Disputes over custody and residency at district courts of Finland

2004 – 2015

Anton Sobolev

University of Helsinki

Department of Mathematics and Statistics

Faculty of Science

Master's Thesis

February 2020

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATISK-NATURVETENSKAPLIGA FAKULTETEN
FACULTY OF SCIENCE

Tiedekunta – Fakultet – Faculty	Koulutusohjelma – Utbildningsprogram – Degree programme
Faculty of Science	Department of Mathematics and Statistics

Tekijä – Författare – Author

Anton Sobolev

Työn nimi – Arbetets titel – Title

Disputes over custody and residency at district courts of Finland 2004 - 2015

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's thesis	January 2020	63

Tiivistelmä – Referat – Abstract

When couples with children split or divorce, they are often unable to come to a mutual agreement concerning their child's place of residency, custody, the child's meetings with the other parent and the frequency of these meetings, or financial aid one parent is obliged to pay the other parent for the child. In many countries, these disagreements quite often lead to long disputes in court. A lot of research has been made (both in Finland and internationally) concerning the court's consideration of disputes about children. This thesis studies the disputes on custody and residency of a child in the district courts of Finland. The objective is to find out which factors play the biggest role in solving these disputes in court.

Nine district courts of Finland have kindly provided the documents of the disputes concerning custody and residency of children from the period of 2004 - 2015. Only the cases where a dispute was solely between the parents of a child (no other relatives) and where the final decision was made by court (no agreement between the parties) are taken into analysis. Disputes are divided into two types - the ones where residency of a child was involved in a dispute (residency disputes) and the ones where it was not involved (custody disputes). The winner of a dispute is a dependent variable. A logistic regression model is applied for the custody disputes, and a cumulative logistic regression model is applied for the residency disputes.

Due to results of the analysis, mothers win more disputes than fathers, but the difference is statistically significant only for the residency disputes. When only father is of a foreign background, it lowers father's winning chances in a custody dispute, but neither father's nor mother's foreign backgrounds are statistically significant for the residency disputes. A substantiated violence of father towards mother again acts negatively for fathers in custody disputes, and so does a non-substantiated accusation regarding alcohol or drug abuse by father. For the residency disputes, the main factors decreasing fathers' probability to win are mother hiring a legal assistant and father receiving legal aid (which takes place when father is not financially capable of hiring a legal assistant). Established conditions of a child at one of the parents increase the winning chances of that parent, but the effect is higher for fathers. All the accusations (both substantiated and non-substantiated in court) act in favor of fathers; these are substantiated mother's mental disorder, non-substantiated alcohol or drug abuse by mother and non-substantiated accusation regarding father's violence towards mother. At the same time, no variables regarding genders of children disputed about, genders of a judge or of legal assistants are statistically significant in the models. The same concerns the parents' demands in court, as well as the ages of parents (and their difference) and of children involved in disputes.

This investigation can be extended by adding the disputes from other years and from other district courts into the analysis.

Avainsanat – Nyckelord – Keywords

custody disputes, residency disputes, district court, cumulative logistic regression

Säilytyspaikka – Förvaringställe – Where deposited

Kumpula science library

Muita tietoja – Övriga uppgifter – Additional information

ACKNOWLEDMENT

I would first like to thank my thesis advisor, university lecturer Pekka Pere, for replying to my numerous calls and emails, whenever I had a question regarding my research and modelling process. He gave me valuable insights concerning both possible improvements of my research methods and the process of writing a research paper in general.

I would also like to thank the district courts of Helsinki, Espoo, Vantaa, Tuusula, Porvoo, Lahti, Hyvinkää, Tammisaari for their collaboration in providing us with research permits granting the right to investigate their past disputes. The courts also kindly provided me with their premises for studying the past disputes.

Finally, I would like to thank Eero Lilja and Aku Leivonen, whom I consulted on certain technical aspects of creating the models employed in this thesis.

Contents

1. INTRODUCTION

Not every family is stable, and quite often married couples or couples living together decide to divorce or split. In case they have a child or several children, this process may become more complicated. While some parents manage to come to mutual agreements, others find themselves in such a conflict that an official dispute becomes unavoidable, and a decision by the district court is needed. Depending on the nature of a dispute, this may take months or years to solve it, and both parents might spend a lot of time and money on making applications to the court and finding the lawyers.

The topic of parents' disputes over children has long been a wide field of research in many countries. Multiple research has been conducted regarding the custody and residency disputes and their consideration in the courts of appeal in Finland. The related researches and publications are among others by Valkama, Litmala (2006), Hautanen (2010), Ervasti (2014), Palo-Repo (2015), Pere, Lilja, Sobolev (2017) and Lilja (2018). The topic of this thesis is close to the topic in the publication by Lilja (2018), but a wider time range and the data from more district courts are taken into analysis, and another statistical method is employed (cumulative logistic regression).

Some of the key questions in disputes over children are: where will the residency of a child be and will the other parent still have a custody? As long as these are agreed upon by parents or a decision is given by court, the other possible sources of disputes are the right of a child to meet the other parent (with whom the child does not live permanently), as well as a parent's obligation to make child support payment to the other parent. The last two issues are, however, out of the scope of this thesis, as the objective is to concentrate solely on the disputes over custody and residency of a child. On what variables, factors and aspects does a decision mainly depend, and how strong is each variable's influence on (or association with) the final decision? In particular, is gender important or are mothers and fathers treated equally in court? Could the gender of a child be decisive? What about the gender of a judge? Or could a foreign background of one parent matter? These are, among others, the issues to be investigated in this thesis.

For the analyses, cases featuring custody and residency disputes during the period of 2004 – 2015 were collected from the district courts of the following cities: Helsinki, Espoo, Vantaa, Tuusula, Hyvinkää, Tammisaari, Lahti, Porvoo and Hämeenlinna. The choice of the courts is based on the short distance to Helsinki, since the data was delivered in a paper format and could

not be taken out of the premises of any district court. The cases considered in the courts are generally classified according to the topic of dispute, therefore the cases with the certain codes were collected, namely: 2510 (decision on a child's custody and a right to meet the other parent), 2520 (changing the decision regarding a child's custody), 2610 (divorce) and 2621 (divorce with the attached application on a child's custody or the right to meet the other parent). The observations collected from each court were studied both separately and in combination. The aim was to explain the probability to win a dispute depending on certain factors. For this purpose, a cumulative logistic regression model was employed.

The logistic regression and the cumulative logistic regression models are explained first in Chapter 2. Chapter 3 includes an overview of how disputes regarding children's custody and residence are processed in the courts. The processes of collecting the data and the key observations are outlined in Chapters 4 and 5, respectively. In Chapter 6, the cumulative logistic regression is applied to model the winning probabilities of mothers and of fathers based on certain factors. This is followed by the diagnostics of the model. The conclusions and evaluations are given in Chapter 7. Descriptions of variables, tables and outputs, as well as additional related theory are in the Appendix.

2. CUMULATIVE LOGISTIC REGRESSION MODEL

The main reason of employing a cumulative logistic regression model in this thesis is the ordinal dependent variable – the winner of the dispute. Depending on a dispute, there can be up to five possible outcomes – mother wins, mother partially wins, no one wins, father partially wins and father wins. Having more than two outcomes does not allow for the use of the general logistic regression, since it suggests a binary variable. Since these outcomes can be clearly listed in an order (in relation to the "strength" of one of the parents' position), a cumulative logistic regression model is a more appropriate solution compared to a multinomial logistic regression model, for which it is sufficient that a dependent variable have more than two classes that cannot be ordered. This chapter closely follows the explanation of logistic regression models using cumulative logits (Agresti, 2010, 44 – 87). As some of the disputes allow only for two outcomes, the basic logistic regression model is used for modelling the winner in these disputes, therefore logistic regression is also briefly explained in the beginning of the next section.

2.1 Logistic regression

A logistic regression can be applied, if a dependent variable Y is binary, meaning that it can get two values, 0 and 1. By denoting the probability of success ($Y = 1$) by π, the vector of explanatory variables by $\mathbf{x}$ and the vector of coefficients of these explanatory variables in the model by $\boldsymbol{\beta}$, the probabilities are calculated as follows (Davison, 2003, 490):

$$P(Y = 1) = \pi = \frac{\exp(\boldsymbol{x^T\beta})}{1 + \exp(\boldsymbol{x^T\beta})} \text{ and}$$

$$P(Y = 0) = 1 - \pi = \frac{1}{1 + \exp(\boldsymbol{x^T\beta})}. \quad (1)$$

The odds of success are calculated as:

$$\frac{P(Y=1)}{P(Y=0)} = \frac{\pi}{1-\pi} = \exp(\boldsymbol{x^T\beta}).$$

Given the independent binary observations $y_1, \dots, y_n$, the likelihood function for beta ($\boldsymbol{\beta}$) is given by:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{n} \left\{\frac{\exp(\boldsymbol{x_j^T\beta})}{1+\exp\left(\boldsymbol{x_j^T\beta}\right)}\right\}^{y_j} \left\{\frac{1}{1+\exp\left(\boldsymbol{x_j^T\beta}\right)}\right\}^{1-y_j} = \frac{\exp(\sum_j y_j \boldsymbol{x_j^T\beta})}{\prod_j \{1+exp(\boldsymbol{x_j^T\beta})\}}.$$

2.2 Cumulative logit

Let us assume that the dependent variable *Y* can obtain *c* different values (later related as categories). We also assume that *Y* is an ordinal variable, which means that its values can be ordered according to a specific feature, as is done for the winner of a dispute later in this thesis. The probability that *Y* obtains a value in a category *j* is marked by π_j. We can then denote:

$$F_j = P(Y \leq j) = \pi_1 + \cdots + \pi_j,\ j = 1, 2, \ldots, c, \text{and}$$

$$0 \leq F_1 \leq F_2 \leq \cdots \leq F_c = 1.$$

The cumulative logit is defined as follows:

$$\text{logit}\,[P(Y \leq j)] = log\,\frac{P(Y \leq j)}{1-P(Y \leq j)} = log\,\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_c},\ j = 1, \ldots, c-1. \quad (2)$$

2.2.1 Odds and odds ratio

Odds of cumulative probabilities are defined as follows:

$$\Omega = \frac{P(Y \leq j)}{1-P(Y \leq j)}.$$

If $\Omega > 1$, it is more likely that $y \leq j$ than $y > j$. In other words, the probability of *Y* to fall into lower categories (1, 2, … , j) is higher than the probability to fall into higher categories (j + 1, j + 2, … , c). Now, let Ω_A be the odds for event "A" and Ω_B the odds for the event "B" respectively. Hence, the odds ratio is:

$$\text{OR} = \frac{\Omega_A}{\Omega_B}.$$

2.2.2 Proportional odds property

For a cumulative logit model, the following property holds:

$$\text{logit}\,[P(Y \leq j \mid \mathbf{x}_1)] - \text{logit}\,[P(Y \leq j \mid \mathbf{x}_2)] =$$

$$= \log \frac{P(Y \leq j \mid \mathbf{x}_1)/P(Y > j \mid \mathbf{x}_1)}{P(Y \leq j \mid \mathbf{x}_2)/P(Y > j \mid \mathbf{x}_2)} = \boldsymbol{\beta}'(\mathbf{x}_1 - \mathbf{x}_2). \quad (3)$$

Odds of the event $Y \leq j$, if $\mathbf{x} = \mathbf{x}_1$, is $\exp[\boldsymbol{\beta}'(\mathbf{x}_1 - \mathbf{x}_2)]$ times the respective odds of the event $\mathbf{x} = \mathbf{x}_2$. The logarithm of the odds ratio is thus directly proportional to the distance between $\mathbf{x}_1$ and $\mathbf{x}_2$. Exactly the same proportionality takes place for all $c - 1$ logits, which gives the name "proportional odds model" to all cumulative logit models satisfying this property (Agresti, 2010, 53).

2.3 Cumulative logit model

The purpose of the cumulative logit model is to explain the variable Y with a set of explanatory variables. Let $\mathbf{x}$ be a k x 1 column vector containing the explanatory variables, and let's denote the category of the dependent variable by j. This leads to the following regression model:

$$\text{logit}\,[P(Y \leq j)] = \alpha_j + \boldsymbol{\beta}'\mathbf{x} = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k\,. \quad (4)$$

where $j = 1, \ldots, c - 1$, and $\boldsymbol{\beta}$ is a column vector containing k parameters that indicate the influence of explanatory variables on the dependent variable. One should notice that α_j depends on category j. Namely, α_j increases with the ordinal number of the category j, since $P(Y \leq j)$ also increases with the increase in j, with fixed $\mathbf{x}$. From formula (4), the following expression for the cumulative probability is obtained (Agresti, 2010, 47):

$$P(Y \leq j) = \frac{\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x})}{1 + \exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x})}, \; j = 1, \ldots, c - 1. \quad (5)$$

The individual probabilities can be solved from the expression above:

$$P(Y = j) = \frac{\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x})}{1 + \exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x})} - \frac{\exp(\alpha_{j-1} + \boldsymbol{\beta}'\mathbf{x})}{1 + \exp(\alpha_{j-1} + \boldsymbol{\beta}'\mathbf{x})}\,.$$

2.4 Fitting the model

The next important task is to estimate the parameter vector $\boldsymbol{\beta}$. It can be done using the maximum likelihood (ML) method.

