sociological science

Family Networks and Childcare Choices: A Predictive Machine Learning Approach

Nicolás Soler, Tom Emery, Agnieszka Kanas

Erasmus University Rotterdam

Abstract: How first-time parents arrange childcare has critical implications for their careers and the child's development. Previous research shows that childcare choices are shaped by family care availability, understood as an additive function of a small set of parental and grandparental characteristics. However, research on family networks suggests that care availability is rather a non-linear, non-additive function of large family networks. We compare the predictive ability of these two perspectives using a machine learning framework and register-based family network data. We find that considering how the child's great-grandparents, aunts, uncles, and cousins shape care availability, and modeling their influence using more flexible models, provides small yet significant improvements in predictive ability, particularly among more disadvantaged parents. Predictions are driven by parents' and grandparents' socioeconomic characteristics, but cousins' age and daycare use are important yet understudied predictors. Other important understudied predictors include parents' self-employment, healthcare spending, and timing of daycare uptake.

Keywords: prediction; machine learning; family network; kinship; childcare

Reproducibility Package: Code to reproduce the results can be found at https://doi.org/10.5281/zenodo.19189668. The data are non-public microdata from Statistics Netherlands that are accessible to accredited researchers under certain conditions (see Statistics Netherlands 2026).

Citation: Soler, Nicolás, Tom Emery, Agnieszka Kanas, 2026. "Family Networks and Childcare Choices: A Predictive Machine Learning Approach" Sociological Science 13: 589-613.

Received: February 18, 2026

Accepted: March 23, 2026

Published: June 2, 2026

Editor(s): Stephen Vaisey, Michael Rosenfeld

DOI: 10.15195/v13.a23

How first-time parents arrange care for their pre-school child has critical implications for the parents' careers, the child's development, and gender equality. In particular, parents who rely exclusively on informal care, not combining it with formal care provided by professionals, face heightened care responsibilities that make it difficult to maintain employment, especially for mothers (Bünning 2017; Compton and Pollak 2014). Relying solely on informal care also entails foregoing the beneficial interaction with peers and caregivers that formal care facilitates (Heckman et al. 2013). Policy-makers and researchers are thus devoting increasing efforts to understanding childcare choices (Attanasio, Cattan, and Meghir 2022; Meyers and Jordan 2006).

A key factor influencing childcare choices is the availability of family care (Meyers and Jordan 2006). For many parents, family care complements rather than substitutes formal care, often by filling gaps in daycare schedules, whereas for other parents, it can altogether replace formal care. Previous work has theorized family care availability as an additive function of a small set of parental, household, and grandparental characteristics (e.g., Biegel, Neels, and Van den Berg, 2021; Morrissey 2008; Zick, Kowaleski-Jones, and Greenwalt 2022). We draw from the literature on family networks to argue that availability also depends on great-grandparents,

uncles, aunts, or cousins (Aeby, Widmer, and De Carlo 2014; Dykstra et al. 2006; Widmer 2016). These family members can influence childcare choices directly by providing care themselves, and indirectly by facilitating or constraining grandparental care supply (Aassve, Meroni, and Pronzato 2012; Kiraly, Humphreys, and Kertesz 2021; Meyers and Jordan 2006). In addition, their influence on childcare choices may be understood as a non-linear multiplicative function that considers the availability and needs of different family members (e.g., Aassve et al. 2012; Mulder and van der Meer 2009).

We thus argue that previous work has overlooked complexity in the relationship between family care availability and childcare choices (Meyers and Jordan 2006; Saraceno 2011). We assess this claim using a data-driven approach that combines predictive machine learning with register-based family network data from the Netherlands. Our analytical strategy consists of four steps. First, we study how our ability to predict the childcare choices of 51,390 first-time parents varies across conceptualizations of care availability, represented by different measurement and modeling choices (Garcia-Bernardo et al. 2025; Salganik et al. 2020; Verhagen 2022). Our first hypothesis is that predictive performance improves as we progressively measure care availability including more family members. Our second hypothesis is that predictive performance improves as we model the relationship with childcare choices by using a more flexible class of models. Second, we validate the best model by evaluating its performance on a held-out set of data. Third, we use interpretability techniques to unpack which dimensions of family care availability drive the predictions (Arpino, Le Moglie, and Mencarini 2022; Molnar 2025; Sun 2024). Lastly, we assess whether our approach differentially improves predictions across parental socioeconomic background to better understand the consequences of overlooking complexity (Garcia-Bernardo et al. 2025).

The results show that the best-performing classifier is an XGBoost model paired with a set of predictors that includes characteristics of parents, grandparents, and other family members (great-grandparents, aunts, uncles, and cousins). Predictive performance is largely driven by the characteristics of parents, especially mothers, but considering grandparental and kin characteristics, as well as using more flexible models, provides small yet significant improvements in predictive performance. Permutation feature importance analysis shows that maternal predictors consistently show higher importance than their paternal counterparts. We also identify important but understudied predictors: whether parents are self-employed, their annual healthcare costs, the child's month of birth (i.e., timing effects), and, in line with our argument about other family members, whether parents' siblings have children of daycare age and whether they use such daycare. Considering these family members improves predictions among a small subsample of parents, including mothers who are younger, less educated, and work fewer hours, and fathers who are less educated and self-employed, suggesting that incorporating complexity may help us better understand socioeconomic inequalities in childcare choices (Attanasio et al. 2022; Emery, 2026). Our results thus indicate that previous research has overlooked subtle yet theoretically relevant complexities in the relationship between family care availability and childcare choices.

Our study contributes to three different bodies of literature. First, we contribute to the literature on family networks (Daw, Verdery, and Margolis 2016; Dykstra

et al. 2006; Widmer 2016) by offering an effective empirical framework to model multidimensionality and showing that incorporating uncles, aunts, cousins, and great-grandparents helps explain variation in childcare choices, even when they are unlikely to be the primary carers. Second, we add to the literature on childcare by providing new evidence on the importance of different family members for parenting strategies (Biegel et al. 2021; Bünning 2017; Zick et al. 2022) and by identifying important but understudied predictors that help close gaps in extant knowledge (Attanasio et al. 2022; Emery 2026). Third, we contribute to the literature on predictive modeling for life course and family research (Hofman et al. 2021; Salganik et al. 2020; Sun et al. 2024; Verhagen 2022) by offering a novel substantive application and leveraging register-based network data (Garcia-Bernardo et al. 2025).

Background

Family Care Availability and Childcare Choices

How to care for their newborn child is a decision that parents make in the context of diverse family support structures (Meyers and Jordan 2006; Uttal 1999). These family structures are typically theorized to form informal care networks with varying degrees of care availability (e.g., Biegel and Maes 2022; Compton and Pollak 2014). Previous work on childcare choices has typically studied such care availability in aggregate terms, in terms of household composition, or in terms of grandparental availability.

The aggregate approach conceptualizes care availability as a resource that either is or is not available to parents, without distinguishing who may provide care or how much care may be available (e.g., Carlin et al. 2019 gauge care availability by asking parents "is there a family member or friend available to care for the child regularly?"; see also Burchinal et al. 2008). The other two approaches aim to understand exactly which family members are available to provide care. The household approach focuses on the number and type of adult family members who reside in the child's household (Crosnoe et al. 2014; Markowitz, Ryan, and Johnson 2014; Morrissey 2008). The grandparental approach assumes that grandparents are the only non-cohabiting family members willing to provide care that is regular and intensive enough to shape parents' childcare choices (e.g., Jappens and van Bavel 2012). Most such studies focus on whether a grandparent resides in proximity to the parental household (Compton and Pollak 2014; Kuhltau and Mason 1996; Zick et al. 2022). A few more detailed studies condition care availability on multiple characteristics of one grandparent, typically the maternal grandmother (e.g., her age, employment, health, and geographical proximity; Biegel et al. 2021; Biegel and Maes, 2022).

Despite their different conceptualizations of care availability, the three approaches theorize its relationship to childcare choices as an additive function. This means that household and grandparental characteristics are generally assumed not to interact in shaping care choices (e.g., Burchinal et al. 2008; Markowitz et al. 2014; Zick et al. 2022). In methodological terms, this implies that most of these studies

use regression models to map this relationship (e.g., OLS, logistic, or multinomial regression; for an event history approach, see Bünning 2017). We argue that treating childcare choices as an additive function of a small set of parental, household, and grandparental characteristics is at odds with the theorization of family members as an informal care network.

Toward a Network Approach

The notion that families constitute informal care networks is also central in the related yet distinct literature on family networks. In this literature, families are understood as groups bound together by biological, emotional, legal, and practical relationships that define the social roles of their members relative to one another (Dykstra et al. 2006; Widmer 2016). These roles impose certain expectations concerning who ought to care for whom. Importantly, the relationships shaping these roles need not be confined to household units or to parents and their children, but often also involve extended kin (Furstenberg 2020; Hünteler et al. 2025). Understanding the availability of family care thus requires considering a large number of family members, as well as studying how their life circumstances and interdependencies condition who is available to provide care (Aeby et al. 2014; Daw et al. 2016; Widmer 2016).

