rejoin their family members in receiving countries (Mato Díaz & Miyar Busto, 2017).

4.5 Stratified log rank test

In survival analyses, it is essential to adjust for potential confounders in order to obtain a valid estimate of the effect of the intervention on the primary event of interest (Moore, 2016). In this regard, the stratified log-rank test is another variation of the log-rank test that allows us to test the equivalence of survival curves by controlling for the stratified variable (Kleinbaum & Klein, 2008). Now, we provide Kaplan-Meier curves with the stratified log-rank tests to compare differences between two sex groups.

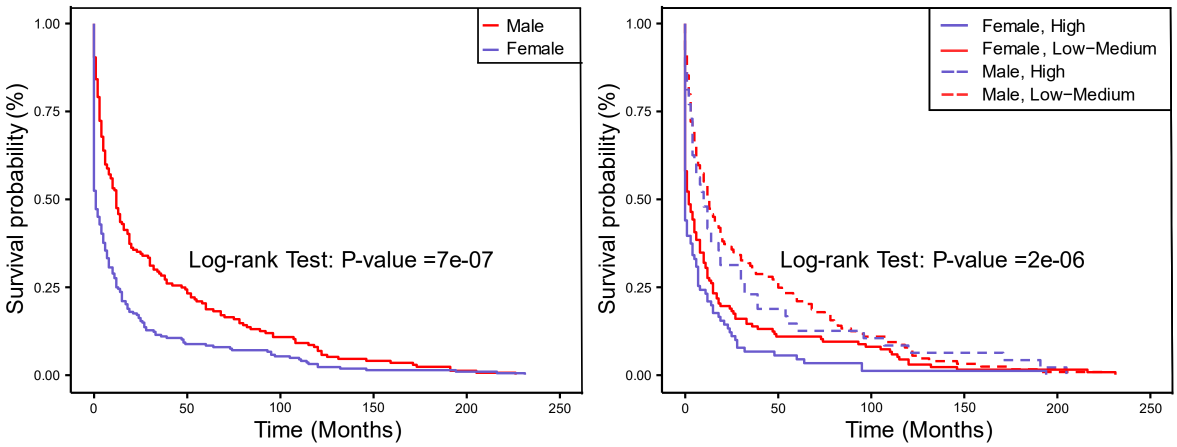


Figure 5. Survival curves for spousal reunification by migrant's gender and stratified by educational level. Comparison of the Kaplan-Meier survival curves for two gender groups ignoring the educational level *(left)* and accounting for it *(right)*. *Source:* Author's own elaboration.

The first control variable to consider is primary migrant's education level, which is a dummy variable divided into two levels: high and low-medium. The Kaplan-Meier survival plots were shown in **Figure 5**. Results from the stratified log-rank test show that the adjusted value (Chisq= 22.8, *p*-value = 2e-06) remain highly significant and differs only slightly from the unadjusted value (Chisq= 24.6, *p*-value = 7e-07). This finding suggests that there are significant differences in the spousal reunification process between gender groups even after adjusting for the confounding effect of migrant's education level.

Similar results were also found for stratified tests with other covariates, as shown in **Figure 6** and **Figure 7**, where, despite the slight increase in *p*-values after controlling for confounding effects, the large Chi-squared value as well as *p*-value < 0.001 suggest that the association between gender groups and survival outcomes remains significant. First, the survival analysis focused on migrant couples' time to reunification based on the primary migrant's gender, while considering education level as a covariate. This significance persisted even after stratifying by education level (*p*-value=1e-06), underscoring gender's continued impact on reunification timing, possibly alongside education.

Second, when accounting for stratification by acquired citizenship status, the analysis still demonstrates a noteworthy disparity in reunification timing between male and female primary migrants. After adding stratification based on acquired citizenship, the significance remains (*p*-value =3e-07), confirming that the gender-based discrepancy in reunification timing persists even when considering citizenship status. Third, considering stratification by the type of union (married/cohabitation), the *p*-value even decreases when the data are stratified, indicating that the effect of gender on survival becomes even more pronounced after accounting for partnership strata.

Fourth, when accounting for stratification by family economic difficulty, the analysis continues to emphasize a significant gender-based distinction in reunification timing for migrant couples (*p*-value=4e-07). The findings underscore the enduring influence of gender on migrant couples' reunification experiences, extending beyond economic factors. Similarly, when incorporating stratification by having or not having a child before migration, the significance remains (*p*-value=1e-05). This underscores that the gender-related divergence in reunification timing persists, regardless of whether couples had children before migration. The implications emphasize gender's lasting impact on migrant couples' reunification experiences, irrespective of parental status prior to migration.

Sixth, accounting for stratification by the partner's educational level, the significance persists (*p*-value=4e-07). This outcome underscores that the gender-related variation in reunification timing remains substantial, even when considering the partner's educational background. Seventh, when accounting for stratification by the partner's age at arrival, the significance remains (*p*-value=6e-08). The findings highlight gender's enduring effect on migrant couples' reunification experiences, irrespective of the partner's age upon arrival.

Eighth, when considering stratification by the couple's region of origin, the significance persists (*p*-value=2e-06). The findings underscore the lasting impact of gender on migrant couples' reunification experiences, irrespective of their regional background.

