IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12989

Efficient Targeting in Childhood Interventions

Alexander Paul
Dorthe Bleses
Michael Rosholm

FEBRUARY 2020

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12989

Efficient Targeting in Childhood Interventions

Alexander Paul
Aarhus University and TrygFonden's Centre for Child Research

Dorthe Bleses
Aarhus University and TrygFonden's Centre for Child Research

Michael Rosholm
Aarhus University, TrygFonden's Centre for Child Research and IZA

FEBRUARY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Efficient Targeting in Childhood Interventions

Many targeted childhood interventions such as the Perry Preschool Project select eligible children based on a risk score. The variables entering the risk score and their corresponding weights are usually chosen ad hoc and are unlikely to be optimal. This paper develops a simple economic model and exploits Danish administrative data to address the issue of efficient targeting in childhood interventions. We define children to be in need of an intervention if they suffer from an socially undesirable outcome, such as criminal behavior, at around age 30. Because interventions are most effective very early in life, we then test if and to what extent indicators available at birth can predict the emergence of these outcomes. We find fair to good levels of prediction accuracy for many outcomes, especially educational attainment, criminal behavior, placement in foster care as well as combinations of these outcomes. Logistic regressions perform as well as other machine learning methods. A parsimonious set of indicators consisting of sex, parental education and parental income predicts almost as accurately as using the full set of predictors. Finally, we derive optimal weights for the construction of risk scores. Unlike the ad hoc weights used in typical childhood interventions, we find that optimal weights vary with the outcome of interest, differ between father and mother for the same predictor and should be disproportionately large when parents are at the bottom of the education and income distribution.

JEL Classification: I18, I28, I38

Keywords: targeting, early childhood intervention, machine learning

Corresponding author:
Michael Rosholm
TrygFonden's Centre for Child Research
Fuglesangs Allé 4
building 2621, 3c
8210 Aarhus V
Denmark
E-mail: rom@econ.au.dk

1 Introduction

Many childhood interventions such as the famous Perry Preschool Project (Weikart 1967) target a selected subset of children rather then applying the intervention universally to the whole cohort. One rationale behind such targeting in childhood interventions is the presence of a budget constraint. Even more important is the idea that only certain children promise a sufficiently large return to justify the intervention's cost. Children who already enjoy a beneficial environment in the absence of the intervention would benefit little or might even be harmed.[1] Although the theoretical motivation for targeting interventions is evident, an important question in practice is *which* children should be targeted. Scholars have thus called for the development of "measures of risky family environments . . . that facilitate efficient targeting" (Heckman 2008, p. 314).

In this article, we approach the problem of efficient targeting from a long-run perspective. We define children as disadvantaged (in need of an intervention) if they are at risk of experiencing undesirable outcomes later in life, thus becoming burdensome to society. Specifically, we define social burden as the presence of adverse outcomes such as criminal behavior or high health care use in individuals wheh they are around 30 years old. Children at risk of such outcomes are likely to gain the most from an early intervention. We then exploit rich register data from Denmark to predict disadvantage, using *only* data available to us at the child's birth. This constraint is motivated by recent literature in human development showing that childhood intervention programs work best if administered very early in life (Heckman 2006; Cunha et al. 2010; Allen 2011). Ideally, children are targeted immediately after birth, as in the prominent Carolina Abecedarian Project (Ramey and And Others 1974).[2]

Typical childhood interventions target children by means of a risk score. Only children that score sufficiently high are eligible for participation. A regular ingredient in defining the risk score is a measure of the family's socioeconomic status (SES), such as household income or parental education, but other variables may enter as well.[3] When combining all the indicators into a one-dimensional score, most interventions place a larger weight on family SES, but they generally make ad-hoc choices about the weights given to each component. As an example, the Carolina Abecedarian Project (p. 65) constructed a "high risk

1. Cornelissen et al. (2018), for example, find that children with immigrant ancestry benefit most from attending child care because their alternative care arrangements are of relatively low quality. Similarly, Havnes and Mogstad (2015) study the effect of child care attendance on earnings in adulthood and show that children from low-income families gain substantially. Children from upper-class families, in contrast, suffer a loss in earnings.

2. Heckman (2012) summarizes this literature: "The highest rate of return in early childhood development comes from investing as early as possible, from birth through age five, in disadvantaged families."

3. The Perry Preschool Project targeted African-American children with low IQ from families that performed poorly on a cultural deprivation scale based on paternal occupation, parental education and density in the home (persons per room) (Weikart 1967). The Carolina Abecedarian Project constructed a high risk index based on parental education and income and additional 10 minor indicators (Ramey and And Others 1974). The Early Training Project considered housing, parental education and parental occupation (Klaus and Gray 1968). Eligibility for Head Start is mainly determined by parental income (U.S. Department of Health & Human Services 2019)

index" that increased by one unit for each year missing from 12 years of parental schooling (separately for both the father and mother). It also increased by 4 units if family income fell short of 5,000 dollars and by one additional unit for each further 1,000-dollar reduction.

There is no indication that such ad hoc weights in constructing risk scores, as applied in many childhood interventions, are optimal in any way. On the contrary, it appears that risk scores can be substantially improved by answering the following basic questions: First, what relative weight should each indicator optimally receive? Does parental schooling matter more than income or vice versa? Second, do paternal and maternal characteristics matter in the same way, e.g. with respect to years of schooling? Third, is the relationship of the outcomes with the predictors non-linear and are there interactions among predictors? E.g., is each additional year of parental schooling equally important?

This study takes an econometric approach to address the problem of optimally selecting and weighting early indicators of disadvantage. We start by performing standard logistic regressions to predict long-run outcomes using predictors measured at birth. Logistic regression has the advantage that it allows for easy computation and interpretation of risk scores, but it is not necessarily best at prediction. We therefore also apply more sophisticated machine learning techniques that are known for good predictive power, in part because they implicitly allow for arbitrary interactions among predictors. We include indicators of family SES (income, education and occupation) and several other parental variables such as hours of work, health status and criminal activity as predictors. Some of these variables potentially correlate with quality time investments, which play an important role in human capital formation (e.g., Del Boca et al. 2014). At the individual level, we include sex, nationality and birth order. We examine whether and which of the individual and parental variables can accurately predict adverse outcomes in adulthood and derive optimal weights for the formation of composite risk scores.

The outcomes we consider are meant to capture the economic cost of different social dimensions, ranging from education and labor market outcomes to health and crime. The cost associated with these outcomes is not spread evenly across all members of society but can vary substantially from one person to another. Indeed, it has been shown that a relatively small fraction of the population accounts for a sizable share of the total economic burden (Caspi et al. 2017; Richmond-Rakerd et al. 2020). In line with this observation, we rank individuals by the outcome-specific cost they generate in adulthood and define the top 20% of the distribution as "at-risk" of the particular outcome. In the case of social benefits, for instance, we define the top 20% recipients in our sample, who account for 76% of total benefit receipt, as "at-risk". For 0-1 outcomes, such as having only compulsory schooling, we simply take the fraction having this outcome, which might be less than 20%. We aim to predict which children are at risk and can potentially be targeted by an intervention. We also predict which individuals will experience combinations of multiple of these outcomes.

We develop a simple theoretical model that demonstrates how prediction can help the policy-maker to assign treatment in a welfare-improving way. We focus on at-risk children as defined above and disregard potential benefits of the treatment to other children. We assume that treatment has a positive and homogeneous effect on at-risk children and does not harm children that are falsely identified as at-risk. In the first stage of the model, the policy-maker faces a budget constraint that limits the fraction of the cohort that can receive the intervention. Given this constraint, the policymaker chooses which children should receive treatment to maximize welfare. We show that this results in maximizing the number of correctly identified at-risk children out of all at-risk children (the true positive rate, TPR), as previously shown by Sansone (2019) using a similar model.

In the second stage of the model, the policy-maker chooses the fraction of the cohort that should receive the intervention. If the policy-maker is completely uninformed about which children are in need of the intervention, marginal improvements in the TPR will be constant irrespective of the fraction of the cohort receiving the intervention. The optimal decision will then be to administer the intervention to either all or none of the cohort members. On the contrary, if prediction is of any value, the improvements in the TPR will be high for small fractions of targeted children and then diminish gradually. This means that prediction can enable the policy-maker to enhance welfare by moving away from these corner solutions, that is by administering the intervention only to a selected fraction of the cohort.

In our prediction exercise, we follow the theoretical motivation above and aim to maximize the TPR under various assumptions about the fraction of a cohort that policy-makers can possibly target. Specifically, we first perform logistic regression or use other prediction methods to generate predicted probabilities of having a particular outcome. We then categorize those children as at-risk who have the highest predicted probability of experiencing the outcome and collectively add up to the fraction of the cohort to be targeted. We also report the so-called area under the curve (AUC) as a summary measure of predictive accuracy across all possible values of targeted fractions.

We find that, first, predictions using register data available at birth are possible and often yield fair to good levels of prediction accuracy. Predictions are most accurate for educational attainment, criminal behavior, placement in foster care and combinations of these outcomes, but are less accurate for health-related outcomes. If the decision-maker wants to target a fixed fraction of the cohort, for example due to a budget constraint, then informed treatment assignment based on predictions will always yield higher welfare than random, uninformed treatment assignment. If the fraction to be treated is instead a choice variable, then informed treatment assignment will improve welfare if it helps the decision-maker move away from assigning treatment to all or no members of the cohort.

Second, we find that logistic regression performs well. Predictions generated by other machine learning methods are generally neither statistically nor economically significantly different from logistic regression. This suggests that interactions among predictors, which some of the machine learning methods flexibly allow for, play a negligible role in predicting outcomes. Moreover, we find that updating the predictors with data from a few years after birth improves predictions very little. It seems as if further improvements are only possible with indicators of the child's behavior and skills, which are much more costly to obtain than the variables generally available from register data.

Third, we find that individual-level predictors (nationality, birth month and birth order) have little predictive power. An exception is sex, reflecting the strong gender bias in some outcomes, especially criminal behavior. In contrast, indicators of parental SES are highly predictive. We find that a parsimonious set of indicators consisting of sex, parental education and income yields predictions that are almost as accurate as using the full set of predictors. Knowledge of an individual's sex and a few variables related to socio-economic background may therefore be sufficient for effectively targeting children in childhood interventions. Many childhood interventions typically include measures of parental SES as a key ingredient in the construction of risk scores. Our study provides support for this practice.

Finally, we derive optimal weights for the formation of risk scores and find that they deviate in important ways from the ad-hoc weights conventionally used in childhood interventions. First, while parental income and education tend to contribute equally to the risk score, there are some differences depending on the outcome of interest: Education influences the risk score more than income when predicting education and hospitalization, but the opposite holds true for predicting psychiatric condition and income. Second, maternal and paternal education affect the risk score in a similar manner, but once again the outcome of interest matters. Maternal income affects the risk score little if she is in the upper 80% of the income distribution, while the relationship with income increases monotonously for fathers. Third, being at the bottom of the education and income distribution substantially raises the risk score. For education, this is equivalent to the parents having only compulsory or only vocational education. For income, these are the bottom 30% of fathers and the bottom 20% of mothers. Ignoring these non-linear relationships when forming risk scores, like in the Perry Preschool Project and the Carolina Abecedarian Project, means that risk is underestimated for children with parents at the bottom of the distribution.

So far, we have been deliberately unspecific about the nature of the intervention that our risk scores can be used for. This is because an intervention can take many forms depending on the context. At the broadest level, the intervention could simply consist of the provision of free or subsidized public child care to parents. A related type of intervention are center-based programs such as the Perry Preschool Project or the Carolina Abecedarian Project that go beyond basic child care by closely involving parents or offering health care. Finally, in the most narrow sense interventions can also be thought of as specific programs aimed at improving the learning environment within the daycare center (e.g., Bleses, Højen, Dale, et al. 2018; Bleses, Højen, Justice, et al. 2018) or at home (e.g., Andersen and Nielsen 2016).

The setting of our study is the Danish welfare state. This has two implications. First, in Denmark all mothers are entitled to free pre- and postnatal care which includes midwife consultations, GP visits, a postpartum hospital stay and home visits from trained nurses (Kronborg et al. 2016). In addition, high public subsidies for child care have led to enrollment rates that are among the highest in Europe: 72% of under-three year olds and more than 95% of three-to-five year olds attend child care (Commission/EACEA/Eurydice 2019). The nearly universal use of perinatal care and child care means that interventions in Denmark are better thought of as focused programs rather than the provision of care per se. It also means that it is realistic for policy-makers to target children found to be at risk based on register data: interventions can easily be implemented by trained nurses during home visits or by caretakers in nurseries and kindergartens. Second, the Danish welfare state already exhibits a high overall spending level on children in childcare and school. This can be expected to alleviate the predictive power of circumstances before or at birth. The predictive patterns uncovered in this study are likely to be even more pronounced in countries with a less generous welfare state. At the same time, the observation that the welfare state has not reduced these disadvantages transmitted from parent to child also suggests the need for more targeting in the institutions of the welfare state.

A study closely related to ours is Caspi et al. (2017). The authors find that a small set of predictors consisting of SES, maltreatment indicator, IQ and self-control could accurately predict adverse outcomes for 1,037 New Zealanders at age 38. Predicting whether individuals experience combinations of multiple adverse outcomes works particularly well. A limitation of their study is that predictors are recorded up until age 11, which is too late for effective early interventions. In addition, obtaining measures of IQ or self-control for the whole population is relatively costly. Our paper improves on Caspi et al. (2017) in that, first, we only use indicators that are inexpensive to measure and available from the Danish registers and, second, we focus on indicators available at birth. As we do here, Chittleborough et al. (2016) use only predictors from around birth. However, they study outcomes at age 5 (before schooling starts), thus missing substantial information on social burden that only a long-run perspective can take into account.

This paper also relates to several other strands of the literature. First, we use machine-learning techniques to predict which children would benefit most from an intervention. A growing number of studies address similar "prediction policy problems" (Kleinberg et al. 2015) in various contexts, e.g. regional allocation of refugees (Bansak et al. 2018), shootings among at-risk youth (Chandler et al. 2011), food-safety inspections (Glaeser et al. 2016), hip and knee replacements (Kleinberg et al. 2015) and judicial bail-or-release decisions (Kleinberg et al. 2017). Second, our paper loosely relates to the theoretical literature on optimal treatment assignment (Bhattacharya and Dupas 2012; Kitagawa and Tetenov 2018; Manski 2004). This literature typically uses experiments or observational studies to estimate covariate-specific heterogeneous treatment effects based on which optimal treatment assignment rules are derived. Our study differs from this, however, since we do not observe treatment effects associated with a particular intervention. Instead, we suggest that treatment should be assigned to individuals who are at risk of an adverse outcome and who thus have the potential to benefit from an appropriately designed intervention.

Finally, our paper adds to the discussion of targeted versus universal programs in that it weakens a typical argument against targeted programs. Targeted programs are a response to limited resources, which is particularly important in the context of the Scandinavian welfare state that continually struggles with the Baumol cost disease: Since increases in productivity are lower in the public than in the private sector (due to its larger share of labor in production), but wages in the public and private sector increase at the same rate (due to, e.g., institutional arrangements and unions' bargaining power), the welfare state will eventually be faced with the problem that the level of services offered cannot be sustained indefinitely. Better targeting may provide an (albeit temporary) solution to this problem. Moreover, targeting may avoid potentially negative effects on subgroups of the population (e.g., Havnes and Mogstad 2015; Cornelissen et al. 2018). At the same time, targeted programs might lead to stigmatization and are less effective when disadvantaged children are hard to identify. We demonstrate that the latter argument against targeted programs is unlikely to hold up in practice because meaningful indicators of disadvantage can be constructed for a wide range of adverse outcomes, through which efficient targeting becomes possible.

This paper is structured as follows. Section 2 presents a simple theoretical model that motivates our analysis. Section 3 deals with the practical aspects of prediction, including the data, the transition from theory to estimation and the specific prediction methods. Section 4 reports the results. Section 5 discusses our findings and concludes.

2 Theory: The policy-maker's problem

2.1 Model setup

Individuals may develop an adverse outcome $Y \in \{0, 1\}$. Fraction α of a cohort is at risk $R \in \{0, 1\}$ of developing Y: $Pr(R = 1) = \alpha$. Individuals not at risk ($R = 0$) will not develop the adverse outcome: $Pr(Y = 1 \mid R = 0) = 0$. An intervention $T \in \{0, 1\}$ targeted at at-risk individuals can prevent Y. In the absence of the intervention, $T = 0$, at-risk individuals are certain to develop the adverse outcome: $Pr(Y = 1 \mid R = 1, T = 0) = 1$. In the presence of the intervention, $T = 1$, the probability of developing the outcome decreases for at-risk individuals. The size of the reduction may depend on $\theta \in \mathbb{R}$, the susceptibility to treatment: $Pr(Y = 1 \mid R = 1, T = 1, \theta) = 1 - \delta(\theta)$ with $\delta(\theta) \in (0, 1]$ and $\delta'(\theta) > 0$. Individuals not at risk are not affected, in particular not harmed, by the intervention. The assumption of zero effects for individuals not at risk appears sensible for outcomes such as ever being criminally charged. It is less plausible for outcomes such as belonging to the top 20% of hospitalized individuals. Treatment effects could also be present below the top, and perhaps even be larger if those at the top are sick for reasons that are not amenable to intervention (e.g., genetic disposition).

The policy-maker's goal is to maximize the social welfare gain from administering a costly intervention that reduces the prevalence of the adverse outcome. The social welfare gain is the expected benefit from avoiding the cost of the adverse outcome minus the cost of administering the intervention. Let β be the fraction of the cohort that receives the intervention. The policy-maker must decide how many members (β) and which members (T) of the cohort should receive the intervention. The policymaker bases his decision on observed characteristics X with support $\mathbb{X}$. T is then a map $T : \mathbb{X} \rightarrow \{0,1\}$ such that $Pr(T = 1) = E(T(X)) = \int T(x) dF_X dx = \beta$. Formally, the problem can be written as follows:

$$\underset{T,\beta}{\text{maximize}} \quad B(T,\beta) - C(\beta),$$

where $B(T,\beta)$ is the expected benefit of the intervention and $C(\beta)$ is its cost. Specifically, the expected benefit of the intervention is the per-person cost of the outcome $Cost_{Outcome}$ (e.g., costs associated with crime or sickness), multiplied with the reduction in the probability that the outcome materializes. This reduction in probability depends on β: $Pr(Y = 1 \mid \beta = 0) - Pr(Y = 1 \mid \beta)$. The cost of administering the intervention is the fraction of the cohort that receives the intervention, β, multiplied with the cost of the intervention per child $Cost_{Intervention}$. Here, we implicitly assume that the marginal per-person cost of the intervention is not only independent of who is treated, but also independent of β. This assumption appears reasonable, even if it holds only true approximately since in practice interventions are typically administered at the classroom, school or day care level.

We can thus rewrite the problem as follows:

$$(1) \qquad \underset{T,\beta}{\text{maximize}} \quad \big(Pr(Y = 1 \mid \beta = 0) - Pr(Y = 1 \mid \beta)\big) Cost_{Outcome} - \beta Cost_{Intervention}.$$

We will divide the policy-maker's decision-problem into two stages. This helps build intuition for the optimal solution and also prepares for the empirical section later on. In the first stage, for a given value of β, i.e. the fraction of the cohort that receives the intervention, the policy-maker must decide which members of the cohort should receive it. In the second stage, he or she must decide about the optimal level of β. Choosing β is not trivial. Marginal reductions in the probability of the adverse outcome are high for low values β, but decrease as we progressively raise β. It is plausible that the optimal value of β lies strictly between 0 and 1, meaning that neither all nor none of the members of the cohort should receive the intervention.

2.2 First-stage problem: Choosing cohort members that should receive the intervention

In the first stage of the decision problem, the policy-maker takes the fraction of the cohort to receive the intervention as given and chooses which members should receive treatment. Problem 1 becomes:

$$\underset{T|\beta}{\text{maximize}} \quad \big(Pr(Y=1 \mid \beta=0) - Pr(Y=1 \mid \beta)\big)Cost_{Outcome} - \beta Cost_{Intervention}.$$

We can gain additional insight by reformulating this problem. First, note that by assumption the cost of the intervention, $\beta Cost_{Intervention}$, is independent of who receives the treatment. In other words, for a fixed fraction of the cohort receiving the treatment, the problem simplifies to maximizing the expected benefit of the intervention: $B(T,\beta) \equiv \big(Pr(Y=1 \mid \beta=0) - Pr(Y=1 \mid \beta)\big)Cost_{Outcome}$.

We derive in section A in the appendix that the expected benefit can be expressed as:

$$B(T,\beta) = \alpha Pr(T=1 \mid R=1,\beta)\bar{\delta}Cost_{Outcome}, \tag{2}$$

where $\bar{\delta} = \delta(\theta)$ is the reduction in the probability of developing the adverse outcome, which we assume to be homogeneous across at-risk children because susceptibility to treatment is typically unknown. Intuitively, the expected benefit of the intervention is the number of identified and targeted at-risk children (as a share of the whole cohort), $\alpha Pr(T=1 \mid R=1,\beta)$, multiplied with the reduction in the probability of the outcome caused by the intervention, $\bar{\delta}$, multiplied with the cost of the outcome that is avoided, $Cost_{Outcome}$. Since all other elements of Equation 2 do not depend on T, the problem simplifies to maximizing the probability of treatment for at-risk children, $Pr(T=1 \mid R=1,\beta)$. For any given β one obtains a maximum expected benefit that we denote by $B^*(\beta)$.

The term $Pr(T=1 \mid R=1,\beta)$ is also called true positive rate (TPR): the share of correctly identified positive instances out of all positive instances. In our setting, it is the share of correctly identified at-risk individuals out of all at-risk individuals. For a given fraction β of individuals that are targeted by the intervention, maximizing the welfare gain of the intervention is thus achieved by maximizing the TPR. This result is intuitive: We have assumed that the intervention does not do any harm, so that we need not be concerned about false positives, i.e. individuals not at risk falsely identified as at risk (type I error). The goal thus becomes to minimize the number of false negatives, i.e. individuals at risk falsely identified as not at risk (type II error), or equivalently to maximize the number of true positives.

2.3 Second-stage problem: Choosing the fraction of the cohort to receive the intervention

In the second stage of the decision problem, the policy-maker must choose the fraction of the cohort that should receive the intervention. Assume that the marginal expected benefit is larger than the marginal cost at $\beta = 0$. Then the policy-maker keeps increasing β as long as the marginal expected benefit of the intervention remains above or at its marginal cost. The marginal expected benefit and the marginal cost of the intervention are:

$$\begin{aligned}
\text{Marginal expected benefit: } \frac{\partial B^*(\beta)}{\partial \beta} &= \alpha \frac{\partial Pr(T=1 \mid R=1,\beta)}{\partial \beta} \bar{\delta} Cost_{Outcome} \\
&\equiv \alpha \frac{\partial TPR}{\partial \beta} \bar{\delta} Cost_{Outcome} \\
\text{Marginal cost: } \frac{\partial C(\beta)}{\partial \beta} &= \frac{\partial \beta Cost_{Intervention}}{\partial \beta} = Cost_{Intervention}.
\end{aligned}$$

If the policy-maker is uninformed about which members of the cohort are at risk, the treatment must be assigned at random. In this case, the probability of treatment will be independent of the individual risk status such that $Pr(T = 1 \mid R = 1, \beta) = Pr(T = 1 \mid R = 0, \beta) = \beta$. The marginal expected benefit will thus be constant: $\partial B^*(\beta)/\partial \beta = \alpha \bar{\delta} Cost_{Outcome}$. This situation gives rise to extreme optimal solutions: If $\alpha \bar{\delta} Cost_{Outcome} \geq Cost_{Intervention}$, then all members of the cohort should receive the intervention ($\beta = 1$). If $\alpha \bar{\delta} Cost_{Outcome} < Cost_{Intervention}$, nobody should receive it ($\beta = 0$). Figure 1 illustrates the two cases, both of which feature a horizontal marginal expected benefit curve. In the former case, the marginal expected benefit curve lies above the marginal cost curve, while in the latter case the marginal expected benefit curve lies below the marginal cost curve.

Figure 1 also illustrates the more realistic scenario in which the policy-maker is at least somewhat informed by receiving a signal about which cohort members are at risk and would thus benefit from the intervention. In this case, the signal allows the policy-maker to assign treatment disproportionately to at-risk children. For low values of β, the marginal expected benefit of the intervention is therefore higher than under uninformed, random assignment, and shrinks as β increases. Compared to the uninformed situation with random assignment, the optimal value of β is more likely to lie between 0 and 1. Indeed, an intervention that is deemed too ineffective in the uninformed case might turn out worthwhile if administered to a positive fraction of the cohort in the informed case. Similarly, the social welfare gain of an intervention that is administered to the whole cohort in the uninformed case can be increased by restricting the intervention to a smaller fraction of the cohort in the informed case.

3 Prediction

3.1 Data

3.1.1 Sample

Our sample consists of the full Danish birth cohorts from the years 1985, 1986 and 1987. This choice is motivated by the fact that most of the Danish register data start in 1980, so a certain period after this year is required to construct powerful predictors, such as parental crime or parental hospital admissions. At the same time, individuals should not be born too recently because otherwise we would not be able to observe socially relevant adult outcomes today, such as educational attainment or disposable income. Because we use parental predictors, we also impose the condition that both parents are known to us and have lived in Denmark continuously since 1980. The final sample contains 149,755 individuals.