We assume that for an observation *i*, $y_{i1}, \dots, y_{ic}$ are binary indicators, for which $y_{ij} = 1$, when response is in the category *j*. In our further analysis, each *i* will indicate a certain dispute in court. Let Y_i be a category to which each observation falls. Now, when $Y_i = j$, $y_{ij} = 1$ and $y_{it} = 0$ for all t = 1, …, c, but different from *j*. The following notation is used:

$$\pi_j(\mathbf{x}_i) = P(Y_i = j \mid \mathbf{X} = \mathbf{x}_i) = P(Y_i \le j \mid \mathbf{x}_i) - P(Y_i \le j - 1 \mid \mathbf{x}_i).$$

The formula above defines the probability than an observation *i* with a vector $\mathbf{x}_i$ of explanatory variables falls into the category *j*. The likelihood function is the product of the probabilities:

$$\prod_{i=1}^{n}[\prod_{j=1}^{c}\pi_j(\boldsymbol{x}_i)^{y_{ij}}] =$$

$$= \prod_{i=1}^{n}\{\prod_{j=1}^{c}[P(Y_i \le j \mid \boldsymbol{x}_i) - P(Y_i \le j - 1 \mid \boldsymbol{x}_i)]^{y_{ij}}\} =$$

$$= \prod_{i=1}^{n}\{\prod_{j=1}^{c}[\frac{\exp(\alpha_j + \boldsymbol{\beta}\prime\mathbf{x}_i)}{1+\exp(\alpha_j + \boldsymbol{\beta}\prime\mathbf{x}_i)} - \frac{\exp(\alpha_{j-1} + \boldsymbol{\beta}\prime\mathbf{x}_i)}{1+\exp(\alpha_{j-1} + \boldsymbol{\beta}\prime\mathbf{x}_i)}]^{y_{ij}}\}.$$

The above likelihood is a function of parameters ($\alpha_1, \dots, \alpha_c$, $\boldsymbol{\beta}$). The log-likelihood function is denoted by $L(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Each likelihood equation is obtained by differentiating $L(\boldsymbol{\alpha}, \boldsymbol{\beta})$ with respect to each parameter and making the result equal to zero. For simplicity, the following notations are used:

$$G(z) = \frac{\exp(z)}{1+\exp(z)}, \quad g(z) = \frac{\exp(z)}{[1+\exp(z)]^2},$$

The log-likelihood function is:

$$\sum_{i=1}^{n}\sum_{j=1}^{c} y_{ij}\log[G(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_i) - G(\alpha_{j-1} + \boldsymbol{\beta}'\mathbf{x}_i)].$$

Let us first differentiate the function above with respect to β_t:

$$\frac{\partial L(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \beta_t} = \sum_{i=1}^{n}\sum_{j=1}^{c} y_{ij}x_{it}\frac{g(\alpha_j + \boldsymbol{\beta}\prime\mathbf{x}_i) - g(\alpha_{j-1} + \boldsymbol{\beta}\prime\mathbf{x}_i)}{G(\alpha_j + \boldsymbol{\beta}\prime\mathbf{x}_i) - G(\alpha_{j-1} + \boldsymbol{\beta}\prime\mathbf{x}_i)} = 0.$$

The differentiation with respect to α_t is more complicated, so Kronecker delta δ_{jt} is introduced for simplicity ($\delta_{jt} = 1$ if j = t, and $\delta_{jt} = 0$ otherwise). The likelihood equations are:

$$\frac{\partial L(\alpha,\beta)}{\partial \alpha_t} = \sum_{i=1}^{n}\sum_{j=1}^{c} y_{ij}\frac{\delta_{jt}g(\alpha_j + \boldsymbol{\beta}\prime\mathbf{x}_i) - \delta_{j-1,k}g(\alpha_{j-1} + \boldsymbol{\beta}\prime\mathbf{x}_i)}{G(\alpha_j + \boldsymbol{\beta}\prime\mathbf{x}_i) - G(\alpha_{j-1} + \boldsymbol{\beta}\prime\mathbf{x}_i)} = 0.$$

2.5 Error estimation

As the parameters are estimated by the ML method, the standard errors of the estimators are calculated. Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ be the complete list of parameters. The Hessian matrix $\mathbf{H}$ is derived as the second order derivative of the log-likelihood function L evaluated at the maximum likelihood estimates $\widehat{\boldsymbol{\alpha}}$:

$$\mathbf{H} = \frac{\partial^2 l\,(\widehat{\boldsymbol{\alpha}})}{\partial \widehat{\boldsymbol{\alpha}} \partial \widehat{\boldsymbol{\alpha}}^{\boldsymbol{T}}}$$

Now, the standard errors can be calculated as follows (Hastie, Tibshirani, Friedman, 2018, 266):

$$\text{se}\,(\widehat{\boldsymbol{\alpha}}) = diag\,(\sqrt{[-\boldsymbol{H}(\widehat{\boldsymbol{\alpha}})^{-\mathbf{1}}]})$$

2.6 Statistical inference

Below is an example of the null hypothesis, which should be tested, in order to examine whether any of the variables x_1 , … , x_k explain the dependent variable *Y*:

$$H_0\text{: } \beta_1 = \ldots = \beta_k = 0.$$

If the null hypothesis stated above is rejected, it means that the coefficients are jointly significantly different from zero, in which case x_t, t = 1, … , k, has explanatory power and should be included in the list of the explanatory variables. Hence the test is crucial in outlining all the significant variables.

The other type of null hypothesis stated in this thesis is:

$$H_0\text{: } \beta_1 = \beta_2,$$

versus the alternative hypothesis:

$$H_1\text{: } \beta_1 \neq \beta_2.$$

The likelihood-ratio (LR) test statistics is:

$$-2(L_0 - L_1)\,. \qquad (6)$$

L_0 is the maximized log-likelihood function under the null hypothesis, and L_1 is the maximized log-likelihood function under the alternative hypothesis. The test statistics follows an

asymptotic reference chi-squared distribution with degrees of freedom equal to the difference in the number of parameters for two models and can be used to test the above hypothesis (Davison, 2003, 139).

2.7 Model diagnostics, measures of fit and predictive power

2.7.1 Goodness of fit

Let **x** be a matrix of the values of the explanatory variables $x_1, \dots, x_k$ with n observations for each variable. For these observations, the respective values of Y fall into each of the response categories $j = 1, \dots, c$. We denote by n_j the number of observations under category j. The estimates of the expected frequencies are:

$$\hat{\mu}_j = n \times P(Y = j \mid \mathbf{x}), \quad j = 1, \dots, c,$$

where $\hat{\mu}_j$ is the expected amount of observations for which Y is in category j.

The test statistics for the likelihood-ratio test are:

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \text{ (Pearson statistic for goodness of fit), and}$$

$$G^2 = 2 \sum_i \sum_j n_j \log \frac{n_j}{\hat{\mu}_j} \text{ (likelihood-ratio statistic).}$$

Under the assumption of null hypothesis (stating that the model holds), X^2 and G^2 both have an asymptotic χ^2-distribution, with the degrees of freedom being the same as the number of parameters in the estimated model (Agresti, 2010, 67).

2.7.2 Residuals

The residuals are widely used for a model's diagnostics; they are generally defined as the difference between the observed and the fitted values. The smaller the residuals are, the better is a model's fit. This definition is, however, more appropriate for the models with continuous dependent variables. For an ordinal dependent variable, this becomes more challenging.

Liu and Zhang (2018) proposed a so-called surrogate residual for an ordinal dependent variable. Let Y be the named variable, and let $f_0(y)$ be its true distribution, while $f_a(y)$ is its assumed

distribution. First, a hypothetical continuous random variable Z is found, with a marginal distribution of Z on Y being $f_a(y)$, meaning that:

$$\int f_a(y, z)\, dz = f_a(y),$$

where $f_a(y, z)$ is a joint distribution. Z can also be referred to as a latent variable. Next, a continuous variable S is defined, following the distribution $\int f_a(z \mid y)\, f_0(y)\, dy / m_c$, with m_c being a normalizing constant.[1] The surrogate residual is now defined as:

$$R_s = S - E(S \mid \boldsymbol{X}), \tag{8}$$

where $\boldsymbol{X}$ is a matrix of explanatory variables. Liu and Zhang (2018) also demonstrate that the standard assumptions of the ordinary least squares model residuals will hold for these derived residuals. They are symmetrical around zero, $E(R_s \mid \boldsymbol{X}) = 0$. They are homogeneous, $\mathrm{Var}(R_s \mid \boldsymbol{X}) = c$, a constant which is independent of $\boldsymbol{X}$. We denote the link function, which transforms the probabilities of the categories of the response variable to a continuous scale, by G. The distribution of residuals thus approaches the distribution of the link function, such that $R_s \sim G(c + \int u dG(u))$.

It is important to distinguish between goodness of fit and explanatory (predictive) power of the model. While a model may fit data well, it may not provide a sufficient explanatory power. A method introduced next measures the predictive power of the model.

2.7.3 Multiple correlation measures

Let the sample marginal proportions for the categories of the dependent variable be $\{\hat{p}_j\}$, where $j = 1, \ldots, c$. Now, the average cumulative proportion in category j is calculated as:

$$v_j = \sum_{k=1}^{j-1} p_k + \frac{1}{2}\hat{p}_j.$$

These scores are referred to as ridit scores. A measure of the model's predictive power is obtained by forming a correlation between the observed outcome category ridit score and the estimated mean score obtained by the model, which is generated by the probability values fitted

[1] The normalizing constant makes the integral of the density function equal to one, in order for the distribution to be statistically valid.

for the model. A higher correlation indicates a better predictive power (Agresti, Tarantola, 2018).

3. DETERMINATION OF CUSTODY

The consideration of a child's custody in courts of Finland has been established by the law regarding children's custody and their right to receive the care from both parents.[2] The main objective of the law is that both parents preserve their custody over their children unless there is a specific reason for the custody of one or both parents to be eliminated.

It may be challenging to state what are the main interests of a child that should be taken into account when investigating the custody cases in court. The main criteria at the moment are child's psychological, health and social development.

Palo-Repo (2015, 19) explains the process of an investigation of dispute regarding a child's custody or residency in court. This chapter is adapted from Palo-Repo's explanations.

3.1 Defining a guardian

When a child is born, both parents become guardians if they are married. In case of a divorce, the custody is not changed unless there is an agreement between parents stating otherwise, or one or both parents apply for the sole custody, leading to a case investigation in court.

Parents have an alternative of making a written agreement regarding their child's custody and residency. The social welfare board approves the agreement, if they do not consider it to be acting against the interests of a child and if at least one of the parties is currently the child's guardian. As the agreement is approved and both parties accept it, the case is not taken to the court.

In case parents do not reach an agreement, the court opens the investigation of the case, and the court judges prepare their final decision. Judges should also pay the most attention to a child's interests, namely which of the parents is likely to provide a more stable and healthy environment for the child, at the same time making sure that the child has enough communication with the other parent through regular meetings.

[2] 315/2014, Laki lapsen huollosta ja tapaamisoikeudesta annetun lain muuttamisesta.

3.2 Rights and obligations of a guardian

The main role of a guardian is to provide safety, welfare and a balanced psychological, educational and social development to a child. Any type of custody should also ensure close connections of the child with both parents, excluding the cases where a child's interaction with one of the parents has been proven to act against the best interest of the child or even to be dangerous for his or her physical or mental health (for instance, the parent's problems with alcohol, mental instability or practicing physical violence). Keeping in mind a child's interests, a guardian is supposed to handle the child's health care, education, place of residency and other personal aspects. A guardian also has the right to receive any information regarding the child from the officials, for instance, a child's health situation from the hospital, grades at school and other related information.

3.3 Consideration of custody and residency cases in court

A child's guardian or the other parent are the only parties eligible for applying for a child's custody or residency. The application is directed to the court of the district where the child is officially located (either permanently or temporarily). In case of divorce, the application regarding custody or residency is considered as part of the divorce application.

In a situation where no agreement is possible to achieve, the process advances as follows. One of the parents prepares an application stating the custody and residency demands. The court informs the other parent about this application and reserves a right for the other parent to prepare a counterclaim with own demands for custody and residency. If the other parent does not respond to the application or accepts the applicant's demands as such, only the first parent's demands are taken into account when making the final decision. If the other parent provides a response with the counterclaim, both parents (and their possible legal assistants) are invited to the court for a preparatory session. During the session, both parents explain their claims in detail. The court may demand additional information from the social welfare board, in which case the latter has to prepare a thorough report on the living conditions of each parent and their relationships with the child. The report has to include a wide scope of aspects, such as: how does each parent communicates and interacts with the child; what is their ability to provide a balanced development of the child. The judge may also want to listen to the child's opinion on the situation and possible willingness to stay with a particular parent. Children are rarely invited

to the court, but if a child's age is 12 years or more (up to 18 years), the decision of the court cannot be executed against the child's will.

After the position of each party was listened in court and after the social welfare board (if necessary) has prepared the report about the living conditions and relationships of both parents with the child, the case is processed to the main investigation. Besides the report, any other documents from the parties involved that are considered to be important for the investigation are accepted. The parents and, if needed, witnesses, are invited for auditions. A final decision is generally done by one judge, and in some cases three judges participate equally in the decision-making process. After the final decision is declared, in case one of the parties is not satisfied, they have a 7-day period to express their willingness to appeal, in which case they are given a month to present their appeal. If no appeal took place or if the appeal was rejected by the court, both parties are obliged to obey the decision of the court.

The custodial law was changed in 2018[3], which has led to certain changes in the consideration of the related cases in court. The disputes studied for this thesis, however, were solved before these modifications, hence they have followed the process described earlier in this chapter.

[3] 257/2018, Laki lapsen huollosta ja tapaamisoikeudesta annetun lain 6 ja 8 a §:n muuttamisesta.

4. DATA COLLECTION

For the collection of data regarding custody and residency cases in courts of Finland, a research permit was applied for each district court. The applications were sent to the courts of the following cities: Helsinki, Espoo, Vantaa, Tuusula, Porvoo, Hyvinkää, Lahti, Tammisaari and Hämeenlinna. All these courts approved the application and provided the required data, which were then combined with the earlier data collected by Eero Lilja for his Bachelor's Thesis (Lilja, 2018). Lilja collected the relevant cases from the courts of three cities – Helsinki, Espoo and Vantaa. The time period of all the cases ranged from the years 2004 to 2015[4] and varied for each court. Table 1 contains the information on the time span for which data on the disputes were collected in the court of each city.

Court (City)	**Time span for which the data was collected completely**
Helsinki	2007
Espoo	2004 – 2015
Vantaa	2004 – 2012
Tuusula	2004 – 2015
Porvoo	2007, 2010, 2013
Hyvinkää	2007, 2010, 2013
Lahti	2007, 2010, 2013
Tammisaari	2007, 2010, 2013
Hämeenlinna	2007, 2010, 2013

Table 1 – Range of years for which the datasets were collected from different courts.

The choice of years for data collection can partially be explained by the availability of data provided by the courts (for example, the court of Helsinki has only provided the cases for the year 2007). There are also three practical reasons for doing this: firstly, there is a sufficient time period between each year, which increases the chance of variability within the data, since, for example, the judges might have changed. Secondly, not all cases started in the later years (2014,

[4] The courts sort the cases according to the starting date, i.e. the date when an application from one of the parents is received.

2015 and 2016) might have been collected, since some of them are first sent to the court of appeal for further consideration and then sent back to the district court. This process may take up to three years for the final decision by the court and thus there most likely were investigations that started in the year 2014 or 2015 and were still in progress at the time of data collection (2015 – 2016). Thirdly, some parents present their renewed applications to the court after getting a decision, if they are not satisfied with this decision, which increases the probability of treating essentially the same case as two separate cases. Taking the data from non-consecutive years reduces the chances of these happening.

The cases taken by court for investigation are given a diary number. Every case contains an application from one of the parties, a response from the other party (if received), possibly additional documents prepared by both parties, a report by the social welfare (if applicable) and the decision of the court. The Legal Register Center (Oikeusrekisterikeskus) provided a list of the diary numbers for the types of investigations that are most relevant for the analysis.

