sociological science

Supplement to:

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

Online supplement for research article 'Family networks and childcare choices: A predictive machine learning approach'

Nicolás Soler[1*], Tom Emery[1], Agnieszka Kanas[1]

[1]Department of Public Administration and Sociology
Erasmus University Rotterdam

[*]soleralvarezmiranda@essb.eur.nl
ORCID: 0009-0001-4239-9326

Published in *Sociological Science*
Spring 2026

Appendix 1: Register data files from Statistics Netherlands

File name	File year (version)	Relevant information
GBAPERSOONTAB	2023 (v.1)	• List of all individuals in the registers • Gender • Date of birth • Country of origin
GBAOVERLIJDENTAB	2023 (v.1)	• Date of death
GBAADRESOBJECTBUS	2023 (v.1)	• Address of residence
VSLGWBTAB	2023 (v.1)	• District and municipality of addresses
SPOLISBUS	2019 (v.7)	• Employment hours
ZELFSTANDIGENMNDBEDRAGBUS	2023 (v.1)	• Self-employed status
HDIPLOMAREGTAB	2019 (v.4)	• Educational attainment
OPLEIDINGSNRREF	n.a. (v.34)	• SOI2016 codes of educational attainment
INPATAB	2020 (v.3)	• Income
VEHTAB	2020 (v.3)	• Homeownership
ZVWZORGKOSTENTAB	2019 (v.2)	• Health expenditures
FAMILIENETWERKTAB	2020 (v.1), 2022 (v.1)	• Family relationships
RTKINDERENTAB	2016-2019 (v.1), 2021-2023 (v.3)	• Childcare benefit receipt
KINDEROPVANGLOCATIES	2020 (n.a.)	• Address of childcare providers

Table A1.1: Register data files used in the study. For more information see Statistics Netherlands (2026).

Appendix 2: Predictor sets and descriptive statistics

Predictor set	Predictor	Operationalization
Parents	Age*	Years (continuous)
(*: predictors with one version for the father and one for the mother)	Education*	None / Low / Middle / High
	Missing: education*	No / Yes
	Income*	Euros (cont., logarithmic, annual)
	Missing: income*	No / Yes
	Employment hours*	Contractual hours (cont., annual)
	Missing: employment hours*	No / Yes
	Self-employed*	No / Yes
	Healthcare costs*	Euros (cont., log., annual)
	Missing: healthcare costs*	No / Yes
	Parents own home	No / Yes (at least one does)
	Child's month of birth	Months (cont.)
	Child's gender	Male / Female
	Municipality supply count	Number of childcare providers in municipality (cont.)
	Municipality supply burden	Number of children 0-4 y-o in municipality / Municipality supply count (cont.)
	District supply count	Number of childcare providers in district (cont.)
	District supply burden	Number of children 0-4 y-o in district / District supply count (cont.)
Grandparents	Type of grandparents	One grandfather / One grandmother / Two grandmothers / One grandfather and one grandmother
(All predictors have one version for paternal grandparents and one for maternal grandparents)	Age (mean)	Years (cont.)
	Employment hours (minimum)	Contractual hours (cont., annual)
	Income (sum)	Euros (cont., log., annual)
	Healthcare costs (sum)	Euros (cont., log., annual)

	Homeownership	No / Yes (at least one does)
	Different partner	No / Yes (at least one grandparent has a different partner)
	Coresident with each other	No / Yes
	Coresident with parent	No / Yes (at least one is)
	Proximity: in municipality	No / Yes (at least one grandparent resides in the same municipality as the child)
	Proximity: in district	No / Yes (at least one grandparent resides in the same district as the child)
Kin	# of great-grandparents	Count
(All predictors have one version for paternal kin and one for maternal kin)	# of (parents') siblings without children	Count
	# of (parents') siblings with children 0-4 y-o	Count
	# of (parents') siblings with children 5-11 y-o	Count
	# of (parents') siblings with children >12 y-o	Count
	The parent has siblings	No / Yes
	# male adults in municipality	Count
	# female adults in municipality	Count
	# children 0-12 in municipality	Count
	% of (parents') siblings that work part-time or are unemployed	Percentage (part-time: <1700 annual employment hours)
	Daycare use by (parents') siblings	None / Minority (<50% of siblings) / Majority (>=50% of siblings)

Table A2.1: Predictors in each predictor set. Parental predictors with an asterisk (*) are calculated separately for the father and the mother. All grandparental and kin (great-grandparents, aunts, uncles, cousins) predictors are calculated separately for the father and the mother.