Previous work on childcare choices portrays families as networks but does not engage with this tradition of network theorizing about care availability. We argue that this has three implications. The first implication is that the focus on household members and grandparents may overlook important family relationships that can shape the availability of care. In addition to grandparents, a child's aunts and uncles can be important sources of in-person care, parenting advice, and normative pressure (Bojarczuk and Mühlau 2018; Kiraly et al. 2021; Uttal 1999). They can also affect childcare choices by competing for grandparental care, especially when they are parents themselves (Aasve et al. 2012; Mulder and van der Meer 2009). A similar argument can be made about great-grandparents, step-kin, and other family members (Furstenberg 2020; Hünteler et al. 2025). Great-grandparents in particular may be important competitors for grandparental care (Xu 2019), but may also be carers.

The second implication is that understanding which family members are available to provide care requires attention to their life circumstances and interdependencies (Aeby et al. 2014; Widmer 2016). This means that care availability is not determined solely by the number of family members parents have but also by their age, gender, health, relationship quality, geographical proximity, workload, and schedules (Bünning 2017; Jarvis 1999; Uttal 1999). The few quantitative studies on childcare choices that consider such characteristics among grandparents find that their employment, civil status, proximity, and educational attainment are related to how parents arrange care (Biegel et al. 2021; Biegel and Maes 2022). Family care availability thus needs to be measured in a multidimensional manner that considers multiple characteristics of family members.

Once we consider additional characteristics of other family members, the third implication is that it is necessary to consider how these may interact to shape childcare choices. For instance, Aassve and colleagues (2012) show that the effect of

grandparental childcare availability on fertility decisions is moderated by the number of grandchildren and the existence of siblings that compete for grandparental care. Mulder and van der Meer (2009) show that care availability depends not only on absolute geographical proximity to grandparents but also on the interaction with proximity to other siblings. These multiplicative effects are typically not considered in work on childcare choices.

A few previous studies incorporate the first implication into their theorization and research design, providing fertile ground to articulate a more complete network perspective. Bojarczuk and Mühlau (2018) estimate the provision of regular childcare by different types of family members of Polish mothers in Dublin, finding that kin have 2.98 times higher odds of providing childcare than friends, but 8.93 times lower odds than grandparents. Similarly, Dykstra et al. (2006) find that 56 percent of parents in the Netherlands rely on grandparental childcare, whereas 20 percent rely on childcare by aunts and uncles. Bünning (2017) finds instead that the number of family members living in the same neighborhood as western German mothers has a larger effect on the likelihood that they will transition into employment than the presence of grandmothers alone, but this does not hold for migrant and eastern German mothers.

These studies suggest that considering kin beyond grandparents can indeed help in understanding childcare choices, but the two other implications we identified remain unaddressed, and the extent to which kin matter is still an open question (Furstenberg 2020; Hünteler et al. 2025).

Predictive Modeling of Complexity in Family Networks

We build on previous work to propose a more multidimensional network approach that models childcare choices as a multiplicative function of family care availability measured by multiple characteristics of parents, grandparents, uncles, aunts, cousins, and great-grandparents. However, shifting to more multidimensional measurement and modeling poses well-known statistical challenges of increased multicollinearity, increased researcher degrees of freedom in model specification (e.g., which interactions to include), and decreased model interpretability, among others (Hofman et al. 2021; Molina and Garip 2019; Verhagen 2022).

To tackle these challenges, we depart from previous work on childcare choices and adopt a predictive rather than explanatory modeling paradigm (Breiman 2001; Hofman et al. 2021). Research designs in the explanatory paradigm typically combine measures of the few factors theorized to explain an outcome with relatively simple models that map their relationship to such an outcome (Shmueli 2010; Verhagen 2022). Both measurement and modeling aim to remain close to theory, tractable, and readily interpretable. The focus is on providing accurate, reliable, and ideally causal estimates of the relationships between factors and outcomes (Shmueli 2010).

In contrast, the predictive paradigm focuses on the capacity to predict outcomes rather than estimating specific effects. The aim is to understand to what degree current theories and research designs can account for variation in that outcome, which is taken as an indication of how much is (not) known about the underlying social process (Shmueli 2010; Verhagen 2022). Predictive approaches thus favor the

inclusion of many predictors and multidimensional operationalizations to capture the complexity of such processes (Sun 2024; Verhagen 2022). This invites the use of machine learning models that can automatically account for multiplicative effects, non-linearities, and complex covariance structures (Arpino et al. 2022; Molina and Garip 2019). The challenge is to make these models interpretable so that they can feed back into theory and lead to inductive discovery (e.g., of previously disregarded predictors, Sun 2024). This can be done, with some caveats, through machine learning interpretability methods (Molnar 2025; Sun 2024).

Hence, the explanatory paradigm seeks parsimony to achieve theoretically grounded estimates, whereas the predictive paradigm seeks complexity to enable aggregate comparisons of conceptual models and to refine them through inductive insights (Breiman 2001; Hofman et al. 2021; Molina and Garip 2019). In our case, predictive modeling allows us to compare how the measurement and modeling choices of previous quantitative work influence our understanding of childcare choices (Shmueli 2010; Verhagen 2022). This is done by comparing how well different models predict the outcome of interest (Arpino et al. 2022; Salganik et al. 2020). For instance, Garcia-Bernardo et al. (2025) study how well the relationship between educational attainment and social context is captured by different operationalizations of social context, finding that more nuanced operationalizations help predict the attainment of more disadvantaged students. We apply the same approach to family support structures and care availability. Based on our comparison of the childcare and family networks literature, we test two hypotheses:

> Hypothesis 1: Our ability to predict childcare choices increases when we measure family care availability in terms of the characteristics of more family members.
>
> Hypothesis 2: Our ability to predict childcare choices increases when we model their relationship to family care availability through more flexible models that account for non-linear and multiplicative effects.

Hypothesis testing will also identify the best-performing combination of measurement and modeling. By unpacking the resulting predictions using interpretability techniques, we aim to "open the black box" to understand which characteristics of family networks appear as most predictive of childcare choices (Arpino et al. 2022; Molnar 2025; Sun 2024). This approach will allow us to inductively search for empirical patterns that may not have been studied yet (Molina and Garip 2019; Sun 2024). In addition, by investigating whether predictions improve more among specific social groups (e.g., younger or lower educated parents), we aim to understand whether overlooking complexity affects our ability to understand social inequalities in childcare choices (c.f., Garcia-Bernardo et al. 2025). In this way, the predictive paradigm allows us to first test our deductive argument about the conceptualization of family care availability by quantifying the implications of methodological choices and then feed back into the literature with new inductive empirical evidence.

Empirical Case: The Netherlands

We test our argument by studying the childcare choices of mixed-gender couples having their first child in the Netherlands.

The Netherlands is a high-income central European country characterized by a relatively low average age at which individuals leave the parental home (23.2 years in 2024, Eurostat 2025), a high average age of entry into motherhood (30.4 years in 2021, Statistics Netherlands 2025a), and a low prevalence of large multigenerational households (Statistics Netherlands 2025b). The majority of parents thus raise their children in their own household. At the same time, family norms favor residing near grandparents and relying on them for childcare (see Dykstra et al. 2006). Grandparental care is facilitated by extensive public transportation and widespread commuting between municipalities. The Netherlands is also characterized by 60 percent of women working part-time and by gendered caregiving patterns (OECD 2019).

Formal childcare is provided for children between four months and four years old by certified centers (*kinderopvang*) or individuals who provide care at home (*gastouders*). We will refer to both services as formal care or daycare. Both services are privately provided but publicly subsidized. The subsidy covers a large proportion of the otherwise very high costs, depending on parental income and as long as parents work and co-reside with the child (see Emery 2026). The subsidy is paid monthly based on an income estimate provided by parents and is subject to reassessment and settlement at the end of the year. As the subsidy is reported in tax returns, it is possible to use register data to study childcare choices (Emery 2026).

Data and Methods

Our analytical strategy compares the predictive performance of different combinations of measurement and modeling choices (c.f., Garcia-Bernardo et al. 2025; Salganik et al. 2020). We will call each of such combinations a "classifier", that is, a model that learns from data to perform a binary classification task. The task is to predict whether parents will rely solely on informal childcare. Each classifier will be characterized by the set of predictors it includes and the type of model used to relate them to the outcome. For each classifier, we need to tune it (i.e., learn hyperparameters), train it (i.e., learn parameters), and evaluate it (i.e., quantify predictive performance). We will then compare their performance, select the best-performing one, and validate it on an independent set of data and through interpretability techniques. We explain each step following the REFORMS reporting guidelines to ensure transparency and reproducibility (Kapoor et al. 2024). The code to reproduce the analysis is available in the following repository: https://doi.org/10.5281/zenodo.19189668.