3.1.2 Outcomes

The age at which we measure the individuals' outcomes ranges from 28 to 32 years, depending on the availability of most recent data. We focus on seven outcomes that capture the economic cost of different societal dimensions, ranging from education, labor market outcomes, health to crime. Reflecting the principle that a small fraction of the population accounts for a disproportionate share of the total economic burden (Caspi et al. 2017; Richmond-Rakerd et al. 2020), we order individuals by the outcome-specific cost they generate in adulthood and define the top 20% of the distribution as "at-risk" of this outcome. For binary outcomes (0-1), such as having only compulsory schooling, we simply take the fraction having this outcome, which might be less than 20%. Table 1 provides an overview of the included outcomes (see also Table D.1 in the appendix for additional details). For binary outcomes, the fraction of the total burden accounted for by at-risk individuals is equal to 100%. For non-binary outcomes, the table shows that the top 20% social benefit recipients in our sample account for 76% of total benefit receipt. Similarly, the top 20% patients with the most hospital admissions account for 55% of all admissions. For income, which is a benefit rather than a burden to society, the pattern is reversed: the bottom 20% of the income distribution receive a disproportionately small share equal to 9% of total income.[4]

Table 2 shows that the various outcomes are not independently distributed across individuals but instead highly correlated. A consequence is that a disproportionately large share of the cohort will have combinations of several outcomes. By the same token, disproportion-

4. A particularly relevant study in this context is by Richmond-Rakerd et al. (2020), who also use Danish administrative data. They show that adult health, crime and social welfare are unequally distributed across people and correlated within people. Our definition of outcomes varies slightly from theirs: we consider the number of hospitalizations rather than the number of hospital days and criminal charges rather than criminal convictions. Richmond-Rakerd et al. (2020) do not attempt to predict outcomes with information available at birth.

ately many individuals will be entirely free of any adverse outcome. This is illustrated in Figure 2, which juxtaposes the actual distribution of the number of negative outcomes with a simulated one based on the assumption that outcomes were uncorrelated. Compared to the simulated distribution, a disproportionately large share of individuals have no adverse outcome (44% vs. 27%) and disproportionately many individuals have combinations of four or more outcomes.

Figure 3 reveals that individuals with combinations of outcomes not only account for a disproportionate share of the population but also for a disproportionate share of the total economic burden. In fact, this share becomes increasingly disproportionate as the number of adverse outcomes rises. Therefore, in addition to predicting each outcome individually, we also aim to predict if individuals are members of the groups with combinations of outcomes. Specifically, we search for a predictive algorithm that can distinguish individuals with 3+ outcomes versus those with 2 or fewer (and similarly for 4+ and 5+ outcomes).

Moreover, we also predict having 3+ outcomes as opposed to zero outcomes. This exercise implicitly assumes that we know a priori that an individual will end up with either 0 or 3+ outcomes. While this assumption is unreasonable and the resulting predictions therefore only of limited practical relevance, we include the results for the purpose of comparison with Caspi et al. (2017).

Finally, in addition to counting the number of adverse outcomes, we also construct a social burden (SB) indicator through confirmatory factor analysis assuming that a single factor underlies the seven individual outcomes. See Table D.2 for additional information.

3.1.3 Predictors

The predictors included at the level of the individual are sex, birth order and nationality. Predictors at the level of the parents are recorded separately for the mother and the father. They include: educational attainment, income, labor market status, weekly hours of work, marital status in the year before birth, criminal charges, placements, age, hospital admissions. All continuous predictors are turned into discrete categorical variables. Missing values are generally assigned to a separate category. We include a dummy for each category (except for the baseline category) in the prediction analysis. See Tables D.3 and D.4 in the appendix for more details and summary statistics.

All of the individual and parental predictors are well-known to be associated with outcomes in adulthood. Here, we will focus on evidence from Scandinavia, i.e. Denmark, Sweden, and Norway. Regarding birth order, there is evidence that first-borns have higher educational attainment than late-borns (Black et al. 2005) and that first-borns are less likely to enter the criminal justice system than late-borns (Breining et al. 2020). The relationship with health is more mixed (Black et al. 2016). As for gender differences, it is well-known that women face an earnings gap on the labor market (Gallen et al. 2019; Kleven et al. 2019), while men account for the vast majority of people serving time in prison (Kriminalvården 2018). Lastly, several studies provide evidence of labor market discrimination with respect to ethnicity in hiring and earnings (Carlsson and Rooth 2007; Rooth 2010; Arai and Skogman Thoursie 2009).

We include predictors at the level of the parents because previous literature has shown that a wide range of socio-economic characteristics and behaviors appear to be passed on from parents to children (see Black and Devereux, 2011, for an overview). The mechanism underlying these intergenerational correlations is not always clear. Parental characteristics such as education might causally affect children's characteristics, but it could also be that both are determined by a third factor, e.g. genes. Intergenerational correlations are not only observed within the same characteristic, but also across different characteristics. Whatever the actual mechanism explaining these correlations may be, what matters for our setting is that parental information can be used to predict children's outcomes.

All of the outcomes considered in this paper have been shown to correlate with parental characteristics in previous studies. Once again focusing on evidence from Scandinavia, intergenerational correlations have been established for education (Andrade and Thomsen 2018; Landersø and Heckman 2017; Hertz et al. 2008), income (Chetty et al. 2014; Landersø and Heckman 2017), welfare receipt (Dahl et al. 2014; Bratberg et al. 2015), crime (Hjalmarsson and Lindquist 2012, 2013), out-of-home care (Mertz and Andersen 2017; Wall-Wieler et al. 2018), and health (Andersen 2019; Björkegren et al. 2019). These studies show intergenerational correlations within the same characteristic, for example between parental income and the child's income. It is clear that the strong correlation among outcomes (see previous subsection) will also give rise to significant cross-correlations, for example between parental education and the child's income.

In the main analysis, all predictors are measured at or before birth. In additional analyses, we extend the time frame to include the first few years after birth. This allows us to update and strengthen parental predictors with more recent data and to include the child's hospitalizations as an additional predictor.

3.2 From theory to estimation

Conducting a benefit-cost analysis as theoretically derived in Section 2 would require information that is not available to us. Specifically, the cost of the outcome and the intervention as well as the impact of the intervention are context-dependent and typically challenging to estimate. For this reason, we do not attempt to determine the optimal fraction of the cohort β that receives the intervention.

However, we can and do examine whether there are treatment assignment mechanisms that improve upon random assignment and thus generate the downward-sloping marginal expected benefit curve in Figure 1. If so, this would be evidence that data available at birth can be gainfully utilized to increase social welfare by administering an intervention to a selected fraction of the cohort. Moreover, we can examine which outcomes can be predicted most accurately and compare the quality of the treatment assignment across several machine learning techniques and sets of predictors.

To determine the shape of the marginal expected benefit curve, we follow the two-stage procedure developed in the previous section. First, for a given value of β, we maximize the expected benefit of the intervention by finding the treatment assignment mechanism that maximizes the true positive rate (TPR). Next, after repeating this step for a grid of β-values, we can study how the TPR changes with β and thus determine the shape of the marginal expected benefit curve.

How do machine learning techniques inform us about the optimal treatment assignment that maximizes the TPR? Using Bayes' rule, we can rewrite

$$\begin{aligned} TPR = Pr(T = 1 \mid R = 1, \beta) &= \frac{Pr(R = 1 \mid T = 1, \beta) Pr(T = 1)}{Pr(R = 1)} \\ &= \frac{Pr(R = 1 \mid T = 1, \beta)\beta}{\alpha}, \end{aligned}$$

where we replaced $Pr(T = 1) = \beta$ and $Pr(R = 1) = \alpha$. We thus get the following expression for the expected benefit:

$$B(T, \beta) = Pr(R = 1 \mid T = 1, \beta)\beta\bar{\delta}Cost_{Outcome}.$$

Hence, for any given β, the children that receive the intervention should be those with the highest probability of being at risk $Pr(R = 1 \mid T = 1, \beta)$. This probability is generally unknown, but can be estimated using appropriate machine learning techniques based on a set of observable characteristics X, yielding estimated probabilities $Pr(R = 1 \mid X)$. Optimal treatment assignment for a given β thus implies administering the intervention to the fraction β of individuals with the largest estimates of $Pr(R = 1 \mid X)$. Treatment T is thus a function $T(X)$ of observable characteristics X as defined in Section 2.

To assess quantitatively how much informed treatment assignment improves upon uninformed treatment assignment, we also derive the widely used receiver operating curve (ROC). The ROC plots the true positive rate (TPR) against the so-called false positive rate (FPR), which is the share of falsely identified positive instances out of all negative instances. Because β is the fraction of individuals identified as positive, both TPR and FPR increase with β. To construct the ROC plot, one can readily derive the FPR as a function of β and its associated TPR (see section B in the appendix).

The area under the ROC provides a summary measure of predictive accuracy (AUC = area under the curve). Intuitively, it measures the probability that a specific machine learning algorithm assigns a higher probability $Pr(R = 1 \mid X)$ to a randomly chosen at-risk child than to a randomly chosen not-at-risk child (Fawcett 2006). The AUC ranges from 0.5 in the case of uninformed treatment assignment up to 1.0 in the case of perfectly informed treatment assignment. Although any value larger than 0.5 – indicating informed treatment assignment – may entail social welfare gains, in practice values are often interpreted as follows (Caspi et al. 2017): worthless (0.5–0.6), poor (0.6–0.7), fair (0.7–0.8), good (0.8–0.9), excellent (0.9–1.0).

3.3 Methods

Predicting the realization of binary outcome variables, as in this study, is known as a classification problem. We start addressing this problem with standard logistic regressions. One advantage of logistic regression is that predictors variables are assumed to combine linearly to form the risk score for each individual. In combination with the exclusive use of dummy variables, the risk score thus becomes easy to construct and to interpret. A disadvantage of logistic regression is that interactions among predictors must be explicitly specified. Another disadvantage is its implicit risk of overfitting and poor out-of-sample performance. Overfitting not only occurs if the number of regressors is large relative to the number of observations, which is not the case in our study, but also if an outcome is so rare that a few values of the predictors spuriously explain it.

To address the issue of overfitting, we split the dataset into a 80% training dataset, on which we perform estimations, and a 20% test data set, on which we evaluate the model fit in terms of TPR and AUC. This guards against overestimating model fit. To actually also improve the predicted probabilities, we additionally employ three modern machine learning methods: (i) logistic LASSO, (ii) random forest, and (iii) gradient boosting.

These models have in common that they allow for different levels of model complexity that are governed by a vector of tuning parameters. A more complex model reduces the bias in representing the relationship between outcome and predictors, but comes at the risk of overfitting. Both overfitting and bias worsen out-of-sample model performance. To find the optimal level of complexity, we tune parameters via 8-fold cross-validation on the training data (following the recommendations by Mullainathan and Spiess 2017). That is, we split the training data set into 8 equally sized folds and set one of the folds aside. We then estimate the model for specified values of the tuning parameters on the remaining seven folds and evaluate the fit on the selected fold. After repeating this step for each of the other 7 folds, we compute the average fit across all folds. The optimal parameter specification is the one that yields the highest average fit in the cross-validation procedure. We use this specification to reestimate the model on the whole training data and evaluate its fit on the test data. See Section C in the appendix for details about the parameter tuning.

The LASSO (Tibshirani 1996) is a method that controls for overfitting by imposing a penalty on the size of the coefficients. As a result, coefficients are shrunk towards zero (odds ratios towards one) as compared with standard logistic regression. Some coefficients will be set to exactly zero and thus drop out of the regression. Constraining the coefficients avoids overfitting, but introduces bias that can lead to inaccurate predictions. The penalty is the tuning parameter that guides the trade-off between overfitting and bias.

Both random forest (Breiman 2001) and gradient boosting (Friedman 2001) are tree-based approaches. They have the advantage that they implicitly allow for interactions among predictors. A single tree is constructed by progressively splitting the data into partitions, which are called nodes. At each step, the data is split using the variable and associated splitting criterion that generates the largest improvement in model fit. The splitting process stops when the tree has reached a specified depth or when further splits would yield nodes that fall below a specified minimum node size.

Because a single tree is prone to overfitting, a random forest computes the average predictions over multiple independently grown trees. Each tree contributing to the forest uses only a random subset of the predictors, which de-correlates the trees and reduces overfitting. The tuning parameters for the random forest that we use are the minimum node size and the fraction of all predictors that is randomly selected to build each tree.

The gradient boosting approach, in contrast, builds trees sequentially. Observations that deviated more from their actual value in the previous tree receive more weight when building the next tree. The final model averages the predictions of all trees. The tuning parameters for the gradient boosting are the maximum depth of the tree, the fraction of random observations used for building each tree and the fraction of random predictors used for building each tree. Building too many trees can once again lead to overfitting, so the number of trees to be grown is another tuning parameter. This paper employs the widely used tree boosting system XGBoost (Chen and Guestrin 2016).

4 Results

4.1 Individual outcomes

Figure 4 presents graphs analogous to Figure 1 from the theoretical section. Based on predicted probabilities from logistic regressions, each graph shows the marginal increase in the TPR for one of the outcomes as we increase the fraction of the cohort to be treated (β). Unlike Figure 1, we cannot draw the marginal expected benefit because we do not know the impact ($\bar{\delta}$) of the intervention and the cost of the outcome ($Cost_{Outcome}$). Note, however, that knowing these constant values would only change the level but not the shape of the curves.[5]

The graphs provide clear evidence that informed treatment assignment based on predictions can substantially improve upon uninformed treatment assignment. For example, targeting 5% of the cohort will reach more than 15% of all individuals at risk for having only compulsory education compared to 5% under random, uninformed assignment. When the problem is to target a given fraction of the cohort, for example in the presence of a

5. The coefficient estimates of the corresponding logistic regressions can be found in Tables D.5 and D.6 in the appendix.

budget constraint, then informed treatment assignment will undoubtedly increase welfare. Informed assignment might also be welfare-enhancing when the fraction to be treated is an endogenous choice variable, but this depends on the specific parameters of the setting (costs of outcome and intervention; impact of the intervention).

In Figure 5, we show ROC plots along with corresponding AUC values to assess the overall accuracy of prediction. Prediction works best for criminal charges, education and foster care placement, with AUC values between 0.75 and 0.81. Health outcomes and income are less predictable by early-life indicators, with AUC values ranging in the low 0.60s. These findings are consistent with earlier evidence that the child's educational outcomes depends strongly on parental education, while intergenerational mobility is larger for income and health.[6] Measures of intergenerational mobility relate a child's outcome to the same outcome of the parent, which might be measured a long time after the child's birth. Our study shows that similar patterns of intergenerational transmission arise when measuring family background already at birth using a combination of multiple indicators. As Landersø and Heckman (2017) pointed out, the Danish welfare state is characterized by large income redistribution through taxes and transfers in addition to wage compression, leading to higher income than educational mobility. Similarly, universal access to tax-financed medical care might explain why family background matters less for health outcomes.[7]

4.2 Combinations of outcomes

Section 3.1.2 demonstrated that outcomes are highly correlated within individuals. A small group of individuals accounts for a disproportionately large share of the total social burden. Predicting (and targeting) these individuals promises large returns and Caspi et al. (2017) – using predictors throughout childhood – showed that prediction works even better for these "high-cost" individuals.

In the left graph in Figure 6, we predict whether individuals have 3+ outcomes versus 2 or fewer (and similarly for 4+ and 5+ outcomes). AUC values are high (0.75-0.81), suggesting that targeting high-cost individuals is indeed easier, even when using only predictors measured at birth. In the right panel, we repeat the analysis for whether individuals have 3+ outcomes versus 0 outcomes (and similarly for 4+ and 5+ outcomes). AUC values are now even higher (0.80-0.87), implying that there is substantial variation between low-cost

6. Hertz et al. (2008) estimate that the intergenerational correlation for education in Denmark is 0.30. Andrade and Thomsen (2018) find values in the range between 0.35 and 0.39. In contrast, intergenerational correlation for income are typically much smaller. For gross income including public transfers Landersø and Heckman (2017) estimate it to be 0.21, while Andersen (2019), studying total income before deductions and taxes, finds values between 0.05-0.06 (maternal income) and 0.13-0.21 (paternal income). Our measure of disposable family income reflects the progressivity of the Danish tax system, so these estimates are probably too large. Intergenerational correlations for health tend to be even smaller, at least with respect to fathers. Andersen (2019) finds values between 0.11-0.12 (paternal health) and 0.13-0.14 (maternal health)

7. Andersen (2019) finds rank-rank slopes for intergenerational health outcomes in Denmark that are only half the size of those found by Halliday et al. (2018) for the U.S., a country with a considerable fraction of uninsured individuals.

and high-cost individuals already at birth. Our estimates are similar to the ones by Caspi et al. (2017). Unfortunately, however, we do not know a priori which individuals will end up with either zero or 3+ outcomes (but not 1 or 2 outcomes), so this latter exercise is only of limited relevance in practice.

4.3 Method comparison

Table 3 compares AUC estimates from different methods. Recall that estimating machine learning methods (LASSO, random forest, gradient boosting) involves finding an optimal set of tuning parameters via cross-validation. For each method and outcome, we therefore present two estimates.

The first estimate (*AUC direct*) comes from choosing the tuning parameters that maximize the AUC. The second estimate (*AUC grid*), allows for varying parameters for different values of the fraction of the cohort to be treated (β). Using a grid of 20 β-values, we first find the set of parameters that maximize the TPR for a given β. Next, we use the resulting TPR values to compute the AUC. This flexible approach potentially increases the AUC. But it might also lower the AUC since we only use a finite grid of β-values and linearly interpolate between grid points. Since the ROC curve has a concave shape, linear interpolation tends to reduce the AUC. This can seen for the logistic regression (columns 1 and 4). Since it involves no parameter tuning, *AUC grid* must be, and indeed is, smaller than *AUC direct*. For the other machine learning techniques, it is an empirical question of whether *AUC direct* or *AUC grid* will be larger.

Table 3 provides two take-aways. First, maximizing *AUC direct* yields better results than indirect optimization over a finite grid of β values. This suggests that the benefits of fine-tuning parameters are rather limited and that a single optimized set of parameters works well for most β values. Second, the LASSO generates AUC values very close to logistic regression. Logistic regression does not seem to suffer from overfitting, which is perhaps unsurprising in light of the large sample size. Interestingly, however, the random forest and the gradient boosting methods do not outperform logistic regression either. This suggests that interactions among regressors play only a limited role and that a linear combination of predictors works well. We will thus focus on logistic regression in the remainder of the paper.

4.4 Post-birth predictors

So far, we have been focusing on predictors available at birth. Can we obtain better predictions by extending the time frame to a few years after birth? Adding years after birth allows us to update parental predictors and to include the child's hospitalizations as an additional predictor. We look at 1, 3 and 5 years after birth.

Figure 7 illustrates how the AUC changes as we extend the time window. As expected, the

AUC increases with more recent predictors. The increase is most pronounced for foster care placement, which is also the outcome that occurs earliest in life and is probably more sensitive to changes in early-life family environment. Overall, however, the marginal improvements in predictive accuracy are rather modest. It seems as if the role that family background plays in shaping long-run outcomes is largely determined by factors set in place at birth. Of course, predictors other than family background, such as performance in cognitive test scores or parenting style, matter as well. But these variables are typically not available in register data and therefore costly to obtain at large scale. If only register data are available, then targeting children already at birth comes at little cost in light of the large benefits of early versus late interventions.

4.5 Parsimonious model

Our predictions are based on a rich set of variables. Do all predictors contribute equally to the quality of the predictions or do few predictors drive our results? If a parsimonious model with a small number of predictors can generate predictions close to the full model, then necessary data collection would be less costly and computing risk scores would be simpler and more transparent.

To investigate the relative importance of the predictors, we run regressions of the outcome on each predictor in isolation and then compare the AUC to the one from the full model. For each predictor, such as parental income, we include the full set of dummies capturing the various values that a predictor may take on. Figure 8 shows the results. Figure (a) reports the actual AUC values, with darker colors indicating larger values. Because the AUC of the full model varies by outcome, Figure (b) also reports the percentage of the AUC of the full model that the simple one-predictor model attains.

Individual-level predictors have little predictive power. Including only nationality, birth month or birth order in the regression generates an AUC of up to 10% of the one from the full model, but mostly less. An exception is sex, reflecting the substantial gender gaps in, for example, criminal behavior. The direction of the gaps is further elucidated in the next section.

Parental predictors have much larger predictive power. Placements, marital status and hospitalizations reach up to 30% of the full-model AUC. Values are slightly higher for crime (up to 40%) and age (up to 50%). Predictors related to SES give rise to the highest AUC. Working hours and wealth reach up to 60% and 70% of the full-model AUC, respectively, while income, education and occupation account for up to 80%.

In the right part of Figures 8 (a) and (b), we use as predictors sex plus various combinations of income, education and occupation. Any two of these three SES predictors (plus sex) generates around 90% of the full-model AUC. The best combination is that of income and education, suggesting that these two variables complement each other most, while occupation captures aspects of both and thus contributes the least independent variation. Adding occupation to income and education (labeled SES) does not yield great additional improvement.

The central role of SES in predicting long-run outcomes does not necessarily mean that SES is the causal driver behind these outcomes. To illustrate, we gauge how predictive performance changes once we exclude certain predictors, including SES, from the full model. To begin with, the left part of Figures 9 (a) and (b) shows the reduction in AUC when leaving out single predictor variables. With very few exceptions, excluding single predictors has close to zero impact on the AUC of the prediction. Even the omission of education, the strongest single predictor besides occupation in Figure 8, is almost entirely absorbed by other predictors, decreasing the AUC by no more than 7 %. In the right part of Figures 9 (a) and (b), we exclude several SES-related predictors at the same time. As expected, reductions in AUC become larger. But even when using the broadest conceivable set of SES indicators (including income, education, occupation, wealth and working hours), the AUC decreases by only 20-30 percent or 0.03-0.07 in absolute value. This means that a sizable portion of the predictive variation in SES is also captured collectively by a set of non-SES predictors. This points to SES acting as good proxy of a latent measure of disadvantage, rather than being itself the causal driver of adverse long-run outcomes.[8]

Overall, this section has demonstrated that a parsimonious model with the predictors sex, parental income and parental education performs almost as well as the full model that includes all predictors. This suggests that collecting data on these few variables suffices for efficient targeting in practice. It also bolsters the way that targeting has been operationalized in many prominent childhood interventions, which typically used indicators of parental SES to define disadvantage. However, a separate but equally important question is how much weight to attach to the different values that an indicator can take on. We address this question next.

4.6 Optimal weights

Efficient targeting based on a one-dimensional risk score requires optimally weighting the selected predictors that contribute to the risk score. The Perry Preschool Project derived a "cultural deprivation rating" in which higher values indicated *better* outcomes (Weikart 1967, p. 3-4). It had three components: Paternal occupation entered the rating with 1 point if the father engaged in unskilled work and 4 points if he engaged in skilled work. Each of the parents' average years of schooling entered with another point. Finally, "density in the home", measured as number of rooms divided number of people, entered after multiplying by one half to give it "a 1/2 weight". These weights already seem rather arbitrary, but each

8. Of course, SES could also causally effect non-SES predictors, which are then correlated with the outcome even if they do not have any direct impact. In general, it is very difficult to draw conclusions about causality without any exogenous variation in the predictors.

component was additionally divided by its standard deviation before aggregating to the final rating, thus giving more weight to components with little absolute variation. Only children with a final rating below 11 were considered further for the experiment.

The Carolina Abecedarian Project (Ramey and And Others 1974, p. 65) constructed a "high risk index". The index increased by one point for each year missing from 12 years of schooling, for both the mother and father. Family income above 5,000 dollars left the risk score unchanged, while income below 5,000 dollars increased it by 4 points, and by another point for each additional 1,000-dollar step downwards. Additional points were assigned for, among others, the absence of the father and low parental I.Q. scores. Only children with an index value of 11 or higher participated in the experiment.

Both the Perry Preschool Project and the Carolina Abecedarian Project assigned weights in an ad hoc way. The motivation behind the chosen weights remains unclear and there is no indication that they were optimal in any way. Thus, at least three questions arise: First, what relative weight should each indicator optimally receive? Second, should indicators such as education enter differently for fathers and for mothers? Third, should indicators such as years of schooling enter linearly? We do not have access to all the indicators used in these two or other studies. Instead, we will use a parsimonious set of predictors consisting of sex, income and education, which we showed to predict quite well in the previous subsection. Income and education are key ingredients in defining early disadvantage in many targeted interventions. Answering the above questions for these two variables is thus highly policy-relevant.

We can address the problem of optimal weighting with the estimated coefficients from the logit regressions. Because our predictors are discretized and a dummy is included for each discrete value, the coefficient directly gives the weight associated with that value of the predictor, relative to the baseline value. For easier interpretation, we rescale all weights such that the lowest possible weight is 0 and the highest possible weight is 100. Note that risk score values are not interpretable as percentiles; their distribution depends on the distribution of predictor values in the population.

Figure 10 illustrates the computation of the weights (see Table D.7 in the appendix for the exact values). The baseline individual – the leftmost circles in each column with a gray edge – relative to which the weights are defined is characterized as: female, master's degree/PhD (both mother and father) and 10th income decile (both mother and father). When the outcome of interest is education, we see that the baseline individual has a score of 1 (=0+0+0+0+1). If the individual's father had only compulsory schooling instead of a master's degree/PhD, her score would increase to 28 (=0+27+0+0+0+1). If instead the maternal income was in the 9th rather than 10th decile, then her score would actually *decrease* to 0. This result is counterintuitive, but note that despite the large sample size, coefficients are estimated with some uncertainty, as indicated by the 95% confidence bands. It is reassuring to see that in general income and education show the expected positively monotone relationship with the outcome.[9]

What lessons about optimal weights can we learn from Figure 10? First, men tend to have higher risk of adverse outcomes, in particular for criminal charges, low education and low income. The gender gap is reversed for health-related outcomes. Income and education seem to contribute equally to the risk score, when measuring contribution as the difference between the lowest and the highest score within a predictor. That said, education contributes more to the risk score than income when predicting education and hospitalization, while the opposite is true for predicting psychiatric condition and income.