Below are the types of collected cases that are not included in the analysis due to their irrelevancy to the main research question. The applicant may cancel the application, after which the case is no longer investigated and no decision is required. The parties may come to an agreement straight away or after some discussion, which in both situations means that the only role of the court is to confirm the agreement if it does not in any way violate the interests of a child involved. All cases where the parties are other than parents are also excluded from the analysis, for instance, mother of a child applying for the grandmother to have a custody as well. A final reason for excluding a case from the analysis is not containing any custody or residency application at all; the cases concerning solely the right of a child to meet the other parent or the financial support from the other parent are excluded from the analyses.

5. DATA OVERVIEW

There were 3 388 documents collected from the nine district courts of Finland, of which 474 cases were relevant for the analysis, and after combining them with the cases collected by Lilja (2018), this number changes to 767.[5] The distribution of the relevant cases for the courts is demonstrated in Table 2:

Court	Number of relevant cases
Vantaa	236
Helsinki	175
Tuusula	122
Espoo	99
Lahti	35
Hyvinkää	30
Porvoo	25
Hämeenlinna	24
Tammisaari	21

Table 2 – District courts and the number of relevant cases collected from each court.[6]

5.1 Winner

The dependent variable of the statistical analyses conducted in the following chapters is the winner of a dispute. This variable takes into account how the final decision of the court is related to initial demands by both parties, in order to find out whether one of the parents' application is satisfied (or at least partially satisfied).

The winner variable can take up to five values – mother wins a dispute (–2), mother partially wins a dispute (–1), no one wins (0), father partially wins a dispute (1) and father wins a dispute (2).

[5] Section 2 of the Appendix presents the most common types of cases that have been collected but are excluded from the analysis.
[6] The number of cases in the table is calculated by combining the cases collected both by the author of this thesis and by Lilja (2018).

In most of the disputes, an applicant has essentially two types of demands: a child should live with the applicant, or the applicant should become the only guardian of a child, or both demands. Besides these, the following types of cases were registered:

- The applicant demands sole custody, but so that the second parent has the right to freely receive information regarding the child from the officials. If the application is accepted by the court, the other parent has no authority to make any decision regarding the child.

- The applicant demands to have the sole right to decide on certain aspects regarding the child, but there are no demands regarding residency or guardianship. These aspects are usually educational (which school the child attends), health-related (which hospital or doctor the child goes to, which medication the child uses, what food the child eats etc.) or related to religion.

- The applicant demands for joint custody, with the child still staying at the other parent. Alternatively, in the same situation, the applicant may just demand to have the right to receive information about the child from the officials, thus allowing the second parent to continue being the only guardian.

The disputes can be classified into two categories: disputes regarding residency and disputes regarding custody. The process of defining a certain case to relate to one of these groups is explained in detail in Section 2 of the Appendix. Briefly, a case is treated as a residency dispute, if at least one of the parties has a demand regarding a child's residency. Otherwise a case is treated as a custody dispute.

The distribution of winners varies between the two types of disputes, as shown in Table 3. Mothers won slightly over half of all custody disputes and fathers won 48 %. For residency disputes, mothers won in over 53 % of the cases, and fathers in 38 %. Mothers become winners more often than fathers in both forms of disputes, according to the data. The differences are statistically significant for residency disputes, as well as for all disputes together, on a 5 % significance level.

	Mother wins completely or partially	Father wins completely or partially	*p*-value
All disputes	51.8 % (398)	43.0 % (330)	< 0.00001
Custody disputes only	50.5 % (167)	49.2 % (162)	0.589
Residency disputes only	53.2 % (232)	38.8 % (169)	< 0.00001

Table 3 – Proportions of winners, overall and for custody and residency disputes separately.[7]

Table 4 demonstrates the proportion of applicants for each type of dispute. The largest difference between applicants by their gender takes place for custody disputes: for each one father-applicant there are at least two mother-applicants. The proportions are almost equal for residency disputes. Overall, the number of mother-applicants for the collected cases is 35 % larger than that of father-applicants.

	Mother-applicant	Father-applicant
All disputes	57.5 % (441)	42.5 % (326)
Custody disputes only	70.6 % (233)	29.4 % (97)
Residency disputes only	47.6 % (208)	52.4 % (229)

Table 4 – Proportions of mother- and father-applicants, overall and for custody and residency disputes separately.

Apart from a larger number of mother-applicants for custody disputes, the demands of mothers are harsher; over 90 % of mother-applicants apply for sole custody and less than 10 % apply for the right to have a sole custody with restrictions (more beneficial for fathers). As for the father-applicants, almost three quarters of them (74 %) apply for the joint custody allowing a child to

[7] The proportion test is explained in Section 3 of the Appendix.

still live with mother (meaning that mothers initially have a sole custody in these cases). Only 23 % of the fathers apply for sole custody. The above mentioned statistics also introduce the fact that the cases with mother-applicants and with father-applicants are different in terms of the initial conditions, namely in over half of the custody disputes (52 %) where father is an applicant, mother initially has sole custody, meaning that a sole custody often sparks new disputes. The more exact figures for all disputes can be found in Table 5.

	Mother applicant	Father applicant
Applies for the sole custody	70.3 % (310)	37.4 % (122)
Applies only for the sole residency[8]	22.2 % (98)	33.4 % (109)
Other[9]	7.5 % (33)	29.2 % (95)

Table 5 – Distribution of applications for mother- and father-applicants.

5.2 Foreign background

In over one-third of the cases (38 %), at least one of the parents is of foreign background. This proportion makes foreign background an important factor, and the possible effects of parents' foreign background on the court's decisions regarding custody and residency of a child have already been studied previously in Finland.[10] The definition of foreign background is challenging, since one of the parents may have a Finnish citizenship, but at the same time it may be obvious (e. g. due to a name) that this parent is not a native Finn. Foreign background is indicated either when the country of origin of a parent was specifically mentioned in the document or when the name and surname of the parent clearly indicated foreign origin.[11] In the

[8] In an application for the sole custody, sole residency is applied automatically, if the applicant does not have sole residency at the moment of application.

[9] In a situation where the other parent initially has sole custody or sole residency, the applicant may apply for a joint custody or joint residency.

[10] E.g. Pere, Lilja, Sobolev (2017)

[11] Names and surnames of Swedish origins made an exception, since about 6 % of the Finnish population are so-called Swedish-speaking Finns. Such a person is indicated as having a foreign background only if it is

latter case, the next step is to define the country of origin. Countries are classified by the scale 0 to 4. The value of 0 means Finnish background, while the values 1 – 4 means foreign background, according to the human development index (HDI)[12] of the country in question. HDI is a continuous variable which gets values from 0 to 1. The classification is explained in detail in the Appendix. For simplicity, we can use the following classification:

1 stands for Scandinavia, USA and Australia;

2 stands for Eastern Europe, Russia and CEE countries that are not part of EU (most applicants and respondents in this group had Russian origin);

3 stands for Asian countries and countries on the South American continent;

4 stands for African countries.

Of the mothers with a foreign background, approximately 33 % belonged to the European countries with the highest HDI, and the same amount belonged to the states of Eastern Europe. Slightly over 15 % came from the Asian countries and 17 % from the African countries with the lowest human development index score.

The corresponding proportions for fathers of foreign background are 34 %, 18 %, 21 % and 27%. Overall, there were 28 cases where only mother had a foreign background, 54 cases where only father had a foreign background, and 98 cases where both parents were not native Finns.

The distribution of winners depending on the foreign background of the parties is recorded in Table 6. For the custody disputes, the proportion of mother winners is the lowest if both parents are native Finns (42 %). The proportion is higher if the mother alone has a foreign background (53 %) and increases (54 %) when both parents have a foreign background. For the families where father alone is a foreigner, the proportion grows to 66 %. The same pattern for the distribution of mother winners depending on the foreign background applies for all the cases in general, regardless of a dispute type: mother wins more often, when she is the only member with the foreign background; even more often, when both mother and father have foreign backgrounds; and most often, when father is the only member having a foreign origin.

otherwise clear from the case that the person is not a Swedish-speaking Finn (for instance, it was mentioned that the person constantly lives in Sweden or has only the Swedish citizenship).

[12] Human Development Index is a measure of a country's economic development and living standards, including level of education and literacy, life expectancy and income per capita. http://hdr.undp.org/en/content/human-development-index-hdi

Pere, Lilja and Sobolev (2017) studied the above mentioned phenomena using logistic regression and found out that the proportions of mother-winners in the four groups of disputes (both parents of Finnish origin, only mother is of a foreign background, only father is of a foreign background and both parents are of a foreign background) are statistically significantly different. The analysis is based on the same data and has similar results, however, the exact percentages in Table 6 below are slightly different. This is explained by the fact that some of the cases that were included as relevant in the article were later excluded from the analysis in this thesis (e. g. respondent not replying to the application).

Foreign background	Mother wins	No winner	Father wins
All disputes			
None	47.6 % (226)	4.4 % (21)	48.0 % (228)
Mother only	52.6 % (41)	7.7 % (6)	39.7 % (31)
Both	59.8 % (58)	11.3 % (11)	28.9 % (28)
Father only	64.3 % (74)	1.7 % (2)	33.9 % (39)
Custody disputes			
None	42.3 % (71)	0 % (0)	57.7 % (97)
Mother only	51.6 % (16)	0 % (0)	48.4 % (15)
Both	51.9 % (28)	1.9 % (1)	46.3 % (25)
Father only	67.1 % (51)	0 % (0)	32.9 % (25)

Residency disputes			
None	50.5 % (155)	6.8 % (21)	42.7 % (131)
Mother only	53.2 % (25)	12.8 % (6)	34.0 % (16)
Both	67.4 % (29)	16.3 % (7)	16.3 % (7)
Father only	59.0 % (23)	5.1 % (2)	35.9 % (14)

Table 6 – Distribution of winners depending on the foreign backgrounds of both parents; all disputes, custody disputes and residency disputes.

Tables 7a and 7b illustrate the distribution of couples having a dispute in terms of their origin. Mothers of Finnish background have a partner also of Finnish background in 69 % of the custody disputes and in 89 % of the residency disputes in the data. For all groups of foreign origin except group 1 (HDI > 0.7), there are a lot of cases where partners come from foreign countries of the same group. In the data, part of these cases included a scenario, where parents originated from the same country, their marriage was registered in Finland, but later father moved abroad (typically to his home country), leaving the mother with a child in Finland, and thus creating a more favorable situation for the mother to win a dispute.

Father's foreign background / Mother's foreign background	0	1	2	3	4
0	68.57 %	12.24 %	3.67 %	5.31 %	10.20 %
1	34.78 %	39.13 %	8.70 %	8.70 %	8.70 %
2	50.00 %	0.00 %	41.18 %	5.88 %	2.94 %
3	26.67 %	0.00 %	0.00 %	73.33 %	0.00 %
4	7.69 %	7.69 %	0.00 %	0.00 %	84.62 %

Table 7a – Distribution of the native origins of parents for the custody disputes, column headers "0", "1", "2", "3" and "4" represent the respective HDI groups.

Father's foreign background / Mother's foreign background	0	1	2	3	4
0	88.76 %	6.92 %	1.44 %	1.73 %	1.15 %
1	69.44 %	19.44 %	2.78 %	5.56 %	2.78 %
2	53.85 %	0.00 %	30.77 %	7.69 %	7.69 %
3	33.33 %	8.33 %	0.00 %	58.33 %	0.00 %
4	25.00 %	6.25 %	0.00 %	0.00 %	68.75 %

Table 7b – Distribution of the native origins of parents for the residency disputes, column headers "0", "1", "2", "3" and "4" represent the respective HDI groups.

5.3 Age, gender and number of children

The number of children in each case defines those children which were included in a dispute, i.e. if a dispute regarded custody or residency of a child, this child was counted. There were few cases, where not all children were involved in a dispute. This situation mostly occurred when parents had previous children from other marriage or partners, therefore these children were not considered as their common children. The form where the information from the courts was collected only suggested the maximum number of children involved in a dispute as four. There still were three cases with a larger number of children. For them, the information about the four oldest children was recorded (gender and age). As Table 8 shows, over half of the cases had only one child involved in a dispute, approximately every third dispute (31 %) was about two children and only one in eight disputes had over two children involved (12 %). Since the disputes with one child involved make the largest part of the data, the disputes are divided into two groups in this section – disputes with only one child and disputes with more than one child.

Number of children	Number of cases	Percentage
One	436	56.9 %
Two	241	31.3 %
Three	76	9.9 %
Four and more	14	1.9 %
All	**767**	

Table 8 – Distribution of disputes according to the number of children involved.

There were in total 621 boys and 587 girls involved in the disputes, making the share of boys equal to 51.4 %, which is not statistically significantly different (*p*-value is 0.83) from the average proportion of boys born in Finland (51.1 %)[13]. Boys account for 49.6 % of all custody disputes and 52.7 % of all residency disputes, therefore these proportions are still not statistically significantly different. Considering only the disputes with one child, 45.4 % of custody disputes and 54.8 % of residency disputes are about boys. This makes a more equal proportion of boys and girls, as compared to the results by Palo-Repo (2015, 41), in which the

[13] Tilastokeskus (2018).

proportion of boys was 55 % overall and 60 % in one-child disputes. It is worth noticing, however, that Palo-Repo (2015) used the data from the courts of appeal, which might make a difference.

Palo-Repo (2015, 43) commented on the different nature of demands by mother- and father-applicants. Fathers tend to apply for a sole custody much more rarely than mothers. The types of applications also seem to be connected to the gender of a child. It is worth noticing that, in the case of disputes concerning one child, when father is an applicant, he applies for a sole residency or a sole custody more often when the child is a boy (72 % versus 59 % for girls). Moreover, if father is a respondent, the proportions are 47 % for boys and only 26 % for girls. At the same time, mothers apply for a sole residency or sole custody in 95 % of the cases where a child is a boy, and in 90 % of the cases where the child is a girl.

In one-child disputes, the gender of a child is not highly associated with the proportion of mother and father winners for both types of disputes (Figures 1a and 1b). It is still worth noticing that mothers are more likely to win a residency dispute if a child is a boy (52.4 % versus 49.5 % when a child is a girl). Even though the result is not statistically significant (*p*-value is 0.66), it contradicts the widely discussed idea that boys are more likely to stay with fathers, while girls with mothers. This phenomenon can be partially explained by the fact that fathers with a foreign background tend to struggle for boys more than for the girls (the proportions of sole residency or sole custody demands by fathers with foreign origin are 38 % for boys and 28 % for girls), and, since fathers with the foreign background have the least probability of winning (as presented earlier), this results in mothers winning most of the disputes. The same can be observed for the demands of fathers with a Finnish origin – 57 % of fathers whose child is a boy apply for a sole custody or a sole residency, while, when the child is a girl, the proportion falls to 51 %. On the other hand, father's behavior applied to both forms of dispute, thus failing to explain the absence of the same phenomenon for the custody disputes.