Dutch Register Data

We use the microdata registers maintained by Statistics Netherlands. The registers contain information on the entire population of the Netherlands compiled from different administrative sources. Data files are accessible to accredited researchers via a secure remote access environment (see Statistics Netherlands 2026). Files can be linked through pseudonymized unique personal identifiers (file names can be found in Appendix 1 in the online supplement).

The Dutch register data offers multiple advantages for our purposes. First, it covers the full population, affording high external validity and providing a large number of cases that facilitate model estimation. Second, it provides rich information on individuals that can be used to derive an extensive list of predictors. Third, it identifies family relationships among individuals in the population, allowing us to study family networks at scale (van der Laan et al. 2023). Lastly, the data allows us to measure childcare choices directly from tax registers rather than relying on self-reports, increasing measurement accuracy, reliability, and validity.

Sample Construction

First, we create a list of all individuals who are alive, have a registered address in the Netherlands, and are 110 years old or younger as of January 1, 2020. These individuals constitute the study population. We create a similar list for January 1, 2022. We process all other data files for each year to exclude individuals not in the respective list.

Second, we identify the family relationships between individuals for the period between January 1, 2020, and January 1, 2022, using the FAMILIENETWERKTAB file. The file identifies family relationships through registries that link parents and children based on birth, recognition of fathers, and legal adoptions, as well as registries of unions. We select parent–child, sibling, and partner relationships, from which we can derive others. We exclude parent–child relationships in which the parent is not more than 10 years older than the child.

Third, we identify all children in the 2022 population list born in 2021 (179,120 children). We then exclude all children who have a full or half-sibling in that year according to the FAMILIENETWERKTAB file, retaining only first-born children (74,850). We then identify the legal parents of each child, retaining only children who have both a father and a mother (70,208).

Finally, we select children that meet the following conditions: (1) both parents must be born in the Netherlands (52,833); (2) both parents must have at least one living parent that resides in the Netherlands in 2020 (51,480), and (3) the child must be registered as resident in the same address as the mother in 2022 (51,390). We exclude immigrants because they tend to have a higher proportion of missing data in the administrative registers, as well as a different underlying missingness mechanism. We exclude parents without at least one living parent because, for them, we cannot trace family relationships reliably. The co-residence condition helps us interpret address-based predictors.

The units of analysis are thus 51,390 children born in 2021 who co-reside with their mother as of January 1, 2022. The children are the firstborn children of the couple and have a father who may (not) co-reside with them.

Outcome

The outcome is a binary measure that identifies children who have not enrolled in publicly subsidized childcare by 2024. We use the KINDERENTAB file to compile a list of all children who were enrolled in publicly subsidized childcare (*kinderopvang* or *gastouders*) at any point between January 1, 2021, and December 31, 2023. We

identify children not present on the list as the positive class. The outcome is imbalanced, with 12.64 percent of children in the positive class (6,495 cases). This measure captures whether children participate in daycare, with its associated benefits for both parents and children, but does not distinguish between non-participation due to choice (e.g., a preference for family care) and lack of access (e.g., administrative or financial barriers).

Predictors

We calculate three nested sets of predictors. Each set is nested within the next, so that the number of predictors increases with each set. All predictors are measured at least one year before birth (January 1, 2020).

The first set of predictors concerns the characteristics of parents and the child. Predictors concerning parental characteristics are disaggregated by parent. The second set of predictors incorporates grandparental characteristics separately for paternal and maternal grandparents. All predictors are measured at the level of the grandparental set (e.g., mean age) to capture grandparental characteristics regardless of the number of living grandparents per parent. The third set of predictors incorporates characteristics of great-grandparents, uncles, aunts, and cousins, differentiating between paternal and maternal kin. If an individual is both paternal and maternal kin, we count her only as maternal kin. Finally, we construct a baseline set of predictors that includes only characteristics of the child and the mother. We use maternal characteristics because of our focus on children who co-reside with mothers and the gendered nature of parenting.

Appendix 2 in the online supplement provides the full list of predictors per set, their operationalization, and descriptive statistics.

Preprocessing

The fragmentation of microdata across domain-specific files results in structural missingness. For instance, parents who do not file a tax return are missing from the INPATAB income file and our income measure. Because such missing values indicate qualitative differences between individuals, and our predictors have meaningful zero values (e.g., annual income), we address structural missingness by imputing a value of zero and additionally flagging missingness using a dummy variable (Shmueli 2010). This approach allows us to explicitly estimate how missingness relates to the outcome (Van Ness et al. 2023). We apply this approach in the parental predictor set.

For predictors in the grandparental and kin sets, structural missings arise when parents do not have the relevant type of family member, or these family members are missing from the microdata files. We thus impute a value of zero in the characteristics of family members (e.g., employment hours) before aggregating them (e.g., before calculating how many of the father's siblings work less than a number of hours). If the predictor is still missing after aggregation (e.g., if the father has no siblings), we input a zero. Tables A2.2 and A2.3 in the online supplement report the counts of structural missing values per predictor.

We also preprocess the predictors to address extreme values. We set negative values of predictors concerning income and health expenditures to zero (767 cases), and then apply a logarithmic transformation. For predictors that concern counts of kin, which have low average values but a few very high values, we clip the maximum at five to reduce dimensionality and tackle outliers (1,409 cases).

We also preprocess data to ensure that models treat all predictors equally. We perform one-hot encoding on all categorical predictors and min-max scaling of all continuous predictors. After preprocessing, the baseline, parental, grandparental, and kin predictor sets contain 26, 44, 72, and 98 predictors, respectively.

Models

We use two types of models to investigate the extent to which predictive performance improves when we consider interaction effects and non-linearities.

The first model is logistic regression, the main model used in the social sciences to map a binary outcome as a function of predictors (e.g., Biegel and Maes 2022; Mulder and van der Meer 2009). This mapping does not account for higher-order interactions and non-linearities (Müller and Guido 2016). Table A3.1 in the online supplement describes the hyperparameter space.

The second model is an XGBoost model, considered to be the gold standard in machine learning tasks (Müller and Guido 2016). The XGBoost model trains an ensemble of decision trees, with the first tree optimizing a loss function and every additional tree adjusting the model parameters to correct previous erroneous predictions. The final prediction is the sum of the predictions of all trees. This approach combines tree weighing, regularization, pruning, and repeated sampling of observations and features to avoid overfitting and optimize computation (Chen and Guestrin 2016). XGBoost can capture higher-order interactions and non-linearities (Müller and Guido 2016). See Table A3.2 in the online supplement for details on the hyperparameter space.

As a baseline model for the baseline feature set, we train a logistic regression without regularization, following standard practice in the social sciences (Hofman et al. 2021).

Model Training and Tuning

Before training and tuning our classifiers, we first separate the data used for training and tuning from the data used for evaluation (Molina and Garip 2019; Müller and Guido 2016; Salganik et al. 2020). We split our sample into a training set (70 percent, 35,973 observations) and a holdout set (30 percent, 15,417 observations). We further split the holdout set into a validation set (30 percent, 4,625 observations) and a test set (70 percent, 10,792 observations). These will be used for different steps of the analytical strategy. We stratify all splits by the outcome to maintain a comparable positive-to-negative class ratio: 12.64 percent (4,547), 12.63 percent (584), and 12.64 percent (1,364) of observations in the train, evaluation, and test sets, respectively, fall into the positive class.

We train seven classifiers (three predictor sets $\times$ two models + one baseline) using stratified nested cross-validation on the train set, following standard practice

in machine learning (Müller and Guido 2016; Wainer and Cawley 2021). This nested procedure consists of an inner and an outer cross-validation. The inner cross-validation uses the training data to learn model hyperparameters (e.g., regularization strength), which then determine how the model learns its parameters in the outer cross-validation (for details, see Müller and Guido 2016; Wainer and Cawley 2021). We use ten outer cross-validation folds to ensure sufficient statistical power for the hypothesis tests (Rainio, Teuho, and Klén 2024) and four inner cross-validation folds to ensure robust tuning while reducing computational costs (Müller and Guido 2016). We train all classifiers using the log-loss as the scoring rule to ensure comparability and propriety (Gneiting and Raftery 2007).

We perform hyperparameter tuning via the Optuna library (Akiba et al. 2019). Optuna uses state-of-the-art optimization algorithms to sample hyperparameters from a predefined hyperparameter space (see Appendix 3 in the online supplement). In each sampling run, we sample hyperparameters to optimize the mean log-loss across the four folds of the inner cross-validation procedure. We perform 150 sampling runs per outer fold of the nested cross-validation. The run that achieves the best mean log-loss is selected as the final one for the outer fold. For XGBoost classifiers, we constrain the tuning process for each outer fold using a Hyperband pruner, a pruning callback that stops a sampling run if the first inner fold is pruned, and setting the early stopping rounds hyperparameter equal to 30. For logistic regression classifiers, we stop tuning for each outer fold if there has been no improvement in the mean log-loss after 20 runs.