Second, maternal and paternal education affect the risk score in a similar manner, but maternal education appears slightly more important for foster care placement and psychiatric conditions. An interesting observation with respect to parental income is that maternal income plays a small role as long as she is in the upper 80% of the distribution, but being in the bottom 20% causes the risk score to spike. The relationship is more monotone for paternal income; i.e. moving from, say, the 8th to the 7th decile increases the score just as moving from the 3rd to the 2nd decile does.

Third, especially values at the bottom of the education and income distribution substantially raise the risk score. For income, these are the bottom 30% of fathers and the bottom 20% of mothers. For education, this is having only compulsory schooling or only vocational education and training. This finding implies that risk scores in which income or years of schooling enter linearly, like in the Perry Preschool Project and the Carolina Abecedarian Project, assign too little weight to children with parents in the bottom of the distribution. Of course, this might be driven by our specific choice of outcomes that also focus on the bottom of the respective distribution. Predictions of, for instance, average hospitalizations rather than top 20% hospitalizations might depend much more linearly on parental SES. However, as we argued above, it is precisely those individuals in the top 20% of the distribution that cause disproportionate burden to the welfare state. Targeting these individuals effectively seems appropriate from an economic point of view.

5 Discussion and conclusion

In this paper, we use prediction to show that efficient targeting in childhood interventions is possible even if only variables available at birth are at the decision-maker's disposal. This applies to interventions addressing a wide range of long-run outcomes, ranging from labor market outcomes to health and crime. We also find that predictions do not improve much by adding post-birth indicators if the decision-maker is restricted to using register data only.

9. The only exception is having low disposable income (second column), which appears to have a – if anything – reversed relationship with parental education. The reason is presumably that children of highly educated parents are themselves highly educated, but earn only relatively little just after graduating and perhaps are more likely to be single. Obtaining vocational schooling, in contrast, guarantees moderate earnings from early on and might also allow for earlier partnership and family formation.

We demonstrate that a parsimonious set of variables consisting of sex, parental education and parental income predicts almost as well as the full set of predictors. Finally, we provide econometrically derived, optimal weights for the formation of risk scores that differ from and improve upon the ad-hoc weights typically used in the literature.

Is our study practically relevant? Should our risk score be employed in targeted childhood interventions? One caveat is that our predictors were recorded more than 30 years ago. To the extent that the relationship between at-birth predictors and long-run outcomes is different today, the risk scores we computed might no longer be optimal. The justification for our approach is that long-run outcomes must be observed to be able to construct meaningful risk scores. The alternative to anchoring risk scores in long-run outcomes is to assign ad hoc weights, which most likely results in less efficient targeting.

Another concern is that targeting high-risk children yields little value if they are not actually the ones benefiting from an intervention. For example, patients with a family history of severe genetically determined chronic disease will not respond to any type of treatment. Unfortunately, information on susceptibility to treatment is typically unknown. Without additional knowledge, we view it as reasonable to target children with a high risk of the outcome.

As a more subtle point, we make the implicit assumption that current policies are not targeted at relevant predictors or are ineffective and do not alter the relationship between predictors and outcomes. If a relevant predictor is already successfully targeted by policy, then its importance will be attenuated; in the extreme, it might not show up as predictive at all. For example, our finding that birth order has little predictive power could be due to interventions effectively targeting children based on this variable. Ignoring birth order in defining risk would then be a grave mistake. Similarly, our risk scores become invalid in the future once they are used for targeting; they would need to be recomputed.

This paper makes a case for targeted interventions by showing that efficient targeting is possible for a wide range of outcomes. That said, an argument against targeting is the possibility for individuals to manipulate their risk score in order to receive treatment. In our setting, the potential for manipulation is small as data were directly taken from the official registers. Survey-based data are more likely to contain false records. Besides misreporting, parents could of course also directly lower their income to become eligible. We consider this to be highly unlikely.

The predictors considered in this article do not include genetic data, which are also fixed at birth and have been shown to predict a wide range of social outcomes (e.g., Belsky et al. 2016). The problem with genetic data is that they are neither widely available in registers nor easy and cheap to collect through surveys, much in contrast to the predictors we are using. So while genetic information may have great potential for predicting long-run outcomes, its usefulness for targeting in real childhood interventions is still rather limited.

Finally, targeting raises ethical issues. Dare (2013) discusses several of these issues within the context of a predictive risk model for child maltreatment. The computation and publication of risk scores might be viewed as reducing the child to a number. Moreover, there is a large difference between a calculated risk score and a realized disadvantage. Intervening of the basis of a perceived risk rises questions concerning e.g. the rights of parents to raise their children as they see fit. Furthermore, if assignment of children identified as "high-risk" to an intervention is considered a stigma, then the positive effects of treatment may be counteracted by the negative effects of the stigma. Resistance to using data-based algorithms in social setting might also come from fears that algorithms are biased and tend to perpetuate existing inequalities (O'Neil 2016), even if our algorithm is actually designed to achieve the opposite. It is obvious that these concerns are very real and should be handled and/or taken into consideration. In the context of child maltreatment, Dare (2013) concludes that the potential gains outweigh any ethical reservations but in our context it is perhaps not so clear cut. Discussing these issues more deeply is beyond the scope of the paper, which aims to demonstrate the extent to which targeting is practically feasible. We thus hope our work will serve as a basis for further discussion, both inside and outside of academia.

References

Allen, Graham. 2011. *Early Intervention: The Next Steps.* Technical report.

Andersen, Carsten. 2019. *Intergenerational Health Mobility: Evidence from Danish Registers.* Working paper, Economics Working Papers 19-04. Aarhus University.

Andersen, Simon Calmar, and Helena Skyt Nielsen. 2016. "Reading intervention with a growth mindset approach improves children's skills." *Proceedings of the National Academy of Sciences* 113 (43): 12111–12113.

Andrade, Stefan, and Jens-Peter Thomsen. 2018. "Intergenerational Educational Mobility in Denmark and the United States." *Sociological Science* 5:93–113.

Arai, Mahmood, and Peter Skogman Thoursie. 2009. "Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings." *Journal of Labor Economics* 27 (1): 127–147.

Bansak, Kirk, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein. 2018. "Improving refugee integration through data-driven algorithmic assignment." *Science* 359 (6373): 325–329.

Belsky, Daniel W., Terrie E. Moffitt, David L. Corcoran, Benjamin Domingue, HonaLee Harrington, Sean Hogan, Renate Houts, et al. 2016. "The Genetics of Success." *Psychological Science* 27 (7): 957–972.

Bhattacharya, Debopam, and Pascaline Dupas. 2012. "Inferring welfare maximizing treatment assignment under budget constraints." *Journal of Econometrics* 167 (1): 168–196.

Björkegren, Evelina, Mårten Palme, Mikael Lindahl, and Emilia Simeonova. 2019. *Pre- and Post-Birth Components of Intergenerational Persistence in Health and Longevity: Lessons from a Large Sample of Adoptees.* Discussion paper 12451. Institute for the Study of Labor (IZA).

Black, Sandra E., and Paul J. Devereux. 2011. "Recent Developments in Intergenerational Mobility." *Handbook of Labor Economics* 4:1487–1541.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2005. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *The Quarterly Journal of Economics* 120 (2): 669–700.

———. 2016. "Healthy(?), wealthy, and wise: Birth order and adult health." *Economics & Human Biology* 23:27–45.

Bleses, Dorthe, Anders Højen, Philip S. Dale, Laura M. Justice, Line Dybdal, Shayne Piasta, Justin Markussen-Brown, Laila Kjærbæk, and E.F. Haghish. 2018. "Effective language and literacy instruction: Evaluating the importance of scripting and group size components." *Early Childhood Research Quarterly* 42:256–269.

Bleses, Dorthe, Anders Højen, Laura M. Justice, Philip S. Dale, Line Dybdal, Shayne B. Piasta, Justin Markussen-Brown, Marit Clausen, and E. F. Haghish. 2018. "The Effectiveness of a Large-Scale Language and Preliteracy Intervention: The SPELL Randomized Controlled Trial in Denmark." *Child Development* 89 (4): e342–e363.

Bratberg, Espen, Øvind Anti Nilsen, and Kjell Vaage. 2015. "Assessing the intergenerational correlation in disability pension recipiency." *Oxford Economic Papers* 67 (2): 205–226.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45:5–32.

Breining, Sanni, Joseph Doyle, David N. Figlio, Krzysztof Karbownik, and Jeffrey Roth. 2020. "Birth Order and Delinquency: Evidence from Denmark and Florida." *Journal of Labor Economics* 38 (1): 95–142.

Carlsson, Magnus, and Dan-Olof Rooth. 2007. "Evidence of ethnic discrimination in the Swedish labor market using experimental data." *Labour Economics* 14 (4): 716–729.

Caspi, Avshalom, Renate M. Houts, Daniel W. Belsky, Honalee Harrington, Sean Hogan, Sandhya Ramrakha, Richie Poulton, and Terrie E. Moffitt. 2017. "Childhood forecasting of a small segment of the population with large economic burden." *Nature Human Behaviour* 1 (1): 1–10.

Chandler, Dana, Steven D Levitt, and John A List. 2011. "Predicting and Preventing Shootings among At-Risk Youth." *American Economic Review* 101 (3): 288–292.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD),* 785–794. ACM Press.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States." *The Quarterly Journal of Economics* 129 (4): 1553–1623.

Chittleborough, Catherine R., Amelia K. Searle, Lisa G. Smithers, Sally Brinkman, and John W. Lynch. 2016. "How well can poor child development be predicted from early life characteristics?" *Early Childhood Research Quarterly* 35:19–30.

Commission/EACEA/Eurydice, European. 2019. *Key Data on Early Childhood Education and Care in Europe – 2019 Edition.* Technical report. Eurydice Report. Luxembourg: Publications Office of the European Union.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2018. "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance." *Journal of Political Economy* 126 (6): 2356–2409.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. 2014. "Family Welfare Cultures." *The Quarterly Journal of Economics* 129 (4): 1711–1752.

Dare, Tim. 2013. *Predictive Risk Modelling and Child Maltreatment: An Ethical Review.* Ministry of Social Development, Wellington, New Zealand.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall. 2014. "Household Choices and Child Development." *The Review of Economic Studies* 81 (1): 137–185.

Fawcett, Tom. 2006. "An introduction to ROC analysis." *Pattern Recognition Letters* 27 (8): 861–874.

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *The Annals of Statistics* 29 (5): 1189–1232.

Gallen, Yana, Rune V. Lesner, and Rune Vejlin. 2019. "The labor market gender gap in Denmark: Sorting out the past 30 years." *Labour Economics* 56:58–67.

Glaeser, Edward L., Andrew Hillis, Scott Duke Kominers, and Michael Luca. 2016. "Crowdsourcing City Government: Using Tournaments to Improve Inspection Accuracy." *American Economic Review* 106 (5): 114–118.

Halliday, Timothy, Bhashkar Mazumder, and Ashley Wong. 2018. *Intergenerational Health Mobility in the US.* Discussion paper 11304. Institute for the Study of Labor (IZA).

Havnes, Tarjei, and Magne Mogstad. 2015. "Is universal child care leveling the playing field?" *Journal of Public Economics* 127:100–114.

Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312 (5782): 1900–1902.

———. 2008. "Schools, Skills, and Synapses." *Economic Inquiry* 46 (3): 289–324.

———. 2012. *Invest in early childhood development: Reduce deficits, strengthen the economy.* The Heckman Equation.

Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina. 2008. "The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends." *The B.E. Journal of Economic Analysis & Policy* 7 (2).

Hjalmarsson, Randi, and Matthew J. Lindquist. 2012. "Like Godfather, Like Son." *Journal of Human Resources* 47 (2): 550–582.

———. 2013. "The origins of intergenerational associations in crime: Lessons from Swedish adoption data." *Labour Economics* 20:68–81.

Kitagawa, Toru, and Aleksey Tetenov. 2018. "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice." *Econometrica* 86 (2): 591–616.

Klaus, Rupert A., and Susan W. Gray. 1968. “The Early Training Project for Disadvantaged Children: A Report after Five Years.” *Monographs of the Society for Research in Child Development* 33 (4): iii–66.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. “Human Decisions and Machine Predictions.” *The Quarterly Journal of Economics.*

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. “Prediction Policy Problems.” *American Economic Review* 105 (5): 491–495.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. “Children and Gender Inequality: Evidence from Denmark.” *American Economic Journal: Applied Economics* 11 (4).

Kriminalvården. 2018. *NORDISK STATISTIK - För kriminalvården i Danmark, Finland, Island, Norge och Sverige 2013-2017.*

Kronborg, Hanne, Hans Henrik Sievertsen, and Miriam Wüst. 2016. “Care around birth, infant and mother health and maternal health investments – Evidence from a nurse strike.” *Social Science & Medicine* 150:201–211.

Landersø, Rasmus, and James J. Heckman. 2017. “The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the US.” *The Scandinavian Journal of Economics* 119 (1): 178–230.

Manski, Charles F. 2004. “Statistical Treatment Rules for Heterogeneous Populations.” *Econometrica* 72 (4): 1221–1246.

Mertz, Mikkel, and Signe Hald Andersen. 2017. “The Hidden Cost of Foster-Care: New Evidence on the Inter-Generational Transmission of Foster-Care Experiences.” *The British Journal of Social Work* 47 (5): 1377–1393.

Mullainathan, Sendhil, and Jann Spiess. 2017. “Machine Learning: An Applied Econometric Approach.” *Journal of Economic Perspectives* 31 (2): 87–106.

O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown.

Ramey, Craig T., and And Others. 1974. *The Carolina Abecedarian Project: A Longitudinal and Multidisciplinary Approach to the Prevention of Developmental Retardation.* Technical report. ERIC Report no. ED 030 490, North Carolina Univ., Chapel Hill. Frank Porter Graham Center.

Richmond-Rakerd, Leah S., Stephanie D’Souza, Signe Hald Andersen, Sean Hogan, Renate M. Houts, Richie Poulton, Sandhya Ramrakha, Avshalom Caspi, Barry J. Milne, and Terrie E. Moffitt. 2020. “Clustering of health, crime and social-welfare inequality in 4 million citizens from two nations.” *Nature Human Behaviour.*

Rooth, Dan-Olof. 2010. "Automatic associations and discrimination in hiring: Real world evidence." *Labour Economics* 17 (3): 523–534.

Sansone, Dario. 2019. "Beyond Early Warning Indicators: High School Dropout and Machine Learning." *Oxford Bulletin of Economics and Statistics* 81 (2): 456–485.

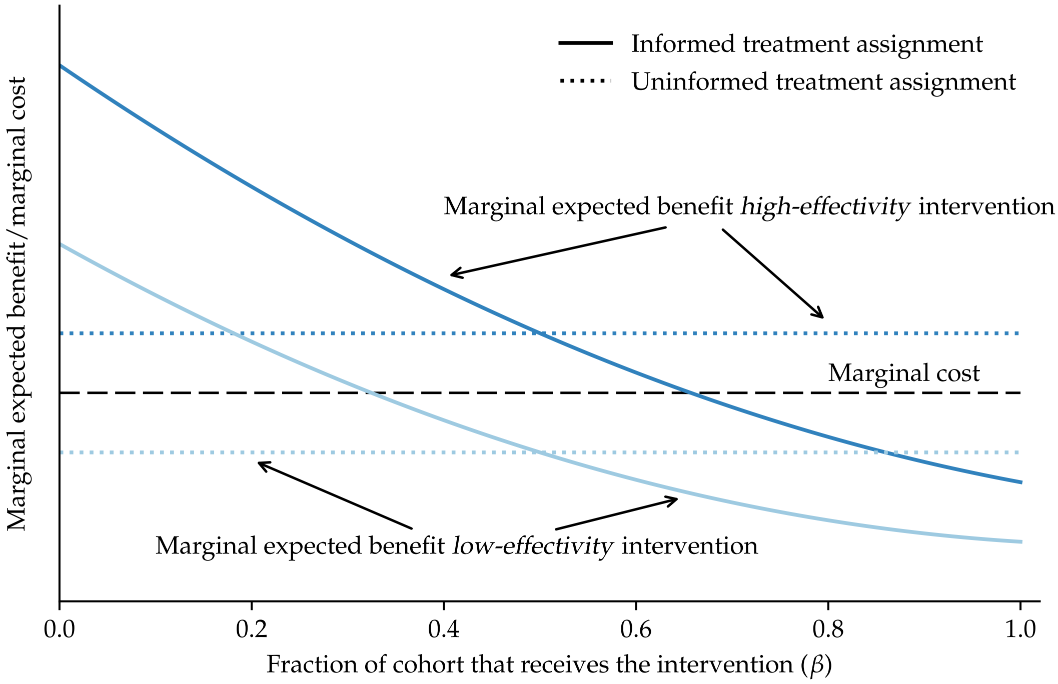
Tibshirani, Robert. 1996. "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–288.

U.S. Department of Health & Human Services. 2019. "Poverty Guidelines and Determining Eligibility for Participation in Head Start Programs." Accessed November 11, 2019. `https://eclkc.ohs.acf.hhs.gov/eligibility-ersea/article/poverty-guidelines-determining-eligibility-participation-head-start`.

Wall-Wieler, Elizabeth, Ylva Almquist, Can Liu, Bo Vinnerljung, and Anders Hjern. 2018. "Intergenerational transmission of out-of-home care in Sweden: A population-based cohort study." *Child Abuse & Neglect* 83:42–51.

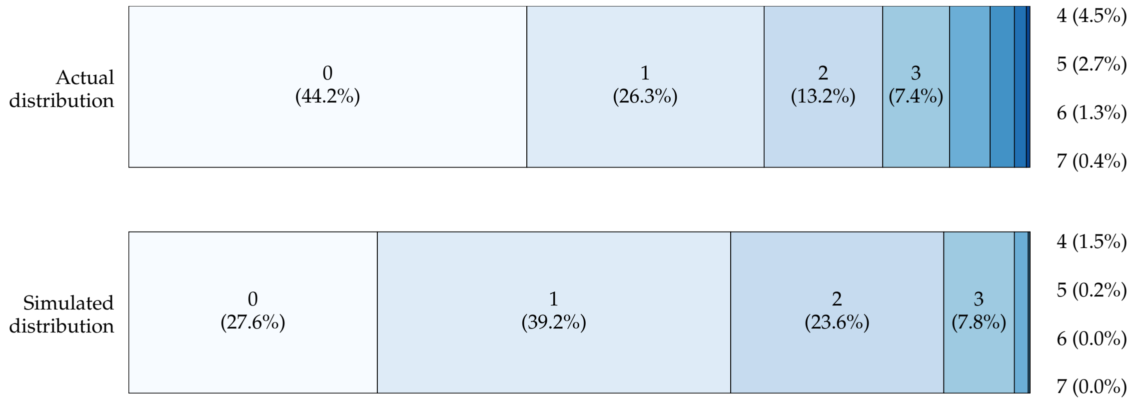
Weikart, David P. 1967. *Preliminary Results From a Longitudinal Study of Disadvantaged Preschool Children.* ERIC Report ED 030 490. presented at the 1967 convention of the Council for Exceptional Children, St. Louis, MO.

Figure 1: Illustration of second-stage problem



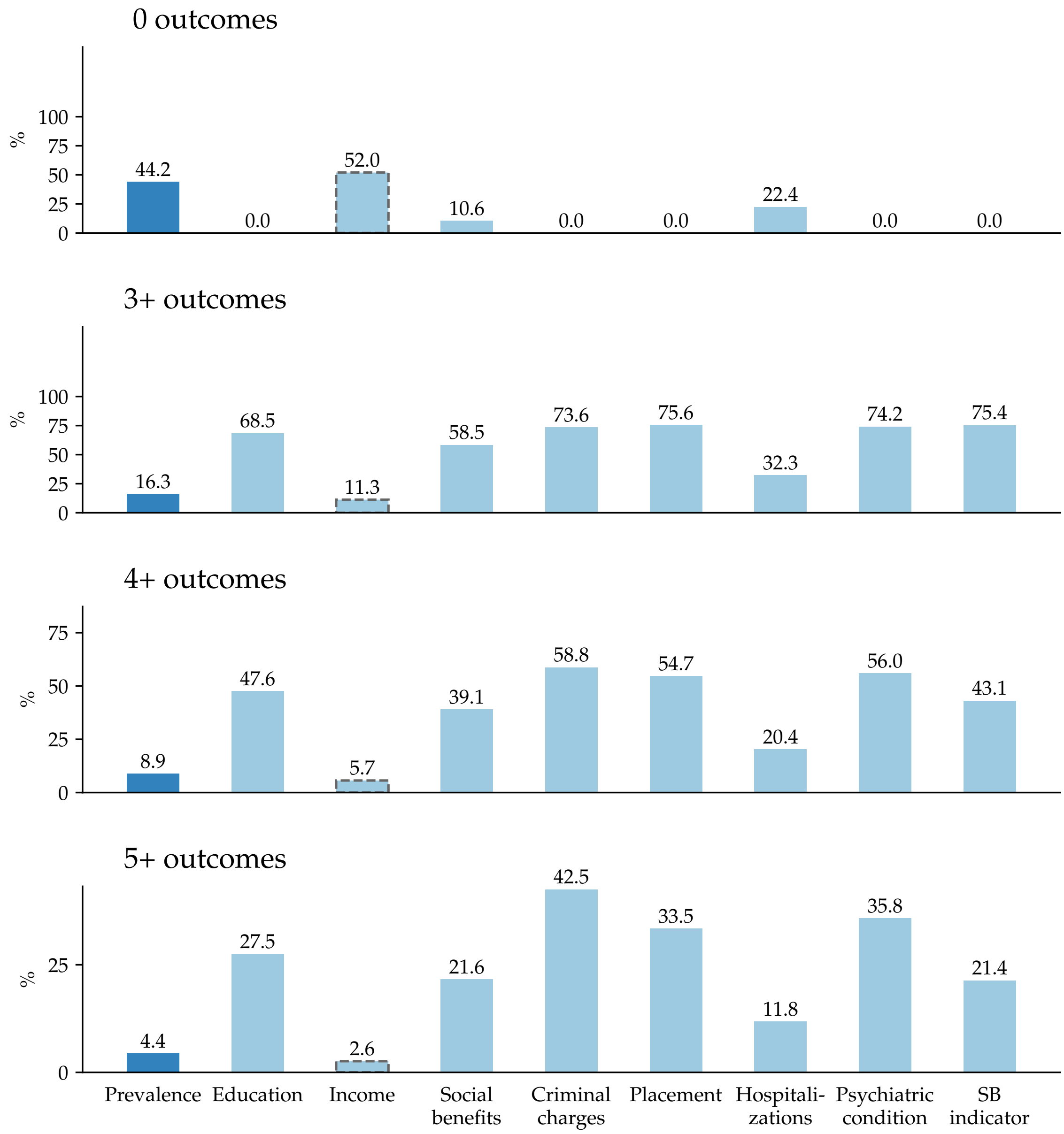
Notes: The figure illustrates the decision-maker's second-stage problem of choosing the fraction (β) of the cohort that should receive the intervention. β should be increased as long as the marginal expected benefit lies above the marginal cost. Two scenarios are depicted: First, a low-effectiveness intervention that would not be administered at all under uninformed treatment assignment. Second, a high-effectiveness intervention that would be administered to the whole cohort under uninformed treatment assignment. In both scenarios, welfare improvements can be realized through informed treatment assignment to a fraction of the cohort only.

Figure 2: Actual vs. simulated distribution of outcomes



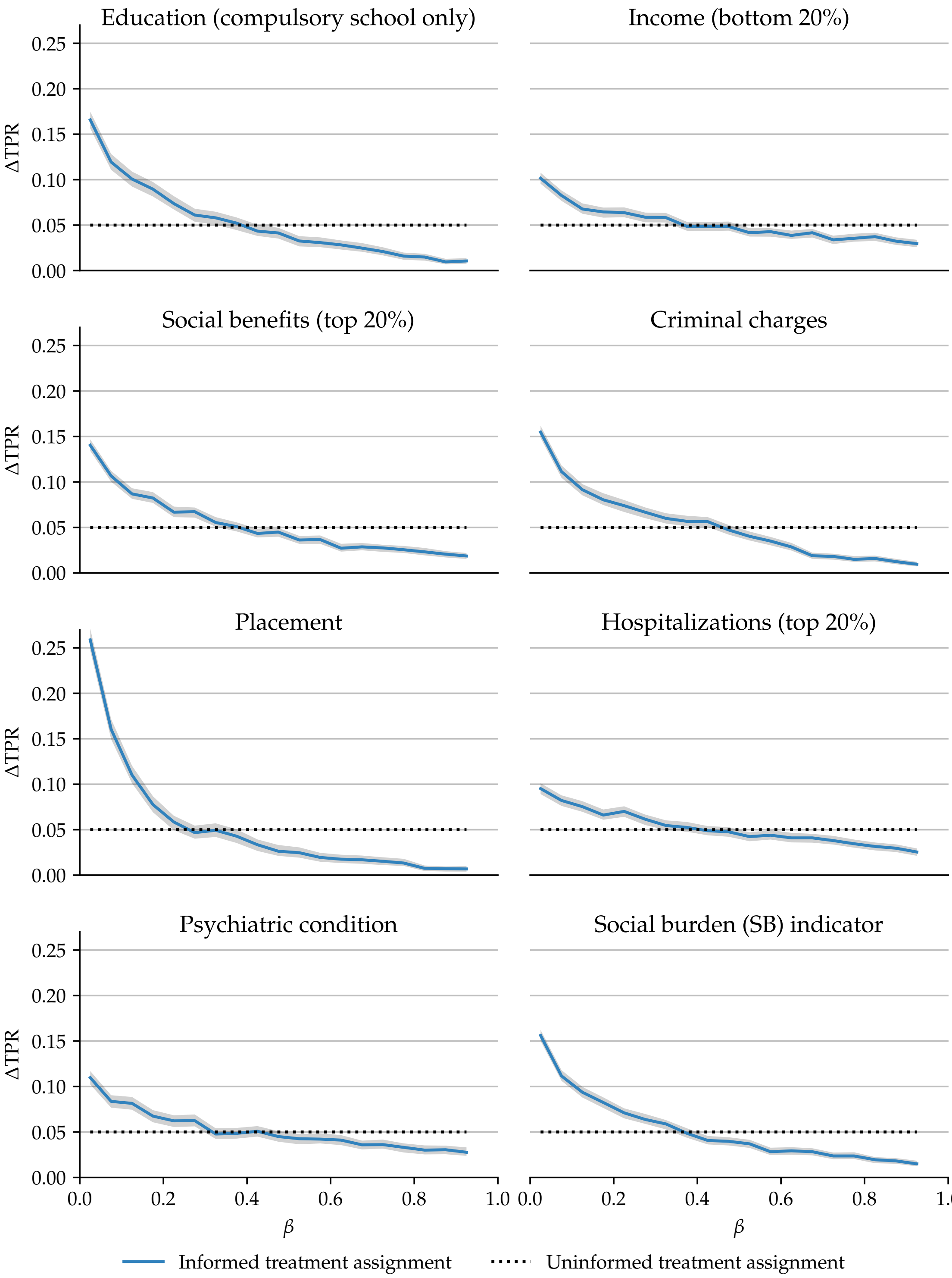
Notes: The actual distribution shows how many individuals have 0, 1, ..., 6 or 7 outcomes. The simulated distribution is the distribution of outcomes that would result if outcomes had the same marginal probability as in the actual distribution but were uncorrelated with each other.