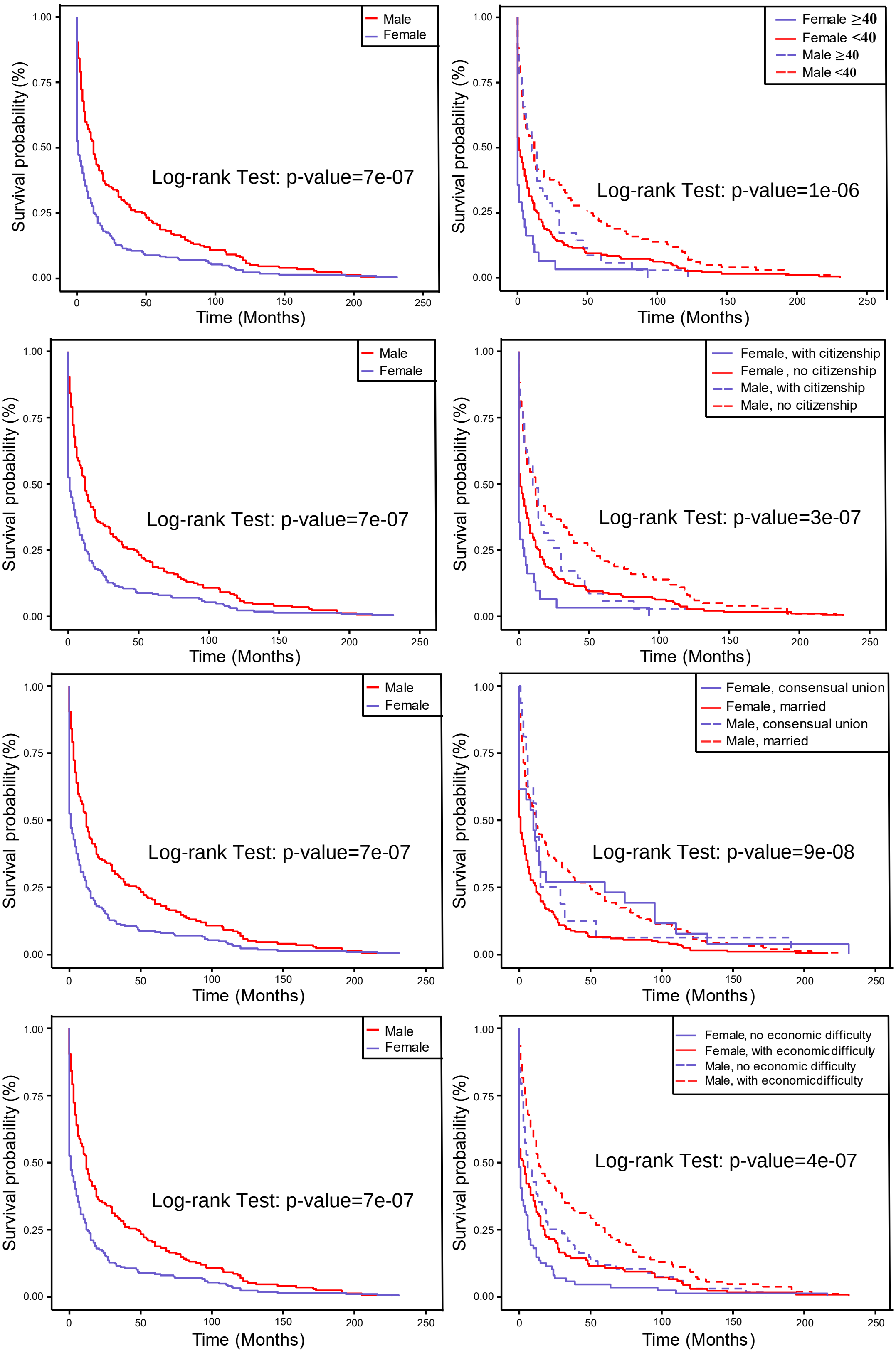


Figure 6. Survival curves for spousal reunification by migrant's gender and stratified by migrant's age at arrival, migrant's acquired Spanish citizenship, partnership status and family economic difficulty. Comparison of the Kaplan-Meier survival curves for two gender groups ignoring covariates *(left)* and accounting for them *(right).*
Source: Author's own elaboration.

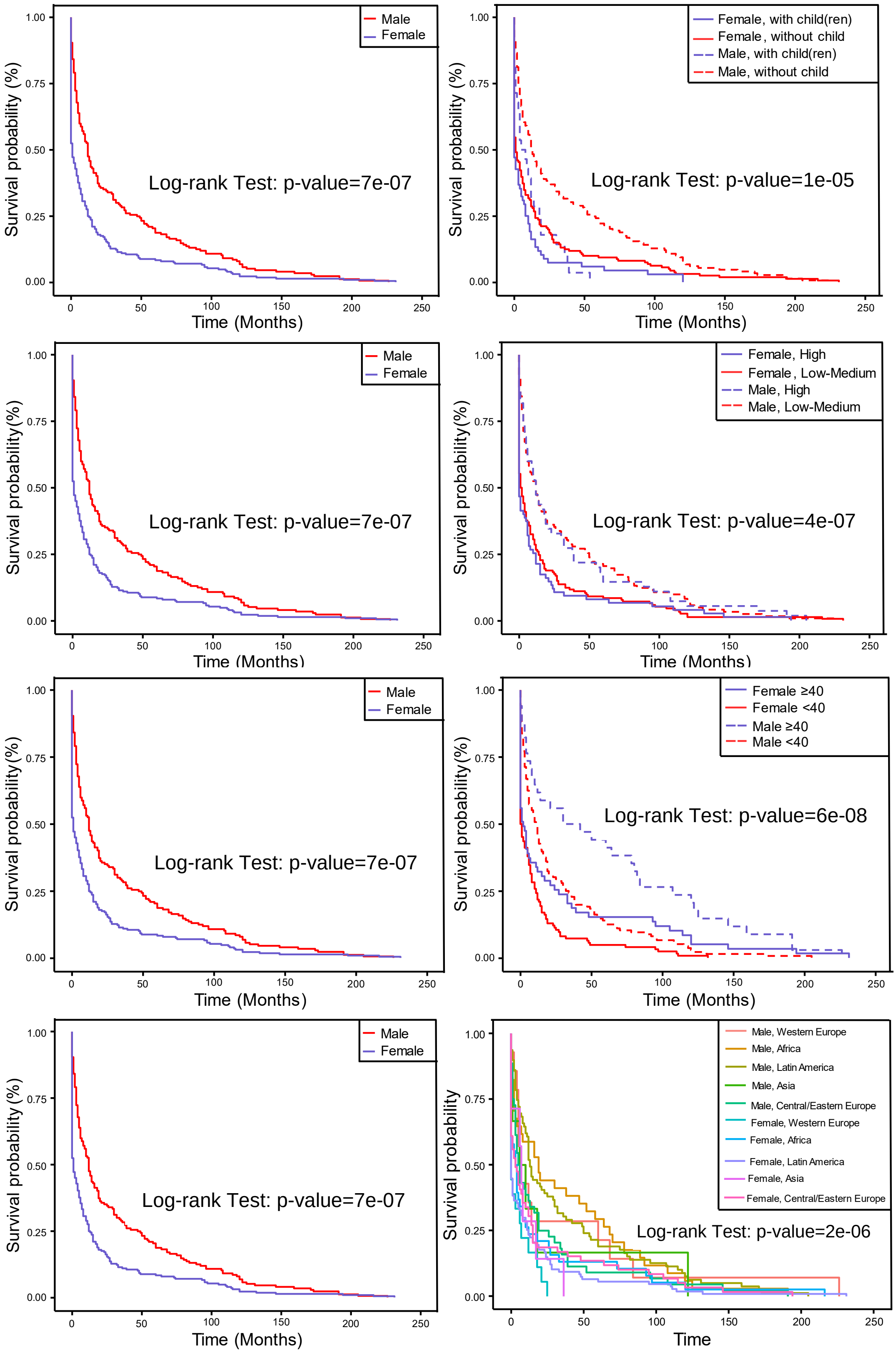


Figure 7. Survival curves for spousal reunification by migrant's gender and stratified by couple's childbearing status, partner's educational level, partner's age at arrival, migrant's acquired Spanish citizenship, partnership status and couple's region of origin. Comparison of the Kaplan-Meier survival curves for two gender groups ignoring covariates (*left*) and accounting for them (*right*).
Source: Author's own elaboration.

4.6 The Cox Proportional Hazards Model

The Cox proportional hazards (or PH) model (Cox, 1972) is a popular mathematical model used for analyzing survival data, which is essentially a regression model commonly used to test the association between the survival time of different groups and one or more predictor variables. It can be estimated as follow:

$$h(t) = h_0(t) \times exp\{b_1x_1 + b_2x_2 + \cdots + b_px_p\} \quad (4.4)$$

where *t* represents the time to occurrence of the event, *h(t)* is the hazard function dependent on a set of p covariates $(x_1, x_2, \ldots, x_p)$, whose impact is measured by the size of the respective coefficients $(b_1, b_2, \ldots, b_p)$. The term h_0 is called the **baseline hazard**, it corresponds to the value of hazard if all the x_i are equal to zero.

In general, a hazard ratio (HR) is defined as the hazard for one individual divided by the hazard for a different individual. The two individuals being compared can be distinguished by their values for the set of predictors, that is, the X's. The hazard ratio can be denoted as the estimate of *h* (*t*,X*) divided by the estimate of *h* (*t*,X), where X* denotes the set of predictors for one individual, and X denotes the set of predictors for the other individual. (Kleinbaum & Klein, 2008).

$$\widehat{HR} = \frac{\hat{h}\,(t,X^*)}{\hat{h}\,(t,X\,)} \quad (4.6)$$

A hazard ratio above 1 indicates a covariate that is positively associated with the event probability, and thus negatively associated with the length of survival. (Bradburn et al., 2003a). In the current study, the hazard ratios provide information about relative risk or relative speed of reunification. The hazard translates the length of time to the event occurrence into a rate. Thus, the hazard ratio can be translated into a relative speed of event occurrence: a group with a relatively higher risk of reunification experiences shorter stay in care and has a higher speed of reunification (Fernandez & Lee, 2011). Therefore, in our Cox proportional hazards model,

- ✓HR = 1: No effect
- ✓HR < 1: Reduced probability of reunification
- ✓HR > 1: Increased probability of reunification

4.7 The Cox Proportional Hazards Assumption

The key assumption of the Proportional Hazards model is that the hazard of the event in any group is a constant multiple of the hazard in any other (Bradburn, Clark, Love, & Altman, 2003). This assumption implies that the hazard curves for the groups should be proportional and cannot cross.

However, if the predictor variables do not vary over time, the hazard ratio comparing any two observations is constant over time (Bradburn et al., 2003). Therefore, for credible estimation and inference, it is important to assess whether the proportional hazards assumption holds for a valid interpretation of the regression coefficients in our Cox model (Clark, Bradburn, Love, & Altman, 2003b).