The nested cross-validation procedure will yield ten estimates of predictive performance (one per outer fold) for each of the seven classifiers, which we will use to test our hypotheses and select the best classifier.

Hypothesis Testing

To test our hypotheses, we compare classifier performance measured on each outer fold of the nested cross-validation procedure. We use the area under the curve of the precision–recall curve (AUC-PR) as the main metric. The precision–recall curve plots precision (the proportion of true positives among all positive predictions) against recall (the proportion of true positives identified correctly) for different thresholds at which predicted probabilities are converted to predicted classes. The larger the area under the curve, ranging from zero to one, the better the predictions. AUC-PR focuses on the accurate prediction of the positive class by quantifying the ability to both correctly predict positive instances and not label negative instances as positive. This is particularly important in imbalanced classification tasks like ours (He and Ma 2013:27,200).

We test our hypotheses using a repeated measures statistical test (Rainio et al. 2024). First, we test for overall differences in performance across classifiers using Friedman's omnibus test with the chi-square distribution. Second, if Friedman's test rejects the null hypothesis that classifiers are identical, we compare performance between pairs of classifiers using one-tailed Wilcoxon signed-rank tests with a Holm–Bonferroni correction for multiple testing.

Evaluation of the Best-Performing Classifier

We select the classifier that (1) has the highest average PR-AUC score and (2) is statistically different from the score of the second-best classifier, according to the Wilcoxon tests (Rainio et al. 2024). We tune and train the selected classifier again on the entire training set using Optuna with four stratified cross-validation folds and 200 trials. Retraining the classifier allows us to fully leverage our data to estimate model (hyper)parameters. The final hyperparameter values are provided in Table A3.3 in the online supplement.

We validate the classifier by evaluating its performance through metrics calculated on the validation and test sets. The confusion matrix summarizes the number of true positives, true negatives, false positives, and false negatives by comparing the predicted classes to the observed outcome. This is useful to assess whether the classifier accurately predicts the positive and negative classes. The true positive rate (the proportion of true positives among true positives and false negatives) and the true negative rate (the proportion of true negatives among true negatives and false positives) help with this. The $F1$ score takes the harmonic mean of precision and recall to provide a balanced assessment of the classifier's ability to predict the positive class. A higher score in its zero-to-one range is better. The Brier score is a strictly proper scoring rule (Gneiting and Raftery 2007) that represents the mean squared error between predicted probabilities and the observed outcome. The Brier score is useful for assessing calibration, that is, the degree to which the probabilities predicted by the classifier reflect the observed probabilities (e.g., approximately 20 percent of observations with a predicted probability of 0.2 should be in the positive class). A lower score in the zero-to-one range is better.

Before calculating these metrics, we tune the threshold at which predicted probabilities are converted into predicted classes. This is necessary to ensure the accurate calculation of performance metrics based on predicted classes (e.g., $F1$). We use scikit-learn's TunedThresholdClassifierCV class with 200 thresholds, scoring on the $F1$ score, and using eight-fold stratified cross-validation on the training set. We then evaluate performance on the validation set to assess whether the model is overfitting the training set and whether the decision threshold tuning has improved performance. If performance is satisfactory, we proceed to the final evaluation on the test set.

Predictor Importance

We investigate the importance of each predictor for the predictions made by the best-performing classifier. We quantify predictor importance through the global model-agnostic permutation feature importance (PFI) method (Fisher et al. 2019; Molnar 2025). PFI consists of reshuffling the values of predictors across observations, one predictor at a time, and recalculating predictive performance. The resulting estimate represents the change in performance caused by the permutation of that predictor, that is, how much the predictor matters (Molnar 2025). We permute each predictor 300 times to derive more precise and reliable estimates. We assess predictor importance on the test set using PR-AUC as the performance metric. Higher values represent a higher importance.

PFI has the advantages of being easily interpretable and incorporating interaction effects into the quantification of importance (Molnar 2025). However, PFI does not provide information on the direction of the relationship between the predictor and the outcome. This is not a large problem for our application because we focus on exploration rather than inference.

Improvements in Predictions

Lastly, we investigate whether more multidimensional measurement and flexible modeling improve our understanding of childcare choices differently across social groups (c.f., Garcia-Bernardo et al. 2025). We select the second-best-performing classifier and re-train it following the same procedure as in the Evaluation of the best-performing classifier section. We then use it to predict the outcome on the test set. We identify parents whose predictions by this second-best-performing classifier are not correct, but whose predictions by the best-performing classifier are correct, that is, parents for whom predictions improve when incorporating multidimensionality. We use a logistic regression model to investigate which parental characteristics correlate with such improvement.

Results

Classifier Performance and Hypothesis Testing

The first step in the analysis is to identify which classifiers, that is, combination of predictors and model type, best predict childcare choices. We hypothesize that classifiers that include information about more family members and/or employ more flexible model types will perform better.

Figure 1 plots classifier performance in terms of the mean PR-AUC score across the ten cross-validation folds. First, we observe that all six classifiers outperform the baseline on average. The baseline classifier scores 0.346 mean PR-AUC (SD: 0.016), whereas the other classifiers score between 0.372 (7.5 percent improvement, SD: 0.014) and 0.465 (34.39 percent improvement, SD: 0.021) mean PR-AUC. Friedman's omnibus test rejects the null hypothesis that there are no differences in performance between classifiers (statistic: 57.39, p-value <0.01). Second, increasing the number of predictors increases performance, regardless of model type. This provides preliminary evidence for Hypothesis 1. Third, increasing the flexibility of the model increases performance, regardless of which predictor set is used. This provides preliminary evidence for Hypothesis 2. However, our results also show that the sociodemographic characteristics of mothers and children, included in the baseline, account for most of the predictive performance. The improvements in performance afforded by increasing complexity in measurement and modeling are small and incremental.

Table 1 reports the results of the hypothesis testing based on one-tailed Wilcoxon tests. We test our two hypotheses by comparing the performance of pairs of classifiers that share model type but differ in predictor set (H1) or share predictor set but differ in model type (H2). In all paired comparisons, the results show that

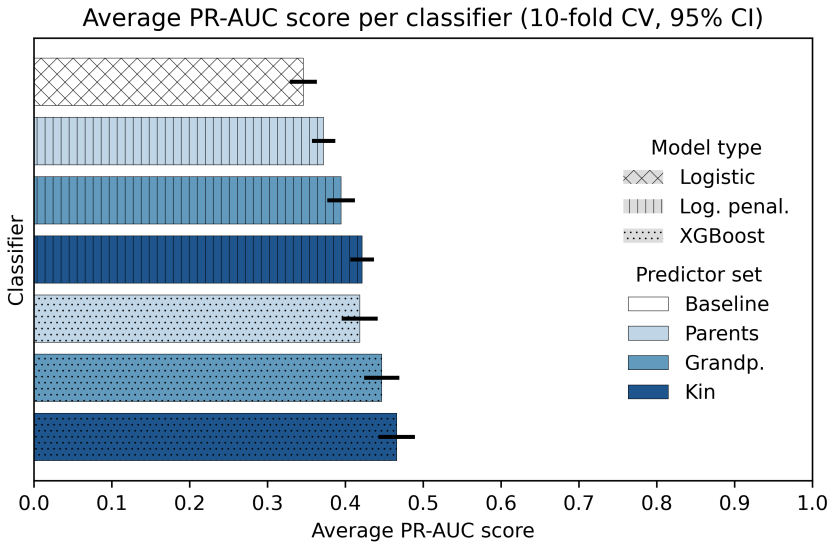

Figure 1: Average PR-AUC score per classifier calculated on the 10 outer folds of the nested cross-validation procedure (95 percent confidence intervals). The confidence intervals have been adjusted to account for the non-independence of the cross-validation estimates, following the procedure proposed by Nadeau and Bengio (2003).

predictive performance is significantly higher for classifiers that use a larger predictor set or a more flexible model. The best-performing classifier combines an XGBoost model with the largest set of predictors. These results provide evidence to reject the null hypothesis for both H1 and H2. In substantive terms, this implies that extant research on the relationship between family care availability and childcare choices fails to map its full complexity. Considering the availability of uncles, aunts, and great-grandparents helps predict childcare choices.