Figure 3: Economic burden by total number of outcomes



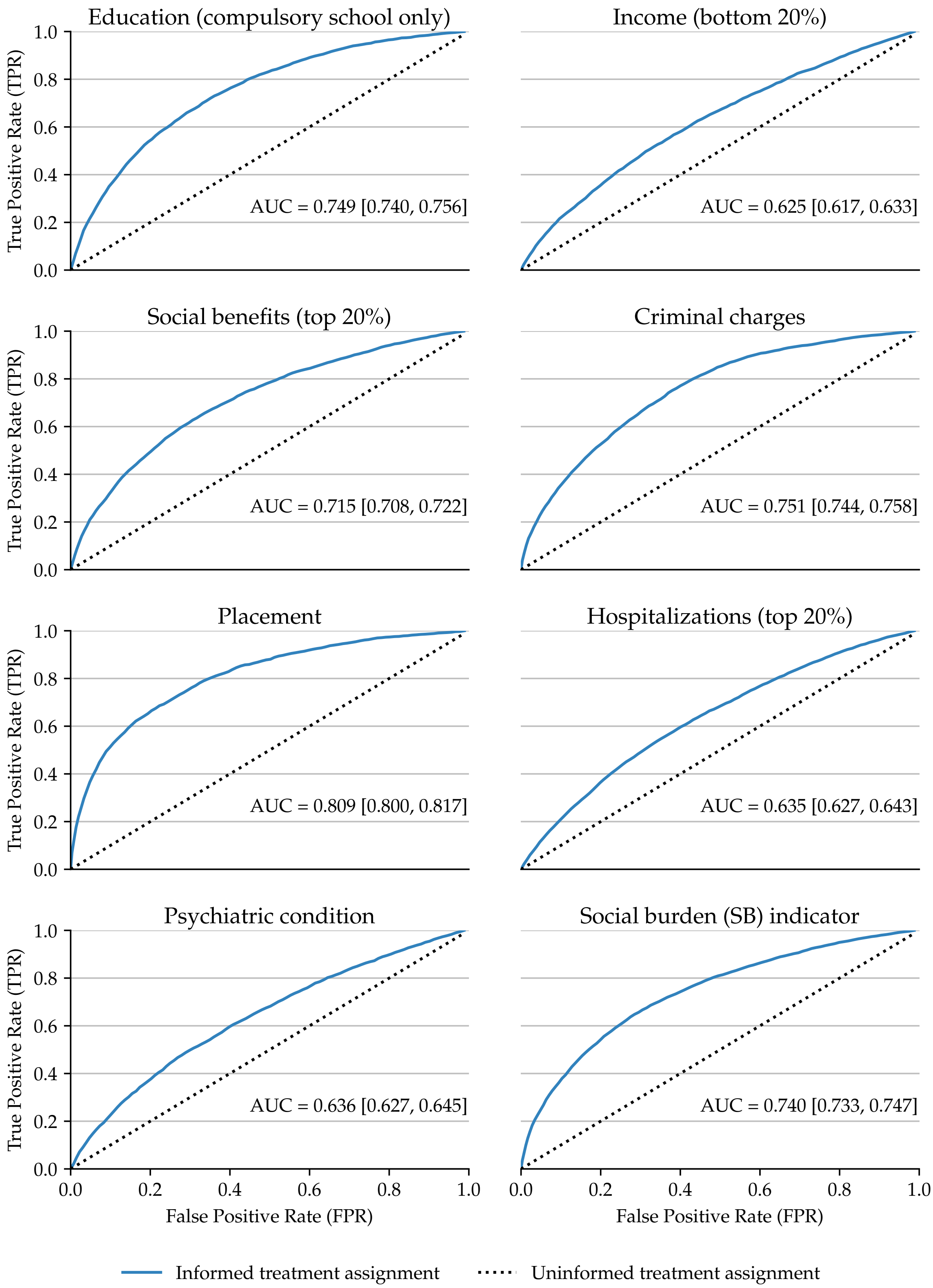
Notes: The first column shows the prevalence of individuals with 0, 3+, 4+ and 5+ outcomes in the population. The other columns show the proportion of the burden of each outcome that these individuals account for. For 3+, 4+ and 5+ outcomes, the scale of the vertical axis has been aligned with respect to prevalence. This makes visible that individuals with combinations of outcomes account for an increasingly disproportionate share of the burden. For income (accentuated by a dashed line), which is a benefit rather than a burden to society, the pattern is opposite in that individuals with combinations of outcomes account for a *decreasingly* disproportionate share of total income.

Figure 4: Prediction - Marginal changes in the true positive rate (TPR)



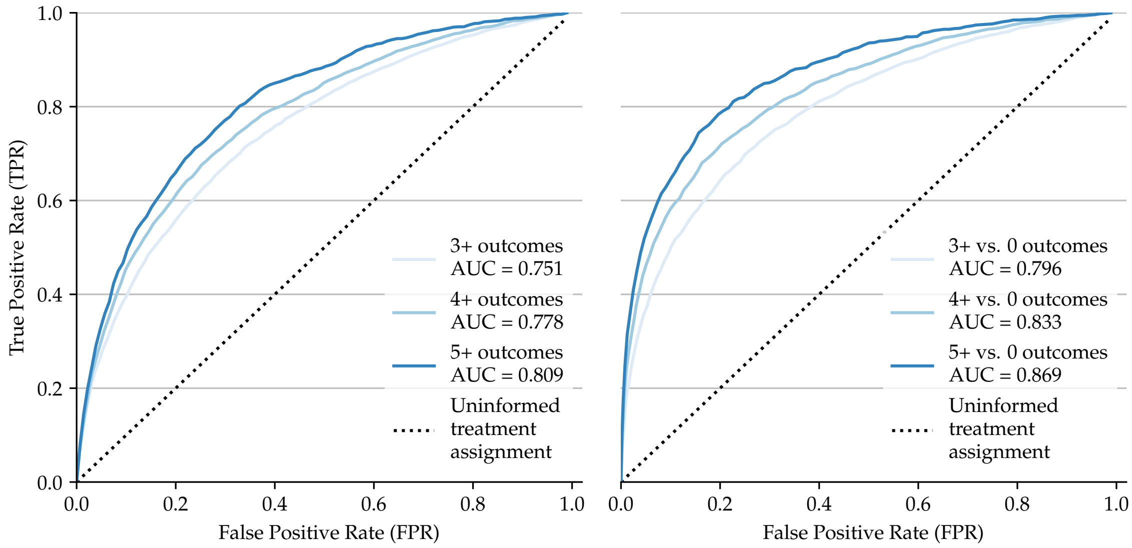
Notes: The solid line is based on predictions from logistic regressions and shows incremental changes in the true positive rate as the fraction of the cohort to be treated is increased in steps of 0.05. The dashed line shows corresponding constant increase under uninformed treatment. The gray area represents 95% bootstrap confidence intervals for the test data set, keeping the prediction function fixed.

Figure 5: Prediction – ROC and AUC



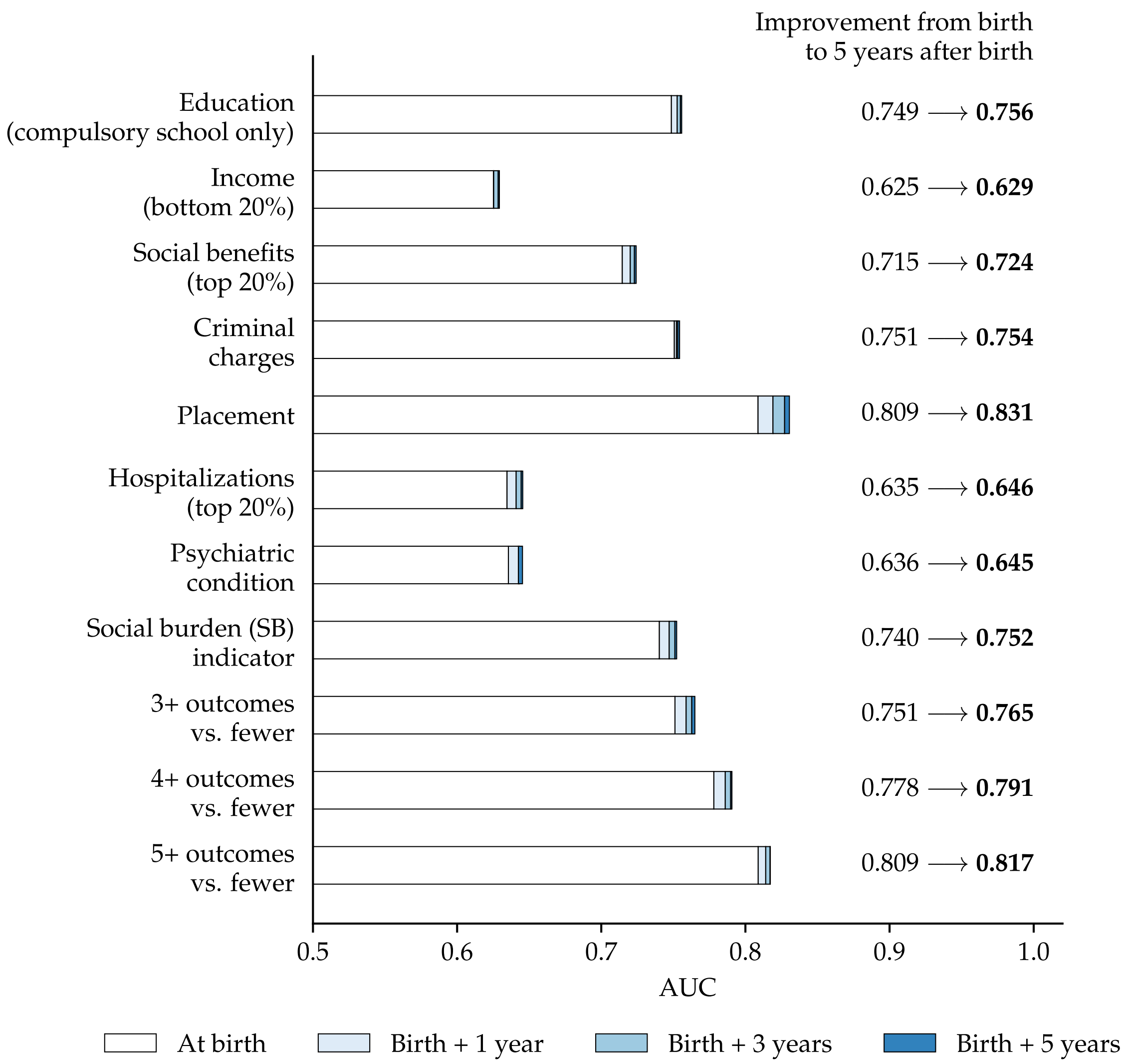
Notes: Receiver operating curve (ROC) and area under the curve (AUC) for predictions from logistic regression. Brackets contain 95% confidence intervals based on 2,000 bootstrap samples from the test data, keeping the prediction function fixed.

Figure 6: Prediction – Combinations of outcomes



Notes: Receiver operating curve (ROC) and area under the curve (AUC) for predictions from logistic regression.

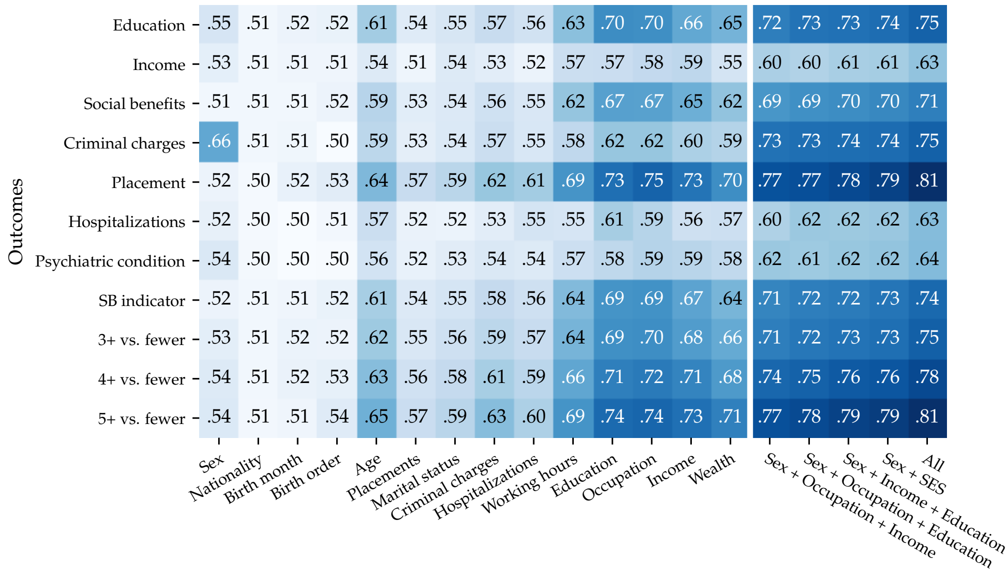
Figure 7: Prediction with post-birth predictors



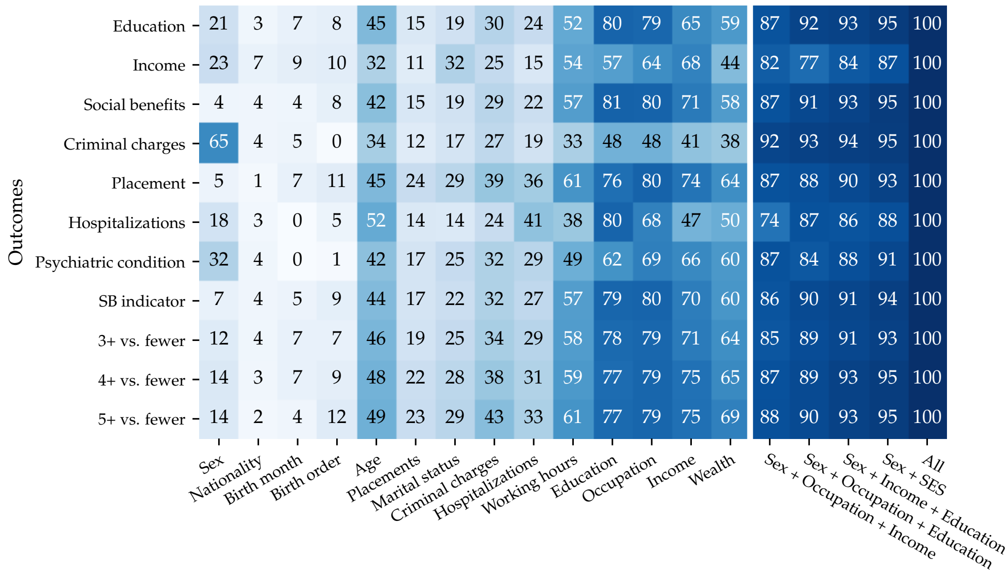
Notes: Improvement in AUC when updating predictors 1, 3 and 5 years after birth and adding the child's hospitalizations as another predictor. All results based on logistic regressions.

Figure 8: Prediction with parsimonious set of predictors – Heatmap

(a) AUC



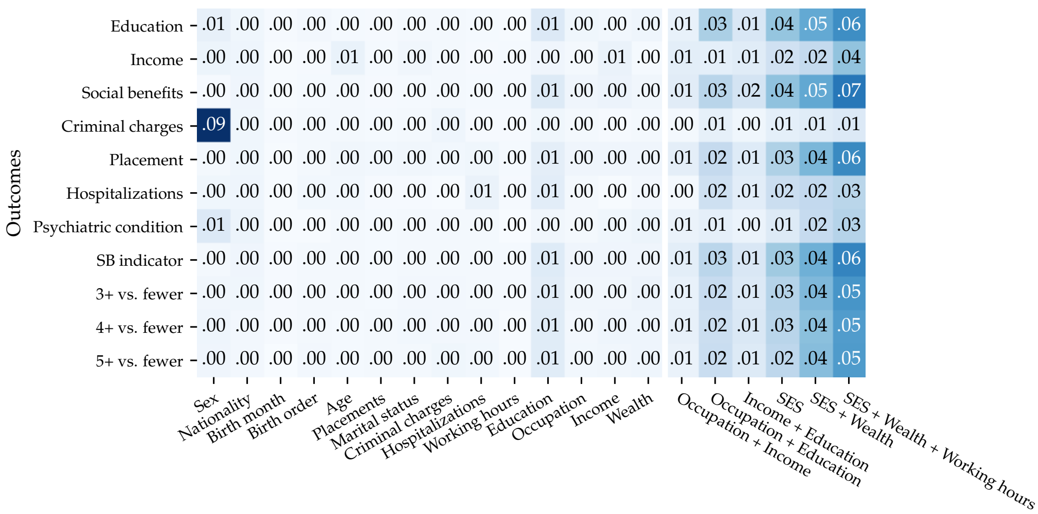
(b) AUC relative to full model (%)



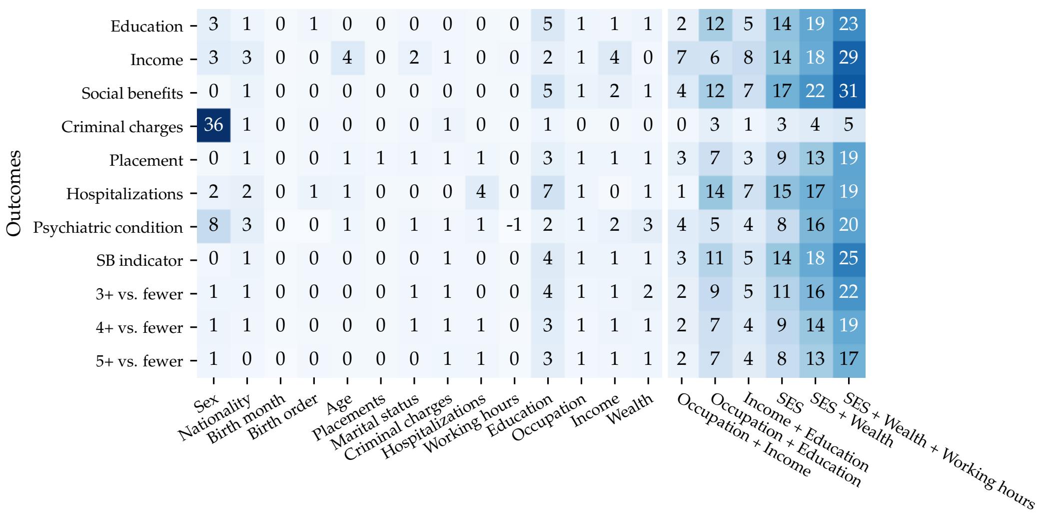
Notes: SES = Income + Education + Occupation. This figure reports the predictive performance of parsimonious models that only take into account a subset of all available predictors. The outcomes are indicated along the vertical axis, while included predictors are indicated along the horizontal axis. All results are from logistic regressions. In the upper panel, the (absolute) AUC is reported. In the lower panel, the ratio of the AUC to the AUC from the full model is reported. Because the AUC ranges between 0.5 and 1.0, we subtract 0.5 from the AUC values before calculating the ratio. Darker cell colors indicate higher (relative) AUC.

Figure 9: Prediction after excluding predictors – Heatmap

(a) AUC reduction

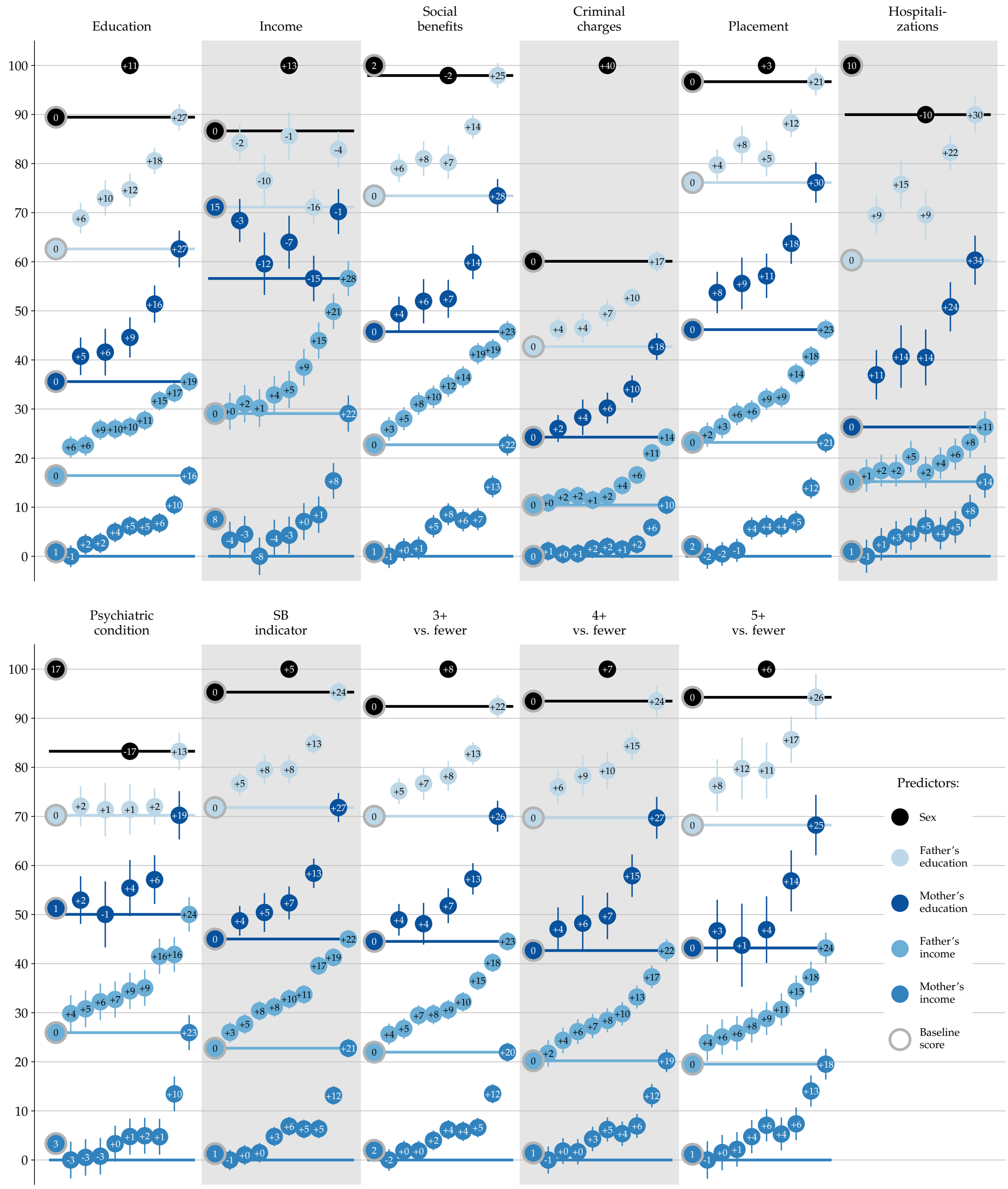


(b) AUC reduction relative to full model (%)



Notes: SES = Income + Education + Occupation. This figure reports how predictive performance changes when when certain predictors are excluded from the full model. The outcomes are indicated along the vertical axis, while excluded predictors are indicated along the horizontal axis. All results are from logistic regressions. In the upper panel, the absolute reduction in AUC is reported. In the lower panel, the reduction in AUC relative to full model's AUC is reported. Because the AUC ranges between 0.5 and 1.0, we subtract 0.5 from the full model AUC before calculating the relative reduction. Darker cell colors indicate higher reductions in AUC.