The greatest proportion of mothers and the smallest proportion of fathers winning a residency dispute (52 % and 38 % respectively) is when a boy is involved. This is the only group which stands for a statistically significant difference between the number of winners for mothers and fathers (*p*-value = 0.021).

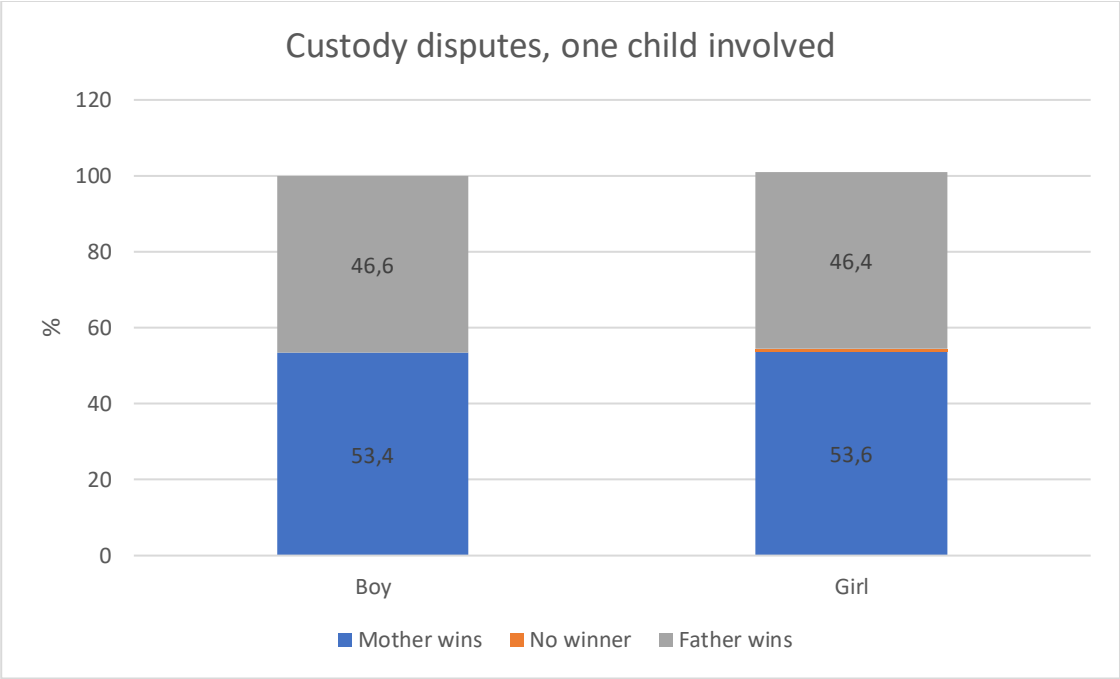

Figure 1a – Distribution of winners in custody disputes with one child involved, according to the gender of a child.

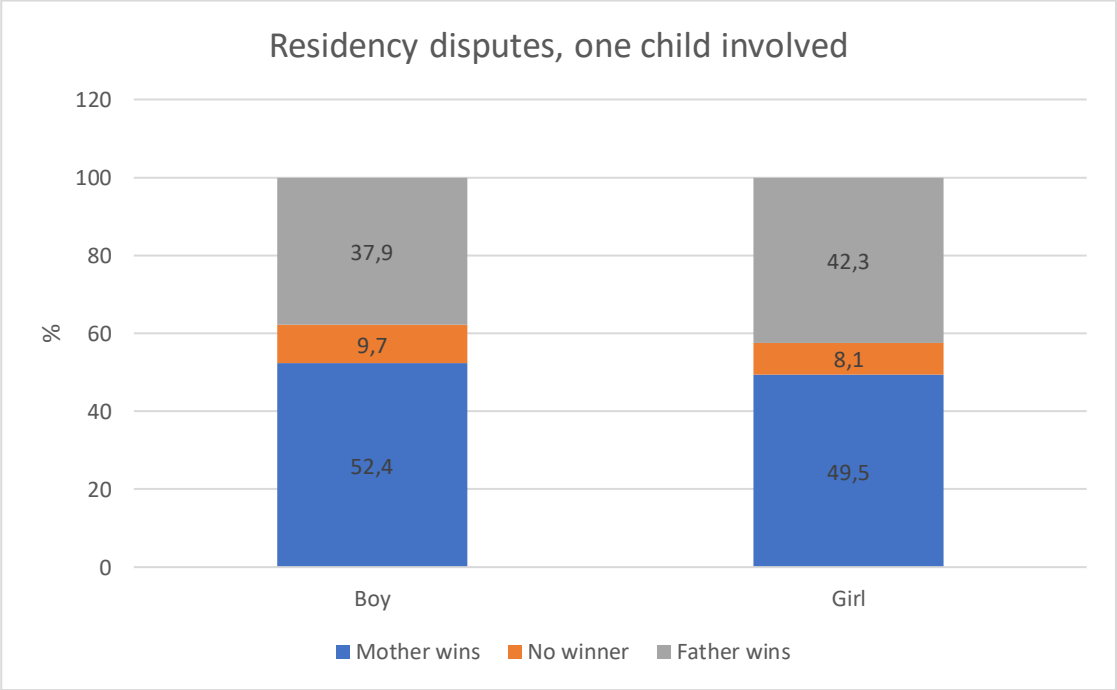

Figure 1b – Distribution of winners in residency disputes with one child involved, according to the gender of a child.

In order to study the possible connection of children's age with the outcome of a dispute, ages are divided into three groups: 0 – 6 years (toddlers and pre-school), 7 – 12 years (primary and

secondary school), 13 – 17 years (teenagers). In a dispute, age group is defined according to the age of the oldest child. The average age of a child is 7.6 years in custody disputes and 8.6 years in residency disputes.[14] As Figure 2a indicates, mothers tend to win custody disputes more often when a child is aged 0 – 6 years (56.2 %), compared to the other age groups of a child (45.1 % for 7 – 12 years old and 51.0 % for 13 – 17 years old). This does not lead to a statistically significant difference between the age groups. For the residency disputes (Figure 2b), in the cases where the ages are 0 – 6 years mothers win 57.9 % of the disputes. As the age group increases to 7 – 12 years, the proportion of mother winners decreases to 55.4 %, and then falls to 43.8 % for the age group 13 – 17 years. The differences are statistically significant, when the youngest and the oldest age groups are compared (*p*-value=0.031). At the same time, the proportions of father winners grow steadily with the increasing age of a child; these are 33.1 %, 36.9 % and 50.0 %.

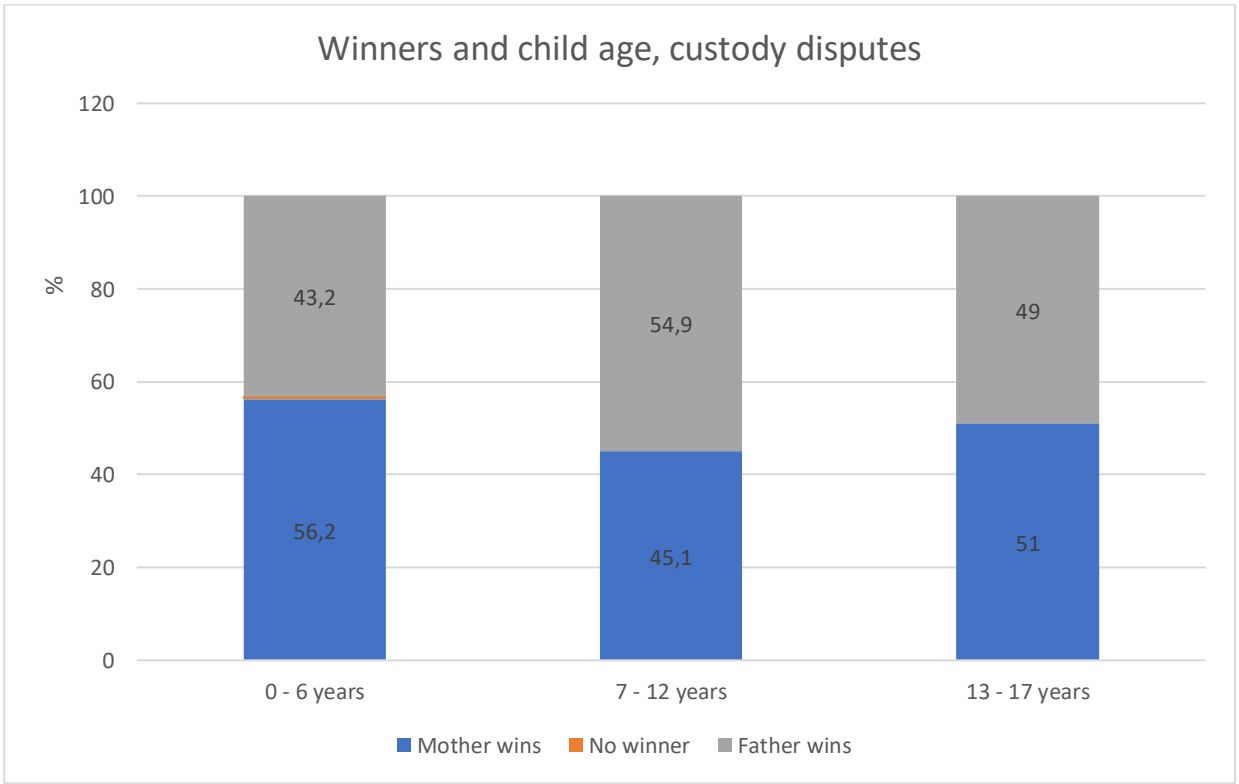

Figure 2a – Distribution of winners in custody disputes, according to the age of the oldest child.

[14] The calculations are made according to the age of the oldest child in the disputes.

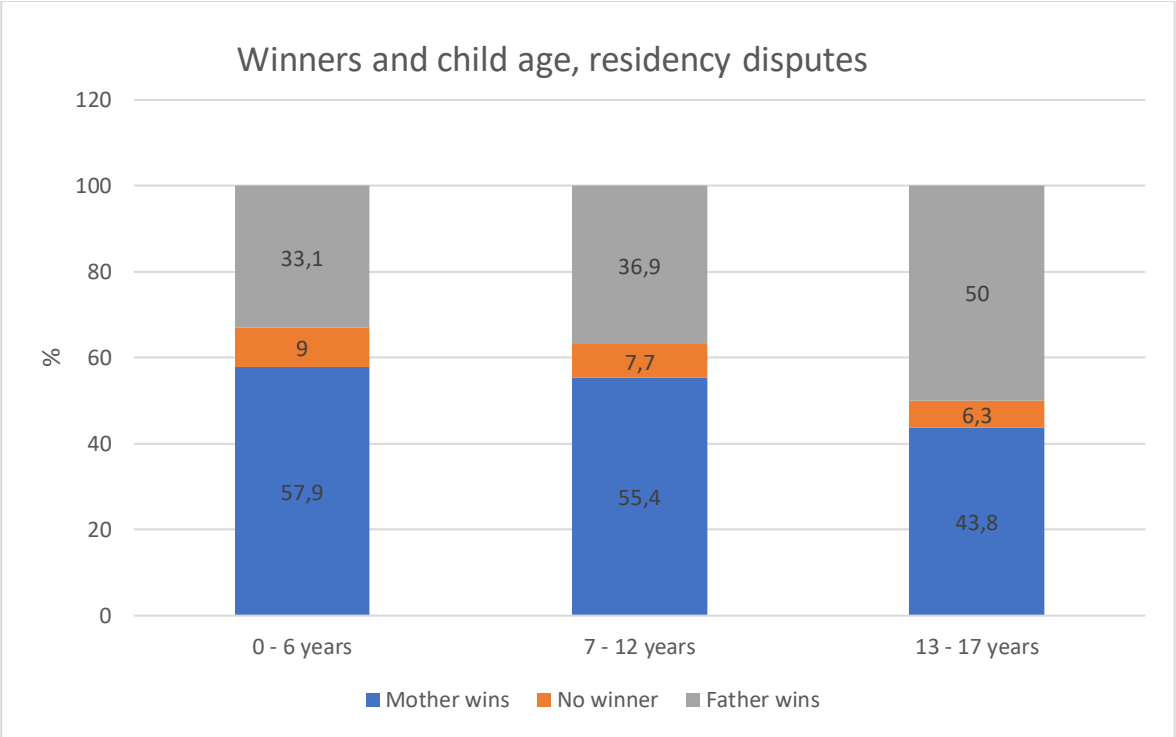

Figure 2b – Distribution of winners in residency disputes, according to the age of the oldest child.

5.4 Age of parents

The average age of two parents in a dispute varies between 21.7 and 61.6 years. The average of the average ages is 38.6 years. The youngest parent is 20 years old and the oldest is 72 years old. Most applicants are in the age group of 30 – 39 years (43 %); the second largest age group is 40 – 49 years (35 %). Slightly over 15 % of the applicants are younger than 30 years old and 7 % are 50 years old and above.

For both dispute types, when applicants are younger, mothers tend to win more often. For custody disputes, the cases with the applicants age between 30 and 39 years lead to the highest proportion of mother winners (54 %), while for residency disputes, the proportions are almost the same for the youngest applicants (less than 30 years old) and for the applicants aged between 30 and 39 years – 55.0 % and 55.2 % respectively. This result might partially be explained by the gender of the applicants from each age group. As Figure 3 indicates, the proportion of mother applicants in all forms of disputes steadily decreases with an increasing age.