Predictor set	Side	Predictor	Structural missings	Mean	Std	10%	50%	90%
Parents	Fa	Age	0	31.02	4.51	26	31	37
	Mo	Age	0	28.87	4.00	24	29	34
	Fa	Missing: Education	0	0.90	.	.	.	.
	Mo	Missing: Education	0	0.04	.	.	.	.
	Fa	Income (log.)	57	4.49	0.58	4.26	4.57	4.82
	Mo	Income (log.)	17	4.39	0.55	4.14	4.48	4.71
	Fa	Missing: Income	0	0.01	.	.	.	.
	Mo	Missing: Income	0	0.01	.	.	.	.
	Fa	Employment hours	6,108	1673.34	762.02	0	1980	2183
	Mo	Employment hours	3,280	1521.55	604.35	447	1683	2076
	Fa	Missing: Employment hours	0	0.12	.	.	.	.
	Mo	Missing: Employment hours	0	0.06	.	.	.	.
	Fa	Self-employed	0	0.14	.	.	.	.
	Mo	Self-employed	0	0.07	.	.	.	.
	Fa	Healthcare costs (log.)	982	2.38	0.64	1.93	2.24	3.22
	Mo	Healthcare costs (log.)	189	2.68	0.62	2.01	2.57	3.57
	Fa	Missing: Healthcare costs	0	0.02	.	.	.	.
	Mo	Missing: Healthcare costs	0	0	.	.	.	.
	Both	Homeownership	0	0.65	.	.	.	.
	Both	Child is female	0	0.49	.	.	.	.
	Both	Municipality supply count	4	253.42	260.62	45	148	673
	Both	Municipality supply burden	4	1.68	0.34	1.23	1.69	2.09
	Both	District supply count	390	27.07	27.41	5	18	58
	Both	District supply burden	391	1.98	2.02	1.00	1.70	2.90
Grandparents	Fa	Age	0	61.12	6.06	53.50	61.00	68.50
	Mo	Age	0	59.43	5.88	52.00	59.50	67.00
	Fa	Employment hours	11,451	920.86	781.61	0	957.00	1981.40
	Mo	Employment hours	14,753	1017.06	757.33	0	1081.65	2001.00

	Fa	Income (log.)	0	4.71	0.46	4.36	4.76	5.06
	Mo	Income (log.)	0	4.73	0.47	4.39	4.78	5.08
	Fa	Healthcare costs (log.)	0	3.29	0.62	2.53	3.29	4.08
	Mo	Healthcare costs (log.)	0	3.26	0.61	2.50	3.25	4.05
	Fa	Homeownership	0	0.76	.	.	.	.
	Mo	Homeownership	0	0.77	.	.	.	.
	Fa	Different partner	43,021	0.16	.	.	.	.
	Mo	Different partner	42,606	0.17	.	.	.	.
	Fa	Co-reside with each other	6,829	0.68	.	.	.	.
	Mo	Co-reside with each other	5,959	0.68	.	.	.	.
	Fa	Co-reside with father	0	0.11	.	.	.	.
	Mo	Co-reside with mother	0	0.12	.	.	.	.
	Fa	Proximity: municipality	0	0.47	.	.	.	.
	Mo	Proximity: municipality	0	0.49	.	.	.	.
	Fa	Proximity: district	0	0.24	.	.	.	.
	Mo	Proximity: district	0	0.29	.	.	.	.
Kin	Fa	# of great-grandparents	23,961	0.90	1.04	0	1	2
	Mo	# of great-grandparents	19,998	1.09	1.11	0	1	3
	Fa	# of siblings without children	15,550	1.10	1.03	0	1	2
	Mo	# of siblings without children	11,878	1.23	1.03	0	1	2
	Fa	# of siblings with children 0-4 y-o	33,592	0.44	0.70	0	0	1
	Mo	# of siblings with children 0-4 y-o	36,731	0.36	0.65	0	0	1
	Fa	# of siblings with children 5-11 y-o	42,738	0.21	0.51	0	0	1
	Mo	# of siblings with children 5-11 y-o	45,675	0.13	0.43	0	0	1
	Fa	# of siblings with children >12 y-o	48,883	0.04	0.25	0	0	0
	Mo	# of siblings with children >12 y-o	49,918	0.03	0.24	0	0	0
	Fa	Has siblings	0	0.94	.	.	.	.
	Mo	Has siblings	0	0.94	.	.	.	.
	Fa	# male adults in municipality	28,426	0.70	0.97	0	0	2

Mo	# male adults in municipality	31,129	0.60	0.90	0	0	2
Fa	# female adults in municipality	27,758	0.75	1.02	0	0	2
Mo	# female adults in municipality	29,945	0.68	0.98	0	0	2
Fa	# children 0-12 in municipality	40,566	0.30	0.85	0	0	1
Mo	# children 0-12 in municipality	43,571	0.28	0.8	0	0	0
Fa	% of siblings that work part-time or are unemployed	3,045	52.17	36.49	0	50.00	100.00
Mo	% of siblings that work part-time or are unemployed	3,281	52.92	38.37	0	50.00	100.00

Table A2.2: Descriptive statistics for continuous and binary predictors. Total observations: 51,390.