Figure 10: Optimal risk scores



Notes: This figure illustrates how to construct optimal risk scores. For each outcome, the baseline individual is indicated by the leftmost circles with a gray edge. The individual is has the following characteristics: female, master's degree/PhD (mother/father) and 10th income decile (mother/father). The score of the baseline individual is the sum of the points in the leftmost circles. For individuals with other characteristics, the score is obtained by adding the points indicated to the right of the baseline individual. The characteristics are given in the following order: master's degree/PhD, bachelor/vocational bachelor, short cycle higher education, high school, vocational education and training, compulsory schooling (mother/father) and 10th decile, 9th decile, ..., 1st decile (mother/father). All figures are based on coefficients from logistic regressions that have been rescaled such that the scores range between 0 and 100. Vertical bars are 95% confidence bands based on robust standard errors.

Table 1: Overview of outcomes

Outcome	Prevalence (%)	Fraction of total burden (%)
1. Education	13.2	100.0
Compulsory education only in 2016		
2. Income	20.0	9.4
Bottom 20% equivalised disposable family income 2014-2015		
3. Social benefits	19.9	76.2
Top 20% social benefit recipients (week 14/2006 - week 13/2017)		
4. Criminal charges	18.9	100.0
Ever criminally charged throughout life (as of 2016)		
5. Placement	10.1	100.0
Ever in placement or preventive measure (as a child)		
6. Hospitalizations	19.3	55.4
Top 20% hospital admissions throughout last 10 years (2003-2016)		
7. Psychiatric condition	15.7	100.0
Whether diagnosed with a psychiatric condition (as of 2016)		
8. Social burden (SB) indicator	20.7	100.0
Top 20% score of the single common factor of all 7 outcomes extracted through confirmatory factor analysis		

Notes: Note that income is a benefit rather than a burden to society. The table shows that the bottom 20% of the income distribution receive a disproportionately *small* share of total income.

Table 2: Correlations

	Education	Income	Social benefits	Criminal charges	Placement	Hospitalizations	Psychiatric condition	SB indicator
Education	1.00							
Income	0.35	1.00						
Social benefits	0.64	0.47	1.00					
Criminal charges	0.44	0.26	0.37	1.00				
Placement	0.56	0.28	0.59	0.41	1.00			
Hospitalizations	0.30	0.09	0.35	0.25	0.30	1.00		
Psychiatric condition	0.46	0.38	0.55	0.28	0.49	0.35	1.00	
SB indicator	0.78	0.56	0.92	0.54	0.77	0.47	0.69	1.00

Notes: Reported correlations are polychoric correlations that assume a normal distribution of the underlying latent variables.

Table 3: Comparison of predictions from various machine learning methods

Outcome	AUC (direct)			
	Logit	LASSO	RF	GB
Education (compulsory school only)	0.749 [0.740, 0.756]	0.749 [0.740, 0.756]	0.744 [0.736, 0.752]	0.748 [0.736, 0.758]
Income (bottom 20%)	0.625 [0.617, 0.633]	0.625 [0.617, 0.633]	0.619 [0.611, 0.627]	0.625 [0.616, 0.638]
Social benefits (top 20%)	0.715 [0.708, 0.722]	0.715 [0.707, 0.721]	0.710 [0.702, 0.717]	0.715 [0.700, 0.721]
Criminal charges	0.751 [0.744, 0.758]	0.751 [0.744, 0.758]	0.747 [0.740, 0.754]	0.751 [0.743, 0.763]
Placement	0.809 [0.800, 0.817]	0.809 [0.800, 0.817]	0.802 [0.794, 0.811]	0.811 [0.804, 0.827]
Hospitalizations (top 20%)	0.635 [0.627, 0.643]	0.635 [0.627, 0.643]	0.631 [0.623, 0.639]	0.635 [0.632, 0.654]
Psychiatric condition	0.636 [0.627, 0.645]	0.636 [0.628, 0.645]	0.632 [0.623, 0.641]	0.637 [0.629, 0.647]
Social burden (SB) indicator	0.740 [0.733, 0.747]	0.740 [0.733, 0.747]	0.736 [0.729, 0.743]	0.742 [0.735, 0.748]
3+ vs. fewer outcomes	0.751 [0.743, 0.759]	0.751 [0.743, 0.759]	0.745 [0.737, 0.753]	0.752 [0.744, 0.760]
4+ vs. fewer outcomes	0.778 [0.768, 0.788]	0.778 [0.768, 0.788]	0.772 [0.762, 0.782]	0.778 [0.768, 0.787]
5+ vs. fewer outcomes	0.809 [0.797, 0.821]	0.809 [0.797, 0.821]	0.802 [0.789, 0.815]	0.811 [0.798, 0.822]
3+ vs. 0 outcomes	0.796 [0.788, 0.803]	0.796 [0.788, 0.803]	0.788 [0.780, 0.796]	0.797 [0.789, 0.804]
4+ vs. 0 outcomes	0.833 [0.823, 0.841]	0.833 [0.823, 0.842]	0.823 [0.814, 0.832]	0.832 [0.823, 0.841]
5+ vs. 0 outcomes	0.869 [0.857, 0.879]	0.869 [0.858, 0.880]	0.860 [0.848, 0.871]	0.869 [0.857, 0.879]

Outcome	AUC (grid)			
	Logit	LASSO	RF	GB
Education (compulsory school only)	0.748 [0.740, 0.756]	0.748 [0.739, 0.755]	0.742 [0.734, 0.750]	0.748 [0.734, 0.756]
Income (bottom 20%)	0.624 [0.616, 0.632]	0.625 [0.616, 0.632]	0.616 [0.608, 0.624]	0.624 [0.616, 0.638]
Social benefits (top 20%)	0.714 [0.707, 0.721]	0.714 [0.706, 0.720]	0.708 [0.700, 0.715]	0.714 [0.697, 0.718]
Criminal charges	0.751 [0.744, 0.758]	0.751 [0.743, 0.757]	0.747 [0.740, 0.754]	0.751 [0.744, 0.764]
Placement	0.806 [0.798, 0.815]	0.806 [0.797, 0.814]	0.800 [0.792, 0.809]	0.809 [0.801, 0.825]
Hospitalizations (top 20%)	0.634 [0.626, 0.642]	0.635 [0.626, 0.642]	0.631 [0.622, 0.638]	0.635 [0.630, 0.652]
Psychiatric condition	0.636 [0.627, 0.645]	0.636 [0.627, 0.645]	0.633 [0.624, 0.641]	0.637 [0.629, 0.646]
Social burden (SB) indicator	0.739 [0.732, 0.746]	0.739 [0.732, 0.745]	0.734 [0.726, 0.741]	0.741 [0.733, 0.747]
3+ vs. fewer outcomes	0.750 [0.742, 0.758]	0.750 [0.742, 0.758]	0.744 [0.736, 0.752]	0.750 [0.742, 0.758]
4+ vs. fewer outcomes	0.777 [0.767, 0.787]	0.777 [0.767, 0.786]	0.770 [0.760, 0.780]	0.777 [0.767, 0.786]
5+ vs. fewer outcomes	0.808 [0.796, 0.819]	0.808 [0.795, 0.819]	0.801 [0.789, 0.813]	0.810 [0.797, 0.821]
3+ vs. 0 outcomes	0.794 [0.786, 0.802]	0.794 [0.786, 0.802]	0.786 [0.778, 0.794]	0.794 [0.786, 0.802]
4+ vs. 0 outcomes	0.832 [0.823, 0.841]	0.832 [0.823, 0.841]	0.822 [0.812, 0.831]	0.830 [0.821, 0.840]
5+ vs. 0 outcomes	0.867 [0.855, 0.877]	0.867 [0.855, 0.878]	0.856 [0.844, 0.868]	0.867 [0.856, 0.878]

Notes: Comparison of estimates from various methods. AUC (direct) comes from directly tuning the models with respect to AUC maximization. AUC (grid) first tunes the models over grid of 20 β values with respect to TPR maximization. The AUC is then constructed from the optimal TPR values. AUC (grid) may be larger than AUC (direct) because the models are fine-tuned along the ROC curve. However, AUC (grid) may also be smaller than AUC (direct) because of approximating a concave function with a finite grid. Brackets contain 95% confidence intervals based on 2,000 bootstrap samples from the test data, keeping the prediction function fixed.

Appendix

A Derivation of the expected benefit $B(T, \beta)$

This section shows that the expected benefit $B(T, \beta) = \big(Pr(Y = 1 \mid \beta = 0) - Pr(Y = 1 \mid \beta)\big)Cost_{Outcome}$ can be written as

$$B(T, \beta) = \alpha Pr(T = 1 \mid R = 1, \beta)\bar{\delta} Cost_{Outcome}. \tag{A.1}$$

First, note that $Pr(Y = 1 \mid \beta)$ can be rewritten as follows, where we only condition on β if the expression depends on it:

$$Pr(Y = 1 \mid \beta) = Pr(R = 1)Pr(Y = 1 \mid R = 1, \beta) + Pr(R = 0)Pr(Y = 1 \mid R = 0).$$

Because of $Pr(R = 1) = \alpha$ and the assumption that $Pr(Y = 1 \mid R = 0) = 0$, it follows that

$$Pr(Y = 1 \mid \beta) = \alpha Pr(Y = 1 \mid R = 1, \beta).$$

Expanding this yields

$$\begin{aligned} Pr(Y = 1 \mid \beta) = \alpha \Big[& Pr(T = 1 \mid R = 1, \beta)Pr(Y = 1 \mid R = 1, T = 1) \\ & + Pr(T = 0 \mid R = 1, \beta)Pr(Y = 1 \mid R = 1, T = 0) \Big]. \end{aligned} \tag{A.2}$$

For $\beta = 0$, note that $Pr(T = 1 \mid R = 1, \beta = 0) = 0$ and $Pr(T = 0 \mid R = 1, \beta = 0) = 1$. Moreover, by assumption, $Pr(Y = 1 \mid R = 1, T = 0) = 1$. Hence, it follows from equation A.2 that

$$Pr(Y = 1 \mid \beta = 0) = \alpha. \tag{A.3}$$

For general β, we proceed by subtracting and adding $Pr(T = 1 \mid R = 1, \beta)Pr(Y = 1 \mid R = 1, T = 0)$ to the right hand side of equation A.2,

$$\begin{aligned} & Pr(Y = 1 \mid \beta) \\ & = \alpha \Big[Pr(T = 1 \mid R = 1, \beta) \Big[Pr(Y = 1 \mid R = 1, T = 1) - Pr(Y = 1 \mid R = 1, T = 0) \Big] \\ & \quad + \Big[Pr(T = 0 \mid R = 1, \beta) + Pr(T = 1 \mid R = 1, \beta) \Big] Pr(Y = 1 \mid R = 1, T = 0) \Big] \\ & = \alpha \Big[Pr(T = 1 \mid R = 1, \beta) \Big[Pr(Y = 1 \mid R = 1, T = 1) - Pr(Y = 1 \mid R = 1, T = 0) \Big] \\ & \quad + Pr(Y = 1 \mid R = 1, T = 0) \Big]. \end{aligned}$$

Because $Pr(Y = 1 \mid R = 1, T = 0) = 1$,

$$\begin{aligned} Pr(Y = 1 \mid \beta) &= \alpha \Big[Pr(T = 1 \mid R = 1, \beta) \big[Pr(Y = 1 \mid R = 1, T = 1) - 1 \big] + 1 \Big] \\ &= \alpha Pr(T = 1 \mid R = 1, \beta) \big[Pr(Y = 1 \mid R = 1, T = 1) - 1 \big] + \alpha. \end{aligned} \tag{A.4}$$

Combining equations A.3 and A.4, we can see that the expected benefit $B(T, \beta)$ of the intervention is

$$\begin{aligned} B(T, \beta) &= \big(Pr(Y = 1 \mid \beta = 0) - Pr(Y = 1 \mid \beta) \big) Cost_{Outcome} \\ &= \alpha Pr(T = 1 \mid R = 1, \beta) \big[1 - Pr(Y = 1 \mid R = 1, T = 1) \big] Cost_{Outcome}. \end{aligned}$$

Note that $Pr(Y = 1 \mid R = 1, T = 1)$ depends on θ, the susceptibility to treatment.

$$\begin{aligned} Pr(Y = 1 \mid R = 1, T = 1) &= \int Pr(Y = 1 \mid R = 1, T = 1, \theta) f(\theta \mid R = 1, T = 1) d\theta \\ &= \int 1 - \delta(\theta) f(\theta \mid R = 1, T = 1) d\theta. \end{aligned}$$

Hence,

$$B(T, \beta) = \alpha Pr(T = 1 \mid R = 1, \beta) \int \delta(\theta) f(\theta \mid R = 1, T = 1) d\theta Cost_{Outcome}, \tag{A.5}$$

where θ is the susceptibility to treatment. Equation A.5 shows that in order to maximize the expected benefit of the intervention, the treatment ($T(X)$) should be administered in such a way that, first, the probability that at-risk children receive it is as large as possible ($Pr(T = 1 \mid R = 1, \beta)$), i.e. the intervention should be targeted at at-risk children. Second, among at-risk children, treatment should be given to those with the largest susceptibility θ and thus the largest reduction in the outcome probability in response to the intervention ($\delta(\theta)$). In practice, however, the treatment susceptibility θ is typically unknown. We therefore make the simplifying assumption that the benefit of the intervention is constant across at-risk children such that $\delta(\theta) = \bar{\delta}$:

$$B(T, \beta) = \alpha Pr(T = 1 \mid R = 1, \beta) \bar{\delta} Cost_{Outcome}. \tag{A.6}$$

B Derivation of the false positive rate $FPR(\beta)$

First, use Bayes' rule to write

$$\begin{aligned} FPR(\beta) &= Pr(T = 1 \mid R = 0, \beta) \\ &= \frac{Pr(R = 0 \mid T = 1, \beta) Pr(T = 1)}{Pr(R = 0)}. \end{aligned}$$

Since $Pr(T = 1) = \beta$ and $Pr(R = 0) = 1 - Pr(R = 1) = 1 - \alpha$, we get

$$\begin{aligned} FPR(\beta) &= \frac{Pr(R = 0 \mid T = 1, \beta)\beta}{1 - \alpha} \\ &= \frac{(1 - Pr(R = 1 \mid T = 1, \beta))\beta}{1 - \alpha} \end{aligned}$$

Finally, we use that $TPR(\beta) = Pr(R = 1 \mid T = 1, \beta)\beta/\alpha$ (see Section 3.2):

$$\begin{aligned} FPR(\beta) &= \frac{(1 - (\alpha/\beta) TPR(\beta))\beta}{1 - \alpha} \\ &= \frac{\beta - \alpha TPR(\beta)}{1 - \alpha}. \end{aligned}$$

C Optimal machine learning tuning parameters

C.1 LASSO

We implement the LASSO using the `glmnet` package in R (Friedman et al. 2010) and specify the following parameters:

Parameter	Description	Values
`lambda`	LASSO penalty	0.15×0.9^x for x in $[0, 1, 2, \ldots, 149]$
`alpha`	Elastic net mixing parameter. `alpha` $= 1$ yields the LASSO model.	1

Parameters not mentioned in the table are left at their default value. We determine the optimal values for `lambda` using 8-fold cross-validation. The table below present the results of the tuning (all numbers are multiplied by 10,000 to enhance readability).

Table C.1: `lambda`

Metric	Education	Income	Social benefits	Criminal charges	Placement	Hospitali-zations	Psychiatric condition	SB indicator	At least 3 vs. fewer	At least 4 vs. fewer	At least 5 vs. fewer
TPR = 0.05	4.108	17.959	9.544	2.995	1.289	3.328	0.001	2.696	0.029	1.160	0.761
TPR = 0.1	8.590	22.171	0.761	9.544	1.592	5.636	2.995	0.007	5.636	0.364	0.083
TPR = 0.15	6.958	7.731	10.604	9.544	1.592	9.544	7.731	14.547	4.565	0.001	0.940
TPR = 0.2	3.328	0.000	4.565	1.592	6.958	0.075	9.544	1.965	9.544	0.026	0.067
TPR = 0.25	2.426	6.958	10.604	19.954	4.565	8.590	4.565	0.001	0.450	1.433	7.731
TPR = 0.3	5.072	3.328	3.328	0.011	2.995	0.103	3.328	0.000	3.698	2.426	1.160
TPR = 0.35	0.846	9.544	0.265	0.239	0.328	5.072	10.604	8.590	4.108	0.007	5.072
TPR = 0.4	3.698	4.108	9.544	9.544	4.565	4.108	0.364	0.001	0.405	3.698	5.636
TPR = 0.45	1.044	10.604	0.617	13.092	6.262	13.092	11.783	4.108	6.262	0.024	0.004
TPR = 0.5	0.157	6.958	0.555	16.163	0.000	16.163	3.328	2.696	4.565	1.592	0.265
TPR = 0.55	6.262	6.958	14.547	9.544	6.958	9.544	10.604	3.328	1.965	0.055	1.044
TPR = 0.6	2.995	16.163	11.783	7.731	0.364	24.635	10.604	2.696	3.698	0.075	3.698
TPR = 0.65	1.592	9.544	13.092	7.731	1.433	24.635	2.696	8.590	6.262	0.024	1.768
TPR = 0.7	10.604	3.698	4.108	8.590	11.783	4.565	0.029	5.636	8.590	3.698	2.696
TPR = 0.75	3.698	13.092	6.262	7.731	9.544	10.604	1.289	11.783	0.103	1.768	0.239
TPR = 0.8	1.965	2.995	0.685	0.846	4.108	22.171	4.565	0.044	1.592	0.450	1.768
TPR = 0.85	0.450	5.636	17.959	16.163	0.194	4.108	6.262	9.544	1.160	9.544	6.262
TPR = 0.9	37.547	1.768	0.215	6.958	11.783	24.635	14.547	27.372	1.289	22.171	14.547
TPR = 0.95	11.783	2.995	2.995	5.072	0.265	2.426	22.171	5.636	0.127	2.696	8.590
AUC	3.328	5.072	3.698	2.696	3.328	9.544	5.636	1.768	2.183	0.499	1.044

C.2 Random Forest

We implement the Random Forest using the `ranger` package in R (Wright and Ziegler 2017) and specify the following parameters:

Parameter	Description	Values
`min.node.size`	Minimum node size	5, 10, 20, 50, 100, 200, 400, 1000, 2000, 4000, 8000
`mtry`	Fraction of random predictors used	0.01, 0.03, 0.05, 0.1, 0.2, 0.3
`replace`	Bootstrap training data	False
`num.trees`	Number of trees	500

Parameters not mentioned in the table are left at their default value. We determine the optimal values for `min.node.size` and `mtry` using 8-fold cross-validation. The tables below present the results of the tuning.

Table C.2: `min.node.size`

Metric	Education	Income	Social benefits	Criminal charges	Placement	Hospitalizations	Psychiatric condition	SB indicator	At least 3 vs. fewer	At least 4 vs. fewer	At least 5 vs. fewer
TPR = 0.05	50	200	100	50	5	400	200	50	10	20	20
TPR = 0.1	200	100	10	400	5	200	100	20	20	50	50
TPR = 0.15	50	5	100	50	5	1000	100	20	50	20	20
TPR = 0.2	200	50	20	200	5	50	100	10	10	10	20
TPR = 0.25	10	400	50	400	20	100	20	20	20	20	50
TPR = 0.3	100	400	50	200	20	200	100	20	5	200	100
TPR = 0.35	50	100	50	200	20	1000	200	20	50	50	100
TPR = 0.4	100	100	50	400	20	400	400	10	20	5	200
TPR = 0.45	20	200	20	2000	10	200	200	10	50	100	50
TPR = 0.5	20	100	10	1000	10	400	400	20	10	100	400
TPR = 0.55	200	400	20	1000	10	400	400	20	10	100	100
TPR = 0.6	400	1000	20	400	5	400	2000	5	400	400	400
TPR = 0.65	200	200	200	8000	5	100	1000	5	400	1000	1000
TPR = 0.7	200	200	400	200	100	50	1000	20	400	400	1000
TPR = 0.75	200	200	100	400	200	100	400	10	400	1000	8000
TPR = 0.8	8000	400	1000	200	8000	400	400	400	400	4000	8000
TPR = 0.85	2000	400	4000	200	8000	400	2000	2000	4000	4000	8000
TPR = 0.9	8000	50	4000	2000	8000	2000	200	8000	8000	8000	5
TPR = 0.95	4000	20	1000	8000	2000	2000	20	8000	8000	20	8000
AUC	100	200	50	200	10	400	400	20	10	10	100

Table C.3: `mtry`

Metric	Education	Income	Social benefits	Criminal charges	Placement	Hospitali-zations	Psychiatric condition	SB indicator	At least 3 vs. fewer	At least 4 vs. fewer	At least 5 vs. fewer
TPR = 0.05	0.05	0.05	0.03	0.20	0.10	0.30	0.03	0.03	0.03	0.01	0.01
TPR = 0.1	0.10	0.05	0.03	0.10	0.03	0.10	0.05	0.05	0.03	0.05	0.03
TPR = 0.15	0.05	0.03	0.03	0.10	0.03	0.05	0.03	0.03	0.05	0.03	0.03
TPR = 0.2	0.10	0.03	0.03	0.20	0.03	0.03	0.05	0.03	0.03	0.03	0.03
TPR = 0.25	0.03	0.05	0.03	0.20	0.03	0.03	0.05	0.03	0.05	0.03	0.03
TPR = 0.3	0.10	0.10	0.05	0.20	0.03	0.05	0.05	0.03	0.03	0.10	0.03
TPR = 0.35	0.05	0.03	0.05	0.20	0.03	0.10	0.05	0.05	0.03	0.03	0.03
TPR = 0.4	0.10	0.03	0.05	0.30	0.03	0.05	0.10	0.03	0.03	0.03	0.05
TPR = 0.45	0.05	0.05	0.03	0.30	0.05	0.03	0.05	0.03	0.10	0.05	0.03
TPR = 0.5	0.05	0.05	0.03	0.30	0.03	0.05	0.10	0.03	0.03	0.10	0.10
TPR = 0.55	0.10	0.30	0.03	0.30	0.03	0.05	0.10	0.03	0.03	0.03	0.03
TPR = 0.6	0.20	0.20	0.03	0.30	0.03	0.05	0.20	0.03	0.20	0.30	0.03
TPR = 0.65	0.05	0.10	0.05	0.20	0.03	0.05	0.10	0.03	0.30	0.30	0.30
TPR = 0.7	0.05	0.20	0.03	0.10	0.03	0.03	0.10	0.03	0.30	0.10	0.30
TPR = 0.75	0.10	0.30	0.03	0.05	0.05	0.03	0.10	0.03	0.20	0.30	0.30
TPR = 0.8	0.30	0.10	0.30	0.05	0.30	0.05	0.05	0.10	0.20	0.30	0.30
TPR = 0.85	0.30	0.10	0.30	0.03	0.30	0.05	0.10	0.30	0.30	0.30	0.30
TPR = 0.9	0.30	0.01	0.30	0.03	0.30	0.30	0.05	0.30	0.30	0.30	0.01
TPR = 0.95	0.30	0.01	0.10	0.03	0.30	0.20	0.03	0.30	0.30	0.03	0.30
AUC	0.05	0.05	0.03	0.20	0.03	0.05	0.10	0.03	0.03	0.03	0.03

C.3 Gradient Boosting

We implement the Gradient Boosting using the `XGBoost` package in R (Chen and Guestrin 2016) and specify the following parameters:

Parameter	Description	Values
`eta`	Shrinkage step size	0.01
`subsample`	Fraction of random observations used	$0.1, 0.3, 0.5, 0.7$
`colsample_bytree`	Fraction of random predictors used	$0.5, 0.75, 1.0$
`max_depth`	Maximum depth of tree	$1, 2, 4, 6, 8$
`nrounds`	Number of trees	$1000, 1500, 2000, 2500, 3000, 3500$

Parameters not mentioned in the table are left at their default value. We determine the optimal values for `subsample`, `max_depth`, `colsample_bytree` and `nrounds` using 8-fold cross-validation. The tables below present the results of the tuning.