At the same time, improvements in predictive performance remain small, even if statistically significant. For instance, classifiers that include grandparents improve on the parental classifiers on average by 0.023 PR-AUC points (logistic regression model) and by 0.028 PR-AUC points (XGBoost model), whereas classifiers that include kin improve on the grandparental classifiers by a further 0.027 and 0.019, respectively. These results suggest that the inclusion of grandparents and kin helps only marginally to understand whether Dutch first-time parents will rely exclusively on informal care. This contrasts with both research focusing on grandparents only (Biegel et al. 2021; Biegel and Maes 2022; Compton and Pollak 2014; Jappens and van Bavel 2012; Kuhltau and Mason 1996; Zick et al. 2022) and research calling for more attention to kin (Furstenberg 2020; Hünteler et al. 2025; Kiraly et al. 2017; Uttal 1999; Widmer 2016). These small, significant effects may nevertheless concern family processes that are of theoretical interest, such as moderation or mediation effects (Aasve et al. 2012; Bojarczuk and Mühlau 2018; Mulder and van der Meer 2009). The rest of the analysis aims to identify such theoretically useful insights through inductive analysis.

Table 1: Hypothesis testing based on Wilcoxon tests with Holm–Bonferroni correction for multiple testing.

Hypothesis	Classifier Comparison	Difference in Average PR-AUC	Wilcoxon's *t*-Statistic
Hypothesis 1	$\text{Log}_{\text{Gpa}} > \text{Log}_{\text{Pa}}$	0.023	55.0†
	$\text{Log}_{\text{Kin}} > \text{Log}_{\text{Gpa}}$	0.027	55.0†
	$\text{XGB}_{\text{Gpa}} > \text{XGB}_{\text{Pa}}$	0.028	55.0†
	$\text{XGB}_{\text{Kin}} > \text{XGB}_{\text{Gpa}}$	0.019	55.0†
Hypothesis 2	$\text{XGB}_{\text{Pa}} > \text{Log}_{\text{Pa}}$	0.047	55.0†
	$\text{XGB}_{\text{Gpa}} > \text{Log}_{\text{Gpa}}$	0.052	55.0†
	$\text{XGB}_{\text{Kin}} > \text{Log}_{\text{Kin}}$	0.044	55.0†

Notes: The classifier comparison column indicates the two classifiers whose average predictive performance is being compared in each one-tailed test (e.g., logistic regression with grandparental features > logistic regression with parental features).
†*p*-Values <0.01.

Performance on the Validation and Test Sets

The second step in the analysis is to validate the best classifier before using it for inductive analysis. We validate the classifier by comparing its performance on the training data to its performance on independent sets of data that were not used during model training. Table 2 shows performance after threshold tuning on the validation set and test set. The results are nearly identical across both sets, so we focus on performance on the test set.

The PR-AUC in the test set equals 0.453 and is within the 95 percent confidence interval of the mean PR-AUC of 0.462 achieved in the training set (0.441–0.483). The similarity of the PR-AUC score and the $F1$ score of 0.457 shows that these results were not dependent on the performance metric. In addition, the true positive rate of 0.470 is sizably lower than the true negative rate of 0.915, implying that the classifier performs much better at predicting childcare choices that include daycare use than those relying solely on informal care. The very low Brier score of 0.088 shows that there are no serious calibration issues.

This is strong evidence that our performance evaluation was neither overfit on the training data nor dependent on the choice of metric. This means that our classifier shows external validity and can be further inspected to inductively explore the relationship between predictors and outcome.

Predictor Importance

The third step is to identify which predictors drive the performance of the best-performing classifier to make the results more interpretable (Molnar 2025).

Figure 2 shows the thirty predictors with the highest permutation feature importance. The most important predictors, by a large margin, concern mothers' income and employment. This is to be expected: these predictors signal mothers' willingness and capacity to be employed after birth, potentially also capturing an early decision to rely solely on informal care. This may also be because daycare subsidies in the Netherlands are conditional on parental employment and income.

Table 2: Performance of the final XGBoost classifier on the validation and test sets.

Performance Metric	Validation Set	Test Set
True negatives	3658 (79.10%)	8631 (79.98%)
False positives	383 (8.28%)	797 (7.39%)
True positives	284 (6.14%)	641 (5.94%)
False negatives	300 (6.49%)	723 (6.70%)
True positive rate	0.486	0.470
True negative rate	0.905	0.915
PR-AUC	0.457	0.453
*F*1	0.454	0.457
Brier score	0.088	0.088
Total observations	4,625	10,792
% in positive class	12.63	12.64

Notes: All predictors included. The *F*1 score is calculated with a tuned decision threshold of 0.266.

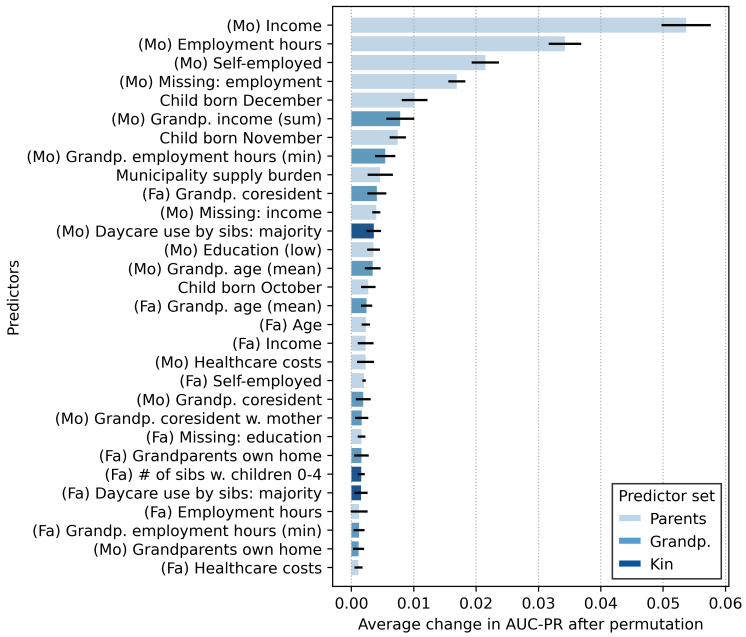

Figure 2: Top 30 predictors in terms of permutation feature importance. The error bar represents the standard deviation. Predictor names indicate in parentheses whether the predictor concerns the father (Fa) or mother (Mo).

These results are in line with the earlier finding that parental characteristics account for the largest share of predictive performance.

The sizably lower relative importance of fathers' income and employment reflects the starkly gendered nature of labor markets and childcare arrangements. Other important predictors that rank higher for mothers than fathers include healthcare costs and educational attainment, as well as grandparental employment hours,

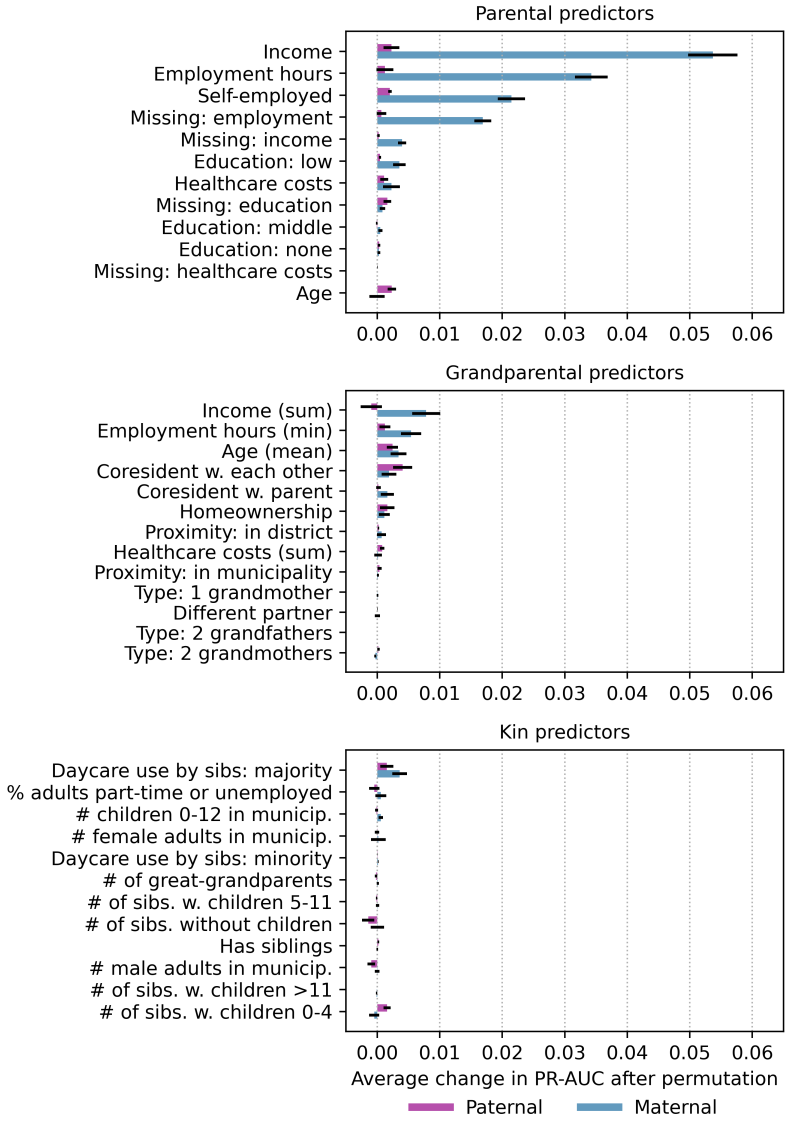

Figure 3: Permutation feature importance disaggregated by parental gender. Only predictors with a paternal and maternal version are plotted. The error bar represents the standard deviation. Predictors that achieve negative values decrease predictive performance (e.g., confounders).

age, and income. Figure 3 shows these gendered differences more clearly by plotting the importances of all predictors with paternal and maternal measures side by side. For most predictors, the maternal version achieves higher importances. The few relevant exceptions concern fathers' missing educational attainment, whether paternal grandparents co-reside, and whether fathers' siblings have children 0–4 years old. These results may partly be specific to the Netherlands, where roughly 60 percent of women work part-time (OECD 2019).