A residuals test provided by Schoenfeld (1982) is particularly useful for assessing the PH assumption, in which residuals can be plotted against time. According to Machin, Cheung and Parmar (2006), a Schoenfeld residual SCH_{kj} can be computed for each explanatory variable x_k and each non-censored observation j, whose failure time is denoted by t_j, by

$$SCH_{kj} = x_{kj} - E\left(x_{kj} | R_j\right) \qquad (4.5)$$

where R_j is the set of observations at risk of failure at the time subject j fails, while $E\left(x_{kj} | R_j\right)$ is the expected value of x_k for subject j , estimated from the log *HR*s from a Cox model. Thus, if proportional hazards hold, the SCH_{kj} should sum to zero over all and should scatter around zero; otherwise, there should be a systematic trend in the SCH_{kj} over time or the rank order of time.

4.8 The univariate Cox regression model fitted to the spousal reunification data

To examine the association between migrant couples' sociodemographic characteristics and spousal reunification process, univariate Cox regression was performed by using the *coxph* function in 'survival' R package (Therneau & Lumley, 2017). Statistical results of univariate analysis of were presented in **Table 5**. First, in accord with the Kaplan-Meier results, the variable primary migrant's sex has highly statistically significant with the given *p*-value <0.0001. The positive coefficient value for female 0.54 and the corresponding HR value of 1.71 indicate that having the female migrate first increases the probability of reunification by 71%.

Second, regarding the effects of the control covariates, we see that the regression coefficients *(coef)* for the variable "primary migrant's age" is -0.3255. Since this value is negative, we can conclude that the age group "≥40" is associated with lower hazards or lower probability of spousal reunification than the reference group "<40". The value of the hazard ratio 0.7222 suggests that migrants in the "≥40" age group reduce the hazard by 28% than the "<40" group. By contrast, the covariate "migrant' s age at arrival" shows different results, in which the estimate of the log hazard ratio in the age group '≥40' 0.2363 indicates that arriving older than 40 years has a positive effect on reunification with respect to migrants who arrived in the age group '<40'. Nevertheless, the *p*-value of 0.0861 provides no statistical significance for this variable, therefore we can drop this covariate from the model.

Third, the education level of forerunner migrants provides strong and statistically significant with a *p*-value of 0.00943. Given that the coefficient 0.2733 is positive, the hazard ratio of 1.31 indicates that the odds ratios of reunification for migrants with high levels of education are 31.4% higher than those with low-medium levels of education. This result is highly consistent with previous studies, suggesting that education is an indicator of the availability of personal resources that may facilitate migrants' couple reunification (Baizán et al., 2014). Fourth, there is not a significant association between migrant's origin region and probability of separation (*p*-value = 0.8), although the negative coefficient values of the stratification from Africa, Latin America, Central/Eastern Europe, and Asia indicate lower odds of reunification compared to migrants from Western Europe. Similarly, the acquisition of Spanish citizenship by the primary migrant has no significant effect on partner reunification in Spain, with a *p*-value of 0.165.

With regard to the socio-economic characteristics of the migrant partner, the covariates in our study show mixed results. First, there is no significant effect of partner's education level on spousal reunification (*p*-value = 0.521), although the odds ratio is slightly higher in the high education category (HR = 1.07). Second, the covariate "partner's age at arrival" is an important determinant of spousal reunification. Specifically, the hazard ratio of 0.6610 indicates that having a partner older than 40 years reduce the probability of reunification by 33.9% compared to the reference level "<40".

In terms of family characteristics, the importance of the household's economic resources is confirmed by the results of subjective financial well-being. Specifically, the probability of reunification is 42% higher for couples without self-reported economic difficulties than for couples with a disadvantaged financial situation. This finding provides additional evidence of the crucial role of economic conditions in shaping family migration decisions (Petroff, 2016; Reher, Requena, & Sánchez-Domínguez, 2013). In addition, it is also interesting to note that the covariate "duration of union", which shows that reunification in the host country is more likely for new couples, as the odds ratio is 41% lower among couples with more than 20 years of union than among couples with less than 20 years of union. A possible explanation for this result could be the prevalence of transnational family life among long-term migrant couples, where living apart-together across countries is a rational and even functional strategy (Caarls & Mazzucato, 2016; Zontini, 2004). Third, partnership status shows no significance in our study (*p*-value = 0.144), with no difference in spousal reunification between married and cohabiting couples. Finally, according to our univariate analysis, children is an important factor influencing the likelihood and timing of couples' migration, with having children prior to migration increasing the odds of reunification by 64% compared to couples without children.

Table 5. Hazard ratios from the univariate Cox proportional hazard model

Covariate	Coefficient (b_i)	HR $exp\ (b_i)$	95% CI	*P*-value
Forerunner migrant's characteristics				
Sex				<0.0001 ***
Male (ref.)	(0.000)	(1.00)		
Female	0.5384	1.7133	(1.41-2.09)	
Age				0.00275**
<40 (ref.)	(0.000)	(1.00)		
⩾40	-0.3255	0.7222	(0.58-0.89)	
Educational level				0.00943**
Low-Medium (ref.)	(0.000)	(1.00)		
High	0.2733	1.3144	(1.07-1.62)	
Region of origin				
Western Europe	(0.000)	(1.00)		0.8
Africa	-0.23992	0.78669	(0.52-1.20)	
Latin America	-0.11669	0.88986	(0.61-1.30)	
Asia	-0.06768	0.93456	(0.49-1.80)	
Central or Eastern Europe	-0.07468	0.92804	(0.62-1.38)	
Age at arrival				0.0861
<40 (ref.)	(0.000)	(1.00)		
⩾40	0.2363	1.2666	(0.97-1.66)	
Acquired citizenship				0.165
No (ref.)	(0.000)	(1.00)		
Yes	-0.1511	0.8598	(0.69-1.06)	

HR =hazard ratio, CI= confidence interval

Signif. codes: 0.0001 '***' 0.001 '**' 0.01 '*'

Source: Author's calculation based on data from the Spanish National Institute of Statistics (INE)

Table 6. Hazard ratios from the univariate Cox proportional hazard model (continued)

Covariate	Coefficient (b_i)	HR $exp\ (b_i)$	95% CI	*P*-value
Migrant partner's characteristics				
Partner's educational level				0.521
Low-Medium (ref.)	(0.000)	(1.00)		
High	0.06842	1.07082	(0.87-1.32)	
Partner's age at arrival				0.00103 **
<40 (ref.)	(0.000)	(1.00)		
⩾40	-0.4141	0.6610	(0.52-0.85)	
Family characteristics				
Partnership status				0.144
Married (ref.)	(0.000)	(1.00)		
Consensual union	-0.2419	0.7851	(0.57-1.09)	
Family economic difficulty				<0.0001 ***
Yes (ref.)	(0.000)	(1.00)		
No	0.3514	1.4211	(1.16-1.74)	
Duration of union				<0.0001 ***
⩽20 years (ref.)	(0.000)	(1.00)		
>20 years	-0.5115	0.5996	(0.46-0.78)	
Having child before migration				<0.0001 ***
No (ref.)	(0.000)	(1.00)		
Yes	0.4925	1.6364	(1.30-2.07)	<0.0001 ***

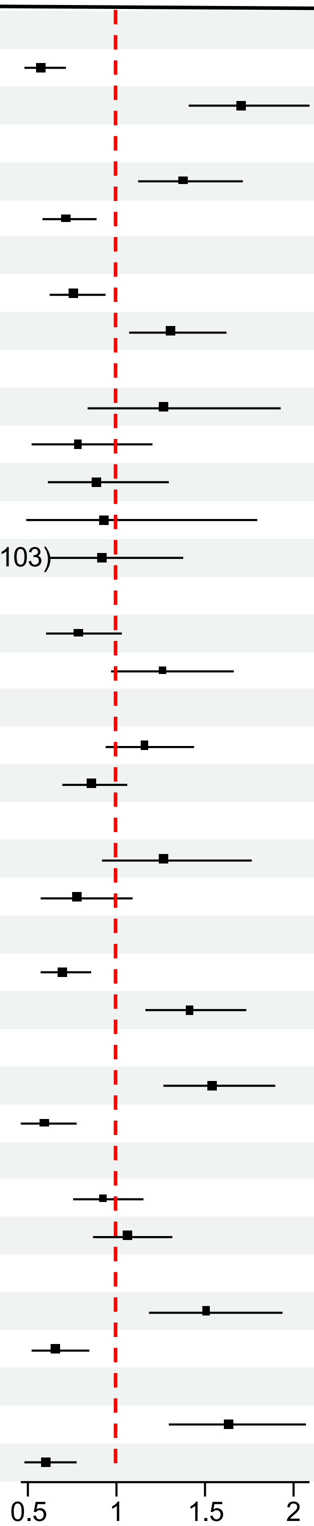
HR =hazard ratio, CI= confidence interval