Table C.4: `subsample`

Metric	Education	Income	Social benefits	Criminal charges	Placement	Hospitali-zations	Psychiatric condition	SB indicator	At least 3 vs. fewer	At least 4 vs. fewer	At least 5 vs. fewer
TPR = 0.05	0.3	0.7	0.1	0.3	0.3	0.1	0.3	0.5	0.7	0.7	0.7
TPR = 0.1	0.3	0.7	0.1	0.1	0.3	0.7	0.3	0.3	0.1	0.1	0.7
TPR = 0.15	0.5	0.5	0.5	0.1	0.5	0.7	0.3	0.5	0.1	0.5	0.5
TPR = 0.2	0.3	0.1	0.3	0.1	0.7	0.1	0.7	0.3	0.1	0.3	0.5
TPR = 0.25	0.1	0.5	0.5	0.1	0.3	0.1	0.1	0.3	0.3	0.3	0.5
TPR = 0.3	0.1	0.1	0.5	0.3	0.7	0.3	0.5	0.5	0.5	0.5	0.5
TPR = 0.35	0.3	0.1	0.7	0.5	0.3	0.1	0.5	0.5	0.1	0.3	0.3
TPR = 0.4	0.3	0.1	0.7	0.3	0.5	0.1	0.7	0.3	0.3	0.5	0.5
TPR = 0.45	0.1	0.1	0.3	0.7	0.3	0.3	0.5	0.5	0.1	0.5	0.5
TPR = 0.5	0.3	0.5	0.5	0.7	0.3	0.3	0.7	0.1	0.3	0.1	0.5
TPR = 0.55	0.5	0.1	0.7	0.7	0.3	0.7	0.3	0.3	0.3	0.1	0.3
TPR = 0.6	0.5	0.7	0.3	0.7	0.1	0.5	0.3	0.3	0.3	0.3	0.1
TPR = 0.65	0.3	0.3	0.3	0.7	0.5	0.1	0.3	0.3	0.1	0.3	0.3
TPR = 0.7	0.3	0.5	0.3	0.5	0.7	0.7	0.3	0.7	0.3	0.5	0.7
TPR = 0.75	0.7	0.3	0.3	0.5	0.7	0.5	0.3	0.5	0.3	0.7	0.5
TPR = 0.8	0.5	0.1	0.5	0.7	0.1	0.7	0.3	0.1	0.3	0.5	0.7
TPR = 0.85	0.3	0.1	0.3	0.5	0.3	0.5	0.7	0.1	0.1	0.3	0.7
TPR = 0.9	0.5	0.1	0.1	0.3	0.7	0.5	0.7	0.7	0.5	0.7	0.1
TPR = 0.95	0.7	0.1	0.3	0.1	0.7	0.5	0.5	0.7	0.7	0.3	0.7
AUC	0.5	0.1	0.5	0.3	0.3	0.7	0.5	0.3	0.1	0.5	0.5

Table C.5: `max_depth`

Metric	Education	Income	Social benefits	Criminal charges	Placement	Hospitali-zations	Psychiatric condition	SB indicator	At least 3 vs. fewer	At least 4 vs. fewer	At least 5 vs. fewer
TPR = 0.05	2	2	6	4	6	4	4	2	4	2	2
TPR = 0.1	4	4	4	4	4	4	4	4	4	4	4
TPR = 0.15	6	4	6	2	8	2	4	6	4	4	4
TPR = 0.2	2	2	4	4	2	2	2	6	4	4	4
TPR = 0.25	2	4	8	2	6	1	4	4	4	2	4
TPR = 0.3	6	4	4	6	4	4	6	2	4	4	4
TPR = 0.35	4	2	4	4	4	2	2	4	4	6	4
TPR = 0.4	4	6	6	6	6	2	6	4	4	4	4
TPR = 0.45	4	4	6	4	4	4	4	6	4	4	4
TPR = 0.5	2	4	4	2	4	4	6	4	6	4	4
TPR = 0.55	2	4	6	4	6	2	4	6	4	4	4
TPR = 0.6	4	2	4	2	4	4	2	6	8	4	2
TPR = 0.65	6	6	4	2	4	4	2	6	4	4	4
TPR = 0.7	6	2	4	1	2	2	4	4	4	6	2
TPR = 0.75	4	2	2	4	4	2	4	6	4	8	2
TPR = 0.8	4	2	6	4	4	1	4	4	6	6	2
TPR = 0.85	4	2	6	4	4	2	2	4	4	8	4
TPR = 0.9	2	6	4	6	4	2	2	4	4	4	6
TPR = 0.95	4	6	2	2	6	4	2	8	6	6	2
AUC	4	2	6	4	4	2	4	4	4	4	2

Table C.6: `colsample_bytree`

Metric	Education	Income	Social benefits	Criminal charges	Placement	Hospitali-zations	Psychiatric condition	SB indicator	At least 3 vs. fewer	At least 4 vs. fewer	At least 5 vs. fewer
TPR = 0.05	0.50	0.75	0.75	0.50	0.75	0.75	0.50	1.00	0.75	0.50	0.50
TPR = 0.1	1.00	0.50	0.50	0.75	1.00	0.75	0.75	1.00	0.50	1.00	0.75
TPR = 0.15	0.50	0.50	0.50	0.75	0.50	0.50	0.75	0.50	0.75	0.50	0.50
TPR = 0.2	0.50	0.50	0.75	0.50	1.00	0.75	1.00	0.75	0.75	0.50	1.00
TPR = 0.25	0.75	0.50	0.50	0.50	1.00	0.75	0.75	0.50	0.75	1.00	0.50
TPR = 0.3	0.50	0.75	0.50	1.00	0.50	0.75	0.50	0.75	1.00	0.50	0.50
TPR = 0.35	0.75	0.50	0.50	0.75	0.50	0.50	0.50	0.75	0.75	0.50	0.50
TPR = 0.4	0.50	0.50	0.50	0.50	0.75	0.50	0.50	1.00	1.00	1.00	1.00
TPR = 0.45	1.00	0.50	0.75	0.75	0.75	1.00	1.00	0.50	0.75	0.50	0.50
TPR = 0.5	0.75	0.50	0.75	0.75	0.75	1.00	1.00	1.00	0.50	1.00	0.50
TPR = 0.55	0.50	0.50	0.50	0.50	0.75	0.50	1.00	0.50	0.50	0.75	0.50
TPR = 0.6	0.50	1.00	0.75	0.75	0.50	0.50	0.50	0.50	0.75	0.75	0.75
TPR = 0.65	1.00	0.50	0.75	0.50	0.75	0.75	0.50	0.75	0.75	0.50	0.75
TPR = 0.7	0.50	0.75	1.00	1.00	0.75	0.75	1.00	0.50	0.50	0.75	0.75
TPR = 0.75	0.50	0.50	0.50	1.00	0.75	1.00	0.50	0.50	0.50	0.50	0.75
TPR = 0.8	0.50	1.00	0.50	0.50	1.00	1.00	0.75	0.50	0.50	0.75	0.75
TPR = 0.85	1.00	0.75	0.50	1.00	1.00	0.75	1.00	0.50	0.75	0.50	0.50
TPR = 0.9	0.50	0.50	0.75	0.75	0.75	0.75	0.50	0.75	0.75	1.00	0.50
TPR = 0.95	0.75	0.50	0.50	1.00	0.50	0.75	0.50	0.50	1.00	0.50	1.00
AUC	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.75	0.50	0.50

Table C.7: `nrounds`

Metric	Education	Income	Social benefits	Criminal charges	Placement	Hospitalizations	Psychiatric condition	SB indicator	At least 3 vs. fewer	At least 4 vs. fewer	At least 5 vs. fewer
TPR = 0.05	3500	3000	1000	1000	1500	1500	1500	2500	1500	3000	3000
TPR = 0.1	1500	1500	2000	1500	1500	1000	2500	3500	2000	2000	2000
TPR = 0.15	1500	1500	1000	1500	1000	3500	1500	2000	2000	2000	2000
TPR = 0.2	3500	3000	2500	1000	3500	3500	2500	2000	2000	2000	2000
TPR = 0.25	3500	1500	1000	3500	1000	3000	1000	3000	2500	3000	1500
TPR = 0.3	1000	2000	2500	1000	2500	2000	1000	3500	3500	2500	2000
TPR = 0.35	3000	2000	2000	1500	3500	3500	3000	1500	2500	2000	2000
TPR = 0.4	1500	2000	1000	1000	1500	1500	1000	2000	3000	1500	1500
TPR = 0.45	2000	1000	1500	1500	2000	1500	2500	1500	2500	2000	2000
TPR = 0.5	3500	2000	2000	3000	3500	1000	1000	2500	1500	3000	2000
TPR = 0.55	2500	1000	1000	2000	1000	2500	1500	1000	3000	3500	2500
TPR = 0.6	1500	3500	3000	3500	3000	2500	3000	1000	1000	2000	3500
TPR = 0.65	1500	1000	2500	3500	2500	1000	3500	1000	1500	2500	1500
TPR = 0.7	1500	3000	1500	3500	2500	1000	2000	1500	3000	1000	3000
TPR = 0.75	1500	3500	3500	3500	1000	1500	2500	1500	2500	1000	3500
TPR = 0.8	3000	3000	2000	3000	3500	3000	2500	2500	1500	1000	3500
TPR = 0.85	3000	3500	1000	1500	3500	3500	3500	2500	3000	1000	2000
TPR = 0.9	1500	2000	1500	1000	1000	2000	2500	2000	3000	1500	1000
TPR = 0.95	1500	1000	3500	2000	3500	1500	3000	1000	2500	2500	3000
AUC	2000	3500	1000	1500	2500	3000	1000	2500	2000	2000	3500

D Tables

Table D.1: Overview of outcomes

Outcome	Prevalence (%)	Data source	Description
1. Education	13.2	Education register (UDDA)	Compulsory education only (including ongoing education) as of October 1, 2016.
2. Income	20.0	Family income register (FAIK)	Bottom 20% equivalised disposable family income within birth year. Income is averaged across years 2014 and 2015. Equivalisation is performed according to the OECD-modified scale.
3. Social benefits	19.9	DREAM database	Top 20% of weeks of social benefit receipt within birth year (weeks 14/2006 – 13/2017). Social benefits included: unemployment benefits, social assistance, early retirement benefit, integration benefit, rehabilitation benefit, vocational rehabilitation programme.
4. Criminal charges	18.9	Preliminary criminal charges (KRSI)	Ever criminally charged. Measured through the individual's birthday in 2016.
5. Placement	10.1	Placements and preventive action registers (BUA/BUFO)	Ever in placement or preventive measure (as a child). Measured at the individual's birthday in 2015.
6. Hospitalizations	19.3	National inpatient register (LPR)	Top 20% hospital admissions within birth year. Measured in the period between the individual's birthdays in 2006 and 2016. We exclude visits related to pregnancy and child birth (ICD-10-codes starting with "Z3" and "O").
7. Psychiatric condition	15.7	Inpatient and outpatient psychiatric registers	Measured in the period between the individual's 10th birthday and the birthady in 2016. Includes ICD-10-codes starting with "F".
8. Social burden (SB) indicator	20.7	-	Top 20% score resulting from confirmatory factor analysis with one factor underlying all 7 outcomes. The corresponding factor loadings are shown in Table D.2 along with several measures of fit. Model fit is very good overall.

Table D.2: Factor loadings

Outcome	Loading	95% confidence interval
Education (compulsory school only)	1.00	-
Social benefits (top 20%)	1.38	[1.36-1.40]
Criminal charges	0.73	[0.72-0.75]
Placement	0.80	[0.79-0.81]
Income (bottom 20%)	0.75	[0.73-0.76]
Hospitalizations (top 20%)	0.60	[0.59-0.62]
Psychiatric condition	0.93	[0.91-0.94]
Measures of fit:		
χ^2 (N=149,755, df=14)	4028.7 (p-value: 0.00)	
RMSEA	0.044, 90% CI: [0.043, 0.045]	
CFI	0.963	
TLI	0.945	

Notes: This table shows factor loadings and measures of fit from the confirmatory factor analysis with a single factor underlying all 7 outcomes. The loading for education is normalized to 1.

Table D.3: Overview of predictors

Predictor	Data source	Description
Sex	Birth register (FTDB)	Sex at birth.
Nationality	Population register (BEF)	Non-Danish persons are individuals who have no parent that is both Danish citizen and born in Denmark. Depending on whether they are born in Denmark or abroad, they are descendants or immigrants, respectively. Measured in the year after birth.
Birth month	Birth register (FTDB)	Born in January, February etc.
Birth order	Birth register (FTDB)	Number of children previously born to the same mother plus 1.
Income	Income register (IND)	Mother's/father's average income in the three years before the child's year of birth.
Wealth	Income register (IND)	Mother's/father's average wealth in the three years before the child's year of birth.
Education	Education register (UDDA)	Mother's/father's highest education (including ongoing education) as of October 1 in the year before the child's year of birth. Levels: master's degree/PhD (2-5 years); bachelor's degree (e.g., economics and business administration; 3 years) / vocational bachelor's degree (e.g. nurse, primary education teacher; 2-4 years); short cycle higher education (e.g., information technologist, real estate agent; 1-2 years); high school (3 years); vocational education and training (e.g., carpenter, electrician; ca. 4 years); compulsory schooling (9 years)
Occupation	Employment register - Persons (IDAP)	Mother's/father's occupation in the year before the child's year of birth.
Working hours	Employment register - Spells (IDAN)	Mother's/father's working hours in main employment in the year before the child's year of birth.
Age	Population register (BEF)	Mother's/father's age at the time of birth.
Marital status	Households and families (FAIN)	Mother's/father's marital status at the beginning of the child's year of birth.
Hospitalizations	National inpatient register (LPR)	Mother's/father's hospitalizations in the 5 years before the child's year of birth. Excludes visits related to pregnancy and child birth (ICD-8-codes starting with "Y6" and with "63", "64", "65", "66", "67", "76", "77").
Placements	Placement register (BUO)	Mother's/father's days spent in placement as a child.
Criminal charges	Preliminary criminal charges (KRSI)	Mother's/father's criminal charges between January 1, 1980 and the child's day of birth. Excludes traffic-related charges.

Table D.4: Descriptive statistics - Predictors

Predictor	Mean	Std.dev.
Nationality		
Non-Danish	0.010	0.098
Danish	0.983	0.131
Sex		
Female	0.485	0.500
Male	0.515	0.500
Birth month		
January	0.075	0.264
February	0.075	0.263
March	0.087	0.283
April	0.091	0.287
May	0.090	0.286
June	0.088	0.283
July	0.091	0.287
August	0.087	0.282
September	0.086	0.280
October	0.080	0.272
November	0.075	0.263
December	0.075	0.264
Birth order		
1	0.450	0.497
2	0.374	0.484
3	0.133	0.340
4	0.032	0.176
5+	0.011	0.104
Income		
Father: 1st decile	0.100	0.300
Father: 2nd decile	0.100	0.300
Father: 3rd decile	0.100	0.300
Father: 4th decile	0.100	0.300
Father: 5th decile	0.100	0.300
Father: 6th decile	0.100	0.300
Father: 7th decile	0.100	0.300
Father: 8th decile	0.100	0.300
Father: 9th decile	0.100	0.300
Father: 10th decile	0.100	0.300
Mother: 1st decile	0.100	0.300
Mother: 2nd decile	0.100	0.300
Mother: 3rd decile	0.100	0.300
Mother: 4th decile	0.100	0.300
Mother: 5th decile	0.100	0.300
Mother: 6th decile	0.100	0.300
Mother: 7th decile	0.100	0.300
Mother: 8th decile	0.100	0.300
Mother: 9th decile	0.100	0.300
Mother: 10th decile	0.100	0.300
Wealth		
Father: 1st decile	0.100	0.300
Father: 2nd decile	0.100	0.300
Father: 3rd decile	0.100	0.300
Father: 4th decile	0.100	0.300
Father: 5th decile	0.100	0.300
Father: 6th decile	0.100	0.300
Father: 7th decile	0.100	0.300
Father: 8th decile	0.100	0.300
Father: 9th decile	0.100	0.300
Father: 10th decile	0.100	0.300
Mother: 1st decile	0.100	0.300
Mother: 2nd decile	0.100	0.300
Mother: 3rd decile	0.100	0.300
Mother: 4th decile	0.100	0.300
Mother: 5th decile	0.100	0.300
Mother: 6th decile	0.100	0.300
Mother: 7th decile	0.100	0.300
Mother: 8th decile	0.100	0.300

Continued on next page

Mother: 9th decile	0.100	0.300
Mother: 10th decile	0.100	0.300
Education		
Father: Compulsory schooling	0.277	0.447
Father: High school	0.039	0.193
Father: Vocational education and training	0.447	0.497
Father: Short cycle higher education	0.036	0.186
Father: Bachelor's or vocational bachelor's degree	0.097	0.296
Father: Master's degree/PhD	0.084	0.278
Mother: Compulsory schooling	0.380	0.485
Mother: High school	0.052	0.222
Mother: Vocational education and training	0.300	0.458
Mother: Short cycle higher education	0.030	0.171
Mother: Bachelor's or vocational bachelor's degree	0.182	0.385
Mother: Master's degree/PhD	0.042	0.200
Occupation		
Father: Employer	0.042	0.200
Father: VAT payer	0.036	0.186
Father: Senior manager	0.102	0.303
Father: Manager	0.113	0.316
Father: Employee	0.149	0.357
Father: Skilled worker	0.219	0.413
Father: Unskilled worker	0.209	0.406
Father: Unspecified employee	0.035	0.183
Father: Unemployed	0.056	0.230
Father: Out of labor force	0.018	0.132
Father: Retired	0.010	0.101
Mother: VAT payer	0.008	0.089
Mother: Helping spouse	0.010	0.098
Mother: Senior manager	0.031	0.172
Mother: Manager	0.150	0.357
Mother: Employee	0.340	0.474
Mother: Skilled worker	0.020	0.138
Mother: Unskilled worker	0.182	0.386
Mother: Unspecified employee	0.031	0.173
Mother: Unemployed	0.146	0.353
Mother: Out of labor force	0.054	0.227
Mother: Retired	0.016	0.125
Working hours		
Father: Not employed, self-employed or unknown	0.169	0.375
Father: Part-time: Fewer than 10 hours	0.017	0.128
Father: Part-time: 10-19 hours	0.027	0.161
Father: Part-time: 20-29 hours	0.071	0.258
Father: Part-time: 30+ hours	0.292	0.455
Father: Full-time	0.389	0.487
Father: Unknown hours	0.036	0.186
Mother: Not employed, self-employed or unknown	0.247	0.431
Mother: Part-time: Fewer than 10 hours	0.022	0.147
Mother: Part-time: 10-19 hours	0.048	0.213
Mother: Part-time: 20-29 hours	0.140	0.347
Mother: Part-time: 30+ hours	0.209	0.406
Mother: Full-time	0.321	0.467
Mother: Unknown hours	0.014	0.119
Age		
Father: <20 years	0.017	0.130
Father: 21-23 years	0.077	0.266
Father: 24-26 years	0.164	0.370
Father: 27-29 years	0.224	0.417
Father: 30-32 years	0.207	0.405
Father: 33-35 years	0.142	0.349
Father: 36-38 years	0.084	0.278
Father: 39-41 years	0.048	0.214
Father: > 42 years	0.037	0.188
Mother: <20 years	0.056	0.230
Mother: 21-23 years	0.158	0.365
Mother: 24-26 years	0.242	0.429
Mother: 27-29 years	0.234	0.424
Mother: 30-32 years	0.163	0.370
Mother: 33-35 years	0.087	0.282

Continued on next page

Mother: 36-38 years	0.040	0.197
Mother: >39 years	0.019	0.135
Marital status		
Father: In partnership	0.883	0.321
Father: Single	0.115	0.320
Mother: In partnership	0.877	0.329
Mother: Single	0.122	0.328
Hospitalizations		
Father: 0 visits	0.777	0.416
Father: 1 visits	0.147	0.355
Father: 2 visits	0.045	0.206
Father: 3 visits	0.016	0.125
Father: 4 visits	0.006	0.080
Father: 5+ visits	0.008	0.090
Mother: 0 visits	0.699	0.459
Mother: 1 visits	0.184	0.387
Mother: 2 visits	0.067	0.249
Mother: 3 visits	0.026	0.159
Mother: 4 visits	0.011	0.107
Mother: 5+ visits	0.013	0.113
Placements		
Father: No placement	0.981	0.136
Father: Less than 2 years in placement	0.009	0.093
Father: 2 years or more in placement	0.010	0.100
Mother: No placement	0.970	0.170
Mother: Less than 2 years in placement	0.016	0.124
Mother: 2 years or more in placement	0.014	0.118
Criminal charges		
Father: No charges	0.891	0.312
Father: 1 charge	0.064	0.244
Father: 2 charges	0.019	0.137
Father: 3 charges	0.009	0.093
Father: 4-5 charges	0.008	0.090
Father: 6+ charges	0.010	0.098
Mother: No charges	0.966	0.181
Mother: 1 charge	0.025	0.156
Mother: 2+ charges	0.009	0.094
N	149,755	

Table D.5: Baseline logistic regression

	Education	Income	Social benefits	Criminal charges	Placement	Hospitalizations	Psychiatric condition	SB indicator
Nationality								
Non-Danish					– omitted –			
Danish	0.32***	-0.12	0.36***	-0.15**	1.23***	0.37***	0.73***	0.45***
	(0.09)	(0.07)	(0.08)	(0.08)	(0.12)	(0.08)	(0.10)	(0.08)
Sex								
Female					– omitted –			
Male	0.53***	0.25***	-0.06***	1.66***	0.20***	-0.22***	-0.35***	0.21***
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Birth month								
January					– omitted –			
February	-0.04	0.04	-0.01	-0.06	-0.11**	-0.02	0.05	-0.02
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
March	-0.01	0.03	0.02	-0.06	-0.06	0.04	0.03	0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
April	0.03	0.09**	-0.02	-0.02	-0.05	-0.00	0.04	-0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
May	0.00	0.00	-0.04	-0.05	-0.02	-0.03	0.03	-0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
June	-0.04	0.02	0.02	-0.02	-0.03	-0.04	0.03	-0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
July	0.00	0.05	0.00	-0.08*	-0.02	0.02	0.07*	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
August	0.00	0.02	-0.08**	-0.06	-0.02	-0.02	0.01	-0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
September	0.03	0.08**	-0.01	-0.05	-0.04	0.00	0.08*	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
October	-0.01	0.11***	-0.01	0.02	-0.01	0.01	0.08*	0.01
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
November	-0.02	0.14***	-0.02	-0.01	0.01	0.01	0.06	0.03
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
December	-0.03	0.11***	-0.11***	-0.02	-0.02	-0.04	0.04	-0.06
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Birth order								
1					– omitted –			
2	0.20***	-0.01	0.09***	0.24***	0.15***	0.18***	0.04*	0.13***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
3	0.32***	-0.03	0.17***	0.26***	0.24***	0.24***	0.10***	0.24***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
4	0.39***	0.00	0.19***	0.29***	0.42***	0.30***	0.08	0.30***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
5+	0.61***	0.03	0.27***	0.33***	0.33***	0.38***	0.06	0.37***
	(0.08)	(0.07)	(0.07)	(0.08)	(0.09)	(0.07)	(0.08)	(0.07)
Income								
Father: 1st decile	0.53***	0.37***	0.51***	0.01	0.59***	-0.02	0.23***	0.47***
	(0.06)	(0.04)	(0.05)	(0.05)	(0.07)	(0.04)	(0.05)	(0.05)
Father: 2nd decile	0.53***	0.33***	0.48***	0.06	0.56***	-0.04	0.17***	0.47***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Father: 3rd decile	0.50***	0.26***	0.48***	-0.01	0.45***	-0.06	0.20***	0.46***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Father: 4th decile	0.35***	0.20***	0.36***	-0.02	0.28***	-0.07*	0.10**	0.29***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Father: 5th decile	0.31***	0.12***	0.32***	-0.07*	0.29***	-0.09**	0.10**	0.28***
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Father: 6th decile	0.31***	0.10***	0.25***	-0.08**	0.17**	-0.01	0.07	0.23***
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Father: 7th decile	0.33***	0.05	0.21***	-0.03	0.16**	-0.05	0.07*	0.21***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Father: 8th decile	0.20***	0.06*	0.12***	-0.01	0.06	-0.03	0.05	0.13***
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Father: 9th decile	0.23***	0.03	0.08*	-0.02	0.02	-0.02	0.06	0.09**
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Father: 10th decile					– omitted –			
Mother: 1st decile	0.33***	0.28***	0.43***	-0.04	0.48***	0.04	0.21***	0.43***
	(0.06)	(0.05)	(0.05)	(0.05)	(0.07)	(0.05)	(0.05)	(0.05)
Mother: 2nd decile	0.19***	0.16***	0.28***	-0.03	0.23***	-0.01	0.09*	0.24***
	(0.06)	(0.04)	(0.05)	(0.05)	(0.07)	(0.04)	(0.05)	(0.05)
Mother: 3rd decile	0.08	0.06	0.11**	-0.09*	-0.01	-0.05	-0.04	0.05
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.05)
Mother: 4th decile	0.06	0.03	0.11**	-0.10**	-0.01	-0.06	-0.01	0.06
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Mother: 5th decile	0.07	-0.04	0.14***	-0.06	-0.00	-0.01	-0.00	0.09**
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Mother: 6th decile	0.05	-0.03	0.10**	-0.04	0.02	-0.03	-0.02	0.05
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Mother: 7th decile	-0.01	-0.08**	-0.01	-0.04	-0.14**	-0.02	-0.05	-0.04
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Mother: 8th decile	0.03	0.01	0.02	-0.01	-0.10	-0.02	-0.03	-0.00
	(0.05)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Mother: 9th decile	-0.04	-0.03	0.00	0.04	-0.07	-0.05	-0.04	-0.02
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.04)
Mother: 10th decile					– omitted –			
Wealth								