Figures 2 and 3 also shed light on the relevance of grandparents and kin. Grandparental resources (income, homeownership) and availability (employment hours,

co-residence, and age) are clearly important, especially on the maternal side. Only three of the top thirty predictors concern other family members: the use of daycare services by siblings of either parent and the presence of children 0–4 years old on the paternal side. The former can signal higher grandparental availability, but together with the latter, also sibling effects through information sharing, role modeling, or support with taking children to daycare. The remaining predictors related to other kin achieve low importance. These estimates are likely conservative compared to those in more familistic countries (Jappens and van Bavel 2012).

This analysis also allows us to inductively identify important predictors that have not received attention in previous work. The first is having self-employed parents, especially mothers. Self-employed parents may have more flexible or irregular schedules and income, creating mismatches between their schedules and daycare schedules (Bihan and Martin 2004). In the Netherlands, irregular income makes it difficult for parents to forecast the amount of childcare subsidy they need, which may also deter them from using daycare (Emery 2026). The second predictor that deserves more attention is parental healthcare costs, which suggests that some parents may need to both provide and receive care, creating complex care arrangements. We have already discussed the third important but novel predictor: daycare use by siblings. Lastly, the fact that the child's month of birth achieves high importance suggests timing effects in formal childcare uptake that could also be studied in detail (Reichert et al. 2026).

Improvements in Predictions across Parental Socioeconomic Background

The results above show that the incorporation of family members beyond grandparents provides small yet statistically significant improvements in predictive performance. We have discussed that these improvements can be small but theoretically relevant, for instance, because they point us to understudied family processes in childcare choices, such as sibling effects. These small improvements can also matter for research on childcare choices if they help account for well-documented social inequalities, that is, if they improve predictions especially for more disadvantaged parents (c.f., Garcia-Bernardo et al. 2025). We assess this by evaluating whether switching from an XGBoost model with only parental and grandparental characteristics to an XGBoost model with also kin characteristics results in more correct predictions for parents of different socioeconomic backgrounds.

Figure 4 plots the standardized odds ratios for improvements in predictions according to a logistic regression model. The improvements concern 293 of the 10,792 couples in the test set (2.72 percent). In a similar study using Dutch register data to predict educational attainment, Garcia-Bernardo et al. (2025) report an even smaller number of cases (247 out of 37,603, or 0.66 percent) for which predictions improved when using more multidimensional measurement and modeling. In our case, predictions are more likely to improve for couples in which the father is less educated and self-employed, and the mother is not highly educated. Predictions are also more likely to improve the fewer hours the mother works and the younger she is. Fathers' age shows a similar effect but does not reach statistical significance.

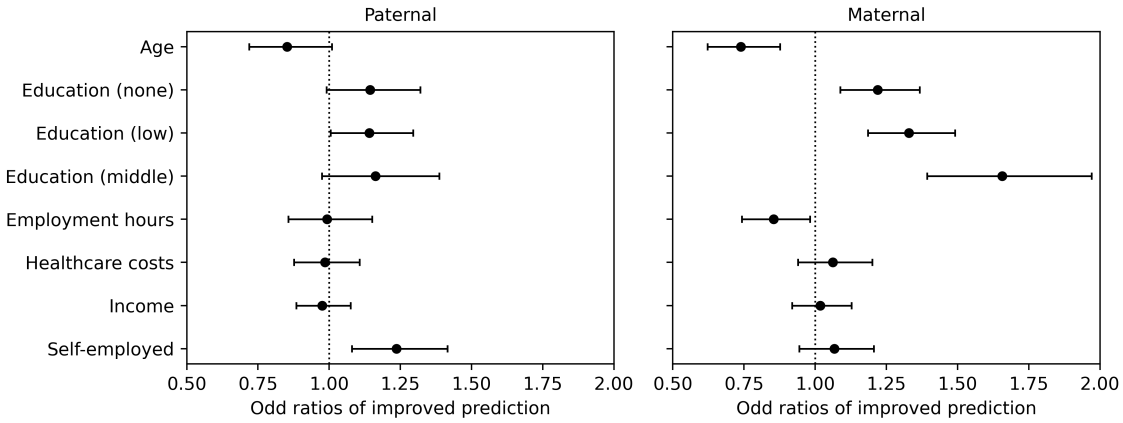

Figure 4: Odd ratios of predictions improving when switching from an XGBoost model with only parental and grandparental predictors to an XGBoost model also including kin predictors. Bars represent 95 percent confidence intervals. The logistic regression model is fit on the test set (10,792 observations). Predictors are z-standardized, and income and healthcare costs are logarithmic transformations. The reference category for educational attainment is highly educated. See Appendix 4 in the online supplement for the full regression output.

In other words, considering family members beyond grandparents helps predict the childcare choices of parents with less human capital, who are at the start of their careers and are less attached to the labor market. These effects are statistically significant but concern a small subgroup of parents, underscoring that incorporating kin may help account for childcare choices and social inequalities therein, but only in an incremental and highly contextualized manner.

Conclusion

We have argued that previous work on the relationship between childcare choices and family care availability relies on measurement and modeling that overlooks its full complexity.

Childcare choices are typically modeled as an additive function of a small set of parental, household, and grandparental characteristics. However, we have shown that our ability to predict childcare choices increases both when we measure family care availability across more family members and model its relationship to the outcome with more flexible models that incorporate non-linear multiplicative effects. Two predictors concerning family members beyond grandparents seem particularly relevant: the presence of young cousins (0–4 years old) on the paternal side and the use of daycare by both paternal and maternal aunts and uncles. Considering such family members improves predictions among a small subset of couples that includes mothers who are younger, less educated, and work fewer hours, as well as fathers who are low educated and self-employed, suggesting that incorporating complexity may help us better understand long-standing socioeconomic inequalities in childcare choices (Attanasio et al. 2022; Emery, 2026).

At the same time, our predictions are driven primarily by the employment and financial status of mothers, as well as the sociodemographics of maternal grandparents, followed by the equivalent predictors for fathers and paternal grandparents. Overall, the improvements in predictive performance we document are statistically significant yet small in size. Previous work has thus rightly focused on parental and grandparental characteristics, but it should also consider other family members and more nuanced modeling to better understand the contextual and stratified nature of childcare choices. How relevant this additional insight is will vary across research questions and settings.

We identify three main avenues for future research. First, researchers of childcare choices could study family care availability in a more multidimensional manner as a form of social capital using network analytical methods (e.g., Aeby et al. 2014; Soler et al. forthcoming). Second, future work could examine the predictors that we have inductively identified. Parental self-employment (Emery 2026) and the timing of daycare uptake (Reichert et al. 2026) can be studied together to extend the literature on atypical working hours and unstable income (e.g., Bihan and Martin 2004). Research on educational inequalities in childcare choices could explore sibling effects (c.f., Biegel and Maes 2022). In other words, our results can be used to refine existing theories of parenting and childcare choices, integrating data-driven discovery with theory-driven inference (Hofman et al. 2021; Verhagen 2022). Here, the nuances potentially captured by machine learning models must be balanced against the need for abstraction when articulating theory (Healy 2017), given that their added predictive value has proven relatively small in our application. Similarly, research on the predictability of life courses would benefit from integrating existing network theories and novel network-based models (e.g., combining Garcia-Bernardo et al. 2025 and Widmer 2016).

The limitations of our study may also inform future work. First, we encounter issues with measurement. Measuring employment hours one year before birth may introduce endogeneity ("leakage", in machine learning parlance), given that (grand)parents may already have adjusted their careers in anticipation of parenthood. There is, however, no straightforward solution, as anticipation could happen even earlier. Moreover, our outcome measure does not distinguish between non-enrollment in daycare due to choice and lack of access. This matters for our discussion of inequalities: if some parents do not enroll out of preference, there are fewer cases for which improvements in predictions address a meaningful inequality.