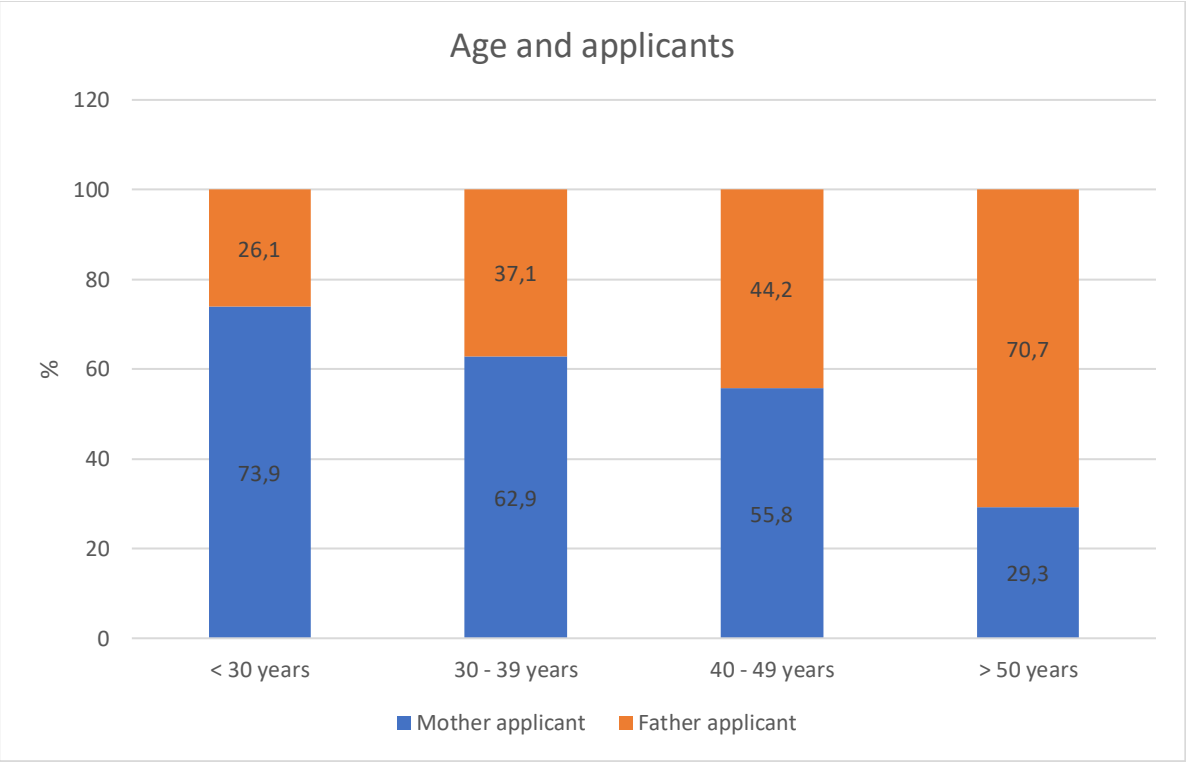

Figure 3 – Distributions of gender of applicants for each age group.

5.5 Legal assistants

The majority of parents involved in the disputes employ a legal assistant to support and defend their claims in court. A legal assistant can be either attorney-at-law or any other assistant who has official rights to provide assistance in court. Over 78 % of applicants employed a legal assistant (81 % of them chose attorney-at-law). Of the respondents, slightly over 70 % had a legal assistant, of which 78 % were attorneys-at-law.

Table 9 demonstrates the proportions of mother and father winners in cases where none of the parties uses a legal assistant and where at least one party does. The share of father winners decreases from 49 % to 38 % and the share of mother winners rises from 51 % to 62 %, as applicants use a legal assistant and respondents do not. This can be explained by the earlier mentioned observation that mothers are more often applicants than fathers. The differences in shares are, however, not statistically significant.

For the residency disputes, the share of mother winners is not statistically significantly different for the cases when no parties use a legal assistant and when both parties have an assistant (48 % and 52 % respectively). When only applicant has an assistant, the proportion of mother winners

grows to 60 %. The share of father winners is the highest (45 %), when no parties hire legal assistant.

According to the data on custody disputes, almost 82 % of mother applicants and slightly over 75 % of father applicants have a legal assistant. The respective shares for the residency disputes are 83 % and 72 %.

	Custody disputes, shares of winning (%)	Residency disputes, shares of winning (%)
No legal assistant	Mother: 51.0 Father: 49.0	Mother: 47.6 Father: 45.1
Only applicant has a legal assistant	Mother: 62.2 Father: 37.8	Mother: 60.4 Father: 35.8
Both parties have a legal assistant	Mother: 48.4 Father: 51.1	Mother: 52.5 Father: 37.7

Table 9 – Distributions of winners and the use of legal assistants, custody and residency disputes.

5.6 Preceding situation with custody and residency

In this section, we will study the custody and residency situations before the parents have started a dispute in each case. A child's current residency and contact with both parents is considered to be an important factor for the judges to pay attention to when considering the application.

The list of possible custody and residency situations is listed in the Appendix. Four of them that take place in the majority of the disputes can be outlined. These are: both parents have custody and child lives with mother, both parents have custody and child lives with father, mother has a sole custody, both parents have custody and child lives with parents in turn – either equal amount of time or more time at one of the parents, if so has been agreed. It is worth noticing that father has a sole custody initially in only 12 cases.

In 50 % of the custody disputes a child is initially living with mother. Almost 95 % of the applicants in these cases are mothers and almost 97 % of their applications are about getting a

sole custody. For the residency disputes, approximately 35 % of cases start with the situations where a child is currently living with mother.

For the custody disputes where mother already has a sole custody, the percentage of mother winners is the highest – 80 % (Figure 4a). It is worth noticing that only fathers are applicants in these cases. Approximately 44 % of the mothers win in the case where a child initially lives with mother; this proportion decreases to 42 % when a child lives with both parents and rises back to 72 %, when a child lives with father. The latter can be explained by the observation that in almost all of these disputes father demanded for a sole custody, which is put in place only if a joint custody acts against the interests of a child. For the residency disputes (Figure 4b), the share of mother winners is 60 % when mother has a sole custody and 54 % when a child lives with mother. The mothers' winning share falls gradually to 25 % when a child is initially living with father, while fathers win 64 % of the disputes in this case. The latter indicates the non-symmetric distribution of mother and father winners according to the initial residency of a child, and the fact that living with father seems to be considered a stronger argument in father's favor than living with mother is considered to be in mother's favor. Still, a statistical test regarding the possible differences in proportions of mother winners, when a child is living with mother, and father winners, when a child is living with father, does not give a significant result (p-value = 0.28). It is worth noticing that the number of cases where a child initially lives with father is small – only 30 % of the number of cases where a child lives with mother.

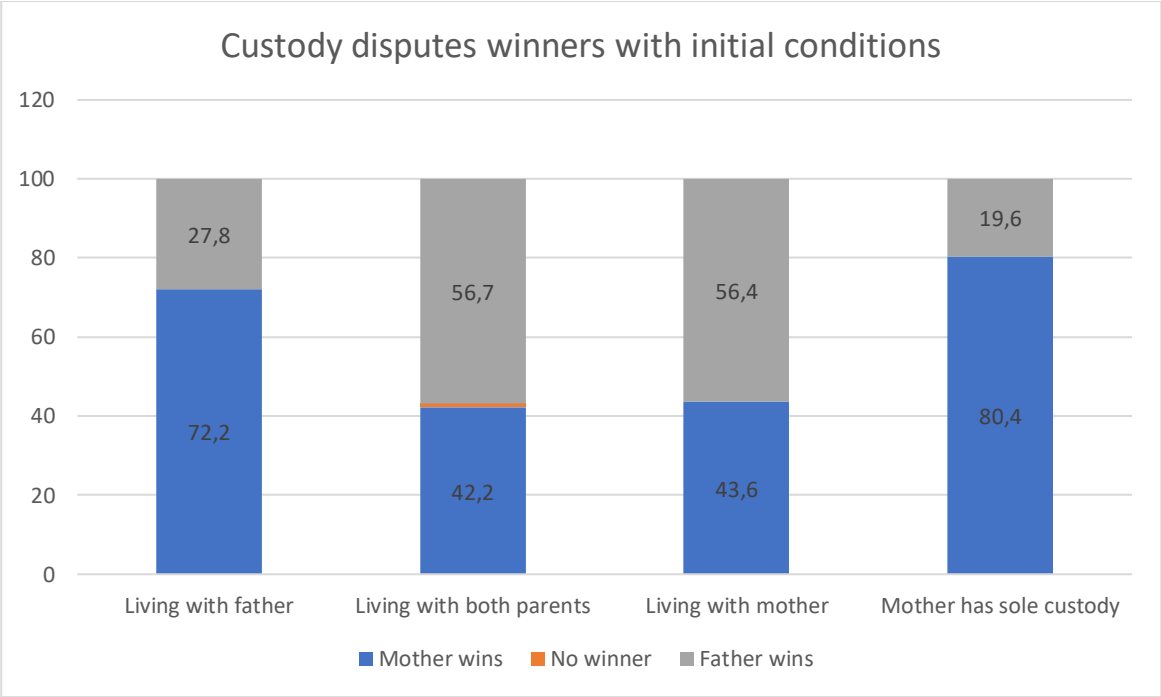

Figure 4a – Distribution of winners in custody disputes, depending on the initial custody and residency situation.

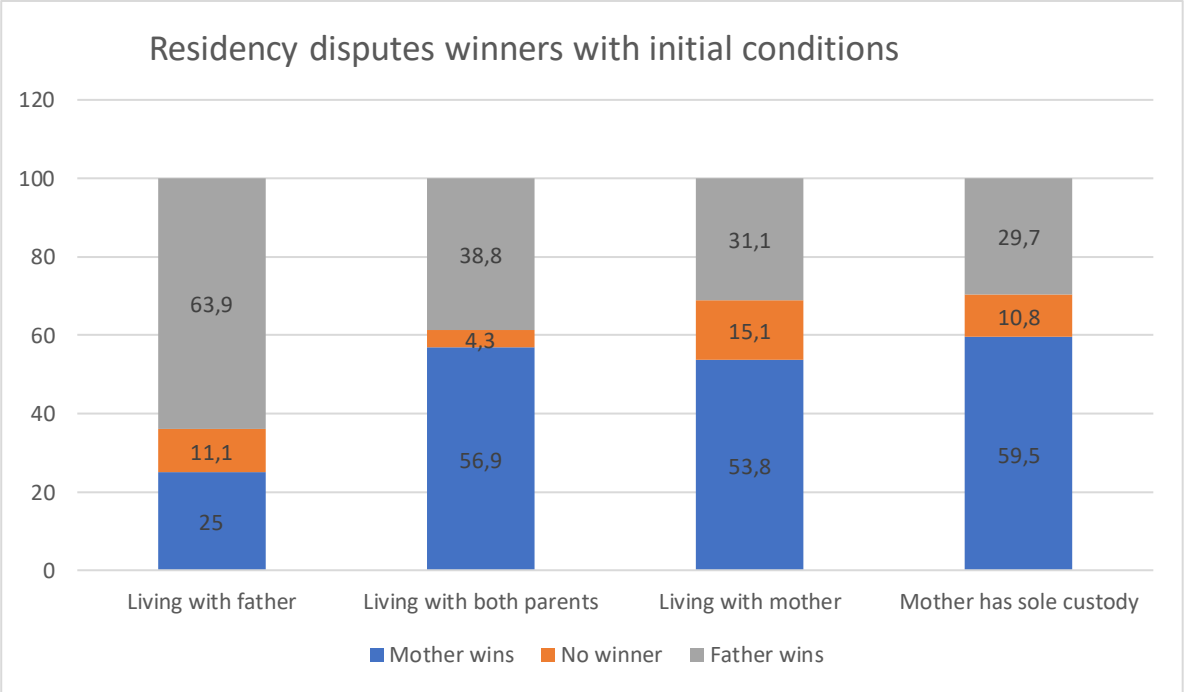

Figure 4b – Distribution of winners in residency disputes, depending on the initial custody and residency situation.

5.7 Earlier decision

All disputes contain information about whether an earlier decision concerning the child was made by the court. If such a decision took place earlier, the previous disputes' documents are also included. Of the 767 cases, 135 contain an earlier decision (18 %). These cases are equally distributed among the two dispute types – 18 % and 17 % for custody and residency disputes respectively. If a child was previously defined to live with mother, the new decision does not change in 40 % of the cases; if mother had sole custody, the respective percentage is over 73 %. For the father, in the cases where a child was living with father, the decision did not change in 44 % of the cases. Finally, there was a small group of cases (17), where father initially had a sole custody, and 65 % of them (11) stayed the same.

It has been claimed[15] that the chances of father to win tend to increase, if the dispute is not the first one. This does not, however, take place for the gathered data. For the custody disputes, the percentage of father winners with and without an earlier decision are 40 % and 51 % respectively. In case of residency disputes, the shares become 43 % and 38 %, which does go along with the earlier research, but the difference cannot be considered statistically significant (p-value = 0.42).

5.8 Judge's gender

Every dispute that does not end with a mutual agreement requires a decision by judge. For the custody and residency disputes there were on average 3 – 5 judges in a district court of each city. Overall, there were almost twice as much cases considered by female judges than by male judges. The greatest difference was in Vantaa (81 % of cases with female judges and 19 % of cases with male judges); the opposite situation took place in Raasepori (33 % and 67 % by female and male judges respectively).

[15] Valkala, Litmala (2006)

	Custody disputes	Residency dispute
Female judges	Mother: 48.6% Father: 50.9 %	Mother: 51.2 % Father: 41.1 %
Male judges	Mother: 54.2 % Father: 45.8 %	Mother: 57.0 % Father: 33.8 %

Table 10 – Gender of judges and winners for custody and residency disputes.

As Table 10 indicates, the highest difference in the distribution of winners between the genders of the judges is for the fathers in residency disputes (57 % of mothers and 34 % of fathers win with male judges). The differences in the shares of mother and father winners for the residency disputes are statistically significant both for female and for male judges.

5.9 Region

The cases were collected from the courts of nine cities, three of which form the capital region (Helsinki, Espoo and Vantaa). The data was divided by two main region groups – the capital region and the other cities. The reason for applying this grouping lies mainly in the larger proportion of disputes involving parents with the foreign background in the capital region and in the use of legal assistants, which will be outlined later. As the number of cases varies between the cities, there is a small chance of obtaining relevant and objective results if comparing the cases by city. At the same time the comparison could be performed for two groups – the capital region and the other cities.

In the set of collected cases, 510 (66 %) came from the capital region and 257 (34 %) from the other cities. Custody disputes make a slightly bigger share in the capital region (45 %) than in the other cities (39 %). Around 13 % of the families in the capital region have disputes concerning more than two children; the same proportion for the other cities is smaller – 9 %. Foreign background takes place more frequently in the capital region (28 % and 32 % for mothers and father respectively) compared to the other cities (12 % for mothers and 20 % for fathers). As for the legal assistant, they are more often involved in non-capital regions (88 % of applicants and 81 % of respondents) than in the capital region (74 % of applicants and 66 % of respondents).

As can be observed from Table 11, mothers tend to win more often in the capital region in both types of disputes. The difference between the regions is higher for the residency disputes, but it is not statistically significant (*p*-value=0.087).

	Capital region	Other cities
All disputes	53.8 %	48.4 %
Custody disputes	50.9 %	49.5 %
Residency disputes	56.2 %	47.8 %

Table 11 – Shares of mother winners by regions (capital region – Helsinki, Espoo, Vantaa – and other cities), overall and for custody and residency disputes separately.