Signif. codes: 0.0001 '***' 0.001 '**' 0.01 '*'

Source: Author's calculation based on data from the Spanish National Institute of Statistics (INE)

Figure 8. Forest plot summarizing the HRs and 95% CIs of the covariates by the univariate Cox regression

Subgroups	HR (95% CI)	P−value
Sex		
Male (n=178)	0.58 (0.48 to 0.71)	<0.0001 ***
Female (n=232)	1.71 (1.41 to 2.09)	<0.0001 ***
Age (years)		
<40 (n=196)	1.38 (1.12 to 1.71)	0.00275 **
≥40 (n=214)	0.72 (0.58 to 0.89)	0.00275 **
Educational level		
Low−Medium (n=269)	0.76 (0.62 to 0.94)	0.00943 **
High (n=141)	1.31 (1.07 to 1.62)	0.00943 **
Region of origin		
Western Europe (n=34)	1.27 (0.84 to 1.93)	0.262
Africa (n=72)	0.79 (0.52 to 1.20)	0.262
Latin America (n=188)	0.89 (0.61 to 1.30)	0.545
Asia (n=13)	0.93 (0.49 to 1.80)	0.838
Central or Eastern Europe (n=103)	0.93 (0.62 to 1.38)	0.714
Age at arrival(years)		
<40 (n=260)	0.79 (0.60 to 1.03)	0.0861
≥40 (n=150)	1.27 (0.97 to 1.66)	0.0861
Acquired citizenship		
No (n=288)	1.16 (0.94 to 1.44)	0.165
Yes (n=122)	0.86 (0.69 to 1.06)	0.165
Partnership status		
Married (n=368)	1.27 (0.92 to 1.76)	0.144
Consensual union (n=42)	0.79 (0.57 to 1.09)	0.144
Family economic difficulty		
Yes (n=253)	0.70 (0.57 to 0.86)	<0.0001 ***
No (n=157)	1.42 (1.16 to 1.74)	<0.0001 ***
Duration of union		
≤20 years (n=71)	1.55 (1.26 to 1.90)	<0.0001 ***
>20 years (n=339)	0.60 (0.46 to 0.78)	<0.0001 ***
Partner's educational level		
Low−Medium (n=312)	0.93 (0.76 to 1.15)	0.521
High (n=198)	1.07 (0.87 to 1.32)	0.521
Partner's age atarrival(years)		
<40 (n=278)	1.51 (1.18 to 1.94)	0.00103 **
≥ 40 (n=132)	0.66 (0.52 to 0.85)	0.00103 **
Having child before migration		
Yes (n=98)	1.64 (1.29 to 2.07)	<0.0001 ***
No (n=312)	0.61 (0.48 to 0.77)	<0.0001 ***



Note: HR: hazard ratio; CI: confidence interval.
Source: Author's own elaboration.

The forest plot, originally developed to display point estimates and CIs for studies in meta-analyses, is also useful for visually displaying the results of the univariate Cox analysis (Bharat, Murray, Cripps, & Hodkiewicz, 2018). **Figure 8** shows the forest plot with HRs and CIs for all variables in the univariate model using the R package '*forestplot*' (Gordon & Lumley, 2022). The first column shows the names of the categorical variables with subgroups, the second column shows a figure of effects and associated confidence bands, the third column shows the values of hazard ratios (HR) and 95% confidence interval. Finally, the fourth column shows the *p*-values for each variable from log-rank tests.

Note that the vertical red line in the middle is known as the "null effect line". Therefore, any horizontal line that crosses the vertical line means that there is no association between the variable and the outcome. Since our primary interest is examining the effect of covariates on the time to reunification, a variable that falls to the right side of the vertical line indicates a positive effect on the probability of reunification, otherwise it indicates a negative factor. Therefore, in **Figure 8** we have identified the following covariates with statistically significant correlations with the survival time to spousal reunification: age of the primary migrant, level of education, subjective financial well-being, duration of union, childbearing status before migration, as well as the age of the partner at arrival. Specifically, being a female primary migrant (HR=1.71), being older than 40 years (HR=1.38), having a high level of education (HR=1.31), not having family economic difficulties (HR=1.42), having a relationship of more than 20 years (HR=1.55), having a partner older than 40 years at arrival (HR=1.51) as well as having a child before migration (HR=1.64) promote a higher probability of spousal reunification in Spain.

4.9 The multivariate Cox regression model fitted to the spousal reunification data

To correctly estimate the effect of the key explanatory variable 'migrant's gender', a multivariable Cox model is required to adjust for potential confounders. We will now perform the multivariate analysis by including only those variables that are statistically significant in the previous univariate analysis in order to avoid the loss of efficiency.

We have included the identified covariates in a sequential manner in order to see the confounding effect that some variables have on the others. First, we constructed a baseline model (Model 1) by including only the gender variable of the primary migrant. In Model 2, we included the age of the first migrant to examine its effect on the association between gender and time to reunification. In Model 3, we additionally account for the educational level of the migrant. Regarding the family characteristics, we incorporated subjective financial well-being, duration of union, and childbearing status prior to migration in Model 4, Model 5 and Model 6, respectively. Finally, the age of partner at arrival was included in Model 7 to control for partner's characteristic.

The results of the multivariate Cox regression analyses are presented in **Table 7**. Our results show that, regardless of the adjustment method used, the female-led migration group remains significantly associated with the increasing occurrence of spousal reunification compared to the male-led migration family. In particular, the HR value of 1.90 in Model 7 suggests that, after controlling for all covariates, the probability of reunification is 90% higher for couples in which women migrate first. Second, age of primary migrant seems to have a significant impact only in Model 7 after controlling for other covariates, where forerunner migrants in the age group "≥40" increase the probability of reunification by 40% compared to the group "<40". Third, with regard to the educational level of the first migrant, while this covariate shows significance in Model 3 and Model 4 with increasing odds ratios among migrants with high educational levels, its effect begins to reduce by controlling for other covariates and finally is shown to be insignificant in the last model.

Considering the effect of family characteristics, first, the results from the multivariate models confirm the previous finding from the univariate analysis that having children prior to migration increases the odds of reunification. This finding is consistent with previous studies showing that parents always tend to migrate together or reunite in the first years of migration to keep the family unit together (Kraus et al., 2019). Second, our results suggest a crucial role of good financial condition in facilitating couples' reunification at the destination countries. Specifically, those with better economic conditions are 46% more likely to experience spousal reunification than those with financial difficulties. This result can be explained in two ways. On the one side, economic resources to large extent shape the capacity of immigrants to bring their partner to live with and to maintain the family life in destination countries (Flahaux, Wayack - Pambè, Soura, Compaoré, & Sanogo, 2020), which result in long periods of unanticipated separation among some couples. On the other side, the politicization of migration in the 1990s and debates about the (in)ability of migrants to integrate have led many states to restrict family reunification by tightening entry conditions (Bragg & Wong, 2016; Kofman, 2004). At this point, migrants' housing, work, and financial situations became the basic criteria for applying for family reunification under family reunification laws, which significantly reduces the chances of reunification for migrants with fewer economic resources (Baizán et al., 2014; Eremenko & González-Ferrer, 2018). Third, the results of Model 6 and Model 7 show that couples with a union duration of more than 20 years have a hazard ratio of 0.64 compared to those with a union duration of less than 20 years. This finding is consistent with previous study suggesting that newly married couples are also more likely to migrate together than couples who have been married for a relatively long time (González-Ferrer, 2007).