Continued on next page

Father: 1st decile	0.37***	0.16***	0.22***	0.28***	0.26***	0.19***	0.19***	0.29***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Father: 2nd decile	0.45***	0.16***	0.27***	0.29***	0.38***	0.20***	0.25***	0.34***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Father: 3rd decile	0.49***	0.19***	0.30***	0.36***	0.38***	0.22***	0.23***	0.37***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Father: 4th decile	0.53***	0.18***	0.30***	0.30***	0.48***	0.22***	0.30***	0.38***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Father: 5th decile	0.55***	0.22***	0.34***	0.40***	0.49***	0.20***	0.29***	0.43***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Father: 6th decile	0.56***	0.25***	0.30***	0.38***	0.59***	0.20***	0.29***	0.43***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Father: 7th decile	0.36***	0.17***	0.20***	0.22***	0.38***	0.06*	0.22***	0.26***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Father: 8th decile	0.35***	0.09**	0.15***	0.20***	0.22***	0.12***	0.15***	0.22***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Father: 9th decile	0.25***	0.03	0.12***	0.14***	0.14**	0.13***	0.14***	0.13***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Father: 10th decile					– omitted –			
Mother: 1st decile	0.23***	0.15***	0.18***	0.16***	0.26***	0.13***	0.14***	0.21***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Mother: 2nd decile	0.23***	0.12***	0.19***	0.13***	0.24***	0.14***	0.15***	0.25***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Mother: 3rd decile	0.28***	0.15***	0.21***	0.22***	0.35***	0.17***	0.20***	0.29***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Mother: 4th decile	0.23***	0.13***	0.20***	0.17***	0.38***	0.17***	0.20***	0.29***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Mother: 5th decile	0.22***	0.07*	0.21***	0.06	0.22***	0.10**	0.12***	0.21***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Mother: 6th decile	0.21***	0.17***	0.23***	0.12***	0.31***	0.12***	0.17***	0.25***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Mother: 7th decile	0.12**	0.08**	0.09**	0.07*	0.12*	0.07**	0.13***	0.13***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Mother: 8th decile	0.01	0.06*	0.06	0.01	0.09	0.02	0.06	0.05
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Mother: 9th decile	-0.03	0.01	-0.04	-0.06	0.08	0.02	0.03	-0.01
	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)
Mother: 10th decile					– omitted –			
Education								
Father: Compulsory schooling	1.00***	-0.14***	0.65***	0.42***	0.68***	0.46***	0.12**	0.68***
	(0.07)	(0.04)	(0.05)	(0.05)	(0.08)	(0.05)	(0.05)	(0.05)
Father: High school	0.49***	-0.02	0.23***	0.20***	0.15	0.15***	-0.01	0.28***
	(0.09)	(0.05)	(0.06)	(0.06)	(0.10)	(0.06)	(0.06)	(0.06)
Father: Vocational education and training	0.73***	-0.27***	0.43***	0.24***	0.46***	0.37***	-0.01	0.44***
	(0.07)	(0.04)	(0.05)	(0.05)	(0.08)	(0.05)	(0.05)	(0.05)
Father: Short cycle higher education	0.47***	-0.17***	0.30***	0.12*	0.37***	0.30***	0.00	0.33***
	(0.09)	(0.05)	(0.07)	(0.06)	(0.11)	(0.06)	(0.06)	(0.07)
Father: (Vocational) bachelor's degree	0.30***	-0.08**	0.22***	0.14***	0.18*	0.19***	0.02	0.21***
	(0.08)	(0.04)	(0.05)	(0.05)	(0.09)	(0.05)	(0.05)	(0.05)
Father: Master's degree/PhD					– omitted –			
Mother: Compulsory schooling	1.05***	0.01	0.78***	0.55***	1.32***	0.53***	0.33***	0.87***
	(0.10)	(0.05)	(0.07)	(0.07)	(0.12)	(0.06)	(0.06)	(0.07)
Mother: High school	0.38***	-0.02	0.25***	0.20***	0.61***	0.23***	0.14**	0.31***
	(0.11)	(0.06)	(0.07)	(0.07)	(0.13)	(0.07)	(0.07)	(0.08)
Mother: Vocational education and training	0.70***	-0.15***	0.50***	0.34***	0.95***	0.43***	0.17***	0.55***
	(0.10)	(0.05)	(0.07)	(0.07)	(0.12)	(0.06)	(0.06)	(0.07)
Mother: Short cycle higher education	0.30**	-0.09	0.28***	0.19**	0.62***	0.28***	0.07	0.30***
	(0.13)	(0.06)	(0.09)	(0.08)	(0.15)	(0.08)	(0.08)	(0.09)
Mother: (Vocational) bachelor's degree	0.24**	-0.02	0.17**	0.08	0.54***	0.16**	0.08	0.20***
	(0.11)	(0.05)	(0.07)	(0.07)	(0.12)	(0.06)	(0.06)	(0.07)
Mother: Master's degree/PhD					– omitted –			
Occupation								
Father: Employer	-0.41**	-0.32**	-0.75***	-0.09	-0.92***	-0.25	-0.29*	-0.66***
	(0.19)	(0.16)	(0.17)	(0.17)	(0.23)	(0.16)	(0.17)	(0.17)
Father: VAT payer	-0.32*	-0.32**	-0.69***	-0.18	-0.80***	-0.32**	-0.24	-0.60***
	(0.19)	(0.16)	(0.17)	(0.17)	(0.22)	(0.16)	(0.17)	(0.17)
Father: Senior manager	-0.17***	-0.02	-0.08*	-0.14***	-0.19**	-0.12***	-0.05	-0.08*
	(0.06)	(0.04)	(0.04)	(0.04)	(0.07)	(0.04)	(0.04)	(0.05)
Father: Manager	-0.21***	0.02	-0.20***	-0.16***	-0.18***	-0.10***	-0.05	-0.19***
	(0.05)	(0.03)	(0.04)	(0.04)	(0.06)	(0.03)	(0.04)	(0.04)
Father: Employee	-0.02	0.00	-0.10***	-0.06**	-0.04	-0.07**	0.05*	-0.05*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Father: Skilled worker					– omitted –			
Father: Unskilled worker	0.16***	0.00	0.07***	-0.02	0.13***	0.04	0.05*	0.11***
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Father: Unspecified employee	0.07	-0.06	-0.05	-0.07	0.08	0.02	-0.05	-0.02
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Father: Unemployed	-0.07	0.04	-0.16	0.18	-0.19	-0.11	0.16	-0.06
	(0.19)	(0.16)	(0.16)	(0.16)	(0.22)	(0.15)	(0.16)	(0.16)
Father: Out of labor force	-0.05	0.03	-0.23	0.20	-0.24	-0.23	0.15	-0.07
	(0.19)	(0.16)	(0.17)	(0.17)	(0.22)	(0.16)	(0.17)	(0.17)
Father: Retired	-0.47**	-0.17	-0.46**	-0.14	-0.62**	-0.33*	-0.04	-0.35*
	(0.22)	(0.17)	(0.18)	(0.18)	(0.25)	(0.18)	(0.18)	(0.18)
Mother: VAT payer	0.01	0.10	0.94	-0.37	-0.02	0.68	-0.19	0.25
	(0.68)	(0.66)	(0.65)	(0.83)	(0.60)	(0.70)	(0.75)	(0.65)
Mother: Helping spouse	-0.34	-0.07	0.80	-0.72	-0.27	0.52	-0.64	0.01
	(0.68)	(0.66)	(0.65)	(0.83)	(0.61)	(0.70)	(0.75)	(0.65)
Mother: Senior manager	-0.25*	-0.02	0.04	-0.10	-0.09	-0.04	0.11	0.08
	(0.13)	(0.08)	(0.09)	(0.09)	(0.15)	(0.09)	(0.09)	(0.09)
Mother: Manager	-0.19**	-0.04	-0.04	-0.12*	-0.32***	0.01	0.00	-0.06

Continued on next page

	(0.08)	(0.06)	(0.07)	(0.07)	(0.10)	(0.06)	(0.07)	(0.07)
Mother: Employee	-0.18***	-0.10*	-0.04	-0.14**	-0.19**	-0.04	-0.02	-0.06
	(0.07)	(0.06)	(0.06)	(0.06)	(0.08)	(0.05)	(0.06)	(0.06)
Mother: Skilled worker					– omitted –			
Mother: Unskilled worker	0.04	-0.04	0.15***	-0.07	0.12	0.03	0.06	0.16***
	(0.07)	(0.06)	(0.06)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)
Mother: Unspecified employee	0.03	0.05	0.16**	-0.04	0.16	-0.04	0.06	0.19***
	(0.08)	(0.07)	(0.07)	(0.07)	(0.10)	(0.07)	(0.08)	(0.07)
Mother: Unemployed	0.21	0.32	1.33**	-0.29	0.56	0.82	-0.04	0.68
	(0.67)	(0.65)	(0.65)	(0.82)	(0.59)	(0.70)	(0.74)	(0.64)
Mother: Out of labor force	0.17	0.28	1.22*	-0.28	0.58	0.77	-0.06	0.62
	(0.67)	(0.65)	(0.65)	(0.82)	(0.59)	(0.70)	(0.75)	(0.64)
Mother: Retired	-0.01	0.21	1.00	-0.28	0.48	0.75	-0.14	0.39
	(0.67)	(0.65)	(0.65)	(0.83)	(0.59)	(0.70)	(0.75)	(0.64)
Working hours								
Father: Not employed, self-employed or unknown	0.36*	0.19	0.39**	-0.06	0.56**	0.26	0.05	0.38**
	(0.20)	(0.16)	(0.17)	(0.17)	(0.23)	(0.16)	(0.17)	(0.17)
Father: Part-time: Fewer than 10 hours					– omitted –			
Father: Part-time: 10-19 hours	0.22**	0.09	0.13*	0.15*	0.31***	0.09	0.08	0.16**
	(0.09)	(0.07)	(0.08)	(0.08)	(0.10)	(0.08)	(0.08)	(0.08)
Father: Part-time: 20-29 hours	0.20**	0.05	0.14**	0.09	0.23**	0.14**	0.11	0.16**
	(0.08)	(0.06)	(0.07)	(0.07)	(0.09)	(0.07)	(0.07)	(0.07)
Father: Part-time: 30+ hours	0.15*	-0.01	0.11*	0.00	0.14	0.10	0.04	0.14**
	(0.08)	(0.06)	(0.06)	(0.07)	(0.09)	(0.07)	(0.07)	(0.07)
Father: Full-time	0.15*	-0.01	0.03	-0.08	0.05	0.06	0.02	0.06
	(0.08)	(0.06)	(0.06)	(0.07)	(0.09)	(0.07)	(0.07)	(0.07)
Father: Unknown hours	0.23**	0.01	0.11	0.00	0.19*	0.18**	0.15**	0.18**
	(0.09)	(0.07)	(0.08)	(0.08)	(0.10)	(0.07)	(0.08)	(0.08)
Mother: Not employed, self-employed or unknown	-0.07	-0.14	-0.93	0.36	-0.34	-0.72	0.16	-0.27
	(0.67)	(0.65)	(0.65)	(0.82)	(0.59)	(0.69)	(0.74)	(0.64)
Mother: Part-time: Fewer than 10 hours					– omitted –			
Mother: Part-time: 10-19 hours	0.00	0.10*	0.09	0.17**	-0.01	0.04	0.02	0.10
	(0.08)	(0.06)	(0.06)	(0.07)	(0.08)	(0.06)	(0.07)	(0.06)
Mother: Part-time: 20-29 hours	-0.01	0.05	0.08	0.03	-0.03	0.06	-0.00	0.08
	(0.07)	(0.05)	(0.06)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)
Mother: Part-time: 30+ hours	0.02	0.10*	0.13**	0.06	-0.08	0.07	0.00	0.13**
	(0.07)	(0.05)	(0.06)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)
Mother: Full-time	-0.03	0.02	0.05	0.02	-0.16**	0.05	-0.07	0.06
	(0.07)	(0.05)	(0.06)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)
Mother: Unknown hours	0.11	0.11	0.26***	0.12	0.13	0.11	0.10	0.29***
	(0.10)	(0.08)	(0.08)	(0.09)	(0.11)	(0.08)	(0.09)	(0.08)
Age								
Father: <20 years	0.14**	0.02	0.05	0.12*	-0.09	0.09	0.01	0.05
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
Father: 21-23 years	0.00	-0.08**	-0.06*	0.12***	-0.19***	0.07*	-0.08**	-0.08**
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)
Father: 24-26 years	0.02	-0.01	-0.04*	0.00	-0.05	0.01	-0.08***	-0.04*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)
Father: 27-29 years					– omitted –			
Father: 30-32 years	0.06**	0.07***	0.08***	0.06**	0.18***	-0.01	0.01	0.09***
	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)
Father: 33-35 years	0.20***	0.17***	0.18***	0.09***	0.26***	0.01	0.05*	0.18***
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Father: 36-38 years	0.19***	0.24***	0.25***	0.08**	0.32***	0.02	0.17***	0.28***
	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)
Father: 39-41 years	0.26***	0.29***	0.35***	0.18***	0.45***	0.05	0.23***	0.41***
	(0.05)	(0.04)	(0.04)	(0.05)	(0.06)	(0.04)	(0.04)	(0.04)
Father: > 42 years	0.31***	0.35***	0.38***	0.20***	0.69***	0.06	0.29***	0.45***
	(0.06)	(0.04)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Mother: <20 years	0.14***	-0.13***	0.07	0.33***	0.08	0.25***	0.15***	0.15***
	(0.05)	(0.04)	(0.04)	(0.05)	(0.06)	(0.04)	(0.05)	(0.04)
Mother: 21-23 years	0.13***	-0.09***	0.06*	0.20***	0.08**	0.18***	0.06*	0.09***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Mother: 24-26 years	0.05*	-0.06**	0.01	0.10***	0.06*	0.09***	0.03	0.04
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
Mother: 27-29 years					– omitted –			
Mother: 30-32 years	-0.04	0.09***	-0.01	-0.08***	-0.07*	-0.06**	0.01	-0.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Mother: 33-35 years	-0.06	0.09***	-0.01	-0.15***	-0.10*	-0.07**	0.00	-0.06*
	(0.05)	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)
Mother: 36-38 years	-0.02	0.20***	0.04	-0.11**	-0.07	-0.26***	0.06	-0.01
	(0.06)	(0.04)	(0.05)	(0.05)	(0.07)	(0.05)	(0.05)	(0.05)
Mother: >39 years	-0.17**	0.16***	-0.04	-0.15**	-0.08	-0.18***	0.04	-0.09
	(0.08)	(0.06)	(0.07)	(0.07)	(0.09)	(0.07)	(0.07)	(0.07)
Marital status								
Father: In partnership					– omitted –			
Father: Single	0.10**	0.11***	0.13***	0.13***	0.21***	0.01	0.06	0.15***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Mother: In partnership					– omitted –			
Mother: Single	0.19***	0.11***	0.04	0.17***	0.22***	0.05	0.15***	0.17***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Hospitalizations								
Father: 0 visits					– omitted –			
Father: 1 visits	0.04*	0.04*	0.04*	0.05**	0.12***	0.15***	0.06***	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Father: 2 visits	0.04	0.05	0.03	0.16***	0.20***	0.24***	0.10***	0.11***

Continued on next page

	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)
Father: 3 visits	-0.11	0.02	-0.05	0.05	0.18**	0.25***	0.18***	0.06
	(0.07)	(0.06)	(0.06)	(0.06)	(0.07)	(0.05)	(0.06)	(0.06)
Father: 4 visits	0.16*	0.15*	0.26***	0.24***	0.28***	0.39***	0.34***	0.31***
	(0.10)	(0.09)	(0.08)	(0.09)	(0.10)	(0.08)	(0.09)	(0.08)
Father: 5+ visits	0.13	-0.08	0.04	0.13	0.33***	0.32***	0.25***	0.10
	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.07)	(0.08)	(0.08)
Mother: 0 visits					– omitted –			
Mother: 1 visits	0.09***	0.02	0.06***	0.08***	0.20***	0.16***	0.15***	0.15***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Mother: 2 visits	0.12***	0.02	0.13***	0.18***	0.38***	0.29***	0.19***	0.23***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Mother: 3 visits	0.21***	0.05	0.26***	0.19***	0.50***	0.49***	0.24***	0.34***
	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)
Mother: 4 visits	0.05	0.09	0.13**	0.16**	0.49***	0.43***	0.30***	0.30***
	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.06)	(0.07)	(0.07)
Mother: 5+ visits	0.34***	0.00	0.24***	0.17***	0.85***	0.59***	0.37***	0.40***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Placements								
Father: No placement					– omitted –			
Father: Less than 2 years in placement	0.14*	0.13*	0.26***	0.15**	0.29***	0.10	0.26***	0.29***
	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
Father: 2 years or more in placement	0.17**	0.22***	0.27***	0.06	0.31***	0.06	0.20***	0.26***
	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Mother: No placement					– omitted –			
Mother: Less than 2 years in placement	0.18***	0.22***	0.26***	0.23***	0.59***	0.16***	0.23***	0.40***
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)
Mother: 2 years or more in placement	0.24***	0.26***	0.29***	0.30***	0.74***	0.08	0.29***	0.44***
	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Criminal charges								
Father: No charges					– omitted –			
Father: 1 charge	0.24***	0.13***	0.23***	0.37***	0.30***	0.13***	0.12***	0.29***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Father: 2 charges	0.32***	0.16***	0.34***	0.55***	0.44***	0.25***	0.16***	0.46***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Father: 3 charges	0.20***	0.18**	0.16**	0.64***	0.36***	0.17**	0.28***	0.25***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)
Father: 4-5 charges	0.25***	0.16**	0.34***	0.74***	0.64***	0.16**	0.27***	0.50***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)
Father: 6+ charges	0.28***	0.18**	0.46***	0.81***	0.54***	0.23***	0.28***	0.52***
	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
Mother: No charges					– omitted –			
Mother: 1 charge	0.12**	0.13***	0.14***	0.34***	0.31***	0.11**	0.10**	0.22***
	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)
Mother: 2+ charges	-0.03	0.12	0.02	0.25***	0.68***	-0.05	0.12	0.17**
	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.07)
N - Training data					119,804			

Notes: Estimates from logistic regressions of various outcomes on the full set of predictors. For predictors with missing information we include an additional dummy indicating missings but do not report the corresponding coeffcient estimates. Robust standard errors in parentheses.

Table D.6: Baseline logistic regression

	At least 3 vs. fewer	At least 4 vs. fewer	At least 5 vs. fewer	At least 3 vs. 0	At least 4 vs. 0	At least 5 vs. 0
Nationality						
Non-Danish			– omitted –			
Danish	0.63***	0.72***	0.80***	0.71***	0.94***	1.12***
	(0.08)	(0.11)	(0.15)	(0.10)	(0.12)	(0.16)
Sex						
Female			– omitted –			
Male	0.35***	0.35***	0.35***	0.52***	0.57***	0.60***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Birth month						
January			– omitted –			
February	-0.05	-0.02	-0.15*	-0.03	-0.02	-0.17*
	(0.04)	(0.06)	(0.08)	(0.05)	(0.06)	(0.09)
March	-0.03	0.00	-0.11	0.00	0.03	-0.06
	(0.04)	(0.05)	(0.07)	(0.05)	(0.06)	(0.08)
April	0.00	0.01	-0.00	0.04	0.06	0.06
	(0.04)	(0.05)	(0.07)	(0.05)	(0.06)	(0.08)
May	-0.02	-0.04	-0.14*	-0.02	-0.06	-0.13
	(0.04)	(0.05)	(0.07)	(0.05)	(0.06)	(0.08)
June	-0.05	-0.03	-0.14*	-0.04	-0.04	-0.14*
	(0.04)	(0.05)	(0.07)	(0.05)	(0.06)	(0.08)
July	0.02	0.02	-0.01	0.03	0.03	0.01
	(0.04)	(0.05)	(0.07)	(0.05)	(0.06)	(0.08)
August	-0.06	-0.05	-0.16**	-0.04	-0.05	-0.17**
	(0.04)	(0.05)	(0.07)	(0.05)	(0.06)	(0.08)
September	-0.01	0.05	0.02	0.01	0.07	0.04
	(0.04)	(0.05)	(0.07)	(0.05)	(0.06)	(0.08)
October	0.01	0.05	0.00	0.05	0.09	0.05
	(0.04)	(0.05)	(0.07)	(0.05)	(0.06)	(0.08)
November	0.06	0.06	-0.03	0.05	0.06	-0.05
	(0.04)	(0.05)	(0.07)	(0.05)	(0.06)	(0.09)
December	-0.04	-0.02	-0.11	-0.03	-0.04	-0.20**
	(0.04)	(0.06)	(0.07)	(0.05)	(0.06)	(0.09)
Birth order						
1			– omitted –			
2	0.17***	0.17***	0.18***	0.24***	0.28***	0.33***
	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)	(0.04)
3	0.26***	0.29***	0.32***	0.34***	0.43***	0.52***
	(0.03)	(0.04)	(0.06)	(0.04)	(0.05)	(0.07)
4	0.34***	0.42***	0.46***	0.45***	0.62***	0.77***
	(0.05)	(0.06)	(0.09)	(0.06)	(0.08)	(0.10)
5+	0.48***	0.43***	0.54***	0.61***	0.67***	0.84***
	(0.07)	(0.09)	(0.12)	(0.09)	(0.11)	(0.15)
Income						
Father: 1st decile	0.50***	0.51***	0.56***	0.62***	0.69***	0.80***
	(0.05)	(0.07)	(0.11)	(0.06)	(0.08)	(0.11)
Father: 2nd decile	0.48***	0.46***	0.51***	0.59***	0.63***	0.71***
	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.11)
Father: 3rd decile	0.39***	0.36***	0.46***	0.49***	0.49***	0.62***
	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.11)
Father: 4th decile	0.27***	0.27***	0.34***	0.33***	0.33***	0.41***
	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.11)
Father: 5th decile	0.23***	0.24***	0.28***	0.26***	0.29***	0.34***
	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.11)
Father: 6th decile	0.21***	0.20***	0.22**	0.23***	0.24***	0.28**
	(0.05)	(0.07)	(0.11)	(0.05)	(0.07)	(0.11)
Father: 7th decile	0.22***	0.17**	0.17	0.24***	0.21***	0.20*
	(0.05)	(0.07)	(0.11)	(0.05)	(0.07)	(0.11)
Father: 8th decile	0.12**	0.10	0.17	0.15***	0.13*	0.19*
	(0.05)	(0.07)	(0.11)	(0.05)	(0.07)	(0.11)
Father: 9th decile	0.11**	0.02	0.16	0.12**	0.03	0.19*
	(0.05)	(0.07)	(0.11)	(0.05)	(0.07)	(0.11)
Father: 10th decile			– omitted –			
Mother: 1st decile	0.44***	0.40***	0.39***	0.48***	0.50***	0.53***
	(0.06)	(0.07)	(0.11)	(0.06)	(0.08)	(0.12)
Mother: 2nd decile	0.27***	0.25***	0.31***	0.23***	0.23***	0.33***
	(0.05)	(0.07)	(0.10)	(0.06)	(0.08)	(0.11)
Mother: 3rd decile	0.04	0.04	0.04	0.01	-0.01	-0.00
	(0.05)	(0.07)	(0.11)	(0.06)	(0.08)	(0.11)
Mother: 4th decile	0.03	-0.01	-0.04	0.02	-0.05	-0.07
	(0.05)	(0.07)	(0.11)	(0.06)	(0.08)	(0.11)
Mother: 5th decile	0.05	0.05	0.08	0.04	0.02	0.06
	(0.05)	(0.07)	(0.10)	(0.05)	(0.08)	(0.11)
Mother: 6th decile	-0.01	0.00	-0.00	0.01	-0.01	0.00
	(0.05)	(0.07)	(0.10)	(0.05)	(0.08)	(0.11)
Mother: 7th decile	-0.05	-0.06	-0.07	-0.06	-0.09	-0.11
	(0.05)	(0.07)	(0.11)	(0.05)	(0.08)	(0.11)
Mother: 8th decile	0.00	0.02	-0.03	-0.02	-0.02	-0.05
	(0.05)	(0.07)	(0.11)	(0.05)	(0.07)	(0.11)
Mother: 9th decile	-0.05	-0.03	-0.05	-0.05	-0.05	-0.06
	(0.05)	(0.07)	(0.11)	(0.05)	(0.08)	(0.12)
Mother: 10th decile			– omitted –			
Wealth						