Second, predictor importance in the permutation feature analysis may have been over- or underestimated for predictors that are relatively highly correlated (e.g., the age of parents and grandparents; Molnar 2025). Importance scores also do not indicate the direction of the effects of predictors, which hinders the interpretation of effects not previously studied (e.g., poor health status may both incentivize and preclude reliance on formal care). The results of our inductive analysis should thus be taken as exploratory evidence. Third, we focus on mixed-gender couples and do not have data about the quality of family relationships, so future work should incorporate diversity in relationship configurations (Widmer 2016), potentially using linked survey data (Soler et al. forthcoming). Lastly, it is possible that our results are affected by COVID-19 effects, although the Netherlands imposed fewer

restrictions on movement and in-person care than other countries, and reimbursed unused daycare hours. Future work could study how our results compare to countries with similar data availability but different family structures, norms, and policies.

References

Aassve, Arnstein, Elena Meroni, and Chiara Pronzato. 2012. "Grandparenting and Childbearing in the Extended Family." *European Journal of Population* 28(4):499–518. https://doi.org/10.1007/s10680-012-9273-2

Aeby, Gaëlle, Eric D. Widmer, and Ivan De Carlo. 2014. "Bonding and Bridging Social Capital in Step- and First-Time Families and the Issue of Family Boundaries." *Interpersona: An International Journal on Personal Relationships* 8(1):51–69. https://doi.org/10.5964/ijpr.v8i1.149

Akiba, Takuya, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. "Optuna: A Next-Generation Hyperparameter Optimization Framework." Pp. 2623–31 in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Anchorage, AK: ACM. https://doi.org/10.1145/3292500.3330701

Arpino, Bruno, Marco Le Moglie, and Letizia Mencarini. 2022. "What Tears Couples Apart: A Machine Learning Analysis of Union Dissolution in Germany." *Demography* 59(1):161–86. https://doi.org/10.1215/00703370-9648346

Attanasio, Orazio, Sarah Cattan, and Costas Meghir. 2022. "Early Childhood Development, Human Capital, and Poverty." *Annual Review of Economics* 14(1):853–92. https://doi.org/10.1146/annurev-economics-092821-053234

Biegel, Naomi and Julie Maes. 2022. "The Educational Gradient in Formal Childcare Use – the Role of Employment Opportunities and (in)Formal Childcare Availability." *Community, Work & Family* 27(4):1–21. https://doi.org/10.1080/13668803.2022.2149387

Biegel, Naomi, Karel Neels, and Layla Van den Berg. 2021. "Uptake of Childcare Arrangements-Grandparental Availability and Availability of Formal Childcare." *Social Sciences* 10(2):50. https://doi.org/10.3390/socsci10020050

Bihan, Blanche Le and Claude Martin. 2004. "Atypical Working Hours: Consequences for Childcare Arrangements." *Social Policy & Administration* 38(6):565–90. https://doi.org/10.1111/j.1467-9515.2004.00408.x

Bojarczuk, Sara and Peter Mühlau. 2018. "Mobilising Social Network Support for Childcare: The Case of Polish Migrant Mothers in Dublin." *Social Networks* 53:101–10. https://doi.org/10.1016/j.socnet.2017.04.004

Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author)." *Statistical Science* 16(3):199–231. https://doi.org/10.1214/ss/1009213726

Bünning, Mareike. 2017. "The Association between Social Support Networks and Maternal Employment: A Comparison of Western German, Eastern German, and Migrant Mothers of Preschool-Aged Children." *Community, Work & Family* 20(3):273–91. https://doi.org/10.1080/13668803.2016.1256270

Burchinal, Margaret, Lauren Nelson, Mary Carlson, and Jeanne Brooks-Gunn. 2008. "Neighborhood Characteristics, and Child Care Type and Quality." *Early Education and Development* 19(5):702–25. https://doi.org/10.1080/10409280802375273

Carlin, Caroline, Elizabeth E. Davis, Caroline Krafft, and Kathryn Tout. 2019. "Parental Preferences and Patterns of Child Care Use among Low-Income Families: A Bayesian Analysis." *Children and Youth Services Review* 99:172–85. https://doi.org/10.1016/j.childyouth.2019.02.006

Chen, Tianqi and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." Pp. 785–94 in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Compton, Janice and Robert A. Pollak. 2014. "Family Proximity, Childcare, and Women's Labor Force Attachment." *Journal of Urban Economics* 79:72–90. https://doi.org/10.1016/j.jue.2013.03.007

Crosnoe, Robert, Kate Chambers Prickett, Chelsea Smith, and Shannon Cavanagh. 2014. "Changes in Young Children's Family Structures and Child Care Arrangements." *Demography* 51(2):459–83. https://doi.org/10.1007/s13524-013-0258-5

Daw, Jonathan, Ashton M. Verdery, and Rachel Margolis. 2016. "Kin Count(s): Educational and Racial Differences in Extended Kinship in the United States." *Population and Development Review* 42(3):491–517. https://doi.org/10.1111/j.1728-4457.2016.00150.x

Dykstra, Pearl A., Matthijs Kalmijn, Trudie C. M. Knijn, Aafke E. Komter, Aart C. Liefbroer, and Clara H. Mulder. 2006. *Family Solidarity in the Netherlands*. Amsterdam: Dutch University Press.

Emery, Tom. 2026. "Inequalities in Early Childcare Strategies: Evidence from Dutch Administrative Data." *Advances in Life Course Research* 67:1–12. https://doi.org/10.1016/j.alcr.2026.100727

Eurostat. 2025. "Estimated Average Age of Young Persons Leaving the Parental Household." https://doi.org/10.2908/YTH_DEMO_030

Fisher, Aaron, Cynthia Rudin, and Francesca Dominici. 2019. "All Models Are Wrong, but Many Are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously." *Journal of Machine Learning Research* 20:1–88.

Furstenberg, Frank F. 2020. "Kinship Reconsidered: Research on a Neglected Topic." *Journal of Marriage and Family* 82(1):364–82. https://doi.org/10.1111/jomf.12628

Garcia-Bernardo, Javier, Weverthon Barbosa Machado, Eva Jaspers, Samuel Plach, and Erik Jan van Leeuwen. 2025. "Prediction Gaps as Pathways to Explanation: Rethinking Educational Outcomes through Differences in Model Performance." *SocArXiV*. https://doi.org/10.48550/arXiv.2506.22993

Gneiting, Tilmann and Adrian E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102(477):359–78. https://doi.org/10.1198/016214506000001437

He, Haibo and Yunqian Ma, eds. 2013. *Imbalanced Learning: Foundations, Algorithms, and Applications*. Piscataway, NJ: Wiley-IEEE Press.

Healy, Kieran. 2017. "Fuck Nuance." *Sociological Theory* 35(2):118–27. https://doi.org/10.1177/0735275117709046

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103(6):2052–86. https://doi.org/10.1257/aer.103.6.2052

Hofman, Jake M., Duncan J. Watts, Susan Athey, Filiz Garip, Thomas L. Griffiths, Jon Kleinberg, Helen Margetts, Sendhil Mullainathan, Matthew J. Salganik, Simine Vazire, Alessandro Vespignani, and Tal Yarkoni. 2021. "Integrating Explanation and Prediction

in Computational Social Science." *Nature* 595(7866):181–88. https://doi.org/10.1038/s41586-021-03659-0

Hünteler, Bettina, Karsten Hank, Diego Alburez-Gutierrez, and Thomas Leopold. 2026. "A Typology of Younger Adults' Nuclear and Extended Family Relations in the United States." *Journal of Marriage and Family* 88(1):284–98. https://doi.org/10.1111/jomf.70014

Jappens, Maaike and Jan van Bavel. 2012. "Regional Family Cultures and Child Care by Grandparents in Europe." *Demographic Research* 27:85–120. https://doi.org/10.4054/DemRes.2012.27.4

Jarvis, Helen. 1999. "The Tangled Webs We Weave: Household Strategies to Co-Ordinate Home and Work." *Work, Employment and Society* 13(2):225–47. https://doi.org/10.1177/09500179922117926

Kapoor, Sayash, Emily M. Cantrell, Kenny Peng, Thanh Hien Pham, Christopher A. Bail, Odd Erik Gundersen, Jake M. Hofman, Jessica Hullman, Michael A. Lones, Momin M. Malik, Priyanka Nanayakkara, Russell A. Poldrack, Inioluwa Deborah Raji, Michael Roberts, Matthew J. Salganik, Marta Serra-Garcia, Brandon M. Stewart, Gilles Vandewiele, and Arvind Narayanan. 2024. "REFORMS: Consensus-Based Recommendations for Machine-Learning-Based Science." *Science Advances* 10(18):1–17. https://doi.org/10.1126/sciadv.adk3452

Kiraly, Meredith, Cathy Humphreys, and Margaret Kertesz. 2021. "Unrecognized: Kinship Care by Young Aunts, Siblings and Other Young People." *Child & Family Social Work* 26(3):338–47. https://doi.org/10.1111/cfs.12814

Kuhltau, Karen and Karen Oppenheim Mason. 1996. "Market Child Care Versus Care by Relatives: Choices Made by Employed and Nonemployed Mothers." *Journal of Family Issues* 17(4):561–78. https://doi.org/10.1177/019251396017004007

Markowitz, Anna J., Rebecca M. Ryan, and Anna D. Johnson. 2014. "Child Care Subsidies and Child Care Choices: The Moderating Role of Household Structure." *Children and Youth Services Review* 36:230–40. https://doi.org/10.1016/j.childyouth.2013.11.009

Meyers, Marcia K. and Lucy P. Jordan. 2006. "Choice and Accommodation in Parental Child Care Decisions." *Community Development* 37(2):53–70. https://doi.org/10.1080/15575330609490207

Molina, Mario and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45(1):27–45. https://doi.org/10.1146/annurev-soc-073117-041106

Molnar, Christoph. 2025. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 3rd ed. Sebastopol, CA.