Finally, Model 7 suggests that having partners whose age at arrival is 40 years or older yields strongly negative impact for reunification in Spain, as the HR indicates that the odds ratio is 41% lower than migrants with partner less than 40 years at migration. The likelihood ratio tests for each model show significant values, indicating that the inclusion of variables in each model provides a better fit for explaining the likelihood of spousal reunification.

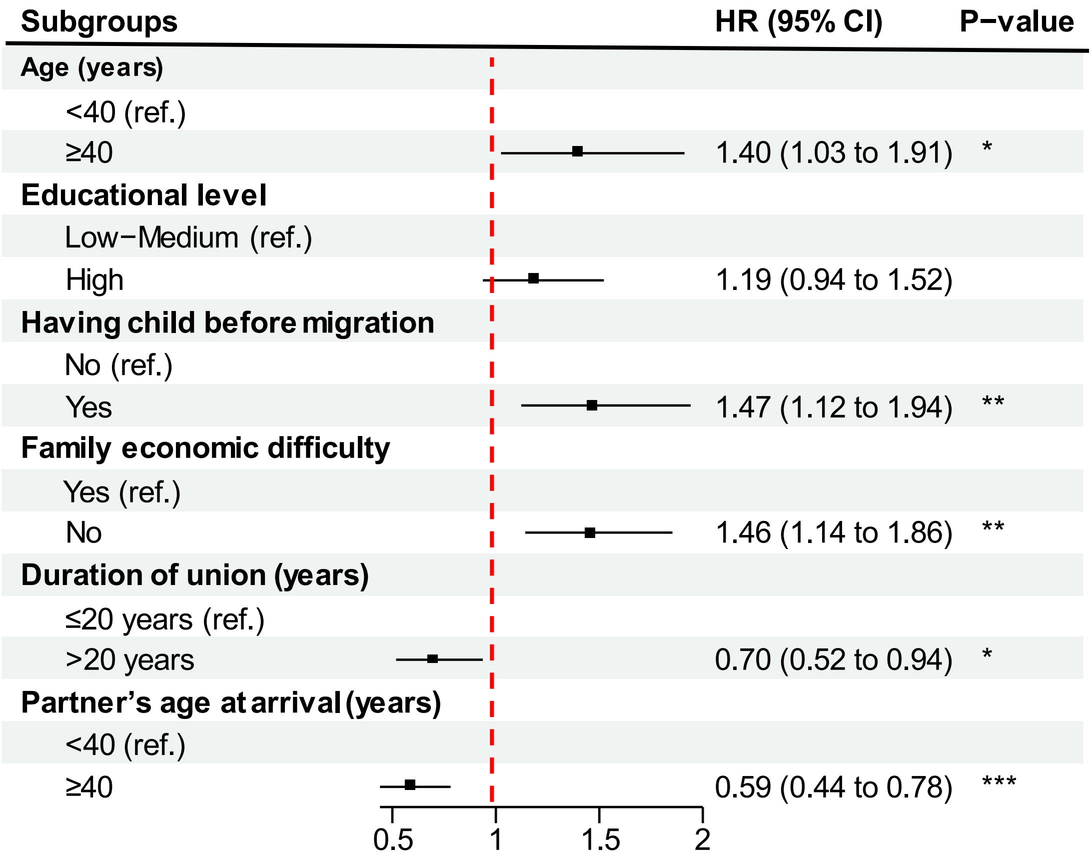
Table 7. Hazard ratios from the multivariate Cox proportional hazard model.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Sex							
Male (ref.)	1	1	1	1	1	1	1
Female	1.71***	1.59***	1.58***	1.55***	1.61***	1.69***	1.89***
Age							
<40 (ref.)		1	1	1	1	1	1
≥40		0.84	0.83	0.87	0.93	1.22	1.40*
Educational level							
Low-Medium (ref.)			1	1	1	1	1
High			1.31*	1.25*	1.21	1.21	1.19
Having child before migration							
No (ref.)				1	1	1	1
Yes				1.31*	1.28	1.25	1.47**
Family economic difficulty							
Yes (ref.)					1	1	1
No					1.44**	1.43**	1.46**
Duration of union							
≤20 years (ref.)						1	1
>20 years						0.64**	0.70*
Partner's age at arrival							
<40 (ref.)							1
≥40							0.59***
Likelihood ratio test	28.7***	25.1***	30.98***	35.25***	45.18***	55.34***	68.51***

Signif. codes: 0.0001 '***' 0.001 '**' 0.01 '*'

Source: Author's calculation based on data from the Spanish National Institute of Statistics (INE)

Figure 9. Forest plot summarizing the HRs and 95% CIs of the covariates by the multivariate Cox regression.



Note: HR: hazard ratio; CI: confidence interval.
Source: Author’s own elaboration.

Figure 9 shows the HR, 95% CI and its statistical significance for all covariates included in the final multivariate model by a forest plot, which helps to interpret the results. Age, childbearing status prior to migration, family economic status, duration of union, as well as partner’s age at arrival are identified as significant factors related to the couple's reunification process, however, “education level” is shown to be insignificant in the final model. Therefore, compared to the reference category of each variable, being less than 40 years old at arrival, having a child before migration, being in a union for less than 20 years, and having a partner less than 40 years old are positively correlated with a rapid pace of reunification of migrant partners.

4.10 Diagnostics for the Cox Model

4.10.1 Testing the proportional hazards assumption

Once the Cox model has been constructed, we will now test the proportionality of the hazards. To assess the validity of the proportional hazards assumption, we used the *cox.zph* function from the 'survival' package (Therneau & Lumley, 2017), which allows us to obtain both the individual x_1^2 statistics for each covariate and the global x_p^2 statistic for each model (Xue & Schifano, 2017).

Table 8. Proportionality test results for final multivariate Cox model

Variable	chisq	df	*p*-value
Sex	0.014	1	0.906
Age	1.266	1	0.261
Childbearing status before migration	1.591	1	0.207
Family economic difficulty	0.542	1	0.462
Duration of union	1.081	1	0.298
Partner's age at arrival	2.334	1	0.127
Global	10.836	6	0.094

Source: Author's own elaboration.

Table 9 summarizes the test results, showing that all covariates pass the individual proportionality test at the 0.05 level, and the model passes the global proportionality test at the 0.05 level. Therefore, we can assume the proportional hazards. Additionally, Schoenfeld residuals for each covariate are plotted in **Figure 10**. The solid line is a smoothing spline fit to the plot, and the dashed lines represent a +/- 2 standard error. If the proportional hazards assumption holds, the Schoenfeld residuals will be close to zero (Schoenfeld, 1982). For all six covariates, the smoothed pointwise confidence bands are all around 0, again confirming that the proportional hazards assumption is supported for all covariates.