Continued on next page

Father: 1st decile	0.34***	0.31***	0.30***	0.48***	0.46***	0.45***
	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.11)
Father: 2nd decile	0.41***	0.41***	0.39***	0.54***	0.57***	0.59***
	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.10)
Father: 3rd decile	0.43***	0.45***	0.52***	0.57***	0.62***	0.72***
	(0.05)	(0.06)	(0.10)	(0.05)	(0.07)	(0.10)
Father: 4th decile	0.48***	0.53***	0.56***	0.61***	0.71***	0.80***
	(0.05)	(0.06)	(0.09)	(0.05)	(0.07)	(0.10)
Father: 5th decile	0.50***	0.54***	0.60***	0.67***	0.76***	0.87***
	(0.05)	(0.06)	(0.09)	(0.05)	(0.07)	(0.10)
Father: 6th decile	0.52***	0.57***	0.61***	0.66***	0.76***	0.88***
	(0.05)	(0.06)	(0.09)	(0.05)	(0.07)	(0.10)
Father: 7th decile	0.31***	0.33***	0.41***	0.41***	0.44***	0.55***
	(0.05)	(0.06)	(0.09)	(0.05)	(0.07)	(0.10)
Father: 8th decile	0.26***	0.24***	0.25**	0.33***	0.33***	0.37***
	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.10)
Father: 9th decile	0.21***	0.22***	0.24**	0.26***	0.27***	0.33***
	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.10)
Father: 10th decile			– omitted –			
Mother: 1st decile	0.24***	0.33***	0.32***	0.32***	0.43***	0.46***
	(0.05)	(0.06)	(0.10)	(0.05)	(0.07)	(0.10)
Mother: 2nd decile	0.23***	0.29***	0.28***	0.30***	0.38***	0.39***
	(0.04)	(0.06)	(0.09)	(0.05)	(0.07)	(0.10)
Mother: 3rd decile	0.30***	0.37***	0.40***	0.39***	0.48***	0.55***
	(0.04)	(0.06)	(0.09)	(0.05)	(0.07)	(0.10)
Mother: 4th decile	0.27***	0.35***	0.34***	0.38***	0.48***	0.52***
	(0.04)	(0.06)	(0.09)	(0.05)	(0.07)	(0.09)
Mother: 5th decile	0.21***	0.25***	0.28***	0.26***	0.35***	0.41***
	(0.04)	(0.06)	(0.09)	(0.05)	(0.07)	(0.10)
Mother: 6th decile	0.27***	0.30***	0.30***	0.36***	0.43***	0.46***
	(0.04)	(0.06)	(0.09)	(0.05)	(0.07)	(0.09)
Mother: 7th decile	0.11**	0.19***	0.16*	0.17***	0.24***	0.22**
	(0.04)	(0.06)	(0.09)	(0.05)	(0.07)	(0.10)
Mother: 8th decile	0.05	0.11	0.08	0.07	0.14**	0.13
	(0.05)	(0.06)	(0.10)	(0.05)	(0.07)	(0.10)
Mother: 9th decile	-0.02	-0.00	0.00	-0.02	0.00	0.02
	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.11)
Mother: 10th decile			– omitted –			
Education						
Father: Compulsory schooling	0.68***	0.82***	1.03***	0.78***	0.99***	1.28***
	(0.06)	(0.09)	(0.15)	(0.06)	(0.09)	(0.15)
Father: High school	0.31***	0.39***	0.52***	0.31***	0.40***	0.53***
	(0.07)	(0.10)	(0.17)	(0.08)	(0.11)	(0.18)
Father: Vocational education and training	0.45***	0.57***	0.78***	0.46***	0.61***	0.86***
	(0.06)	(0.09)	(0.15)	(0.06)	(0.09)	(0.15)
Father: Short cycle higher education	0.28***	0.38***	0.60***	0.30***	0.42***	0.68***
	(0.08)	(0.11)	(0.19)	(0.08)	(0.12)	(0.19)
Father: (Vocational) bachelor's degree	0.22***	0.27***	0.46***	0.24***	0.30***	0.50***
	(0.06)	(0.09)	(0.15)	(0.06)	(0.10)	(0.16)
Father: Master's degree/PhD			– omitted –			
Mother: Compulsory schooling	0.92***	1.14***	1.11***	1.06***	1.36***	1.41***
	(0.08)	(0.12)	(0.19)	(0.09)	(0.13)	(0.20)
Mother: High school	0.36***	0.40***	0.22	0.39***	0.42***	0.21
	(0.09)	(0.13)	(0.21)	(0.09)	(0.14)	(0.22)
Mother: Vocational education and training	0.59***	0.81***	0.76***	0.64***	0.90***	0.91***
	(0.08)	(0.12)	(0.19)	(0.09)	(0.13)	(0.20)
Mother: Short cycle higher education	0.27***	0.42***	0.15	0.29***	0.47***	0.21
	(0.10)	(0.15)	(0.25)	(0.11)	(0.16)	(0.26)
Mother: (Vocational) bachelor's degree	0.26***	0.28**	0.32*	0.28***	0.33**	0.38*
	(0.08)	(0.12)	(0.20)	(0.09)	(0.13)	(0.20)
Mother: Master's degree/PhD			– omitted –			
Occupation						
Father: Employer	-0.52***	-0.44**	-0.72**	-0.80***	-0.82***	-1.29***
	(0.18)	(0.22)	(0.33)	(0.21)	(0.26)	(0.36)
Father: VAT payer	-0.43**	-0.43**	-0.66**	-0.74***	-0.86***	-1.25***
	(0.18)	(0.22)	(0.32)	(0.21)	(0.26)	(0.36)
Father: Senior manager	-0.07	-0.13*	-0.06	-0.14***	-0.21***	-0.13
	(0.05)	(0.07)	(0.11)	(0.05)	(0.08)	(0.12)
Father: Manager	-0.16***	-0.08	-0.21**	-0.22***	-0.17***	-0.31***
	(0.04)	(0.06)	(0.09)	(0.05)	(0.06)	(0.09)
Father: Employee	-0.04	-0.03	-0.02	-0.08**	-0.08*	-0.07
	(0.03)	(0.04)	(0.06)	(0.04)	(0.05)	(0.06)
Father: Skilled worker			– omitted –			
Father: Unskilled worker	0.10***	0.08**	0.12**	0.11***	0.09**	0.13**
	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.06)
Father: Unspecified employee	-0.05	0.09	0.13	-0.07	0.08	0.14
	(0.05)	(0.06)	(0.08)	(0.06)	(0.07)	(0.09)
Father: Unemployed	0.10	0.13	-0.06	0.02	0.01	-0.32
	(0.17)	(0.21)	(0.31)	(0.20)	(0.25)	(0.35)
Father: Out of labor force	0.06	0.09	-0.05	-0.13	-0.16	-0.45
	(0.18)	(0.21)	(0.31)	(0.21)	(0.26)	(0.35)
Father: Retired	-0.25	-0.31	-0.56	-0.55**	-0.66**	-1.08***
	(0.19)	(0.25)	(0.36)	(0.22)	(0.29)	(0.40)
Mother: VAT payer	0.03	0.23	0.03	0.23	-0.10	-0.26
	(0.75)	(0.74)	(0.98)	(1.03)	(0.71)	(0.85)
Mother: Helping spouse	-0.41	-0.38	-0.62	-0.32	-0.86	-1.07
	(0.76)	(0.75)	(1.01)	(1.03)	(0.72)	(0.88)
Mother: Senior manager	0.16	0.15	-0.03	0.12	0.16	0.03
	(0.11)	(0.15)	(0.24)	(0.11)	(0.16)	(0.25)
Mother: Manager	-0.04	-0.02	-0.27*	-0.11	-0.07	-0.26

Continued on next page

	(0.08)	(0.10)	(0.15)	(0.08)	(0.11)	(0.16)
Mother: Employee	-0.07	-0.10	-0.20*	-0.13*	-0.16*	-0.25**
	(0.06)	(0.08)	(0.12)	(0.07)	(0.09)	(0.13)
Mother: Skilled worker	– omitted –					
Mother: Unskilled worker	0.13**	0.18**	0.14	0.10	0.14	0.14
	(0.06)	(0.09)	(0.12)	(0.07)	(0.09)	(0.13)
Mother: Unspecified employee	0.16**	0.15	0.25*	0.17*	0.22*	0.35**
	(0.08)	(0.10)	(0.14)	(0.09)	(0.12)	(0.15)
Mother: Unemployed	0.48	0.60	0.54	0.74	0.35	0.36
	(0.75)	(0.73)	(0.96)	(1.02)	(0.69)	(0.83)
Mother: Out of labor force	0.41	0.55	0.52	0.70	0.33	0.42
	(0.75)	(0.73)	(0.96)	(1.02)	(0.70)	(0.83)
Mother: Retired	0.22	0.42	0.40	0.40	0.02	0.11
	(0.75)	(0.73)	(0.97)	(1.02)	(0.70)	(0.84)
Working hours						
Father: Not employed, self-employed or unknown	0.25	0.21	0.51	0.51**	0.64**	1.12***
	(0.18)	(0.22)	(0.32)	(0.21)	(0.27)	(0.37)
Father: Part-time: Fewer than 10 hours	– omitted –					
Father: Part-time: 10-19 hours	0.15*	0.23**	0.32**	0.25**	0.43***	0.58***
	(0.08)	(0.10)	(0.14)	(0.10)	(0.13)	(0.17)
Father: Part-time: 20-29 hours	0.17**	0.18*	0.29**	0.28***	0.38***	0.53***
	(0.07)	(0.09)	(0.13)	(0.09)	(0.11)	(0.16)
Father: Part-time: 30+ hours	0.11	0.10	0.18	0.19**	0.27**	0.34**
	(0.07)	(0.09)	(0.13)	(0.08)	(0.11)	(0.15)
Father: Full-time	0.06	0.04	0.18	0.11	0.18	0.33**
	(0.07)	(0.09)	(0.13)	(0.08)	(0.11)	(0.15)
Father: Unknown hours	0.19**	0.17	0.29**	0.29***	0.35***	0.47***
	(0.08)	(0.10)	(0.15)	(0.10)	(0.13)	(0.17)
Mother: Not employed, self-employed or unknown	-0.11	-0.20	-0.25	-0.27	0.18	0.09
	(0.74)	(0.73)	(0.96)	(1.02)	(0.69)	(0.82)
Mother: Part-time: Fewer than 10 hours	– omitted –					
Mother: Part-time: 10-19 hours	0.13*	0.14	0.07	0.21**	0.21**	0.18
	(0.07)	(0.09)	(0.12)	(0.08)	(0.11)	(0.14)
Mother: Part-time: 20-29 hours	0.10	0.08	0.03	0.16**	0.14	0.09
	(0.06)	(0.08)	(0.11)	(0.08)	(0.10)	(0.13)
Mother: Part-time: 30+ hours	0.14**	0.08	0.06	0.21***	0.18*	0.14
	(0.06)	(0.08)	(0.11)	(0.07)	(0.10)	(0.13)
Mother: Full-time	0.06	0.00	-0.06	0.10	0.04	-0.02
	(0.06)	(0.08)	(0.11)	(0.08)	(0.10)	(0.13)
Mother: Unknown hours	0.30***	0.32***	0.33**	0.30***	0.33**	0.40**
	(0.09)	(0.12)	(0.16)	(0.10)	(0.14)	(0.18)
Age						
Father: <20 years	0.04	0.13*	0.17**	0.06	0.12	0.11
	(0.06)	(0.07)	(0.09)	(0.08)	(0.09)	(0.11)
Father: 21-23 years	-0.08**	0.01	-0.01	-0.08*	-0.01	-0.05
	(0.04)	(0.04)	(0.06)	(0.04)	(0.05)	(0.07)
Father: 24-26 years	-0.05*	-0.01	0.00	-0.07**	-0.05	-0.06
	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.06)
Father: 27-29 years	– omitted –					
Father: 30-32 years	0.06**	0.11***	0.07	0.10***	0.16***	0.16***
	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.06)
Father: 33-35 years	0.17***	0.25***	0.18***	0.24***	0.32***	0.28***
	(0.03)	(0.04)	(0.06)	(0.04)	(0.05)	(0.07)
Father: 36-38 years	0.24***	0.28***	0.28***	0.34***	0.37***	0.42***
	(0.04)	(0.05)	(0.07)	(0.05)	(0.06)	(0.08)
Father: 39-41 years	0.36***	0.38***	0.32***	0.48***	0.54***	0.51***
	(0.05)	(0.06)	(0.09)	(0.05)	(0.07)	(0.10)
Father: > 42 years	0.46***	0.50***	0.36***	0.63***	0.71***	0.62***
	(0.05)	(0.07)	(0.09)	(0.06)	(0.08)	(0.11)
Mother: <20 years	0.16***	0.10*	0.11	0.26***	0.21***	0.23**
	(0.05)	(0.06)	(0.07)	(0.06)	(0.07)	(0.09)
Mother: 21-23 years	0.13***	0.15***	0.12**	0.19***	0.22***	0.25***
	(0.03)	(0.04)	(0.06)	(0.04)	(0.05)	(0.07)
Mother: 24-26 years	0.07**	0.05	0.10**	0.08***	0.10**	0.17***
	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.06)
Mother: 27-29 years	– omitted –					
Mother: 30-32 years	-0.05	-0.05	-0.06	-0.04	-0.06	-0.08
	(0.03)	(0.04)	(0.06)	(0.03)	(0.05)	(0.07)
Mother: 33-35 years	-0.06	-0.12**	-0.19**	-0.08*	-0.14**	-0.25***
	(0.04)	(0.06)	(0.08)	(0.05)	(0.06)	(0.09)
Mother: 36-38 years	-0.03	-0.09	-0.03	-0.06	-0.17**	-0.17
	(0.06)	(0.07)	(0.10)	(0.06)	(0.09)	(0.12)
Mother: >39 years	-0.13*	-0.17*	-0.14	-0.21**	-0.33***	-0.37**
	(0.08)	(0.10)	(0.14)	(0.09)	(0.12)	(0.16)
Marital status						
Father: In partnership	– omitted –					
Father: Single	0.16***	0.21***	0.22***	0.19***	0.26***	0.28***
	(0.04)	(0.05)	(0.06)	(0.05)	(0.06)	(0.08)
Mother: In partnership	– omitted –					
Mother: Single	0.15***	0.15***	0.11*	0.28***	0.30***	0.31***
	(0.04)	(0.05)	(0.06)	(0.05)	(0.06)	(0.08)
Hospitalizations						
Father: 0 visits	– omitted –					
Father: 1 visits	0.10***	0.12***	0.14***	0.13***	0.15***	0.18***
	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.05)
Father: 2 visits	0.15***	0.07	0.16***	0.27***	0.24***	0.33***

Continued on next page

	(0.04)	(0.05)	(0.06)	(0.05)	(0.06)	(0.08)
Father: 3 visits	0.11*	0.10	0.11	0.16**	0.18*	0.18
	(0.06)	(0.07)	(0.10)	(0.07)	(0.09)	(0.12)
Father: 4 visits	0.36***	0.42***	0.44***	0.50***	0.55***	0.58***
	(0.09)	(0.10)	(0.13)	(0.11)	(0.14)	(0.16)
Father: 5+ visits	0.18**	0.16*	0.31***	0.27***	0.33***	0.54***
	(0.08)	(0.09)	(0.11)	(0.10)	(0.12)	(0.16)
Mother: 0 visits	– omitted –					
Mother: 1 visits	0.17***	0.18***	0.18***	0.21***	0.24***	0.28***
	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)	(0.04)
Mother: 2 visits	0.28***	0.26***	0.28***	0.37***	0.40***	0.41***
	(0.03)	(0.04)	(0.05)	(0.04)	(0.05)	(0.06)
Mother: 3 visits	0.43***	0.41***	0.43***	0.56***	0.60***	0.66***
	(0.05)	(0.06)	(0.07)	(0.06)	(0.07)	(0.09)
Mother: 4 visits	0.37***	0.33***	0.43***	0.48***	0.51***	0.62***
	(0.07)	(0.08)	(0.10)	(0.09)	(0.11)	(0.14)
Mother: 5+ visits	0.51***	0.60***	0.56***	0.68***	0.83***	0.93***
	(0.06)	(0.07)	(0.09)	(0.07)	(0.09)	(0.11)
Placements						
Father: No placement	– omitted –					
Father: Less than 2 years in placement	0.28***	0.25***	0.26***	0.41***	0.42***	0.46***
	(0.07)	(0.08)	(0.09)	(0.10)	(0.12)	(0.14)
Father: 2 years or more in placement	0.27***	0.26***	0.23***	0.36***	0.37***	0.32**
	(0.07)	(0.07)	(0.09)	(0.10)	(0.11)	(0.13)
Mother: No placement	– omitted –					
Mother: Less than 2 years in placement	0.37***	0.37***	0.36***	0.59***	0.63***	0.65***
	(0.05)	(0.06)	(0.07)	(0.08)	(0.09)	(0.10)
Mother: 2 years or more in placement	0.46***	0.46***	0.42***	0.62***	0.72***	0.72***
	(0.06)	(0.06)	(0.07)	(0.08)	(0.09)	(0.11)
Criminal charges						
Father: No charges	– omitted –					
Father: 1 charge	0.29***	0.29***	0.31***	0.44***	0.48***	0.53***
	(0.03)	(0.04)	(0.05)	(0.04)	(0.05)	(0.06)
Father: 2 charges	0.46***	0.43***	0.44***	0.67***	0.68***	0.78***
	(0.05)	(0.06)	(0.07)	(0.07)	(0.07)	(0.09)
Father: 3 charges	0.32***	0.41***	0.47***	0.56***	0.66***	0.80***
	(0.07)	(0.08)	(0.09)	(0.10)	(0.11)	(0.13)
Father: 4-5 charges	0.55***	0.50***	0.49***	0.86***	0.89***	0.95***
	(0.07)	(0.08)	(0.09)	(0.11)	(0.12)	(0.14)
Father: 6+ charges	0.48***	0.57***	0.66***	0.85***	0.95***	1.09***
	(0.07)	(0.08)	(0.09)	(0.11)	(0.12)	(0.14)
Mother: No charges	– omitted –					
Mother: 1 charge	0.21***	0.22***	0.25***	0.41***	0.46***	0.54***
	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)	(0.09)
Mother: 2+ charges	0.19**	0.21***	0.20**	0.46***	0.57***	0.61***
	(0.07)	(0.08)	(0.09)	(0.11)	(0.12)	(0.14)
N - Training data	119,804					

Notes: Estimates from logistic regressions of various outcomes on the full set of predictors. For predictors with missing information we include an additional dummy indicating missings but do not report the corresponding coeffcient estimates. Robust standard errors in parentheses.

Table D.7: Optimal risk score – Regressions

Predictor	Education	Income	Social benefits	Criminal charges	Placement	Hospitalizations	Psychiatric condition	SB indicator
Father's education								
Compulsory schooling	26.8	-3.8	24.5	17.4	20.5	29.6	13.0	23.5
	[24.1, 29.6]	[-7.5, -0.2]	[22.0, 27.0]	[15.3, 19.5]	[17.8, 23.3]	[26.0, 33.3]	[9.3, 16.8]	[21.3, 25.7]
Vocational education and training	18.0	-15.5	14.1	10.0	12.1	21.9	1.8	13.0
	[15.2, 20.7]	[-19.0, -12.0]	[11.6, 16.6]	[7.9, 12.0]	[9.3, 14.9]	[18.4, 25.5]	[-1.9, 5.5]	[10.9, 15.2]
High school	12.1	-1.1	6.8	6.8	4.9	9.2	1.2	7.8
	[8.7, 15.5]	[-5.9, 3.8]	[3.4, 10.2]	[4.1, 9.6]	[1.3, 8.5]	[4.2, 14.2]	[-4.0, 6.3]	[4.9, 10.7]
Short cycle higher education	10.4	-10.1	7.6	3.8	7.7	15.5	1.1	7.7
	[6.8, 14.0]	[-15.5, -4.7]	[4.0, 11.2]	[0.8, 6.8]	[3.9, 11.5]	[10.4, 20.5]	[-4.3, 6.6]	[4.6, 10.8]
(Vocational) bachelor's degree	6.3	-2.4	5.7	3.6	3.6	9.2	1.9	4.9
	[3.2, 9.4]	[-6.2, 1.4]	[2.8, 8.5]	[1.3, 5.9]	[0.4, 6.8]	[5.1, 13.2]	[-2.2, 6.0]	[2.4, 7.4]
Master's degree/PhD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mother's education								
Compulsory schooling	27.0	-0.9	27.7	18.4	30.0	34.0	18.9	26.8
	[23.3, 30.8]	[-5.5, 3.7]	[24.3, 31.1]	[15.7, 21.2]	[25.8, 34.1]	[29.0, 39.0]	[14.0, 23.9]	[23.8, 29.8]
Vocational education and training	15.8	-14.6	14.2	9.8	17.6	24.5	5.8	13.4
	[12.0, 19.6]	[-19.2, -9.9]	[10.7, 17.6]	[7.0, 12.6]	[13.4, 21.8]	[19.5, 29.5]	[0.9, 10.8]	[10.4, 16.4]
High school	9.0	-7.2	6.7	5.9	11.0	14.2	4.1	7.4
	[4.9, 13.1]	[-12.6, -1.8]	[2.8, 10.6]	[2.8, 9.1]	[6.5, 15.5]	[8.5, 19.9]	[-1.6, 9.8]	[4.0, 10.7]
Short cycle higher education	6.0	-11.6	6.2	4.0	9.4	14.4	-1.2	5.4
	[1.2, 10.8]	[-17.9, -5.2]	[1.7, 10.7]	[0.4, 7.7]	[4.1, 14.7]	[8.0, 20.7]	[-8.0, 5.5]	[1.5, 9.4]
(Vocational) bachelor's degree	5.2	-2.7	3.6	1.8	7.6	10.6	1.7	3.7
	[1.3, 9.0]	[-7.1, 1.6]	[0.2, 7.1]	[-1.0, 4.5]	[3.3, 11.8]	[5.6, 15.7]	[-3.2, 6.5]	[0.7, 6.7]
Master's degree/PhD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Father's income								
1st decile	19.2	27.5	23.1	13.8	23.0	11.1	24.1	22.2
	[17.2, 21.2]	[24.0, 31.1]	[20.9, 25.2]	[12.0, 15.6]	[20.9, 25.0]	[7.9, 14.3]	[20.6, 27.6]	[20.4, 24.1]
2nd decile	16.9	20.8	19.4	10.6	17.5	8.0	15.9	18.5
	[14.9, 18.8]	[17.2, 24.4]	[17.3, 21.6]	[8.8, 12.4]	[15.4, 19.6]	[4.9, 11.2]	[12.4, 19.5]	[16.6, 20.4]
3rd decile	15.3	14.9	18.6	6.1	14.0	5.6	15.6	16.8
	[13.3, 17.3]	[11.2, 18.6]	[16.4, 20.7]	[4.3, 8.0]	[11.8, 16.1]	[2.4, 8.8]	[12.0, 19.1]	[14.9, 18.7]
4th decile	11.3	9.5	13.8	4.0	9.3	3.8	9.1	11.0
	[9.2, 13.3]	[5.7, 13.2]	[11.6, 16.0]	[2.1, 5.9]	[7.1, 11.5]	[0.6, 7.0]	[5.5, 12.8]	[9.0, 12.9]
5th decile	9.9	4.9	12.0	1.8	8.8	1.9	8.5	10.0
	[7.9, 12.0]	[1.1, 8.7]	[9.8, 14.3]	[-0.1, 3.7]	[6.6, 11.1]	[-1.3, 5.1]	[4.8, 12.2]	[8.1, 12.0]
6th decile	9.5	3.8	9.8	1.0	6.4	5.1	6.8	8.4
	[7.4, 11.6]	[-0.0, 7.6]	[7.5, 12.1]	[-0.9, 2.9]	[4.1, 8.7]	[1.9, 8.3]	[3.0, 10.5]	[6.4, 10.4]
7th decile	9.4	1.1	8.4	1.9	5.8	2.3	6.3	7.6
	[7.3, 11.4]	[-2.7, 5.0]	[6.1, 10.7]	[-0.0, 3.8]	[3.5, 8.1]	[-1.0, 5.5]	[2.6, 10.0]	[5.6, 9.6]
8th decile	6.2	2.0	5.3	1.6	3.2	2.2	4.8	4.9
	[4.0, 8.4]	[-1.8, 5.8]	[3.0, 7.7]	[-0.3, 3.5]	[0.8, 5.6]	[-1.0, 5.5]	[1.1, 8.6]	[2.9, 6.9]
9th decile	5.9	0.5	3.3	0.4	1.6	1.2	3.9	3.2
	[3.7, 8.1]	[-3.3, 4.2]	[0.8, 5.7]	[-1.5, 2.4]	[-0.9, 4.1]	[-2.0, 4.5]	[0.1, 7.6]	[1.2, 5.3]
10th decile	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mother's income								
1st decile	15.6	21.5	21.8	10.4	21.2	14.2	22.6	21.5
	[13.6, 17.5]	[17.8, 25.2]	[19.6, 24.0]	[8.5, 12.4]	[19.1, 23.3]	[10.9, 17.5]	[19.0, 26.2]	[19.6, 23.4]
2nd decile	9.6	7.8	13.3	5.9	11.9	8.3	10.1	11.9
	[7.6, 11.6]	[4.2, 11.4]	[11.1, 15.5]	[4.0, 7.8]	[9.8, 14.1]	[5.0, 11.6]	[6.6, 13.7]	[10.0, 13.7]
3rd decile	5.9	0.9	6.7	2.5	5.0	5.0	1.4	5.1
	[3.9, 7.9]	[-2.8, 4.6]	[4.4, 8.9]	[0.6, 4.4]	[2.8, 7.2]	[1.7, 8.3]	[-2.3, 5.0]	[3.2, 7.0]
4th decile	5.2	-0.5	6.4	1.4	4.2	3.7	1.6	5.0
	[3.1, 7.2]	[-4.2, 3.3]	[4.2, 8.7]	[-0.5, 3.3]	[1.9, 6.4]	[0.4, 7.0]	[-2.1, 5.3]	[3.1, 7.0]
5th decile	5.3	-3.3	7.6	2.0	4.1	5.3	1.4	5.6
	[3.3, 7.3]	[-7.1, 0.5]	[5.4, 9.9]	[0.0, 3.9]	[1.9, 6.4]	[2.0, 8.5]	[-2.2, 5.1]	[3.7, 7.5]
6th decile	4.0	-3.9	5.2	1.5	3.7	3.6	-0.0	3.5
	[2.0, 6.1]	[-7.7, -0.1]	[2.9, 7.5]	[-0.4, 3.4]	[1.4, 5.9]	[0.3, 6.9]	[-3.7, 3.7]	[1.5, 5.4]
7th decile	1.8	-7.6	0.8	0.5	-0.8	2.9	-2.6	0.2
	[-0.3, 3.9]	[-11.4, -3.7]	[-1.6, 3.1]	[-1.4, 2.5]	[-3.1, 1.6]	[-0.4, 6.2]	[-6.3, 1.2]	[-1.8, 2.2]
8th decile	1.6	-3.1	0.4	0.4	-1.5	1.5	-2.8	-0.3
	[-0.5, 3.7]	[-6.8, 0.7]	[-1.9, 2.8]	[-1.5, 2.4]	[-3.9, 0.9]	[-1.9, 4.8]	[-6.5, 0.9]	[-2.3, 1.7]
9th decile	-0.9	-4.3	-0.9	1.1	-2.0	-1.0	-3.3	-1.3
	[-3.1, 1.4]	[-8.0, -0.6]	[-3.3, 1.5]	[-0.9, 3.0]	[-4.5, 0.6]	[-4.4, 2.4]	[-7.1, 0.4]	[-3.4, 0.8]
10th decile	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Sex								
Male	10.6	13.3	-2.1	39.9	3.3	-10.0	-16.7	4.7
	[9.8, 11.3]	[11.7, 14.9]	[-2.9, -1.2]	[39.0, 40.8]	[2.6, 4.1]	[-11.4, -8.7]	[-18.2, -15.2]	[4.0, 5.4]
Female	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Notes: Estimates from logistic regressions of various outcomes on sex, education (mother/father) and income (mother/father). Coefficients have been rescaled such that the difference between the highest and the lowest prediction (or risk score) is equal to 100 in line with Figure 10. 95% confidence intervals based on robust standard errors in brackets.

References

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 785–794. ACM Press.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33 (1): 1–22.

Wright, Marvin N., and Andreas Ziegler. 2017. "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." *Journal of Statistical Software* 77 (1).