5.10 Accusations

In part of the collected cases, one or both parents made an accusation towards the other parent aimed at demonstrating that the other parent is not capable to take care of the child in a proper way. The accusations concern alcohol issues, taking drugs, mental illnesses, violence towards child or towards the other parent and preventing the other parent from seeing a child.

Overall, there are 204 cases (27 %), where mother is accused of one or more of the aspects mentioned above. The respective amount for fathers is 293 (38 %). In 70 cases (9 %) both parents have accusations towards each other. Of these, 90 accusations against mother (44 %) and 124 accusations against fathers (42 %) were considered by the court as valid.

If an accusation against mother is considered valid, father wins the majority (69 %) of the cases, and if the accusation against father is valid, mothers win in 73 % of the cases. It should be noted that the numbers of cases for each type of accusation are quite low, so the results might not necessarily be statistically significant, even if the differences seem high. Another aspect to be taken into account is the fact that several accusations towards one parent could be claimed and substantiated within the same case. The proportions of mother and father winners based on a substantiation of one of the accusations are shown in Tables 12a and 12b.

	Accusation towards mother is substantiated	Accusation towards father is substantiated
Alcohol & drugs	25.6 %	74.4 %
Mental problems	28.1 %	61.5 %
Violence towards a child	33.3 %	68.0 %
Violence towards the other parent	0 %	78.4 %

Table 12a – Share of mother winners when one or more accusations is substantiated.

	Accusation towards mother is substantiated	Accusation towards father is substantiated
Alcohol & drugs	74.4 %	23.3 %
Mental problems	71.9 %	30.8 %
Violence towards a child	58.3 %	16.0 %
Violence towards the other parent	100 %	13.5 %

Table 12b – Share of father winners when one or more accusations is substantiated.

It is worth noticing that if an accusation of mother's violence towards father is substantiated, mother does not have a chance of winning a dispute at all. However, it can be easily explained by the fact that only one dispute had a substantiated accusation of mother being physically violent towards father. As for the violence towards a child, a substantiated accusation of mother's violence towards the child is not so critical for mothers as a substantiated accusation of father's violence towards the child is for fathers. One-third of mothers (33 %) practicing violence towards a child still win a dispute, while the respective share for fathers is only 16 %.

5.11 Judges

The 330 custody disputes have been considered by 104 judges, and the 437 residency disputes by 113 judges. Overall the disputes have been solved by 129 judges. Approximately one third of them (40) have been involved in just one dispute, while only five judges for custody disputes and ten judges for residency disputes have been involved in ten or more cases. The maximum number of disputes considered by one judge was 28 for custody disputes and 30 for residency disputes.[16]

In the modelling process presented in the next chapter, the number of disputes solved by a judge will be tested as a potential explanatory variable.

[16] Most of the judges have been involved in both types of disputes. In particular, 79 judges are in both groups.

6. MODELLING THE WINNER

In the previous chapter we investigated how some of the variables of the data may explain the winner of the disputes. Still, the main objective of this thesis is to create a model for explaining the simultaneous associations of several variables, including the ones not discussed earlier.

It has been shown in Chapter 5 that the statistical results often differ for custody and residency disputes. Moreover, the nature of these two dispute types is clearly different – for residency disputes, it is more crucial to clarify which of the parents is more able to provide a healthier environment for the child's living, while for custody disputes the question is whether a sole custody of one parent is of the best interest of the child. Another important reason is the classification of the response variable – the winner of a dispute. While custody dispute may only result in mother winning, father winning, or no winners, a residency dispute also allows for partial winners. For instance, in a residency dispute, where one or both parents also apply for a sole custody, they become partial winners, if the residency demand is satisfied and the sole custody demand is not satisfied by the court. For the mentioned reasons, custody and residency disputes are analyzed separately in this chapter. As mentioned earlier, a dispute is classified as a residency dispute, if it involves residency issues, otherwise it is denoted as a custody dispute.

The dependent variable is the winner of a dispute. One parent in a dispute is a winner, if the court has completely or partially fulfilled his or her demands. If a court's decision does not differ from the initial situation, i.e. there is no change in custody or residency of a child, even though both parents applied for some change, there is no winner. For the custody disputes, mothers won 51 % and fathers 48 % of the cases. The respective shares for the residency disputes are 53 % and 39 %.

6.1 Choice of explanatory variables

The variables used in the models described later can be divided into two groups – the ones recorded during the process of data collection and derived (new variables that have been formed from the former variables). The detailed information about the recorded variables can be found in Section 1 of the Appendix.

The following basic variables have been taken account of in the construction of the models: gender of applicant and respondent, genders of children, ages of parents and children (in years),

established living condition of children at the moment of application, initial custody and residency situation, demands of both parents, gender of a judge, existence of an earlier court decision regarding the same dispute, using a legal assistant, gender of legal assistants of the applicant and of the respondent (if they are involved), foreign origin of parents, accusation of violence towards the other parent, accusation of violence towards the child, accusation of usage of alcohol or drugs, accusation of mental illness and accusation of preventing the other parent from seeing the child, possible substantiation of any of these accusations, legal aid to any of the parents, city of the district court.

The reasons for deriving new variables were based on previous research as well as on intuition. The prerequisites for deriving some of the variables will be stated in this chapter. A complete description of every variable is in the Appendix.

As outlined in Section 5.11, there was a small number of judges who have been involved in ten or more disputes. It was therefore decided to create indicator variables[17] for the judges, in order to separate the "less experienced" from the "more experienced". One variable defined whether a judge had been involved in 1 – 9 disputes. For every judge whose name was met in ten or more cases, a separate indicator variable was created.

As the data contains the disputes from a range of years (2004 – 2015), time scale was also considered to be important, since the decision-making process might have been changing throughout this period. Every dispute was given a time-scale value, which was calculated as the number of months passed from the earliest occurrence in the data.

Age and age difference were other interesting aspects, both for parents and for children. The idea was to test an association of father's winning chances with his age and with the age difference between the parents (including the direction of this difference). According to the intuition, older parents should be preferred over the younger ones, due to their supposedly more stable financial and psychological conditions. Two respective variables were created for this purpose. The analogical variables were made regarding the ages of children. The youngest child's age, as well as the age difference between the youngest and the oldest children were

[17] An indicator variable is a binary variable that takes a value of 1, if a certain condition is satisfied and a value of 0 otherwise.

calculated.[18] If all children involved in a dispute were boys, this was recorded as a separate indicator variable in the model.

There were five types of accusations presented in court: violence towards the other parent, violence towards the child, abuse of alcohol or drugs, mental health problems and preventing the other parent's meeting with the child. The court may or may not find some of the accusations valid, so the following indicator variables were introduced for every accusation: firstly, if an accusation took place from at least one of the parties, secondly whether it was substantiated. This results in ten variables regarding accusations, all of them were taken into the model initially. If such variable turned out to be statistically insignificant, it was removed and replaced by the two variables – whether this accusation concerns mother or father. Alternatively, if a variable kept being significant in the final version of the model (after removing all insignificant variables), it was still deleted and "split" into the same two variables in the end. This helped to find out, if the final decision of the court was associated with investigating a certain type of accusation and with particular gender.

Some disputes included an earlier decision regarding the custody or the residency of a child. In addition, there was evidence that the presence of an earlier decision increases the chances of fathers to win.[19] In order to test this, two indicator variables were created: one telling whether father had sole custody or residency based on the court's previous decision, and the other indicator variable telling the same information for mother. If both variables turned out to be significant, gender equality test was performed to find whether one variable has a larger association with the dependent variable in the model than the other. In the presence of an earlier decision, one of the parents appealed to the court for a change in decision. This also led to a creation of two indicator variables and a gender equality test. The same approach was used for the established living conditions of a child; two variables – for established living conditions at mother's place and established living conditions and father's place – were made, in order to test whether one has a more statistically significant association with the court's final decision.

[18] If there is only one child, he or she is considered as both the youngest and the oldest, and the age difference is 0.
[19] Valkala & Litmala (2006).

6.2 Choice of models

All the models described in this section are obtained using the statistical software R (R Core Team, 2019) and the package "ordinal" installed into R . The dependent variable can obtain up to five numerical values: –2 (mother wins a dispute), –1 (mother partially wins a dispute), 0 (no winner), 1 (father partially wins a dispute), 2 (father wins a dispute). For the residency disputes, the winner variable gets all the five mentioned values, due to a lot of residency disputes concerning both residency and custody, and thus allowing for partial winners. For the custody disputes, it is impossible to have a partial winner, therefore the response variable is limited to three values – mother wins, no winner and father wins. In addition, since there is only one dispute with no winner for the custody disputes, it is excluded from the analysis, and the response variable is further limited to two values – mother wins (0) and father wins (1) (the response variable is ordered with the same logic as for the residency disputes – it gets the highest value when father wins and the lowest value when mother wins).

The different nature of the response variables implies two different models. For the custody disputes, the logistic regression model (1) is employed, explained in Section 2.1. The model for the residency disputes requires a multinomial regression model, so the cumulative logistic regression model (2) is employed, due to the ordinal nature of the dependent variable. The model is explained in Section 2.2. As outlined in Section 5.1, a dispute is treated as a custody dispute, if only the custody of a child is disputed about; otherwise a dispute is a residency dispute. Each dispute used in the analysis therefore falls in exactly one type of a dispute. More information on how a dispute type is defined according to the demands by both parents can be found in the Appendix.

Most of the explanatory variables are so-called indicator variables – the ones that get a value of 1, if some condition is satisfied, and a value of 0 otherwise. Some of the variables are discrete variables with numeric values (e.g. age in years). To suit better for the logit-type models, the values of these variables are subtracted from their average values (for the custody and for the residency disputes separately).

Separate models were calculated for the custody disputes and for the residency disputes. The procedure of selecting the explanatory variables was identical for both models: first, a model including all potential explanatory variables was calculated, after which a variable with the largest p-value was eliminated. A new model was estimated not containing this variable, and the process was repeated, until the p-values of all the remaining variables were less than 0.05. As

mentioned in the previous section, when an indicator variable for an accusation was eliminated, it was replaced by the two variables stating whether the accusation concerned mother or father. If an accusation variable was still statistically significant after eliminating all non-significant variables, it was replaced in the same way. The same procedure ("splitting" the variable by two – for mother and for father) was performed with the variable regarding established conditions at one of the parents' place and with the variable regarding whether one of the parents is appealing to the court on the basis of the previous decision.

Interaction variables made an exception to the model selection process. The usual practice is that variables from which the interaction variable is formed are not eliminated unless the interaction variable itself becomes statistically insignificant. This gives a possibility for models, where an interaction variable is significant, while the original variables are not, but they all stay in the model (Fox, 2015, 144). For the models described in this section, if such a situation took place, the variables forming the interaction variable were not eliminated unless the interaction variable itself was eliminated. If the interaction variable stayed as significant, the joint statistical significance of the main variables and an interaction term was tested using LR-test. The decision on keeping or omitting the main variables was made based on both the interpretation of the model and the test. The models described in the next two sections have been obtained after the LR-tests were performed.

6.3 Model for custody disputes

As the custody disputes have only two categories of the outcome, 1 (father wins) and 0 (mother wins), the basic logistic regression (1) was used to build the model. The procedure outlined in the previous section yielded the following model for the custody disputes:

	Estimate	Standard error	Z	*p*-value
(Intercept):	–0.82	0.19	–4.30	
Father applicant	–1.26	0.28	–4.50	< 0.001
Only father is of a foreign background	–0.84	0.30	–2.79	0.005

Father is accused of alcohol / drug abuse	–0.98	0.33	–2.96	0.003
Substantiated violence towards mother	–2.09	0.46	–4.50	< 0.001
Porvoo	1.90	0.92	2.08	0.038
Female judge in one of the courts	1.34	0.52	2.60	0.009

Table 13 – Logistic regression model for the custody disputes.

According to the above mentioned classification of the response variable, its higher value indicates a higher winning probability for fathers. A positive value of a coefficient from the "Estimate"-column means that the father's winning probability increases with an increasing value of the respective explanatory variable and vice versa.

A dispute where father is an applicant reduces the chances of the father to win, according to the model. Most of these disputes are, however, continuation to the earlier disputes, where mothers have won. Namely, over 65 % of the father's demands state that both parents should have a custody, meaning that at the moment of application mother has a sole custody. Fathers lost 70 % of these disputes. On the other hand, mother's application of a sole custody increases the winning chances of fathers. According to the data, fathers won in 48.6 of the cases where mothers applied for a sole custody, and only in 38.5 % of the cases where mother's application did not concern sole custody. This can be explained by the fact that in over half of the disputes where mother demanded for a sole custody (54.5 %) a child (or children) was already living at mother's place and fathers demanded that the situation stayed unchanged. As sole custody is a serious change concerning the rights of fathers and should only be defined if a joint custody acts against the interests of a child, it was not accepted by the court in many cases, thus making fathers winners.

Among the nine cities where the district courts were located, Porvoo is the only significant one in the model, suggesting a higher chance for fathers to win a dispute, if applying to Porvoo's

court. Still, there were only eight observations from the district court of Porvoo (2 % of the data, of which fathers won six disputes).

The model includes two accusations. Both act against fathers, but only one of them – violence towards mother – was substantiated. The other concerned accusation regarding alcohol or drug abuse by father, and it also decreases the father's winning chances, even without the substantiation of the accusation. It is worth noting that the magnitude of the negative coefficient of the indicator variable for the substantiated accusation is over two times larger than the magnitude of the indicator variable for the non-substantiated accusation, meaning that the former is worse for a father. Of the 61 of these accusations against fathers, slightly less than half (28) were substantiated. Fathers accused of alcohol or drug abuse won 23 disputes (37 %) out of 61, which is a smaller proportion compared to the one for all the custody disputes (48.3 %)

Of the foreign background variables, only father's foreign background is statistically significant. It has a negative impact on father's winning chance, as was also outlined in the article by Pere, Lilja, Sobolev (2017) referred to in Chapter 5. Since the effects of mother's foreign background were also found to be significant in the article, an alternative model, for which two remaining foreign background variables were added (mother's foreign background and both parents' foreign backgrounds), was tested against the original model. The original model was found to be more suitable (*p*-value of a joint test being 0.436), so the other foreign background variables were not added to the model.

Of the judges involved in ten or more disputes, only one judge's decisions are significantly different. A dispute considered by this judge increases the changes of fathers to win. According to my supervisor, this judge is unofficially regarded by some judges and lawyers as an independent thinker. She was also involved in the highest number of custody disputes (28); fathers won 22 of these disputes. Still, in most of them, a child was initially living with mother, mother demanded for a sole custody, father demanded that the situation stays the same, and the final decision stated that neither residency nor custody changes, thus being in accordance with father's demand. Such a decision can be actually thought more as to reject mother's demand than to act according to father's demand, since the final situation stays the same; these fathers are not as "strong" winners as fathers who demanded a change and won a dispute.