Morrissey, Taryn W. 2008. "Familial Factors Associated With the Use of Multiple Child-Care Arrangements." *Journal of Marriage and Family* 70(2):549–63. https://doi.org/10.1111/j.1741-3737.2008.00500.x

Mulder, Clara H. and Marieke J. van der Meer. 2009. "Geographical Distances and Support from Family Members." *Population, Space and Place* 15(4):381–99. https://doi.org/10.1002/psp.557

Müller, Andreas C. and Sarah Guido. 2016. *Introduction to Machine Learning with Python*. Sebastopol, CA, USA: O'Reilly.

OECD. 2019. *Part-Time and Partly Equal: Gender and Work in the Netherlands*. Paris: OECD Publishing. https://doi.org/10.1787/204235cf-en

Rainio, Oona, Jarmo Teuho, and Riku Klén. 2024. "Evaluation Metrics and Statistical Tests for Machine Learning." *Scientific Reports* 14(1):6086. https://doi.org/10.1038/s41598-024-56706-x

Reichert, Max, Alzbeta Bartova, and Tom Emery. 2026. "Social Inequalities in the Timing of Childcare Use in Europe." *SocArXiV*. https://doi.org/10.31235/osf.io/ju75w_v1

Salganik, Matthew J., Ian Lundberg, Alexander T. Kindel, Caitlin E. Ahearn, Khaled Al-Ghoneim, Abdullah Almaatouq, Drew M. Altschul, Jennie E. Brand, Nicole Bohme Carnegie, Ryan James Compton, Debanjan Datta, Thomas Davidson, Anna Filippova, Connor Gilroy, Brian J. Goode, Eaman Jahani, Ridhi Kashyap, Antje Kirchner, Stephen McKay, Allison C. Morgan, Alex Pentland, Kivan Polimis, Louis Raes, Daniel E. Rigobon, Claudia V. Roberts, Diana M. Stanescu, Yoshihiko Suhara, Adaner Usmani, Erik H. Wang, Muna Adem, Abdulla Alhajri, Bedoor AlShebli, Redwane Amin, Ryan B. Amos, Lisa P. Argyle, Livia Baer-Bositis, Moritz Büchi, Bo-Ryehn Chung, William Eggert, Gregory Faletto, Zhilin Fan, Jeremy Freese, Tejomay Gadgil, Josh Gagné, Yue Gao, Andrew Halpern-Manners, Sonia P. Hashim, Sonia Hausen, Guanhua He, Kimberly Higuera, Bernie Hogan, Ilana M. Horwitz, Lisa M. Hummel, Naman Jain, Kun Jin, David Jurgens, Patrick Kaminski, Areg Karapetyan, E. H. Kim, Ben Leizman, Naijia Liu, Malte Möser, Andrew E. Mack, Mayank Mahajan, Noah Mandell, Helge Marahrens, Diana Mercado-Garcia, Viola Mocz, Katariina Mueller-Gastell, Ahmed Musse, Qiankun Niu, William Nowak, Hamidreza Omidvar, Andrew Or, Karen Ouyang, Katy M. Pinto, Ethan Porter, Kristin E. Porter, Crystal Qian, Tamkinat Rauf, Anahit Sargsyan, Thomas Schaffner, Landon Schnabel, Bryan Schonfeld, Ben Sender, Jonathan D. Tang, Emma Tsurkov, Austin van Loon, Onur Varol, Xiafei Wang, Zhi Wang, Julia Wang, Flora Wang, Samantha Weissman, Kirstie Whitaker, Maria K. Wolters, Wei Lee Woon, James Wu, Catherine Wu, Kengran Yang, Jingwen Yin, Bingyu Zhao, Chenyun Zhu, Jeanne Brooks-Gunn, Barbara E. Engelhardt, Moritz Hardt, Dean Knox, Karen Levy, Arvind Narayanan, Brandon M. Stewart, Duncan J. Watts, and Sara McLanahan. 2020. "Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration." *Proceedings of the National Academy of Sciences of the United States of America* 117(15):8398–403. https://doi.org/10.1073/pnas.1915006117

Saraceno, Chiara. 2011. "Childcare Needs and Childcare Policies: A Multidimensional Issue." *Current Sociology* 59(1):78–96. https://doi.org/10.1177/0011392110385971

Shmueli, Galit. 2010. "To Explain or to Predict?" *Statistical Science* 25(3):289–310. https://doi.org/10.1214/10-STS330

Soler, Nicolás, Tom Emery, and Agnieszka Kanas. Forthcoming. "Families, Households, Networks: Rethinking the Relational Structure of Families through Large-Scale Network Data." *Demography*. https://doi.org/10.31235/osf.io/wxqyt

Statistics Netherlands. 2025a. "Households; Size, Composition, Position in the Household, 1 January." https://www.cbs.nl/en-gb/figures/detail/82905ENG?q=households

Statistics Netherlands. 2025b. "How Old is the Average First-Time Mother in the Netherlands?" https://longreads.cbs.nl/the-netherlands-in-numbers-2025/how-old-is-the-average-first-time-mother-in-the-netherlands

Statistics Netherlands. 2026. "Microdata: Conducting Your Own Research." https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research

Sun, Xiaoran. 2024. "Supervised Machine Learning for Exploratory Analysis in Family Research." *Journal of Marriage and Family* 86(5):1468–94. https://doi.org/10.1111/jomf.12973

Uttal, Lynet. 1999. "Using Kin for Child Care: Embedment in the Socioeconomic Networks of Extended Families." *Journal of Marriage and Family* 61(4):845–57. https://doi.org/10.2307/354007

Van Ness, Mike, Tomas M. Bosschieter, Roberto Halpin-Gregorio, and Madeleine Udell. 2023. "The Missing Indicator Method: From Low to High Dimensions." Pp. 5004–15 in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '23*. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3580305.3599911

van der Laan, Jan, Edwin de Jonge, Marjolijn Das, Saskia Te Riele, and Tom Emery. 2023. "A Whole Population Network and Its Application for the Social Sciences." *European Sociological Review* 39(1):145–60. https://doi.org/10.1093/esr/jcac026

Verhagen, Mark D. 2022. "A Pragmatist's Guide to Using Prediction in the Social Sciences." *Socius* 8:1–17. https://doi.org/10.1177/23780231221081702

Wainer, Jacques and Gavin Cawley. 2021. "Nested Cross-Validation When Selecting Classifiers Is Overzealous for Most Practical Applications." *Expert Systems with Applications* 182:115222. https://doi.org/10.1016/j.eswa.2021.115222

Widmer, Eric D. 2016. *Family Configurations: A Structural Approach to Family Diversity*. London: Routledge.

Xu, Hongwei. 2019. "Physical and Mental Health of Chinese Grandparents Caring for Grandchildren and Great-Grandparents." *Social Science & Medicine* 229:106–16. https://doi.org/10.1016/j.socscimed.2018.05.047

Zick, Cathleen D., Lori Kowaleski-Jones, and Benjamin Greenwalt. 2022. "All in the Family? The Community Context of Childcare Options and Parents' Childcare Choices." *Community, Work & Family* 25(2):210–31. https://doi.org/10.1080/13668803.2020.1719977

Acknowledgments: The authors thank Flora Zhou, Javier Garcia-Bernardo, Malte Lüken, Maximilian Reichert, and participants of the ODISSEI Summer Incubator 2025 for helpful discussions during the conceptualization of this article. The authors also thank Editor Stephen Vaisey and an anonymous reviewer for their constructive feedback. This work was financed by the European Research Council under grant agreement 101042662—Complex Contagions of Childcare Strategies.

Nicolás Soler: Department of Public Administration and Sociology, Erasmus University Rotterdam. E-mail: soleralvarezmiranda@essb.eur.nl.
Tom Emery: Department of Public Administration and Sociology, Erasmus University Rotterdam. E-mail: tom@odissei-data.nl.
Agnieszka Kanas: Department of Public Administration and Sociology, Erasmus University Rotterdam. E-mail: kanas@essb.eur.nl.