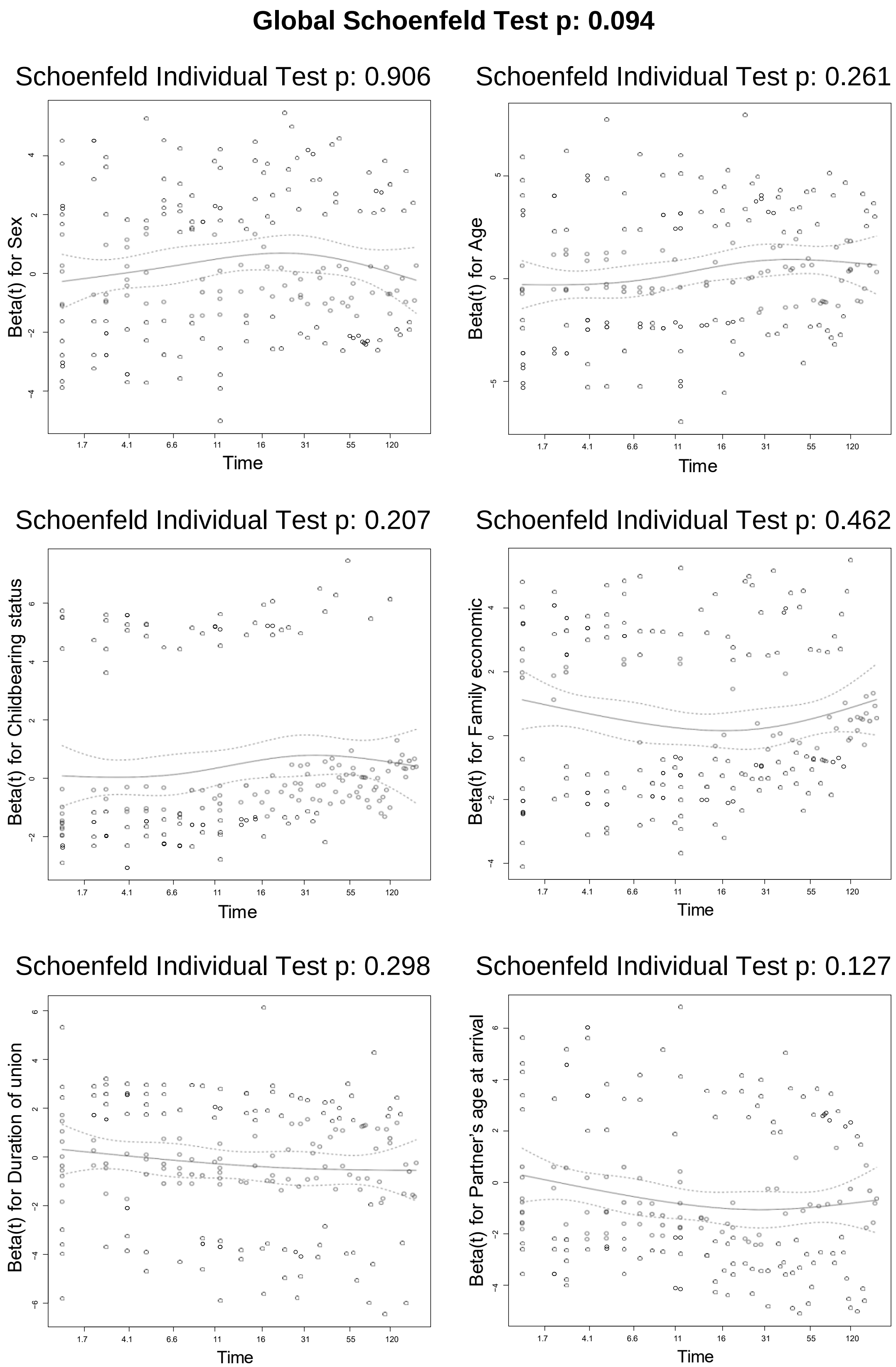


Figure 10. Schoenfeld residuals for each covariate against survival time.

Source: Author's own elaboration.

4.10.2 Testing influential observations

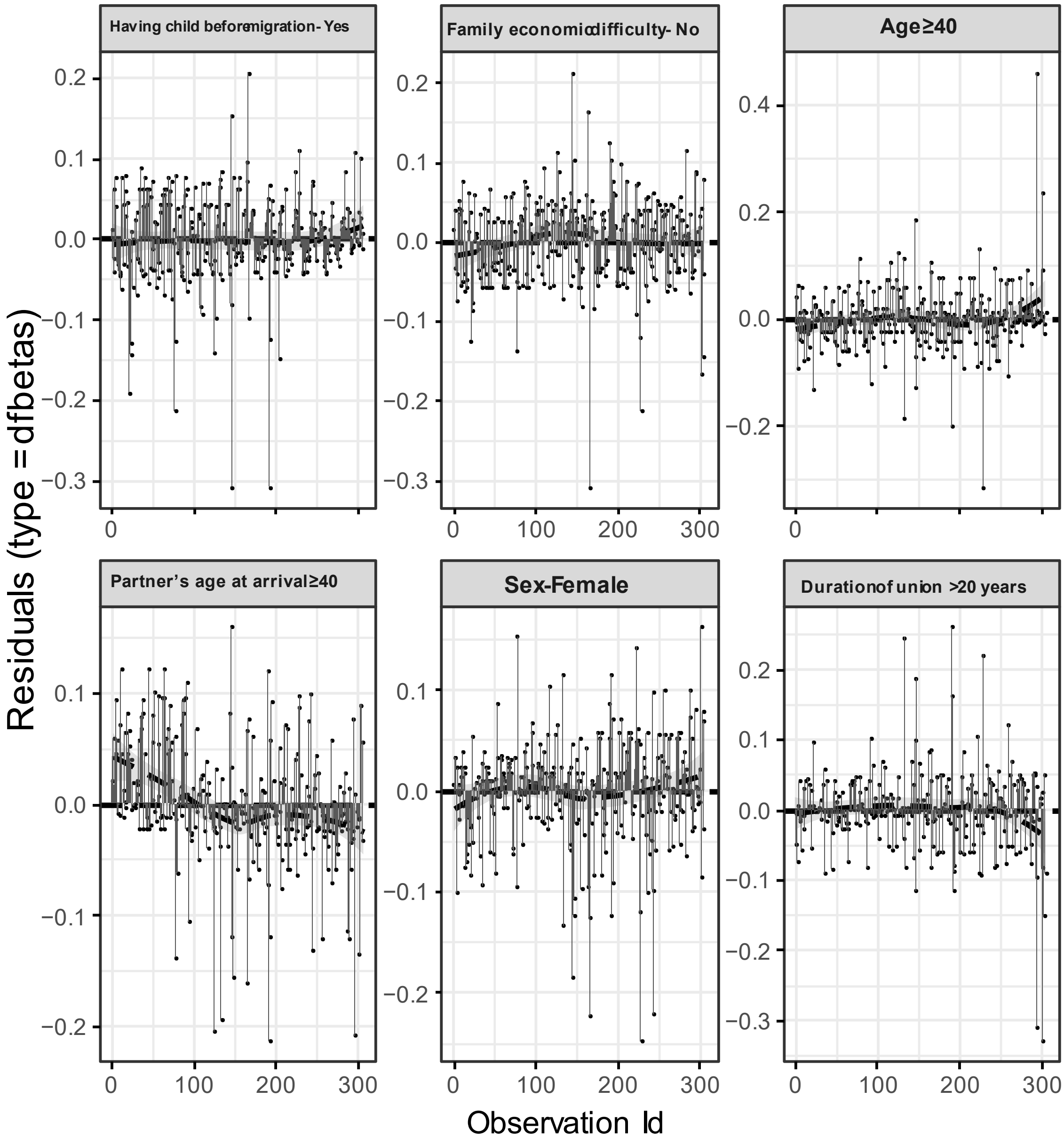


Figure 11. *Dfbetas* residuals for significant covariates of Model 7.
Source: Author's own elaboration.

It is also important to assess whether a fitted Cox regression model adequately describes the data. The *dfbetas* residual, which is a transformation of the score residual, allows us to check the impact of dropping a single observation on the parameter estimates. For simplicity, we only present the results of our final model by using the function *ggcoxdiagnostics* in 'survminer' package; for a detailed formulation of the score residual, see *Xue & Schifano* (2017). As illustrated in **Figure 11**, no observation had any *dfbetas* residuals greater than 1, which indicates that there are no significantly influential observations.

5

Discussion and Concludison

Our results show that couple reunification appears to be relatively common among migrants in Spain with a high propensity for joint migration. 50 months after separation (due to migration), about 84.6% of migrants had reunited with their spouses. Although a small group of migrants endure rather long separations, most couples seem to reunite relatively quickly. With respect to the gender of the primary migrant, the results show that women migrated first yields strong positive impact on spousal reunification; compared to male migrants, the odds of reunification are 71% higher for female migrants. This result challenges the notion of “secondary migration” which positions women as dependents of male migrants (Mato Díaz & Miyar Busto, 2017). This finding, however, can be explained by the traditional migration pattern embedded in gender inequalities, where living apart-together across countries or family dispersion is a functional strategy for male-led migrant families, where women are more likely to be left behind living with their children (Baizán et al., 2014; Mazzucato, Schans, Caarls, & Beauchemin, 2015); in contrast, female-led migration may indicate a higher intention for permanent settlement, with left-behind family members tending to join them quickly through stepwise migration (Francisco-Menchavez, 2020). As such, this finding may be relevant for future research and public policy to adequately support migrant families.