The model does not contain any variables regarding the situation before the dispute – namely, the initial custody and residency situation and established conditions at one of the parent's place. These can, however, be partially interpreted by the variable "Father applicant". As

pointed out earlier, 65 % of father-applicants in custody disputes apply for a joint custody, meaning that mothers initially have a sole custody in these cases.

6.4 Model for residency disputes

The residency model was obtained using the cumulative logistic regression (2). It was first tested with a 5-step response variable (including partial winners) but proved to violate the proportional odds property (3). One possible explanation for this is the fact that part of the residency disputes included both residency and custody applications, meaning that there could be a partial winner in case the residency application is approved and the custody application is not approved by the court. At the same time, a residency dispute containing only residency applications does not allow for partial winners. A 3-step response variable is used instead, treating partial winners as winners, so that the response variable gets the following values: –1 (mother wins), 0 (no winner) and 1 (father wins). The list of explanatory variables included in the residency model is noticeably different from the custody model:

	Estimate	Standard error	Z	*p*-value
(Intercept):1	0.30	0.18	1.66	
(Intercept):2	0.73	0.18	3.95	
Established conditions at mother	–0.55	0.14	3.79	< 0.001
Established conditions at father	0.88	0.17	4.93	< 0.001
Only mother has a legal assistant	–0.94	0.38	–2.47	0.014
Substantiated mental disorder for mother	1.41	0.51	2.77	0.006
Mother is accused of	1.00	0.32	3.13	0.002

alcohol / drug abuse				
Father is accused of violence towards a child	0.75	0.35	2.18	0.029
Initial custody and residency situation	–0.15	0.06	–2.63	0.009
Father gets legal aid	–1.15	0.47	–2.46	0.014
Mother appeals for a new decision	0.93	0.40	2.30	0.021

Table 14 – Cumulative logistic regression model for residency disputes. (Intercept):1 and (Intercept):2 are the intercepts (thresholds) for the cumulative logits $logit[P(Y_i \leq 0)]$ and $logit[P(Y_i \leq 1)]$, respectively - α_j in the formula (4).

Established conditions at one of the parents (meaning that the child has been living mainly with that parent before the initial application) lead to a higher chance for that parent to win a dispute. The test for equal magnitude of the coefficients in the model reveals that the coefficients for these two variables are significantly different. The established conditions at father have a higher effect on the father's winning chance than the established living conditions at mother have on the mother's winning probability. It appears to be more common for a child to stay with mothers before a dispute (151 cases for mothers and 62 for fathers), meaning that if a child's living was established at father's place, there might be a specific reason for this, such as mother's inappropriate behavior (e.g. mentioned in accusations). This could explain why a child's initial living at father's place might have a stronger positive effect for fathers. Indeed, for the disputes with the established living conditions at father's place, the share of substantiated accusations against mothers was twice as large as that against fathers in the group with the established living conditions at mother's place (25 % and 12 % respectively).

All the accusation variables involved in this model increase the winning chances of fathers. One of them is substantiated mental disorder for mother; it has the highest coefficient, as the only substantiated accusation towards mother. The other two accusations are not substantiated – one of them concerns alcohol or drug abuse by mother and the other concerns father acting violently towards the child. Of 56 former accusations, slightly below half (25) were substantiated, while 12 out of 44 were substantiated for the latter ones. Both accusations increased the fathers' chances to win and can be explained by a relatively higher proportion of mothers for whom alcohol or drug abuse was substantiated (supposedly positive indicator for fathers) and a majority of fathers for whom the accusation regarding violence towards the child was not substantiated (also a positive indicator for fathers). As most of the accusations regarding father's violence towards the child were not substantiated, it appears that mothers more often tend to present faulty accusations. The motivation is clearly to increase their chances to win, since this seems to be one of the strongest arguments against fathers. Still, it is also one of the most investigated accusations, since it would be irresponsible from the court's side to rely on mother's claim without further study of the case, and it appears that a non-substantiated accusation acts against mothers. There might be a straight connection of this variable with the decision of the court, namely that presenting faulty accusations decreases chances of mothers to win. Alternatively, mothers presenting these accusations may initially be in a weaker position and, understanding it, make their accusations, meaning that the accusation itself does not act negatively for mothers.

A father getting legal aid has a smaller winning chance. Legal aid is provided mainly to a parent (or both parents) for whom it is financially difficult to hire and pay for a legal assistant. A father without a proper financial stability may be considered not capable of taking care of a child, thus explaining this result of the model. The respective variable for a legal aid for mothers is not significant, possibly indicating a weaker association of the final decision with mother's financial stability.

If only mother employed a legal assistant, it lowers the probability of fathers to win. This fact goes according to the intuition, since a legal assistant is supposed to have more professional experience to win a dispute. It is worth noticing that only the indicator for a legal assistant in general is significant, without specifying whether it was attorney-at-law – this indicator was eliminated from the model (for both parents).

The initial custody and residency situation is also a statistically significant variable, but its sign is negative, meaning that the more favorable the initial situation is for fathers the less chances they have to win. This was not, however, confirmed after a quick check of the proportions of father winners – they tend to increase with the more favorable initial situations for fathers (29 % when mother has a sole custody, 30 % when a child lives with mother, 39 % when a child lives with both parents and 67 % when a child lives with father). A possible explanation is that the initial custody and residency situation may strongly correlate with other variables, such as the established conditions at one of the parents' place, and thus it is challenging to investigate how initial situation itself influences the winning chances of parents. According to the model, however, the coefficient for this variable is relatively small, but this can be explained by the fact that in contrast to the indicator variables, this variable can get 11 values (from –5 in 5). All variables used in the models, as well as their values, are explained in the Appendix.

In cases where the earlier decision was made and one of the parents appeals to the court for a new decision, a mother appealing to the court increases the winning changes of fathers. When a mother appeals, it is most likely that father has won the previous dispute, supposedly meaning that the courts are reluctant to change their earlier decision, in particular for the cases where father wins.

While the p-values of 0.05 or smaller were taken as criteria of variable's significant in the models, most of the p-values in both models were much smaller (less than 0.01). This further emphasizes the significance of these variables.

Foreign background is not statistically significant for the residency disputes. It goes according to the expectation that the effect of a foreign background will be higher for the custody disputes, since a sole custody allows a parent to take the child abroad and thus increases the risk of abduction. On the other hand, this contradicts the results by Pere, Lilja, Sobolev (2017), meaning that their result might demonstrate a problem of omitted variable bias, since after including other variables in the model in this thesis, the potential effects of the foreign background variables are proved to be statistically insignificant.

Based on the research results by Pere, Lahti, Sutela (2018), it was also tested whether the genders of judge and of the legal assistants are associated with the winner. For this, the four indicator variables were created: a judge and both legal assistants are men, a judge is a man and both legal assistants are women, a judge is a woman and both legal assistants are men, and a

judge and both legal assistants are women.[20] None of these variables, however, proved to be statistically significant in the model above.

Compared to the results for the custody model, where a dispute investigated in the court of Porvoo had a strong association with the winner-variable, location of a district court was not statistically significant in a residency model. A model was also tested when instead of the city indicator variables, the so-called "female judge index" -variables were used, indicating the proportion of female judges in each district court. These variables did not turn out to be significant, either.

Both custody and residency models do not contain any variables indicating the parents' demands. These could, however, be partially associated with the "Father applicant" -variable for custody disputes and initial custody and residency situation as well as established conditions at one of the parents' place for residency disputes, since the typical applications in different situations were discussed in Section 5.1. In addition, none of the age variables are statistically significant in the model, thus contradicting the earlier introduced intuition regarding higher winning chances for an older parent.

6.5 Evaluating the models

The methods described in Section 2.7 are now applied for the validation of the previously obtained models and for investigating their explanatory power. Let us denote the models in the following way:

Model 1 – model for custody disputes.

Model 2 – model for residency disputes.

As mentioned above, the "final" models were tested against the other possible models with the LR-test and proved to be the best models in terms of goodness of fit. Table 15 presents the log-likelihood measures for the Models 1 – 2 compared to the intercept-only models. As can be observed, both Model 1 and Model 2 have higher log-likelihood values and thus outperform the intercept-only models.

[20] These variables were only applicable for the disputes where both parents hired a legal assistant.

	Log-likelihood for the obtained model	Log-likelihood for the intercept-only model
Model 1	–189.11	–227.33
Model 2	–342.77	–396.38

Table 15 – Goodness of fit for the models indicated as log-likelihood values of the Model 1 and Model 2 versus the intercept-only models.

Figures 5a and 5b illustrate the surrogate residuals for the two models. The graphs are obtained with the help of the R-package "sure". Greenwell, McCarthy, Boehmke, Liu (2018) demonstrate a usage of the "sure"-package on the data with one explanatory variable x, in which case the residuals are plotted against x. As our models have more than one explanatory variables, the idea is extended so that the surrogate residuals are plotted against the sum of the values of all explanatory variables.

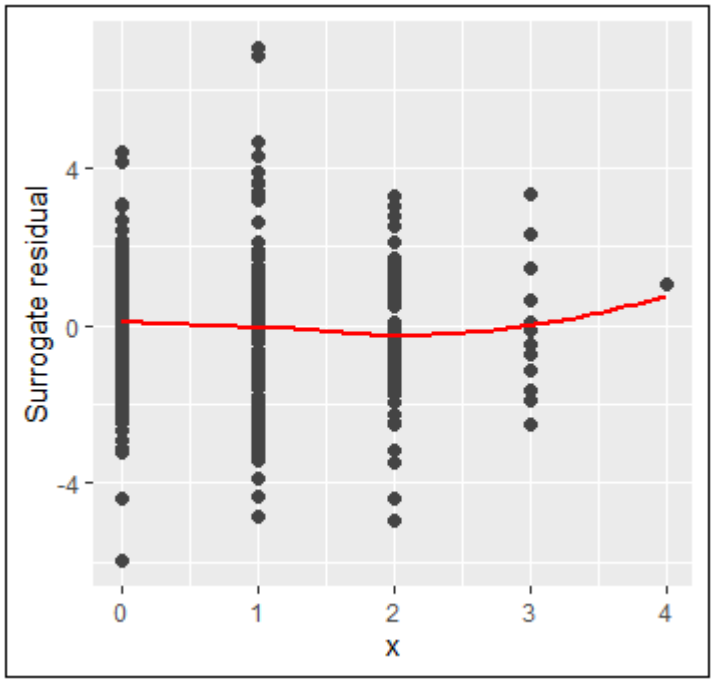

Figure 5a – Surrogate residuals for Model 1, x indicates the sum of the values of all explanatory variables in the model for custody disputes, points on the graph indicate the distribution of the surrogate residuals.

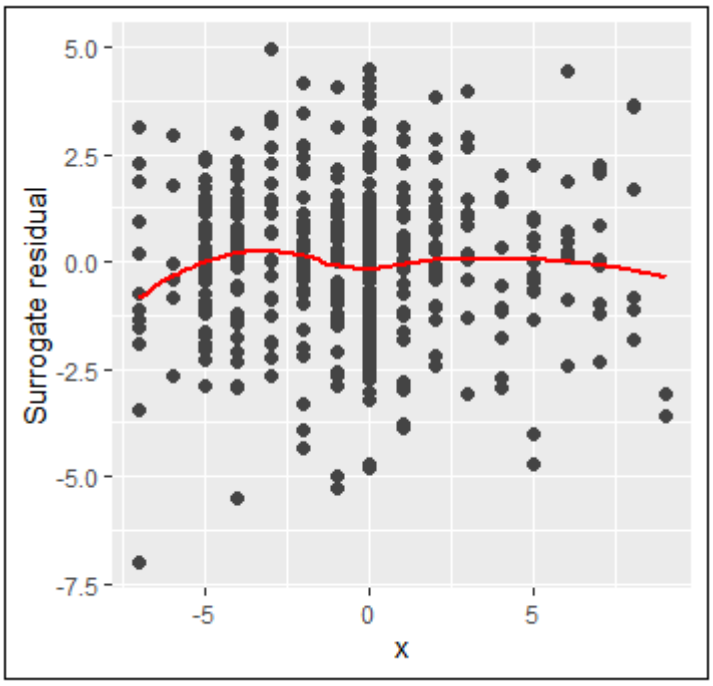

Figure 5b – Surrogate residuals for Model 2, x indicates the sum of the values of all explanatory variables in the model for residency disputes, points on the graph indicate the distribution of the surrogate residuals.

As can be seen from the tables, both models have a good fit in terms of residuals, since the average values for the residuals tend to be closer to zero.

The explanatory power of the models is determined with the help of the R-squared type measure and the multiple correlation measure, described in Section 2.7. Table 16 summarizes the results for the two models.

	Model 1	Model 2
Multiple correlation measure	0.838	0.797

Table 16 – Multiple correlation measure for the two models.

As can be seen, both models have high values for the multiple correlation measures, meaning that their predictive power is sufficiently high.

7. CONCLUSION

The main criteria in making decisions on a child's custody and residency in court is that the decision is not acting against the interests of the child, and judges also aim at clarifying both parents' abilities to take care of a child's mental, psychological and physical development. In the disputes investigated for this thesis, mothers have won more often than fathers, especially in the residency disputes, where the difference is statistically significant.

Initial custody and residency situation was statistically significant only for the residency disputes. A more favorable initial situation for fathers decreased the winning chances of fathers involved in custody disputes. At the same time, established conditions at one of the parents increased the probability of fathers to win, meaning that these two variables, being positively correlated, act in an opposite way. This could be explained by a correlation of initial custody and residency situation with other variables, such as established conditions, or the gender of the applicant.

Foreign backgrounds of parents turned out to be significant only in the custody models. A dispute where only father has a foreign background decreased a father's winning chance the most. Dispute with only mother having a foreign background or both parents having a foreign background did not have a statistically significant effect in the models, but Pere, Lilja, Sobolev (2017) have demonstrated that the proportion of father winners in these disputes was significantly smaller than in disputes where both parents were native Finns.

Various accusations parents have claimed on each other were significant in both models. For custody disputes, these were substantiated violence towards mother and a non-substantiated use of alcohol or drugs by father. Both accusations act against fathers, decreasing their winning chances, with the substantiated accusation having a greater negative effect. As for the model for residency disputes, the accusations, on the contrary, increased the winning probabilities of fathers. Of the three accusations, only one was substantiated – mother's mental disorder. The other two were accusations regarding mother having alcohol or drug problem and father's violence towards the child. The latter accusation proved to be substantiated in only one quarter of the disputes, which partially explains why it does not have a negative association with father's winning chances.