Predictor set	Parent / Side	Predictor	Structural missings	Number of cases (%)
Parents	Father	Education	4,932	None = 4,932 (9.60) Low = 4,411(8.58) Middle = 22,525 (43.83) High = 19,522 (38.00)
	Mother	Education	1,827	None = 1,827 (3.56) Low = 3,109 (6.05) Middle = 21,003 (40.87) High = 25,451 (49.53)
	Both	Child's month of birth	0	January = 4,093 (7.97) February = 3,731 (7.26) March = 4,151 (8.08) April = 3,998 (7.78) May = 4,062 (7.90) June = 4,108 (7.99) July = 4,591 (8.93) August = 4,686 (9.12) September = 4,702 (9.15) October = 4,667 (9.08) November = 4,381 (8.53) December = 4,220 (8.21)
Grandparents	Father	Type of paternal grandparents	0	One grandmother = 4,873 (9.48) Two grandmothers = 13 (0.03) One grandfather = 1,956 (3.81) One of each = 44,548 (86.69)
	Mother	Type of maternal grandparents	0	One grandmother = 4,313 (8.39) Two grandmothers = 19 (0.04) One grandfather = 1,646 (3.20) One of each = 45,412 (88.37)
Kin		Daycare use by father's siblings	3,045	None = 34,178 (66.51) <50% of sibs. = 3,547 (6.90) >=50% of sibs. = 13,665 (26.59)
		Daycare use by mother's siblings	3,281	None = 37,956 (73.86) <50% of sibs. = 2,974 (5.79) >=50% of sibs. = 10,460 (20.35)
Total observations				51,390

Table A2.3: Descriptive statistics for categorical predictors. Counts are not disaggregated by the outcome to prevent disclosure risks in line with the privacy policy of Statistics Netherlands.

Appendix 3: Model hyperparameters

Hyperparameter	Possible values	
	Baseline model	Other models
solver	["saga"]	["saga"]
max_iter	{1500}	{1500}
class_weight	["balanced"]	["balanced"]
penalty	["None"]	["None", "L1", "L2", "elasticnet"]
C_param	N.A.	{0.0001, 1000.0}, sampled from range in logarithmic scale, only when penalty ≠ "None"
l1_ratio	N.A.	{0.01, 0.99}, only when penalty = "elasticnet"

Table A3.1: Hyperparameter space for logistic regression models. Hyperparameter values are sampled during tuning from the range of values defined in brackets. Square brackets define the possible values for hyperparameters that take one categorical value, while curly brackets define the possible values for hyperparameters that take one numerical value sampled from the range in brackets.

Hyperparameter	Possible values
booster	["gbtree"]
early_stopping_rounds	{30}
n_estimators	{100, 1100}, sampled from range in steps of 25
max_depth	{2, 15}
learning_rate	{0.01, 0.4}, sampled from range in logarithmic scale
max_delta_step	{1, 8}
tree_method	["approx", "hist"]
min_child_weight	{1, 15}
max_bin	{128, 512}, sampled from range in steps of 32
subsample	{0.7, 1.0}
colsample_bytree	{0.7, 1.0}
reg_alpha	{0.000001, 10.0}, sampled from range in logarithmic scale
reg_lambda	{0.000001, 10.0}, sampled from range in logarithmic scale
gamma	{0.001, 5.0}, sampled from range in logarithmic scale
scale_pos_weight	{0, 15}

Table A3.2: Hyperparameter space for XGBoost models. Hyperparameter values are sampled during tuning from the range of values defined in brackets. Square brackets define the possible values for hyperparameters that take one categorical value, while curly brackets define the possible values for hyperparameters that take one numerical value sampled from the range in brackets.

Hyperparameter	Final value
booster	“gbtree”
early_stopping_rounds	30
n_estimators	775
max_depth	4
learning_rate	0.06
max_delta_step	6
tree_method	“hist”
min_child_weight	13
max_bin	160
subsample	0.73
colsample_bytree	0.74
reg_alpha	0.40
reg_lambda	1.27
gamma	0.00
scale_pos_weight	1.11

Table A3.3: Hyperparameter values for the final best-performing XGBoost model.

Appendix 4: Logistic regression of improvements in predictions

	Coefficient	Standard error	Odd ratios
Constant	-4.01†	0.08	0.02
(Fa) Age	-0.16	0.09	0.85
(Fa) Education (none)	0.13	0.07	1.14
(Fa) Education (low)	0.13*	0.07	1.14
(Fa) Education (middle)	0.15	0.09	1.16
(Fa) Employment hours	-0.01	0.08	0.99
(Fa) Healthcare costs (log.)	-0.02	0.06	0.99
(Fa) Income (log.)	-0.03	0.05	0.98
(Fa) Self-employed	0.21†	0.07	1.24
(Mo) Age	-0.30†	0.09	0.74
(Mo) Education (none)	0.20†	0.06	1.22
(Mo) Education (low)	0.29†	0.06	1.33
(Mo) Education (middle)	0.51†	0.09	1.66
(Mo) Employment hours	-0.16*	0.07	0.85
(Mo) Healthcare costs (log.)	0.06	0.06	1.06
(Mo) Income (log.)	-0.03	0.05	1.02
(Mo) Self-employed	0.07	0.06	1.07
Observations		10,792	
Log-Likelihood		-1221.3†	
Pseudo R-squared		0.09	

Table A4.1: Logistic regression of improvements in predictions when switching from an XGBoost model with only parental and grandparental predictors to an XGBoost model with also kin predictors on parental characteristics. P-values: * < 0.05, † < 0.01.