Self-selection in the process of reunification is also confirmed by socioeconomic variables among primary migrants and their partners. Specifically, the higher the educational level of first migrants, the more likely their partners are to join them in Spain. This finding is not surprising, as educational attainment is associated with the ability of migrant couples to successfully integrate into their host societies, both economically and culturally

(Schmalzbauer, 2008). Likewise, the availability of economic resources is crucial for couple reunification, with the probability of reunification for families without self-reported financial difficulty is 42% higher than for couples with a disadvantaged financial situation. This may be particularly true for migration flows from the global South to the global North, due to the restrictive immigration policies on family reunification as well as the additional costs to support family life in destination countries (Straut-Eppsteiner, 2021).

The propensity of reunification tends to decrease with the duration of the partnership, with couples of more than 20 years duration showing a 30% lower probability of reunification, which is consistent with previous studies (González-Ferrer, 2007). Indeed, transnational family strategy or living apart-together across countries is a common practice among long-term couples to reduce economic costs (Baizán et al., 2014; Beauchemin et al., 2015). According to the New Economics of Labor Migration (NELM) theory, by placing different household members in different countries where employment conditions are weakly or not correlated, families can diversify the sources of risk and better provide for their economic well-being (Stark, 1991), therefore, living apart across national borders becomes a rational decision, especially among partners with longer relationship duration. Finally, the relatively rapid reunification among couples older than 40 years at arrival may be partly explained by their better economic conditions, employment status or social networks when compared to their counterparts with young age (Fresnoza-Flot, 2015).

To conclude, this thesis utilized Cox regression models and survival estimation techniques to examine migrant couples' reunification process in Spain by using data from the 2018 Spanish Fertility Survey. We argue that this thesis advances our understanding of the use of advanced statistical methods in other disciplines, and the novel empirical findings contribute to the literature on family dynamics among immigrant populations in Spain. Such an understanding would be very useful in elaborating more realistic migration policies that take into account the needs of immigrants and contribute to a better integration of migrants' family members in destination countries.

Bibiography

Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology*, *13*, 61–98.

Baizán, P., Beauchemin, C., & González-Ferrer, A. (2014). An Origin and Destination Perspective on Family Reunification: The Case of Senegalese Couples. *European Journal of Population*, *30*(1), 65–87. https://doi.org/10.1007/s10680-013-9305-6

Battistella, G. (1995). Family reunification: Policies and issues. *Asian and Pacific Migration Journal*, *4*(2–3), 233–252. https://doi.org/10.1177/011719689500400204

Beauchemin, C., Nappa, J., Schoumaker, B., Baizan, P., González-Ferrer, A., Caarls, K., & Mazzucato, V. (2015). Reunifying versus Living Apart Together across Borders: A Comparative Analysis of sub-Saharan Migration to Europe. *International Migration Review*, *49*(1), 173–199. https://doi.org/10.1111/imre.12155

Bharat, C. I., Murray, K., Cripps, E., & Hodkiewicz, M. R. (2018). Methods for displaying and calibration of Cox proportional hazards models. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *232*(1), 105–115. https://doi.org/10.1177/1748006X17742779

Bland, J. M., & Altman, D. G. (2004). The logrank test. *Bmj*, *328*(7447), 1073. https://doi.org/10.1136/bmj.328.7447.1073

Bledsoe, C. H., & Sow, P. (2011). Family Reunification Ideals and the Practice of Transnational Reproductive Life among Africans in Europe. In C. H. Browner & C. F. Sargent (Eds.), *Reproduction, Globalization, and the State: New Theoretical and Ethnographic Perspectives* (pp. 175–191). New York, USA: Duke University Press. https://doi.org/https://doi.org/10.1515/9780822393948-017

Bonizzoni, P. (2015). Uneven Paths: Latin American Women Facing Italian Family Reunification Policies. *Journal of Ethnic and Migration Studies*, *41*(12), 2001–2020. https://doi.org/10.1080/1369183X.2015.1037257

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *British Journal of Cancer*, *89*(3), 431–436. https://doi.org/10.1038/sj.bjc.6601119

Bragg, B., & Wong, L. L. (2016). “Cancelled Dreams”: Family Reunification and Shifting Canadian Immigration Policy. *Journal of Immigrant and Refugee Studies*, *14*(1), 46–65. https://doi.org/10.1080/15562948.2015.1011364

Bueno, X., & Vidal-Coso, E. (2018). Vulnerability of Latin American Migrant Families Headed by Women in Spain During the Great Recession: A Couple-Level Analysis. *Journal of Family Issues*, 0192513X1880458. https://doi.org/10.1177/0192513X18804584

Caarls, K., & Mazzucato, V. (2015). Does International Migration Lead to Divorce? Ghanaian Couples in Ghana and Abroad Sociodemographic. *Population*, *70*(1), 135–161. https://doi.org/10.3917/popu.1501.0135

Caarls, K., & Mazzucato, V. (2016). Transnational relationships and reunification: Ghanaian couples between Ghana and Europe. *Demographic Research*, *24*(21), 587−614. https://doi.org/10.4054/DemRes.2016.34.21

Cerrutti, M., & Douglas S. Massey. (2001). On the Auspices of Female Migration from Mexico to the United States. *Demography*, *38*(2), 187–200. https://doi.org/10.1353/dem.2001.0013

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003a). Survival analysis part I: basic concepts and first analyses. *British Journal of Cancer*, *89*(2), 232–238. https://doi.org/10.1038/sj.bjc.6601118

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003b). Survival Analysis Part IV: Further concepts and methods in survival analysis. *British Journal of Cancer*, *89*(5), 781–786. https://doi.org/10.1038/sj.bjc.6601117

Collett, D. (1994). *Modelling Survival Data in Medical Research.* (2nd Editio; C. & Hall, Ed.). London. https://doi.org/http://dx.doi.org/10.1007/978-1-4899-3115-3

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society Series B-Methodological*, *B*(74), 187–220.

Del Boca, D. (2002). The effect of child care and part time opportunities on participation and fertility decisions in Italy. *Journal of Population Economics*, *15*(3), 549–573. https://doi.org/10.1007/s001480100089

Del Rey, A., & Grande, R. (2017). El efecto de las condiciones de llegada en la reproducción de las mujeres migrantes. *Revista Española de Investigaciones Sociológicas*, *159*, 113–134. https://doi.org/10.5477/cis/reis.159.113

di Belgiojoso, E. B., & Terzera, L. (2018). Family reunification - Who, when, and how? Family trajectories among migrants in Italy. *Demographic Research*, *38*(1), 737–772. https://doi.org/10.4054/DemRes.2018.38.28

E.L. Kaplan, P. M. (1958). Nonparametric Estimation from Incomplete Observations. *American Statistical Association*, *53*(282), 457–481.