Established conditions at mother's place or at father's place were only significant for the residency disputes. A parent at whom the conditions were established had a higher chance to win, but according to the model, the impact of established conditions at father's place on father's winning

chances was higher than that of established conditions at mother's place on mother's probability to win.

According to the models for residency disputes, if father gets a legal aid, it significantly lowers his winning chances. If a mother gets legal aid, this is, however, not significant. If only mother hires a legal assistant (regardless of the type of assistant), this lowers the chances of fathers to win, which could be explained by a more professional approach by the legal assistant.

If a custody dispute was applied to the district court of Porvoo fathers had a higher chance to win. The number of disputes was not, however, sufficiently high to form a conclusion. At the same time, a female judge in one of the courts judged the most collected custody disputes. A dispute with this judge involved acted as a significant variable for the custody models, increasing fathers' winning chance.

In approximately 17 % of all the disputes, one of the parents appealed to the court based on the previous decision. If mother appealed, meaning that the father has won the previous dispute, father wins again with a higher probability, which goes along with the previous research.

Apart of the genders of the parents involved in disputes, no other genders were significant in any models. The potential variables were genders of children and genders of both judge and legal assistants (separate models were made for the disputes where both parents hired a legal assistant). No age variables turned out to be significant, either.

It is possible that due to a limited number of disputes collected for this thesis and due to certain variables being recorded, some of the important variables may be left not investigated. In addition, of the district courts involved in the research, only one (located in Porvoo) had a statistically significant role in the model, therefore this investigation can also be extended by collecting the cases from other years and from the district courts of other cities.

REFERENCES

Agresti A. 2010. *Analysis of ordinal categorical data*. Second edition. Wiley, New Jersey.

Christensen R. H. B. 2018. *Cumulative Link Models for Ordinal Regression with the R Package ordinal.* Technical University of Denmark & Christensen Statistics.

Davison A. C. 2003. *Statistical Models.* Cambridge University Press, New York.

Ervasti K. T. 2014. *Lasten kuuleminen huoltoriidoissa*. Kauppakamari, Helsinki.

Fox J. 2016. *Applied Regression Analysis and Generalized Linear Models.* Third edition. Sage Publications, California.

Greenwell B., McCarthy A. & Boehmke B. 2017. *sure: Surrogate Residuals for Ordinal and General Regression Models.* R-package version 0.2.0.

Greenwell B., McCarthy A., Boehmke B. & Liu D. Residuals and Diagnostics for Binary and Ordinal Regression Models: An Introduction to the sure Package. *The R Journal Vol. 10/1, July 2018, 381-394.*

Hautanen T. 2010. *Väkivalta ja huoltoriidat*. Tampereen yliopistopaino, Tampere.

Lilja E. 2018. Huolto- ja asumisriidat pääkaupunkiseudun käräjäoikeuksissa 2004-2013. Helsingin yliopisto, Helsinki.

Liu D. & Zhang H. 2018. Residuals and Diagnostics for Ordinal Regression Models: A Surrogate Approach. *Journal of the American Statistical Association, 113:522, 845-854.*

Montgomery D. C. & Runger G. C. 2014. *Applied Statistics and Probability for Engineers*. Third edition. Wiley, New York.

Palo-Repo M. 2015. Lasten huolto- ja asumisriidat Helsingin hovioikeudessa 2003 – 2006. Helsingin yliopisto, Helsinki.

Pere P. J., Lahti T., & Sutela M. 2018. Sentences and prosecutors' demands for aggravated drunk driving in Finland. *International Journal of Comparative and Applied Criminal Justice, 42*(2-3), 157-175.

Pere P., Lilja E. & Sobolev A. Maahanmuuttajien mielikuva naisten suosimisesta tuomioistuimissa. *Lakimies 115 (2017) : 1, 90-99.*

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (cited on 2.2.2020).

Suomen virallinen tilasto (SVT): Syntyneet [verkkojulkaisu]. Tilastokeskus, Helsinki. http://www.stat.fi/til/synt/tau.html (cited on 2.2.2020).

United Nations Development Programme, Human Development Reports, Human Development Index (HDI). http://hdr.undp.org/en/content/human-development-index-hdi (cited on 2.2.2020).

United Nations Development Programme, Human Development Reports, Global Development Indicators. http://hdr.undp.org/en/countries (cited on 2.2.2020).

Valkala E., Litmala M. 2006. Lasten huoltoriidat käräjäoikeuksissa. Oikeuspoliittinen tutkimuslaitos, Helsinki.

APPENDIX

Section 1 – Classifications of variables

This section contains a list of explanatory variables involved in the analysis with the explanation of possible values they were assigned. If a variable is an indicator variable, it gets a value 1 if a certain condition is fulfilled and a value 0 otherwise.

1. Established conditions at one of the parents. –1 (at mother), 0 (non-established), 1(at father).
2. Initial custody and residency situation:
   –5 = mother has a sole custody
   –4 = mother has a sole custody and father has a right to acquire the information about the child from the officials
   –3 = both parents are guardians, and the child is constantly living with mother
   –2 = both parents are guardians, but mother has an authority in deciding on certain aspects regarding the child (usually educational or health-related issues)
   –1 = both parents are guardians, child's books are at mother's home
   0 = both parents are guardians, child is living equal time with both of them and child's books are shared
   1 = both parents are guardians, child's books are at father's home
   2 = both parents are guardians, but the father has an authority in deciding on certain aspects regarding the child
   3 = both parents are guardians, and the child is constantly living with father
   4 = father has a sole custody and mother has a right to acquire the information about the child from the officials
   5 = father has a sole custody.
3. Mother's demand (same classification as for "Initial custody and residency situation")
4. Father's demand (same classification as for "Initial custody and residency situation")
5. Duration of the dispute consideration in court (measured in days from the date of application to the date of final decision by the court, averaged afterwards).
6. Location of the district court in the capital area (Helsinki, Espoo, Vantaa), indicator variable.
7. Age of the youngest child (in years, averaged).
8. Number of children involved in a dispute.
9. Father is an applicant, indicator variable.

10. Father's age (in years, averaged).
11. Difference between the ages of mother and father, calculated as father's age minus mother's age.
12. A judge is a male, indicator variable.
13. Only mother hires a legal assistant, indicator variable.
14. Only father hires a legal assistant, indicator variable.
15. Only mother hires attorney-at-law, indicator variable.
16. Only father hires attorney-at-law, indicator variable.
17. All the children involved in a dispute are boys, indicator variable.
18. District court of Espoo, indicator variable.
19. District court of Vantaa, indicator variable.
20. District court of Hyvinkää, indicator variable.
21. District court of Hämeenlinna, indicator variable.
22. District court of Lahti, indicator variable.
23. District court of Porvoo, indicator variable.
24. Only mother is of a foreign background, indicator variable.
25. Only father is of a foreign background, indicator variable.
26. Both parents are of a foreign background, indicator variable.
27. Mother gets legal aid, indicator variable.
28. Father gets legal aid, indicator variable.
29. Difference in the ages of the oldest and the youngest child, calculated as the age of the oldest child minus the age of the youngest child, gets value 0, if only one child is involved in a dispute.
30. An accusation regarding violence towards a child, –1 (accusation towards father), 0 (no accusation), 1 (accusation towards mother).
31. An accusation regarding violence towards the other parent, –1 (accusation towards father), 0 (no accusation), 1 (accusation towards mother).
32. An accusation regarding alcohol or drug abuse, –1 (accusation towards father), 0 (no accusation), 1 (accusation towards mother).
33. An accusation regarding mental disorder, –1 (accusation towards father), 0 (no accusation), 1 (accusation towards mother).
34. An accusation regarding preventing the child from seeing the other parent, –1 (accusation towards father), 0 (no accusation), 1 (accusation towards mother).

35. Substantiated accusation regarding violence towards a child, –1 (accusation towards father), 0 (no accusation), 1 (accusation towards mother).
36. Substantiated accusation regarding violence towards the other parent, –1 (accusation towards father), 0 (no accusation), 1 (accusation towards mother).
37. Substantiated accusation regarding alcohol or drug abuse, –1 (accusation towards father), 0 (no accusation), 1 (accusation towards mother).
38. Substantiated accusation regarding mental disorder, –1 (accusation towards father), 0 (no accusation), 1 (accusation towards mother).
39. Substantiated accusation regarding preventing the child from seeing the other parent, –1 (accusation towards father), 0 (no accusation), 1 (accusation towards mother).
40. One child is involved in a dispute, this child is a boy, indicator variable.
41. Time difference of the date of the first application in a dispute from the date of the earliest application of all the collected disputes, measured in total months.
42. Indicators for certain judges, who were involved in ten or more disputes, differs for custody and for residency disputes, indicator variables
43. Mother won the previous dispute, indicator variable.
44. Father won the previous dispute, indicator variable.
45. Mother has a sole custody initially, indicator variable.
46. Both parties have legal assistants, both legal assistants are female and a judge is female.
47. Both parties have legal assistants, both legal assistants are male and a judge is female.
48. Both parties have legal assistants, both legal assistants are female and a judge is male.
49. One of the parents appeals, only applicable for the cases with the previous decision by the court, –1 (mother appeals), 0 (no one appeals), 1 (father appeals).

Section 2 – Processing of collected data

3 388 cases were explored from the courts of Helsinki, Espoo, Vantaa, Tuusula, Porvoo, Hyvinkää, Lahti, Tammisaari and Hämeenlinna. These courts provided the documents on the following types of disputes: 2510 (Decision on a child's custody and a right to meet the other parent), 2520 (Changing the decision regarding a child's custody), 2610 (Divorce) and 2621 (Divorce with the attached application on a child's custody or a right to meet the other parent). Only 474 cases were chosen for the analysis. In order to merge these cases with the earlier collected cases by Lilja (2018), it was important that the same criteria for relevancy was used. The following are the mandatory conditions for each document to be treated as relevant:

1. The dispute is about custody or residency of a child (children).
2. Only mother and father of a child should be the participants of a dispute (applicant and respondent).
3. No agreement between the parents should take place, i.e. the dispute should be solved in court with a final decision carried out by a judge(s).
4. Both parties should participate in the dispute and provide their demands.

Below are the most typical cases collected from the court, for which at least one of the conditions stated above did not hold and which were therefore excluded from the analysis.

- The parents managed to come to an agreement during the process, so that the court's decision just confirmed their agreement (assuming the agreement is not acting against the best interests of a child).
- The applicant cancelled the application, after which the dispute was no longer investigated.
- The parents applied for divorce and either had no children at all or had agreed on their residency and custody. Such cases were collected under the code 2610 (Divorce).
- The respondent did not reply to the initial application of the applicant, in which case the court's decision follows the applicant's demand, if it is not considered to be against the interests of a child.
- One or both parties were not the parents of a child. The most frequent application was to assign custody to another relative in addition to one of the parents (for instance, mother as

the only guardian demanded that her close relative also becomes a guardian). The other parent was not involved in the dispute, e.g. due to death, serious illness, imprisonment or due to a loss of any contacts and connections with this parent. The same applied to the couples of the same gender, where one of the parties, being the only guardian, applied that his or her spouse becomes the other guardian.

- The application was made by the social welfare board. This typically happened when both parents of the child had died, there was no connection with them, or the social welfare board had taken the child away from the family, since living with parents is against the child's best interests, according to the social welfare board. When someone capable of becoming a child's guardian is found, the social welfare board makes an appropriate application to the court.
- The case initially included the parties' disputes over residency or custody and also financial support or the right of a child to meet the other parent. Then, the parents managed to come to a partial agreement, which included residency or custody issues, but the financial issues or the right of a child to meet the other parent had to be solved in court, and therefore the dispute became irrelevant for the analysis.

The cases were divided into two groups – residency disputes and custody disputes. A case was registered as a custody dispute only if it did not contain a dispute on residency of a child. This means that all cases where both custody and residency were disputed upon were regarded as residency disputes. Below is the list of the applications combinations, which did not include a dispute on the residency and were therefore classified as custody disputes. The first code in the bracket refers to the applicant's demand (explained in the previous section of the Appendix) and the second code refers to the respondent's demand.

(1, 2)

(1, 3)

(1, 4)

(2, 1)

(2, 3)

(2, 4)

(3, 1)

(3, 2)

(3, 4)

(4, 1)

(4, 2)

(4, 3)

(7, 8)

(7, 9)

(7, 10)

(8, 7)

(8, 9)

(8, 10)

(9, 7)

(9, 8)

(9, 10)

(10, 7)

(10, 8)

(10, 9)

All cases with other combinations of demands were classified as residency disputes.

Section 3 – The z-test for proportions

When analyzing whether applicant-mothers tend to win with a different proportion than applicant-fathers, the z-test for proportions is applied.

Let n_1 be the total number of applicant-mothers and n_2 the total number of applicant-fathers. Let the proportions of applicants who win the dispute be p_1 for mothers and p_2 for fathers. The null hypothesis states that the equal proportions of applicants win the disputes:

$H_0: p_1 = p_2$

$H_1: p_1 \neq p_2$ for a two-sided alternative hypothesis, or $p_1 > p_2$ ($p_1 < p_2$) for a one-sided alternative hypothesis.

It is worth noticing that the test requires the assumption that the two populations (mothers and fathers in this case) are independent. This should hold, since there is always one applicant per case and each case is assumed to be independent. The only exception could be a situation where two cases were collected where one is a continuation of the other, with the same parties involved. However, this was not the case for the collected data.

The next step is to calculate the z-statistics, with the following formula (Montgomery, Runger, 2014, 362):

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$.

Large values of z-statistics indicate a statistically significant difference between the proportions p_1 and p_2 and lead to rejection of the null hypothesis. *P*-value is an important indicator of the test significance; it tells the probability to get the same or a more extreme result given that the null hypothesis holds. The null hypothesis is rejected with too small *p*-values (usually less than 0.05 or less than 0.01). The calculation of *p*-value is performed with the help of statistical software.

The other test is performed when testing for the differences between two proportions within the same population. Let n be the total number of disputes, p_1 be the proportion of disputes where

mother is a winner, p_2 be the proportion of disputes where father is a winner. In this case, the two proportions are not independent, so previously described test cannot be applied. Instead, the z-test for one population proportion should be used. The idea is to take a larger (smaller) proportion of these two and test the null hypothesis that it is equal to 0.5 against the alternative hypothesis that it is larger (smaller) than 0.5:

H_0: $p_1 = 0.5$

H_1: $p_1 >$ (<) 0.5.

The z-statistics is:

$$\frac{p_1 - 0.5}{\sqrt{\frac{0.5\ (1-0.5)}{n}}},$$

where any other hypothesized value of p_1 than 0.5 could be used instead.