Eremenko, T., & González-Ferrer, A. (2018). Transnational families and child migration to France and Spain. The role of family type and immigration policies. *Population, Space and Place*, *24*(7), 1–13. https://doi.org/10.1002/psp.2163

Estévez-Abe, M., & Caponio, T. (2022). Badante or Bride? Patterns of Female Migration in Italy, Japan, Korea, and Spain. *International Migration Review*, *56*(4), 1167–1194. https://doi.org/10.1177/01979183211070296

Fernandez, E., & Lee, J. S. (2011). Returning children in care to their families: Factors associated with the speed of reunification. *Child Indicators Research*, *4*(4), 749–765. https://doi.org/10.1007/s12187-011-9121-7

Flahaux, M.-L., Wayack ‐ Pambè, M., Soura, A. B., Compaoré, Y., & Sanogo, S. (2020). Reunifying and separating: An analysis of residential arrangements of migrant couples in Ouagadougou, Burkina Faso. *Popul Space Place*, *26*, 1–16. https://doi.org/10.1002/psp.2287

Francisco-Menchavez, V. (2020). All in the family: Transnational families and stepwise migration strategies. *Geographical Research*, *58*(4), 377–387. https://doi.org/10.1111/1745-5871.12416

Fresnoza-Flot, A. (2015). The Bumpy Landscape of Family Reunification: Experiences of First- and 1.5-generation Filipinos in France. *Journal of Ethnic and Migration Studies*, *41*(7), 1152–1171. https://doi.org/10.1080/1369183X.2014.956711

Galeano, J., & Gerber, R. (2023). The Economic Performance of Reunited Families in Switzerland, 2013–2018. *Journal of International Migration and Integration*, (0123456789), 2013–2018. https://doi.org/10.1007/s12134-023-01047-3

Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research*, *1*(4), 274–278. https://doi.org/10.4103/0974-7788.76794

González-Ferrer, A. (2007). The process of family reunification among original guest-workers in Germany. *Journal of Family Research*, *19*(1), 10–33. https://doi.org/10.20377/jfr-335

González-Ferrer, A. (2013). Spousal reunification among recent immigrants in Spain: Links with undocumented migration and the labour market. In A. Kraler, E. Kofman, M. Kohli, & C. Schmoll (Eds.), *Gender, Generations and the Family in International Migration* (pp. 193–218). Amsterdam University Press. https://doi.org/10.1515/9789048513611-009

González-Ferrer, A., Baizán, P., Beauchemin, C., Kraus, E., Schoumaker, B., & Black, R. (2014). Distance, transnational arrangements, and return decisions of senegalese, ghanaian, and congolese migrants. *International Migration Review*, *48*(4), 939–971. https://doi.org/10.1111/imre.12148

González-Ferrer, A., Hannemann, T., & Castro-Martín, T. (2016). Partnership formation and dissolution among immigrants in the Spanish context.

Demographic Research, *35*(1), 1–30. https://doi.org/10.4054/DemRes.2016.35.1

Gordon, M., & Lumley, T. (2022). *Forestplot: Advanced Forest Plot Using “grid” Graphics*. Retrieved from https://cran.r-project.org/web/packages/forestplot/index.html

Gupta, P. (2003). *Marriage at a distance: Spouse separation and the migrant family.* Philadelphia, PA: Population Studies Center, University of Pennsylvania.

Heering, L., van der Erf, R., & van Wissen, L. (2004). The role of family networks and migration culture in the continuation of Moroccan emigration: A gender perspective. *Journal of Ethnic and Migration Studies*, *30*(2), 323–337. https://doi.org/10.1080/1369183042000200722

Hervías Parejo, V. (2015). Políticas migratorias y dificultades de género en los reagrupamientos familiares. *Investigaciones Feministas*, *5*(0), 390–415. https://doi.org/10.5209/rev_infe.2014.v5.48301

Jasso, G., & Rosenzweig, M. R. (1986). Family Reunification and the Immigration Multiplier: U.S. Immigration Law, Origin- Country Conditions, and the Reproduction of Immigrants. *Demography*, *23*(3), 291–311. https://doi.org/https://doi.org/10.2307/2061432

Kanaiaupuni, S. M. (2000). Reframing the Migration Question: An Analysis of Men, Women, and Gender in Mexico. *Social Forces*, *78*(4), 1311–1347. https://doi.org/10.2307/3006176.

Kleinbaum, D. G., & Klein, M. (2008). *Survival Analysis: A Self-Learning Text* (3rd ed.). New York, NY: Springer. https://doi.org/10.1007/978-0-387-49979-6_25

Kofman, E. (2004). Family-related migration: A critial review of European studies. *Journal of Ethnic and Migration Studies*, *30*(2), 243–262. https://doi.org/10.1080/1369183042000200687

Kraus, E. K., Sauer, L., & Wenzel, L. (2019). Together or apart? Spousal migration and reunification practices of recent refugees to Germany. *Journal of Family Research*, *31*(3), 303–332. https://doi.org/https://doi.org/10.3224/zff.v31i3.04

Kreyenfeld, M., Diehl, C., Kroh, M., & Giesecke, J. (2021). Female employment and migration in European countries: Introduction to the Special Issue. *Journal of Family Research*, *33*(2), 230–251. https://doi.org/10.20377/jfr-700

Lahav, G. (1997). International versus national constraints in family-reunification migration policy. *Global Governance*, *3*(3), 349–372. https://doi.org/10.1163/19426720-00303010

Lübke, C. (2015). How Migration Affects the Timing of Childbearing : The Transition to a First Birth Among Polish Women in Britain. *European*

Journal of Population, *31*, 1–20. https://doi.org/10.1007/s10680-014-9326-9

Machin, D., Cheung, Y. B., & Parmar, M. (2006). *Survival Analysis: A Practical Approach*.

Mato Díaz, F. J., & Miyar Busto, M. (2017). Immigration policy and economic cycle effects on spousal reunification in Spain. *Revista Internacional de Sociología*, *75*(3), 071. https://doi.org/10.3989/ris.2017.75.3.15.138

Mazzucato, V., Schans, D., Caarls, K., & Beauchemin, C. (2015). Transnational families between Africa and Europe. *International Migration Review*, *49*(1), 142–172. https://doi.org/10.1111/imre.12153

Moore, D. F. (2016). *Applied Survival Analysis Using R*. Springer International Publishing.

Morris, E. (2015). Family Reunification and Integration Policy in the EU: Where Are the Women? *Journal of International Migration and Integration*, *16*(3), 639–660. https://doi.org/10.1007/s12134-014-0363-3

Petroff, A. (2016). Turning points and transitions in the migratory trajectories of skilled Romanian immigrants in Spain. *European Societies*, *18*(5), 438–459. https://doi.org/10.1080/14616696.2016.1139158

Reher, D., Requena, M., & Sánchez-Domínguez, M. (2013). Divided families among Latin American immigrants in Spain: Just how level is the playing field? *History of the Family*, *18*(1), 26–43. https://doi.org/10.1080/1081602X.2012.755931

Requena, M., & Sánchez-Domínguez, M. (2011). Las familias inmigrantes en España. *Revista Internacional de Sociologia*, *69*(M1), 79–104. https://doi.org/10.3989/ris.2011.iM1.387

Rey, A., & Stanek, M. (2022). Long goodbyes : pathways of leaving home by gender and destination in Spain. *European Sociological Review*, (December), 1–19.

Schmalzbauer, L. (2008). Family divided : the class formation of Honduran transnational families. *Global Networks*, *8*(3), 329–346. https://doi.org/https://doi.org/10.1111/j.1471-0374.2008.00198.x

Schoenfeld, D. (1982). Partial Residuals for The Proportional Hazards Regression Model. *Biometrika*, *69*(1), 239–241.

Stark, O. (1991). *The migration of labour*. Basil, Blackwell.: Oxford.

Straut-Eppsteiner, H. (2021). Undocumented Mothers and Work–Family Conflict in Restrictive Policy Contexts. *Journal of Marriage and Family*, *83*(3), 865–880. https://doi.org/10.1111/jomf.12737

Therneau, T. M., & Lumley, T. (2017). *survival: Survival analysis. R package. (Version 2.41-3)*.

Toma, S., & Vause, S. (2013). On Their Own? A Study of Independent Versus Partner-Related Migration from the Democratic Republic of the Congo and

Senegal. *Journal of Intercultural Studies*, *34*(5), 533–552. https://doi.org/10.1080/07256868.2013.827832

Toma, S., & Vause, S. (2014). Gender differences in the role of migrant networks: Comparing congolese and senegalese migration flows. *International Migration Review*, *48*(4), 972–997. https://doi.org/10.1111/imre.12150

Xue, Y., & Schifano, E. D. (2017). Diagnostics for the Cox model. *Communications for Statistical Applications and Methods*, *24*(6), 583–604. https://doi.org/10.29220/CSAM.2017.24.6.583

Yamaguchi, K. (1991). *Event history analysis.* Newbury Park, CA: SAGE Publications Ltd.

Zontini, E. (2004). Immigrant Women in Barcelona: Coping with the Consequences of Transnational Lives. *Journal of Ethnic and Migration Studies*, *30*(6), 1113–1144. https://doi.org/10.1080/1369183042000286278